SENTIMENT ANALYSIS OF DIALECTICAL ARABIC SOCIAL MEDIA CONTENT USING A HYBRID LINGUISTIC-MACHINE LEARNING APPROACH

GHADAH ALWAKID

Department of Computer Science & Technology
Nottingham Trent University

A thesis submitted in partial fulfilment of the requirements of Nottingham Trent University for the degree of Doctor of Philosophy

January 2020

Abstract

Despite the enormous increase in the number of Arabic posts on social networks, the sentiment analysis research into extracting opinions from these posts lags behind that for the English language. This is largely attributed to the challenges in processing the morphologically complex Arabic natural language and the scarcity of Arabic NLP tools and resources. This complex task is further exacerbated when analysing dialectal Arabic that do not abide by the formal grammatical structure. Based on the semantic modelling of the target domain's knowledge and multi-factor lexicon-based sentiment analysis, the intent of this research is to use a hybrid approach, integrating linguistic and machine learning methods for sentiment analysis classification of dialectal Arabic. First, a dataset of dialectal Arabic tweets was collected focusing on the unemployment domain, which is annotated manually. The tweets cover different dialectal Arabic in Saudi Arabia for which a comprehensive Arabic sentiment lexicon was constructed. This approach to sentiment analysis also integrated a novel light stemming mechanism towards improved Saudi dialectal Arabic stemming. Subsequently, a novel multi-factor lexicon-based sentiment analysis algorithm was developed for domain-specific social media posts written in dialectal Arabic. The algorithm considers several factors (emoji, intensifiers, negations, supplications) to improve the accuracy of the classification.s. Applying this model to a central problem of sentiment analysis in dialectical Arabic, these operational techniques were deployed in order to assess analytical performance across social media channels which are vulnerable to semantic and colloquial variations. Finally, this study presented a new hybrid approach to sentiment analysis where domain knowledge is utilised in two methods to combine computational linguistics and machine learning; the first method integrates the problem domain semantic knowledgebase in the machine learning training features set, while the second uses the outcome of the lexicon-based sentiment classification in the training of the machine learning methods. By integrating these techniques into a single, hybridised solution, a greater degree of accuracy and consistency was achieved than applying each approach independently, confirming a pragmatic solution to sentiment classification in dialectical Arabic text.

Copyright Statement

Declaration

I declare that the presented work in this thesis is the original work of the author except where explicitly stated otherwise in the text. I declare that this thesis as well as the materials contained in the thesis have not been used in any other submission for an academic award.

Parts of the presented work in the thesis have been published in:

Paper 1: "Challenges in sentiment analysis for Arabic social networks", published in "2017. *Procedia Computer Science*, *117*, pp.89-100. ACLing 2017: The Third International Conference on Arabic Computational Linguistics", Published by Elsevier B.V.

Paper 2: "Towards Improved Saudi Dialectal Arabic Stemming", published in "2019. *International Conference on Computer and Information Sciences (ICCIS)* (pp. 1-5)", Published by IEEE.

The data that support most of the findings of this study are openly available in [GitHub.com]. The dataset is available at [Ghadah-Alwakid/Unemployment_dataset]. The sentiment lexicon is available at [Ghadah-Alwakid/Unemployment_Lexicon]

Acknowledgement

First, all praises and thanks be to Allah, the Almighty, for helping me finish my study and for giving me the strength, ability and opportunity to undertake this research study and to persevere and complete it satisfactorily. Without his blessings, this achievement would not have been possible.

I am grateful to the Saudi Government and for Jouf University for the scholarship which enabled me to undertake a PhD research at the Nottingham Trent University. My deepest gratitude and honour to my supervisor, Dr. Taha Osman for his enduring support, patience and for their encouragement. His valuable comments have made a great impact on the progress of my research during all stages of my PhD research. Without his richness in experience, encouragement and continuous follow-up; this research would never have been. No word is enough to say ‗Thank you ‘. I would like also to express my deep gratitude to my second supervisor, Dr. Thomas Hughes-Roberts who was abundantly helpful and offered valuable feedback and suggestions; always prepared to help and always open to discussion.

I owe special thanks and love to my family for their steadfast support during my stay at UK. Words cannot express the debt of gratitude and admiration I feel for my parents Dr. Naif Alwakid and Nawal Alghalib for their support, sacrifice and encouragement during my study. Their impact on me has been greater than I could say, they are my ideals. I would like to express my gratitude to my beloved sisters Wafa and Wiam, who sharing me my journey in UK, for being very supportive. Many thanks to them for their constant support and encouragement. Special and lovely thanks to my sister Safa, my brothers and my husband Abdulhadi for their encouragement and support even I am abroad far away from them. Thank you for giving me strength to reach for the stars and chase my dreams. I dedicate this achievement to my little angel Talal, Indeed, I am truly grateful for having this lovely family.

Table of Contents

List of Tables

List of Figures

Chapter 1

1 Introduction

1.1 Research Background

Over the years, surveys have been utilised as the primary method for collecting and exploring opinions about a given topic. A selected sample of participants and a formatted questionnaire constitute the standard means for gauging opinion (Rubin and Babbie, 2016). However, the survey method has evolved with the growing popularity of social media and Internet access, such as Twitter and Facebook. Social media, in recent years, has played an important role in the interactions of users. A novel approach to learning about and analysing people's opinions has emerged through social media, particularly people's opinion regarding popular topics or products. This approach is called Sentiment Analysis (SA), which is what will be explored in this thesis. The aim is to identify opinions and emotions from a given text, with an emphasis on social media, in particular the Twitter medium. Twitter messages vary, ranging from politics to retail reviews. The aim of sentiment analysis is to clarify emotions represented in these messages with a polarity range of negative, neutral and positive.

Over the last decade, sentiment analysis has become a highly popular topic, as well as a speculative industry. LexisNexis[1], for instance, explores consumer attitudes and brand awareness through news outlets. Further examples within this industry include IBM SPSS[2], which forwards quantitative sentiment analysis summaries of data in an attempt to assist businesses with understanding consumer preferences. Major

---

[1] http://www.lexisnexis.com/risk/data-analytics.aspx

[2] http://www-01.ibm.com/software/analytics/spss/

social media news outlets, such as Politico[3] and The Washington Post[4], forward statistics and opinions regarding popular political figures. Global economic powerhouses, such as Wall Street, use sentiment analysis in their algorithmic analysis of trade. For example, they use OpFine[5], which allows for cutting-edge analysis of financial developments (Olson, 2012.).

Early research of sentiment analysis concentrated on product reviews, such as comments on Amazon.com[6], conducting sentiment analysis in a subjective manner. This approach allows for labelled data ratings; star ratings were utilised as indicative of quantitative expressions of opinion. After that, annotated datasets for general types of writing (blogs, news articles, and web pages) were created and became a popular method. For example, the popularity of Twitter ensured rich data tracking and sentiment analysis for a variety of applications, such as monitoring earthquakes (Sakaki et al., 2010). Although instrumental in the diversity of their applications, most sentiment analysis studies have concentrated on a singular source like customer reviews, and then researchers began adapting approaches to a variety of texts, such as social media content. Recently it has been usefully applied in a variety of political areas particularly in elections and voters' sentiments. However, analysis of political and social issues is a challenge, and there are questions as to whether sentiment analysis approaches, which are primarily designed for mining product evaluations, are suitable for analysis of complicated emotions like social media content.

Although initially lagging behind research surrounding other languages, in recent years, Arabic sentiment analysis has gained increasing popularity, covering trending problems in different domains Some of the earliest research on this topic was conducted by Ahmad et al. (2007) and Almas and Ahmad (2007). They apply sentiment analysis to a collection of news articles about finance, using a grammatical approach. Subsequently, publications referencing Arabic sentiment analysis increased year after year, as shown in Figure 1.1.

---

[3] http://news.cnet.com/8301-13772 3-57358111-52/politico-to-mine-facebook-forinsight-into-voter-sentiments/

[4] http://www.washingtonpost.com/politics/mention-machine

[5] http://www.opfine.com/

[6] https://www.amazon.com/

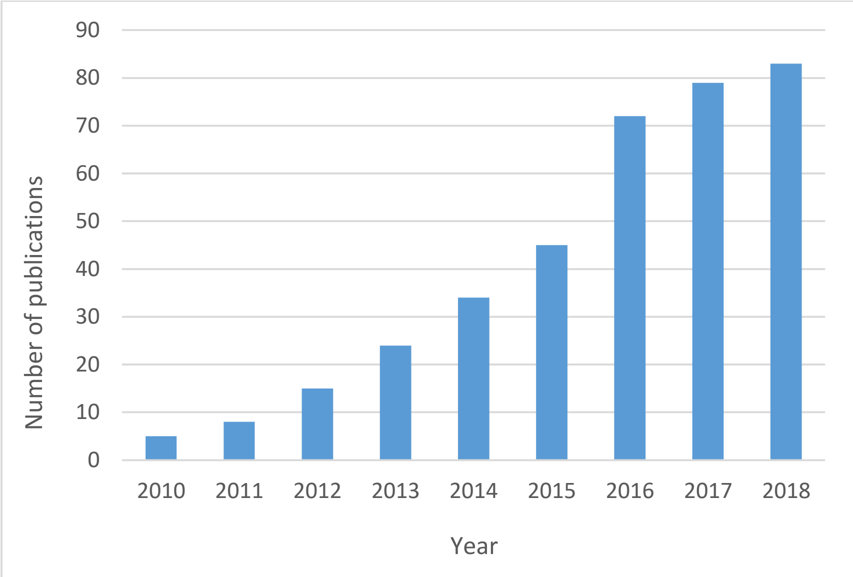


Figure 0.1: The number of papers published on Arabic sentiment analysis per year (Created for Study)

1.2 Research Motivation

Sentiment analysis has increasingly gained interest in both academia and industry. There has been clear progress in models of sentiment analysis, and the topic is an active area of research in spite of the breadth and diversity of global lanugages. As a critical justification for research in this field, Arabic remains the fifth most widely used language globally and the fourth most frequently used on the internet (Statista, 2020) Over the past five years there has been an increasing number of people using Twitter in Arabic countries to freely express their opinions about various issues that impact their daily lives (e.g. Arab Spring, Elections). This presents public authorities with the opportunity to deploy sentiment analysis on Twitter feeds to examine the impact of policies on the citizens.

Arabic sentiment analysis is an active research area, but it is still open to many obstacles. The Arabic language is morphologically entrenched and ambiguous; in other words, it has many irregular forms, complex morph syntactic alignment rules and a high degree of dialectal variants with limited writing rules. In comparison to English, there is a limited amount of freely available tools and resources for Arabic sentiment analysis, particularly for dialectical Arabic variations that do not abide by the syntactic rules of the Modern Standard Arabic (MSA).

The overall objective of this research effort was to develop a sentiment analysis system that can capture the public sentiments expressed on social media about social issues, particularly those that are expressed in non-standard dialectical Arabic; achieving this goal would benefit various groups, including citizens, policy makers, journalists and civic organizations.

1.3 Research Aim and Main Questions

The aim of this study was to propose a technique for achieving high sentiment analysis accuracy for tweets written in non-standard dialectical Arabic extracted from social media (Twitter). The problem domain selected for this research was the trending topic of unemployment in Saudi Arabia. Lexicon-based analysis and machine learning are the most common approaches for opinion classification; therefore, the hypothesis of this research has predicted that combining the two approaches in a hybrid sentiment approach that takes into consideration multiple factors impacting the precision of the lexicon analysis, such as dialectical Arabic NLP and considering negation, would result in a sentiment analysis system capable of opinion classification in dialectical Arabic text with high degree of accuracy. Achieving this research aim required addressing the following two primary research questions:

Primary Question 1: Can a hybrid approach combining domain semantic knowledgebase features with machine learning improve the performance of sentiment analysis?

Primary Question 2: Can a hybrid approach combining multi-factor lexicon-based sentiment analysis scores with machine learning improve the performance of sentiment analysis?

In addition to answering these two central research questions, the following 6 sub-questions were answered over the course of this study.

RQ1. What are the main challenges in utilising the methods and tools designed for MSA in the NLP of dialectal Arabic?

RQ2. Can a domain specific framework support a knowledge-based approach to dialectical Arabic sentiment analysis?

RQ 3. Which linguistic features of the Arabic language can impact lexicon-based sentiment analysis, and how can these be collectively considered to improve the accuracy of the analysis?

RQ4. Can the Semantic knowledgebase improve the accuracy of the feature extraction task? How can the semantic modelling of the domain knowledge further contribute to improving lexicon-based sentiment analysis?

RQ5. What is the impact of Arabic language light stemming on the performance of machine learning sentiment classification?

RQ6. What is the optimum algorithm and features set for utilising machine learning in dialectal Arabic sentiment analysis?

1.4 Thesis Contributions

The work described in this thesis has provided the following contributions to this field of study:

- Developed a novel stemming approach for dialectical Arabic that integrates the Information Science Research Institute (ISRI) stemmer and a rule-based stemmer, which was developed in-house. The new approach addresses the challenges of dialectical Arabic stemming. The proposed stemmer was found to provide improved accuracy compared to other stemming algorithms.

- Developed a gold-standard corpus[7] for multi-dialects Saudi Arabic sentiment analysis is generated by the manual annotation of tweets.
- Created a comprehensive multi Saudi dialects for Arabic sentiment lexicon[8]. The lexicon construction process includes sentiments, negation, emoji and special phrases, including supplications, proverbs and interjections.
- Experimented with a novel phrase-based method for handling supplications in dialectal Arabic in an attempt to extract the sentiment from tweets as accurately as possible.

---

[7] https://github.com/GhadahAlwakid/Unemployment_dataset/blob/master/Tweets_Unemployment%20dataset.csv

[8] https://github.com/Ghadah-Alwakid/Unemployment_Lexicon

- Developed a novel, multi-intensity lexicon-based sentiment analysis algorithm that considers several factors to improve the accuracy of classification, including emojis, intensifiers, negations, and special phrases (supplications, proverbs, and interjections).
- Modelled the optimum machine learning classifiers for sentiment analysis of social media content in dialectical Arabic social media. Determined the most suitable features, such as N-grams and TF-IDF, that function at an increased rate with dialectal Arabic sentiment analysis in alternative machine learning classification. This will aid new researchers to provide a baseline within this area.
- Presented a novel, linguistic-machine learning hybrid approach for sentiment analysis of social media content in Saudi dialectical Arabic alongside hybrid methods for sentiment analysis of social media content in dialectical Arabic, a hybrid semantic knowledgebase-machine learning approach and hybrid lexicon based-machine learning approach, which resulted in significant improvement of the accuracy of sentiment classification of dialectical Arabic text.

1.5 Research Methodology

The research methodology adopted for this project was based on standard research activities that included a literature review, requirement analysis and refinement, incremental and iterative development, and evaluation.

1.5.1 Literature Review

This study involved an evaluative literature review of Arabic natural language processing, corpus construction, the semantic knowledgebase to assisted sentiment analysis, lexicon-based sentiment analysis, machine learning algorithms and hybrid approaches for sentiment analysis. The literature assessment ensured the originality of the study to avoid any repetition of existing studies. All relevant fields were processed through an iterative approach throughout the progress of the research. Related theses were assessed and prior experimental models yielded substantial input during the requirement analysis, refinement procedure and overall analysis.

1.5.2 Requirement Analysis

Throughout this research, specifications for the methodology and research approaches were refined, analysed, and examined to determine their relevance in giving adequate responses to research questions. This study was undertaken to study the sentiment analysis of dialectical Arabic social media content. Due to the lack of open dataset resources for the Arabic language, it was not possible to find a dataset available publicly for Saudi dialect Arabic regarding social issues. To resolve this issue, a gold-standard corpus for sentiment analysis was created by manually annotating native Arabic tweets. The advantages of other pre-developed tools and approaches were adopted in order to account for tasks within the framework's terms, such as Natural Language Processing, lexicon construction, machine learning algorithms, and sematic knowledgebase tools. Due to the complexity of dealing with dialectal Arabic and after investigation of the tools and techniques, it became necessary to develop a novel tool or technique for achieving the main objective of providing adequate answer(s) to the underlying research motivation and central research questions.

1.5.3 Incremental and Iterative Development

The process of adapting the proposed solutions is central to an incremental and iterative progression in a field of study that has historically neglected the Arabic language. Incremental progression involves adapting various stages of the framework incrementally, gradually, and persistently filling in framework gaps and omissions. Iterative development is a revision of a strategy to adapt and enhance independent phases of the framework. The dataset is incrementally and iteratively developed in order to assess required tools and approaches for implementing the proposed framework.

1.5.4 Evaluation

It is well established that evaluative performance of sentiment analysis classification systems uses the following four indices), see Table 1.1:

- **Accuracy**: portion of all true anticipated instances compared with predicted instances
- **Precision**: the number of positive predicted occurrences compared with positive predicted instances
- **Recal**l: the number of accurate positive predicted occurrences against actual positive findings
- **F1-score**: a harmonic mean of recall and precision

$$P = \frac{TP}{TP + FP}$$

$$R = \frac{TP}{TP + FN}$$

$$Acc = \frac{TP + TN}{TP + FP + FN + TN}$$

$$F1 = \frac{2PR}{P + R}$$

Table 0.1: Confusion Matrix

	Predicted Positives	Predicted Negatives
Actual positive instances	Number of True Positive (TP) instances	Number of False Negative (FN) instances
Actual negative instances	Number of False Positive (FP) instances	Number of True Negative (TN) instances

1.6 Thesis Organisation

This chapter has provided a comprehensive overview of the core research aim and underlying objectives, outlining the primary research questions and justifying the targeted orientation of this study towards meaningful and beneficial outputs for Arabic sentiment analysis and language processing. The remainder of the thesis is organised as follows:

Chapter 2 presents background of sentiment analysis approaches and introduces the use-case motivation scenario. Then overview of natural language processing (NLP) as the main enabling technology for Arabic sentiment analysis with illustrate the characteristics of the Arabic language and challenges of Arabic text in social media content.

Chapter 3 presents literature review of main approaches and methodologies in this research, this chapter present literature review of arabic natural language processing and tools, arabic stemming tools, lexical resources for dialectical arabic language processing. Also, the literature review of sentiment analysis approaches lexicon-based, machine learning and hybrid sentiment analysis.

Chapter 4 introduces details of the collected resources and developed tools for dialectical Arabic for the benefit of sentiment analysis.

Chapter 5 explains a novel approach for multi-factor lexicon-based sentiment analysis of social media content in dialectical Arabic. It includes an evaluation and discussion of the experimental results.

Chapter 6 presents a new algorithm by using machine learning approach for sentiment analysis of social media content in dialectical Arabic social media. It includes an evaluation and discussion of the experimental results.

Chapter 7 presents the architecture framework of the proposed hybrid approach for sentiment analysis of social media content in dialectical Arabic social media. It includes discussion of the experimental results. Then explains the usability of a sentiment analysis approach to aid government and decision makers.

Chapter 8 concludes this research and summarises the main outcomes of this work and outlines suggested further works.

Chapter 2

2 Overview of Sentiment Analysis and Natural Language Processing

2.1. Introduction

As Web 2.0 technology expands, so does the number of web forums and social media platforms; current internet users contribute their opinions, ratings, and reviews on a multitude of web sites, whether they are commercial, or news related. Analysing expressed positive and negative reviews is cumbersome and time consuming and, in turn, leads to the need for new techniques for extracting opinions in relation to listed topics, which is labelled sentiment analysis. Sentiment analysis (SA) involves references to Natural Language Processing (NLP) and text analysis in the extraction of sentiment derived from a text related to specific topic (Yi et al., 2003; Ghadeer et al., 2017). Turban et al. (2014) defines sentiment analysis as an approach for finding positive and negative views towards products and services via a plethora of textual data sources. It is imperative to analyse texts and data mining in sentiment analysis fields since opinions, sentimentality and personal viewpoints that undergird texts are of high significance within this field. The success of blogs and social media sites only confirms and enhances the importance of SA.

2.2. Background to Sentiment Analysis

Sentiment analysis involves creating a system for collating and assessing opinions posted in blog posts, reviews, tweets, or the comments section of various websites. The majority of users express opinions and ideas via social media. These textual inputs are vital for determining decisions making processes for research work, industry, and for individuals. In marketing, the success of an advertising campaign or an original product launch determines which product or service is popular, and it likewise determines the demographic responses regarding the pros and cons of a particular feature (Zhou et al., 2014; Tsytsarau and Palpanas, 2012). SA can be utilised in differing aspects of society and interests such as business, public concerns, finance and politics. Within the business sector, a multitude of studies have been carried out with the aim to review consumer services and products. There are internet sources offering automated precis and evaluation of product reviews, e.g. Google Product Search[9]. In a business context, sentiment analysis is also utilised to explore and enhance brand reputation and online advertising and commerce. It is applied in the monitoring of reputed brands on Facebook and/or Twitter.

A further function of sentiment analysis in the business world is the advancement of e-commerce. The given premise is that consumers take note of others' opinions regarding travel, restaurants and retail outlets, through online tracking and research, hence resulting in Bing/Google star quality rating. An influential study within this field was developed by (Kang et al., 2012), providing a senti-lexicon for culinary reviews. Whether positive or negative, opinions have a distinct influence, cause and effect. In the modern-day digital world in which we live, published and shared opinions can enhance or destroy a brand's hitherto established and accepted reputation. Statistics portray that 40% of consumers derive an opinion of a business or company after access to 1-3 reviews online and 64% of potential software buyers gain access to a minimum of 6 online reviews before coming to a decision whether to buy, indicating the importance of company awareness of public opinion and ascertain the true feelings behind expressed views (Source: BrightLocal[10]).

2.3. Social Media and Sentiment Analysis

In respect to the politic arena, voting advice and feedback apps are a vital indication of sentiment analysis. Their analysis allows campaign advisers to track and influence public opinion concerning a variety of issues and monitor how speeches and

[9] https://www.google.com/shopping?hl=en

[10] https://www.brightlocal.com/research/local-consumer-review-survey/

activities of candidates may affect the vote itself. An in-depth analysis of tweets relating to the U.S. presidential elections. The sentiment analysis includes measuring the popularity of both major presidential candidates (Donald Trump and Hillary Clinton) on Twitter. Daily average sentiment score of tweets containing each candidate is calculated. All twitter attributes are utilised along with tweet text to better understand the sentiment displayed and the content generated or shared by the users (Buccoliero et al., 2020). Within this context, sentiment analysis is also utilised to clarify the public stance and viewpoint of politicians, including which issues they support or oppose, hence enhancing the quality and accuracy of information that voters can access. From social issues to political opinions to consumer sentiment, the field of sentiment analysis spans a broad and diversified spectrum of insights and interpretations.

In Arabic countries, interaction via social media is increasingly popular as these users deem it to be a vital tool for openly and freely sharing their views. Facebook is one of the most recognised social media platforms in Saudi Arabia. With 42,400,000 Facebook users in Egypt in January 2020, which accounted for 40.4% of its entire population (Napoleoncat, 2020). The effectiveness of Twitter as a social media tool was most noted during the Arab Spring uprisings in Syria, Yemen, Tunisia and Egypt. Specifically, during this period of vocalization and civilian uprising, citizens expressed their opinions freely through Twitter. By communicating across informal, socially connected media channels, protesters were able to not only organise movements and coordinate protests, but to raise awareness across other regional and global populations (Shearlaw, 2016). Summarizing the importance of this form of social activism, Alhindi (2012) reported a tweet frequency of approximately 40–45 posts per minute in Egypt alone on the 25th of January, 2011, precipitating the conflagration and social mobilisation that would form the basis for the Arab Spring.

The total number of Arabic users accessing twitter has increased by well over 100% since 2017. Referencing usage statistics, Crowd Analyzer (2019) claims that the number of active Twitter users within Saudi Arabia eclipsed 11 million between 2018-19, proving to be more popular than Egypt and UAE as shown in Figure 2.1. Saudi Arabia currently has a population of 33.85 million, 23 million (68%) of whom are active social media users. Further highlighting the importance of this high participation rate, GMI Blogger (2019) has reported that Saudi Arabia has the largest worldwide social media presence, with an estimated 43.8 million mobile subscribers. Indicative of multiple devices used by multiple users, this finding suggests that Saudi Arabian consumers are continuing to depend upon social media channels for cross-network communication and socialisation. The Hoot Suite (2019) reports that globally, Saudis are the largest community of active users on Twitter in the Arab vicinity. On average, Saudis spend an average of 2 hours and 50 mins daily on a variety of social media devices (Hoot Suite, 2019). Wonder (2019), has further confirmed that across all platforms, Saudi Arabian users predominately expressed an interest in nationalism, religion, culture and social development, with interactions mainly completed in Arabic. The total number of users preferring to communicate in Arabic is 3 million, with a mere 663 thousand opting for English.

Recently, According to the Saudi Ministry of Communications and Information Technology report in 2020, Over 18 million users of social media applications in Saudi Arabia. Due to the interest of the Saudi people in social media has grown, their impact on their daily lives has increased. The number of Saudi users of social applications and programs has doubled in the Kingdom during the recent years, from 8.5 million to 12.8 million users, and most recently, the number reached 18.3 million users, equivalent to 58 % of the population of Saudi Arabia. Smart phones constitute the largest platform in logging into social networks, with 260 minutes a day as an estimated average of logging per person using smart phones. Twitter dominate with the largest number of social media users in the Kingdom, where the number of Twitter users amounted to 9 million users (MCIT - Media Center , 2020).

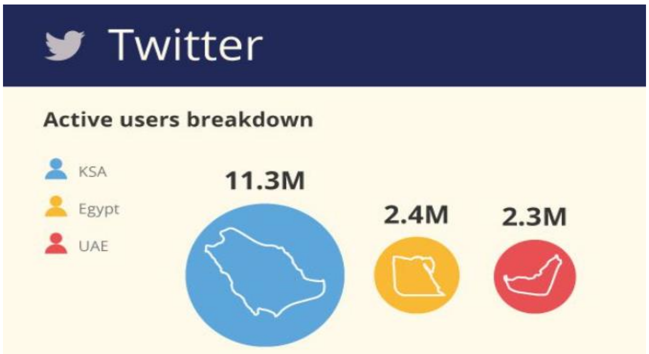


Figure 2.1: The number of active Twitter users in Saudi Arabia, UAE and Egypt
(Source: State of Social Media 2019 reported by Crowed Analyzer)

2.4. Unemployment Problem Domain in Saudi Arabia: A Case Study

An increasing number of Saudis use Twitter to freely express individual viewpoints on issues that impact their daily lives (Al-Harbi and Emam, 2015). This presents local authorities with the opportunity to use sentiment analysis on Twitter feedback in order to assess the impact of implemented policies (Government policies) on the Saudi populace (Aldayel and Azmi, 2016). A report distributed by the General Authority for Statistics ascertained the full extent of the Saudi Arabian workforce in 2016, and they discovered that the unemployment level was 12.1%; this is an increase from earlier in the same year, when unemployment was at a rate of 11.6%. The report also notes that an excess of 1/3 of unemployed Saudis are aged from 24 to 29 years old; women accounted for 63.4% of the unemployed and men 36.6% (General Authority for Statistics, 2016). Lack of job opportunities, particularly for young people of both sexes, exacerbates this problem, with an excess of 7,000 university graduates with doctoral or master's degrees who struggle to find jobs (Al-smayel, 2016).

In an attempt to address the problem of unemployment, Khalife (2019) notes that Saudi Arabia has launched a projects to create in excess of half a million jobs by the year 2030. These jobs will be in the private sector, and plans were released as part of the Saudi agenda to digitise and modernise the labor market.

Due to the significance of this issue for the Saudi Kingdom amongst workers and employers, of the issue of domestic unemployment is often a trending topic on Twitter in Saudi Arabia. Thus, the fundamental aim of this research was to develop a sentiment analysis system that could capture and explore public sentiments expressed on social media platforms, concentrating on the issue of unemployment in Saudi Arabia. A successful outcome of a Twitter-derived sentiment analysis would potentially benefit multiple research, political, and consumer groups including, journalists, civic organisations, citizens and policy makers. However, the sentiment analysis of Saudi dialects tweets is a challenging task for the following reasons

- The complexity of dealing with the dialectical Arabic; many tweets are written in non-standard dialectal Arabic, including orthographic errors, slang and spelling mistakes.

- Arabic NLP is complex in terms of morphology and structure. Indeed, Arabic grammar is a highly complex entity. Differing sentence structures can frequently be found within the Arabic language: a sentence may begin with a nominal or verb phrase, and, in some cases, it can begin with a noun phrase. Additionally, Arabic allows for different variations such as syntactical variations within all types of sentences. Many different parts of speech exist that are found only in Arabic. Additionally, Arabic is highly derivative and inflectional, containing many diacritics and word strands (Al-Shalabi, and Obeidat, 2008; Alhajjar et al., 2009). For instance, the same three-letter root can create many different words that have differing meanings. The same word can also exist in different forms, with added suffixes, affixes, and prefixes.
- Performing sentiment analysis is a challenge when exploring Twitter posts due to the fact that each tweet is limited to just 140 characters, however, Twitter's doubling of character count from 140 to 280 in 2018 increased the linguistic complexity of the posts. Twitter also makes use of URLs, hashtags, and user references (mentions). Additionally, users express their views in many ways, and their language may contain abbreviations and slang words, and there may also be repetition of letters as a means of showing emotion and emphasis. One major problem that many tweets are in non-standard dialectal Arabic, and it contain orthographic errors or spelling mistakes. The next chapter, then, will provide a more detailed examination of this issue.

2.5. Sentiment Analysis Approaches

There are three main approaches to the sentiment analysis process. These can be labelled and presented as Lexicon-Based, Machine Learning, and Hybrid approaches. Table 2.1 illustrate the strengths and weaknesses of sentiment analysis approaches.

2.5.1. Lexicon-Based Approaches

The lexicon-based approach is divided into two techniques: Corpus Based and Dictionary Based. Lexicon based techniques fundamentally concentrate on analysing the sentiment lexicon (i.e. the collation of words where each one contains a mark that indicates the negative, neutral or positive tone of the text to be explored). For the chosen text information, marks for the subjective words are assessed and inputted separately, and the maximum score will decide the overall polarity. The text is analysed via this sentiment lexicon (Kang et al., 2012).

2.5.1.1. Corpus-Based Approach

The corpus-based approach deals with the construction of a list of opinion seed words and is expanded and enlarged by extracting the information from the corpus text. Seed opinion words represent those general concepts and common words drawn from a pool of information that is domain-specific and directly linked to the topical or contextual origins of the discussion (Keshtakar and Inkpen, 2013). The corpus based approach typically involves a combination of sentiment analysis and statistical discrimination, weighing the accuracy of the output on a contextual basis (Keshtakar and Inkpen, 2013).

2.5.1.2. Dictionary-Based Approach

The dictionary-based approach is focuses on determining the opinion seed words from the chosen text through a dictionary search for synonyms and antonyms. Initially, a seed list is created by manually extracting opinion words. This is a complex process, since related opinion words are limited. Context oriented texts, such as a thesaurus and dictionary, are then explored to seek out and analyse antonyms and synonyms. At a later stage, synonyms are included in the list of seed words and the process is repeated until a sufficiently robust representation is collated from the overlapping compendium of similar words (Keshtakar and Inkpen, 2013).

2.5.2. Machine Learning Based Approaches

The machine learning approach consists of a supervised, semi-supervised and unsupervised learning. The features are utilised and extracted to perform classification through a multi-stage process which reconciles and analyses large scale datasets automatically. This technique is most popular in text classification because no human interaction is needed. There are three different approaches to machine learning including supervised, unsupervised, and semi-supervised.

2.5.2.1. Supervised Learning Approach

Supervised machine learning-based is a leading technique that has been broadly applied to sentiment analysis classification. Two sub-sets are needed, the first is the labelled data set for training and the second is test set data for comparative purposes. The accuracy and effectiveness of these techniques depend upon the accuracy of the training data; therefore, if there is any wrongly labelled data within the training set inaccuracies are likely to be observed. The primary algorithms within this category include Decision Trees (DT), Naïve Bayes (NB), and Support Vector Machines (SVM) (Badaro et al., 2019).

2.5.2.2. Unsupervised Learning Approach

Unsupervised learning techniques do not take advantage of the labelled or training set of data. In situations where it is difficult to label the input data, this technique is useful, allowing outputs to be derived from scalar results, rather than a more dependent, results-limited training set. These include algorithms such as K-means clustering and Word2Vector and have been commonly applied to large scale social media analyses. However, this process requires a large data repository in order to fit the model and generate accurate results. Model failure can result in incomprehensible or erroneous results that lead to a loss of time and retreat to another, more rigorous machine learning model (Zhang and Yu, 2017).

2.5.2.3. Semi-Supervised Learning Approach

Semi-supervised learning is a combination of the advantages found in both supervised and unsupervised learning approaches. Designed to compensate for a lack of labelled data, this solution allows for learning protocols to be expanded, emphasising multiple datasets (Chapelle et al., 2009). . Within this learning technique, the model is trained with the combined assistance from both labelled and unlabelled data (Chen et al., 2014).

2.5.3. Hybrid approach

The hybrid approach is a combination of the lexicon-based and machine learning approaches; hence it increases overall performance. For sentiment analysis, hybrid techniques draw upon more complex features of machine learning (e.g. ontologies, semantic networks) and analytical techniques from lexicon-based approaches to reconcile semantic variances that complicate the results (Mumtaz and Ahuja, 2016). Utilising both approaches improves the accuracy and performance of the sentiment analysis task (Mala and Devi, 2017).

Table 2.1: The strengths and weaknesses of sentiment analysis approaches

Sentiment analysis approaches	Strengths	Weaknesses
Lexicon-based Approach	• Wider term coverage • Simple to understand and implement • No training necessary • High speed of classification	• Requires large-scale external lexical resources. • Context not considered • Accuracy dependent upon size and quality of lexicon.
Supervised Machine learning approach	• The capability of adapting new cases • Creating a trained dataset for precise contexts and purposes. • N-grams representation of sentence. • Uses high order of n-grams including context.	• Costly in in terms of labelling data and time • An increase of sparsity with increase of order of n-grams. • Limited availability of NLP tools for differing dialectal Arabic.
Hybrid Approach	• High performance • Detection and measurement of sentiment analysis at concept level • Weak sensitivity to alterations in domain	• Expensive, time consuming and takes up space

2.5.4. Deep Learning Approach to Sentiment Analysis

The ubiquitous and persistent use of social media for cross-platform idea sharing and user communication has created datasets of an unprecedented scale which

Sharath and Tandon (2017) acknowledge must be analysed through automated systems capable of topically-specific sentiment analysis. By training a convolutional neural network (CNN) as a deep learning solution to tweet-based sentiment analysis, the researchers developed a corpus of more than 14,000 words, 10 overarching domains, and 370 topics that could be used to train the solution (Sharath and Tandon, 2017). By applying a conditional modelling solution to the CNN training, sentiment was then classified as either positive, negative, or neutral. The network itself was decomposed into two specific blocks, the sentence block and the topic block, creating bidirectional layers that were designed to reduce cross-entropy during the analytical exercise (Sharath and Tandon, 2017).

Central to the challenges of analysing diversified Twitter posts are the variations that manifest as a result of slang, emoticons, and contextual domains (Asghar et al., 2019). Deep learning models such as the hybrid solution proposed by Asghar et al. (2019) allow for each of these variables to be classified according to its sentiment polarity. Through deep learning and system training, variations and domain-specific indicators can be used to further subdivide the output, resulting in more accurate, comprehensive results (Asghar et al., 2019). By applying this technique to what Magumba et al. (2018, p.7) describe as a 'recurrent neural network', sequential information is retained and the computational units are subdivided into multiple layers. By applying this approach to complex linguistic challenges (e.g. native language, medical terminology), the deep learning process involves updating the weight of the layer before subsequent forward passes are conducted (Magumba et al., 2018).

Although Stojanovski et al. (2018) acknowledge the importance and value of machine learning approaches for prior sentiment analysis techniques, the deep learning solution relies upon a form of iterative classification approach which translate CNN outputs into sentiment-grouped outputs. This solution includes several central stages including pre-processing of the tweets, embedding, convolutional operations, pooling of output features, and classification of the fixed size vector (Stojanovsski et al., 2018). Due to the complexity of linguistic traits and incongruities, Stojanovski et al. (2018, p. 32220) have proposed that intuitive, neural language models are needed to generate word representaitons that can yield scalable feature vectors that 'encode syntactic and semantic regularities of the words'.

2.6. Levels of Analysis

Sentiment analysis is typically categorized into the following three levels

2.6.1. Document Level Analysis

The document concentrates on a single topic. Hence, texts concerning comparative learning are not relevant within the document level. This level classifies whether the document in question portrays the tone of a positive or negative sentiment (Pang et al., 2002; Turney, 2002).

2.6.2. Sentence Level Analysis

The sentence level expresses factual information from the sentences that portrays subjective opinions. i.e. good/bad. The sentence level analysis is a sentence by exploring the sentiment indicators and decides if each sentence conveys an opinion that could be categorized as negative, positive, or neutral (Wiebe et al. 1999).

2.6.3. Entity/Aspect Level Analysis

Entity/Aspect levels are adopted throughout the analysis. The central purpose of the entity level is to identify constructs, while the aspect level identifies and clarifies the opinion or sentiment. This approach is centrally based on the concept of an opinion residing of an opinion and an attitude (Liu, 2012).

2.7. Applications of Sentiment Analysis

Sentiment analysis can be utilised to explore underlying tone of a variety of different data sources (e.g. e-mails, memos, transcripts). The text can be in any format, including feedback, tweets, Facebook posts or comments, and all of these formats can announce their associated sentiments. In today’s highly competitive marketplace, businesses are compelled to closely monitor customers’ sentiments in order to gauge public reaction. Positive feedback is a reflection of customer satisfaction and may help the business grow, while negative and neutral feedback may indicate areas of deficiency or underperformance. The accuracy of such systems is low, however, since the machines are unable to comprehend and translate sarcasm and other complex or unreliable statements. Despite this limitation, the scale of text generated across social media is extremely large, requiring automated, autonomous solutions that are capable of making meaningful interpretations out of overlapping datasets. The following sections offer insights into several practical applications for enterprise purposes.

2.7.1. Social Media Monitoring

Social media monitoring yields business intelligence by extracting sentiment, meaning, and patterns from textual mining. Opinions from millions of tweets or posts float throughout social media networks daily. Mining outputs yield an overview of consumer brand perceptions, trending patterns, and potential threats. This type of analysis can be completed in minutes electronically due to the high-efficiency solutions offered via automated mining technologies (Nogueira and Tsunoda, 2018).

2.7.2. Product Management

Products evolve and are managed by their creators, but sales can be better influenced by evaluating consumer opinions and their change or patterns over time. Product managers can electronically merge data from various sources and analyse the data in order to identify specific patterns and models that are relevant to understanding consumer perceptions and needs. Customer-derived insights improve decision making and help to develop new strategies and initiatives to improve the product, the experience of the customer, and enhance product performance over time (Suchdev et al., 2014).

2.7.3. Government Policy Review

As instruments responsible for service and support of their civilian constituents, governments across the world issue new policies, laws, and guidelines that are derived from an intimate knowledge of civilian expectations and priorities. Once implemented, civilian discourse offers meaningful discussion of government policies, highlighting areas of deficiency or opportunity that can be resolved to improve service outputs. Both negative and positive feedback are vital to the assessment of the overall impact of any policy or scheme. Many applications in use are based on sentiment analysis such as legal assessments, brand value monitoring, and enterprise search (Mridula and Kavitha, 2018).

2.8. Overview of Natural Language Processing (NLP)

Natural Language Processing (NLP) is the main technology for data extraction from text documents. In recent years, research in Arabic Natural Language Processing (ANLP) enjoyed increasing attention, and several cutting-edge systems have been created for a wide range of uses, including speech synthesis and recognition, machine translation, data retrieval and extraction, text-to-speech conversion, tutoring, and localisation and multilingual data retrieval systems. These applications deal with a range of complex problems inherent in the style and structure of Arabic. Derived from its complex linguistic structure, Arabic creates unique problems for NLP solutions, requiring specialised modules and systems that are capable of reconciling these variances and incongruities (Habash, 2010). The ANPL applications must therefore cope with several complex practical problems aligned with the structure and nature of the Arabic language. As a result of such hurdles, ANLP systems are deemed null and void if they fail to consider specific linguistic features of the Arabic language. Therefore, successful solutions must be capable of reconciling the morphological aspects of Arabic in order to yield effective, consistent language processing outcomes.

2.8.1. Arabic Natural Language Processing

NLP, otherwise termed as Computational Linguistics, is a facet of computer science and dovetails the science of Artificial Intelligence (AI). NLP tools analyse texts as an automatic function, so there is no need for human involvement (Ghosh, 2009). The aim of NLP is to allow a machine to understand human expressions and language. The main NLP techniques relevant to the Arabic language are as follows:

2.8.1.1. Character Agreement

In Arabic, there are 8 characters that can be utilised as additions; they form additional primary characters dependent upon their placement in the word, for example the vowel letters such as letter Alef has several shapes: for example, (ا,أ,آ,إ),

Alef maksoura “ ى”, mirrors Alef, but is constantly confused when written alongside the letter ya “ ي”. Taa Marbota “ة” is regularly confused with ha “ ه”. Many forms of Hamza: “ ؤ“,” ئ ”, and “ ء ”, are interchangeable with parts of a word within sentences (Darwish et al., 2012).

2.8.1.2. Tokenization

This is the process of splicing a text; it places each word in isolation, which distinguishes the following word via the initial space; each division is then labelled as a token (Alhanjouri, 2017).

2.8.1.3. Named Entity Recognition (NER)

This is the process of identifying names of persons, data, expressions of times, phone numbers, organizations, locations, percentages and quantities etc. (Ghosh, 2009). NER aids in the location of isolated text to extract information and knowledge.

2.8.1.4. Part of Speech Tagging (POS)

This is the process of identifying each individual word based on its location and appearance location in common expressions, such as verbs, adverbs, nouns, and adjectives (Habash and Rambow, 2005).

2.8.1.5. Stemming

Stemming removes prefixes and suffixes from the word, returning it to its root state. There are four examples of affixes: Suffixes, Postfixes, Antefixes, and Prefixes, which can be applied to words (Froud et al., 2010).

2.8.1.6. Arabic Stop Word Removal

Stop words are those that need to be filtered out prior to processing the given text. Stop words should be deleted since it may misrepresent and skew the results, so they need to be ignored in order to enhance the research process (Alhanjouri, 2017).

2.9. Characteristics of the Arabic language

2.9.1. The Arabic Language Features

The Arabic language is considered Semitic (a language that is complex and uncommon in terms of its morphology) and is the officially recognised language of 22 countries worldwide (Boudad et al., 2018). An excess of 400 million people speaks it worldwide. It is deemed 4th most-used language on the internet (Boudad et al., 2018). It is also one of the 10 most-used languages on the web (Alison,2018). Arabic has 28 letters in its alphabet, and it is written from right to left; it uses a free-word order where several specific rules are in place. The morphology of Arabic consists of many intonations of root words (Abdelali et al., 2004).

Arabic exists in two formats: Modern Standard Arabic (MSA), which is the formal and most recognised language commonly used in the media and literature throughout the Arab world. MSA adheres to the grammatical rules of the Quran and has a vocabulary in excess of 1.5 million words. The second format is dialectal Arabic (slang), which is the common socially used language in Arab countries. Although dialectal Arabic is derived from MSA, it may involve variations in word choice and grammar, dependent upon the dialectal Arabic used (McCarus, 2008). ANLP, as a result, is a challenging process due to its lack of applicative referencing; instead, it derives context and meaning from a specific origin. For instance, a particular root such as كتب (katab) 'write' represents the source of words such as 'she writes' تكتب (taktob) or 'she wrote' كتبت (katabat), but this is not the case in English and other languages. However, communication in an Arabic social media context is carried out using dialectical Arabic rather than MSA Arabic. Dialectal Arabic substantially differs from MSA in terms of phonology, morphology, lexical choice, and syntax. Dialectal Arabic can be subdivided into six main groups: Gulf, Maghrebi, Egyptian, Iraqi, Levantine and others dialectical Arabic (Guellil et al., 2019).

2.9.2. The Difference Between English and Arabic Language

There are many differences between English and the Arabic language including lexis variations, grammatical distinctions, and syntax errors. The lexis variations include deletion non-vocalisation, inadequate lexicon, multiple meaning, and connotation and collocation. The grammar and syntax variations include distinctions arising from word order, gender and reference, incorrect analysis of input, tense and aspect, prepositions, definite articles, coordinators and conjunctions. Most Arabic spellings are phonetic since each Arabic letter represents a particular sound, and hence there are no silent letters like those in English. Furthermore, the Arabic language does not combine letters to come up with a specific sound. For instance, in the word ‘thing’ the ‘th’ sound in the English language is reduced to the ث character in Arabic. Arabic does not have the linking verb ‘to be’, and it also lacks an indefinite article (ARTC). Arabic lacks distinction between upper and lower case when words are written from right to left.

As a result, the letters vary in their form depending on whether they appear at the beginning, centre or end of a sentence. The letters that have the capability of being connected can be joined both in written and printed forms. The Arabic language has either natural or grammatical gender, and all nouns are either feminine or masculine. Grammatical gender applies to lifeless objects (O) while natural gender is used for living things. The productive gender masculine yields the feminine through the addition of the particular suffix ‘ة’ to the last part of the masculine word (Salem, 2009). Due to these intrinsic differences between English and the Arabic language, it is difficult to apply the English NLP tools to Arabic text, requiring the application of an ANLP solution.

2.9.3.Examples of Arabic Natural Language Processing

2.9.3.1. Free Word Order

Arabic language utilises free word order. According to Al Aqad (2013), there are multiple word orders used in the Arabic language. Due to this flexibility, Arabic is rich in grammatical structure and features multiple free word order solutions including subject-verb-object (SVO), verb-subject-object (VSO), verb-object-subject (VOS) and object-verb-subject (OVS) (Abu Shquier, 2014). The word order in the Arabic language (MSA and dialectal Arabic) is very different from the order of words in English, as can be seen in the following Table 2.2 and Figure 2.2. where the green checkmark represents the traditional arrangement with SVO agreement in English, followed by three variations. The red checkmarks indicate improper word ordering in the English language, but in Arabic, all four variations are identified as appropriate (green checkmarks) due to free word order associations.

Table 2.2: The Arabic Language Features - Free word order examples

Sentence Form	English	Arabic
SVO	The girl walks slowly √	البنت تمشي ببطء √ *albento tamshi bebote'*
VSO	Walks the girl slowly x	تمشي البنت ببطء √ *tamshi albento bebote'*
VOS	Walks slowly the girl x	تمشي ببطء البنت √ *tamshi bebote' albento*
OVS	Slowly walks the girl x	ببطء تمشي البنت √ *bebote' tamshi albento*

The Arabic language allows agreement between the subject and verb as a suitable morphological marking on the words to distinguish the subject from the object, employing a free word order. The position of the actor creates the difference between the sentences. In Arabic, the sentences bear similar meaning, whereas English has a standard (SVO) sentence form. In the other words, unlike English, the Arabic language has a free word order that does not restrict its arrangement of words (Alduais, 2012). For example, in English, adverbs come before the verbs they describe, whereas in Arabic the adverbs may come before or after the verbs they describe. Thus, the morphological syntactic analysis of the Arabic language is complex compared to English or other languages.

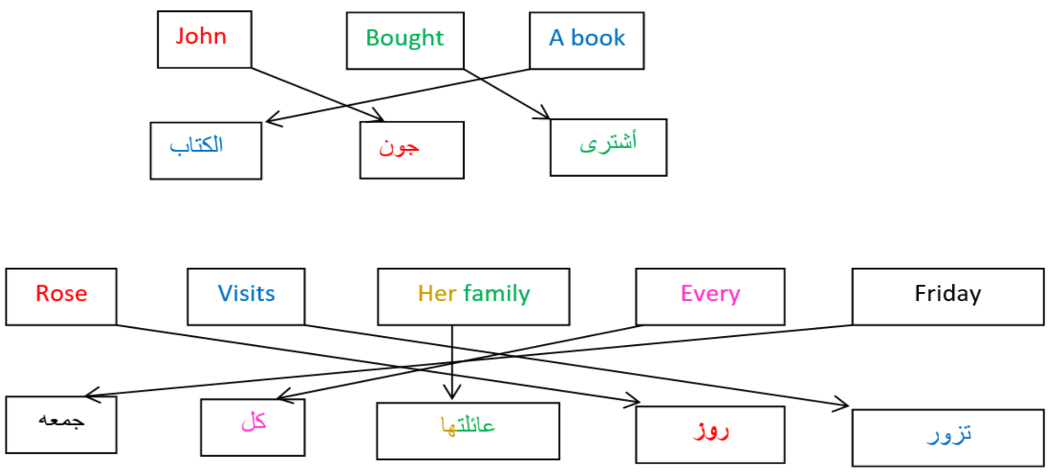


Figure 2.2: Examples of Arabic Free word order

2.9.3.2. Negation:

Negation is an English concept entailing the use of words such as ‘do not’, ‘does not’, ‘did not’ and ‘no’ as shown in Table 2.3.

Table 2.3: The Arabic Language Features – Negation examples

	English	Arabic (MSA and dialectal Arabic)
Do not, does not = لا	Do not give up	لا تستسلم *la tastaslem*
Did not = لم	He did not sleep all night	لم ينم طوال الليل *lam yanam twal allayl*
No = ممنوع / لا / ليس	No smoking	ممنوع التدخين *mamnoo‘ altadkheen*

2.9.3.3. Singular and Plural:

In Arabic (MSA and dialectal Arabic) non-imperative sentences, when the gender of the referent changes, there are morphological verbal changes (Adawood and Mohammed, 2008) as in Tables 2.4, 2.5 and 2.6.

Table 2.4: The Arabic Language Features - Singular and Plural examples

English	Arabic	Gender
He wrote the report	هو كتب التقرير *howa kataba altaqreer*	Masculine
She wrote the report	هي كتبت التقرير *hiyya katabat altaqreer*	Feminine

Sentences phrased by male and female speakers in different tenses

Table 2.5: The Arabic Language Features - Sentences phrased by male and female speakers in different tenses examples

English	Arabic	Tense	Gender
I write the report	أكتب التقرير *aktob altaqreer*	Present simple	Masculine / feminine
I wrote the report	كتبت التقرير *katabt altaqreer*	Past simple	Masculine / feminine

Sentences phrased by male and female speakers in plural

Table 2.6: The Arabic Language Features - Sentences phrased by male and female speakers in plural examples

English	Arabic	Tense	Gender
We write the report	نكتب التقرير *naktob altaqreer*	Present simple	Masculine / feminine
We wrote the report	كتبنا التقرير *katabna altaqreer*	Past simple	Masculine / feminine

2.9.3.4. Gender:

Nouns may classify into gender classes if they fall under a language with a 'grammatical gender' system (Badr et al., 2009). The grammatical gender may have an influence on the basis of a word's morphological or phonological features. This leads to difficulties in translating grammatical gender to the Arabic language (MSA and dialectal Arabic).

The literature indicates that a gender problem occurs from generalisation in the English language with elements such as 'I', which occasionally take the form of '**أنا**' in Arabic. Sensitive gender treatment is of great concern in the Arabic language (Holes, 2004), see examples in Tables 2.7, 2.8, 2.9 and 2.10.

Table 2.7: The Arabic Language Features – Gender examples

English	Arabic	Gender
I am a tourist (male)	أنا سائح *ana sae'h*	Masculine
I am a tourist (female)	أنا سائحه *ana sae'hah*	Feminine

'Grammatical gender' terminology is a two-level semantic component. Usually, it refers to the biological gender (male and female). In Arabic, this may represent as (man =**رجل** ragol) and (women = **امرأه** amra'ah), indicating gender specificity. Antagonistically, for such nouns as 'doctor' and 'driver', gender is generalised. Many Arabic words are changed according to their gender. However, if an Arabic word ends with (**ى ,ى**), (**ـه,ه,ة**) or (**ـاء, اء**) it is considered as feminine (Badr et al., 2009).

2.9.3.5. Suffixing

Suffixing – The linked taa' (ـه)

Table 2.8: The Arabic Language Features - Suffixing – The linked taa' (ـه) examples

English	Arabic (male)	Arabic (female)
Lawyer	محامي *mohami*	محاميه mohamiyah
Secretary	سكرتير *secretaire*	سكرتيره secretairah

Suffixing – al Alif al Maqsūra (ى)

Table 2.9: The Arabic Language Features - Suffixing – al Alif al Maqsūra (ى) examples

English	Arabic	Gender
The oldest brother	الاخ الاكبر *al'akh al'akbar*	Masculine
The oldest sister	الاخت الكبرى *al'okht alkobra*	Feminine

Suffixing – al Alif al Mamdūdah (ـاء)

Table 2.10: The Arabic Language Features - Suffixing – al Alif al Mamdūdah (ـاء) examples

English	Arabic (male)	Arabic (female)
Single	أعزب *a'azab*	عزباء 'azbaa'

Nonetheless, there are nouns in the Arabic language that are formally treated as feminine even though they are functionally masculine, such as (Hamza, حمزة) and (Moawya, معاويه).

2.9.3.6. Proper Nouns in the Arabic Language:

Closer consideration of changes in the gender of the referent will lead to changes at the phrasing and sentential levels, as illustrated Table 2.11.

Table 2.11: The Arabic Language Features - Proper Nouns in the Arabic Language examples

English		Arabic
A big castle	In Arabic, castle = masculine, which makes big = masculine	قصر كبير *qaser kabeer*
A big bus	In Arabic, bus = feminine, which makes big = feminine	حافلة كبيرة *hafelah kabeerah*

2.9.3.7. Changing Verbs According to Gender:

Wightwich and Gaafar (2005) highlight that, in the Arabic language, the verb will change according to gender. The example shows not only a change in the verb but also a gender-based change in three other words, as in Table 2.12.

Table 2.12: The Arabic Language Features - Changing Verbs According to Gender examples

English	Arabic	Gender
One of my students did not attend the session	لم يحضر احد طلابي الجلسة *lam yahdor ahad tolabbi algalsah*	Masculine
One of my students did not attend the session	لم تحضر احدى طالباتي الجلسه *lam tahdor ehda talebati algalsah*	Feminine

2.9.3.8. Using a Nominal Phrase with the Pronouns 'He' or 'She':

Consider the phrase, 'a smart manager'. In English, this could refer to either a male or a female. In fact, it is not essential to inflect the determinative and the adjective (ADJ) to make them agree with the head noun of the phrase, which remains unchanged (Ryding, 2005). However, in the Arabic language this should be changed to inflect the determinative and the adjective with the feminine head noun, see Table 2.13.

Table 2.13: The Arabic Language Features - Using a Nominal Phrase with the Pronouns 'He' or 'She' examples

English	Arabic	Gender
He is a smart manager	هو مدير ذكي *howa modeer dhaki*	Masculine
She is a smart manager	هي مديرة ذكية *hiyya modeerah dhakiyah*	Feminine

2.9.4. Morphological differences between MSA and the dialectal Arabic

The morphological differences between MSA and dialectal Arabic can be found in many aspects, as exemplified in Tables 2.14 and 2.15.

- **Future proclitic:**

Table 2.14: The Arabic Language Features - future proclitic examples

In English		MSA		Dialectal Arabic	
Will	He will go	سـ s سوف sawf	سيذهب *sayadhhab* سوف يذهب *sawf yadhhab*	حـ h بـ b رح rh	حيروح *hyrwh* بيروح *bayruh* رح يروح *rh yaruh*

- **The substitution of the pronouns:**

Table 2.15: The Arabic Language Features - the substitution of the pronouns examples

In English		MSA		Dialectal Arabic	
This That	This teacher That teacher	هذا *hadha* هذه *hadhih* ذاك *dhak* تلك *tilk* هؤلاء *hwla'*	هذا المعلم *hdha almuelam* هذه المعلمه *Hadhih almuelimuh* ذاك المعلم *hadha almaelam* تلك المعلمه *tilk almuelimuh* هؤلاء المعلمين *hwla' almuealimin*	هـ h هاك hak هذيك hdhik هذاك hadhak ذولاك dhulak هذولاك hadhulak هذيليك hadhilik هذاهمك hdhahimk	هالمعلم *halmelm* هاك المعلم *hak almaelam* هذيك المعلمه *hdhyk almuelamih* هذاك المعلم *hadhak almaelim* ذولاك المعلمين *dhulak almuealimin* هذولاك المعلمات *hdhwlak almuealamat* هذيليك المعلمين *hadhilik almuealimin* هذاهمك المعلمين *hadhahimik almuealimin*

- **Prepositions of Time and place**

Another observation is the use of one letter (abbreviated) such as (ع) letter in place of the (على), see Table 2.16.

Table 2.16: The Arabic Language Features - Prepositions of Time and place examples

In English		MSA		Dialectal Arabic	
In On	In the hotel On the table	في fi على ealaa	في الفندق *fi alfunduq* على الطاوله *ealaa altaawilih*	فـ f بـ b ع a	فالفندق *falfunduq* بالفندق *bialfunduq* عالطاوله *aaltawulh*

2.9.5.Particulars of Arabic Social Media Content - Challenges with Analysing Social Media Output

Text used in social media, particularly micro-text such as tweets, presents numerous challenges when compared to formally structured text, such as text that is presented in newspapers and scientific journals. As explained by Nabil et al. (2015), the challenges related to the sentiment analysis systems applied to Twitter arise due to several features of tweets. Tweets can contain unstructured language, numerous orthographic mistakes, slang words, ironic sentences, contractions, colloquial expressions, abbreviations, or idiomatic expressions.

Analysing tweets composed in Arabic is a particularly challenging task due to spelling inconsistencies, the use of connected words, and a lack of capitalisation, which would otherwise be used to identify features. In addition, most people write tweets as they speak; for instance, it’s important to consider the emotional character of Arabic tweeters and the frequent tendency to repeat letters for exaggeration, examples of which include ‘sorrrrrrrry’ and ‘noooooo’ as shown in Figure 2.3. Moreover, some characters have more than one form, an issue that highlights the need for normalisation, i.e., the unification of Arabic characters, as demonstrated in Table 2.17.

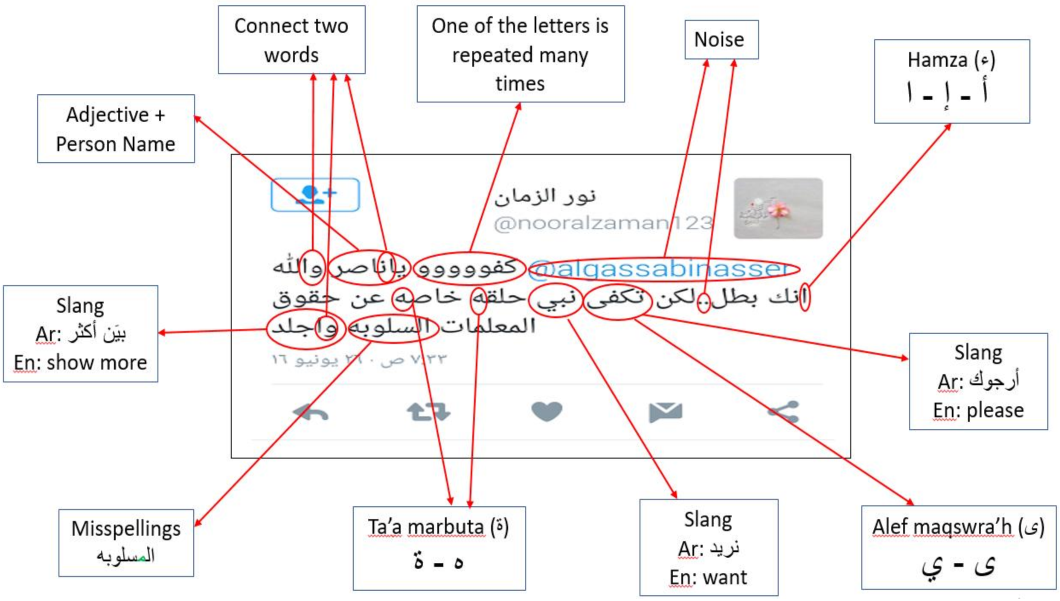


Figure 2.3: Example of challenges with analysing social media output

Table 2.17: Arabic normalisation

Letter	After normalisation	Example
ي *ya'a* ى *alef maqsura*	ي ya'a	علي Ali — male name على On top of
ة *ta'a marbuta* ه *ha'a*	ه *ha'a*	حلوة Beautiful حلوه Beautiful
إ *alef hamza'h* أ *alef hamza 'h* ا *alef wasel*	ا *alef wasel without hamza'h*	افضل Best أفضل Best

To summarise, Arabic social media content is associated with the following challenges:

- Unstructured language
- Orthographic mistakes
- Slang words
- Ironic and colloquial expressions, contractions, abbreviations
- Spelling inconsistencies
- Lack of capital letters in Arabic, which would otherwise be used to identify features
- Emoticons
- The tendency to repeat letters in writing to convey feelings

2.10. Characteristics of the Arabic Language Relevant to Sentiment Analysis

Ahmed et al. (2013) explained that the Arabic language is a morphologically rich language (MRL), in which a substantial quantity of information regarding syntactic elements and relations is articulated at the level of a single word. Certain sentiment analysis systems were developed for the English language at the word level; however, the English language has less morphological disparity. Therefore, the direct application of lexical features to the sentiment analysis of the Arabic language will lead to insufficient data. The reason for this is that one word in the Arabic language can have several different surface forms as examples in Table 2.18.

Table 2.18: Example of multiple forms of Arabic verbs

English word	Arabic word	Forms in English	Forms in Arabic
Love	حب (Root) *hub*	I love	أحب *'uhiba*
		He loves	يحب *yuhibu*
		She loves	تحب *tuhibu*
		They love	يحبوا *yuhbuu*
		We love	نحب *nahab*

Hence, in the English language, the verb 'love' has few forms and can assume individual features. In the Arabic dataset, however, there is a high probability that every word could have a substantial number of forms. Additionally, the majority of Arabic first names (and, to a lesser extent, family names) are derived from Arabic adjectives that can be easily confused for sentiments (Mohammed, 2016), see Table 2.19. This is a challenging problem in Arabic natural language processing (NLP) that is traditionally addressed by applying pattern analysis of POS tags of words in text to determine whether the word is a proper noun or adjective. However, this solution is more difficult to apply to dialectal Arabic where the accuracy of the POS tagger is especially poor.

Table 2.19: Examples of Arabic names

Arabic name	Adjective
نبيل *Nabil*	Noble
سعيد *Saeid*	Happy
جميله *Jamiluh*	Beautiful

Due to the use of diacritics and rich morphology, Arabic words with the same root can have incompatible emotional orientations. This poses a significant challenge when applying stemming mechanisms to identify the polarity of sentiments. Some instances of inconsistent sentimental words with the same Arabic roots are displayed in Table 2.20.

Table 2.20: Arabic is morphologically rich

Arabic word	In English	Sentiment	Root
تلاعب *talaeub*	Manipulate	Negative -1	لعب *laeib*
يلعب *yaleab*	Plays	Positive +1	
تمييز *tamyiz*	Discrimination	Negative -1	ميز *miz*
إمتياز *'iimtiaz*	Excellent	Positive +1	

The challenges presented in this section suggest that the effective sentiment analysis of Arabic tweets requires linguistic processing to clean the data. After deep understanding of the characteristics of the Arabic language and all the challenges of analysing social media content, it is evident that the NLP tools of MSA Arabic language may not work efficiently with dialectal Arabic. Also, the dialectal Arabic are different from country to country, so the Arabic NLP tool for the Egyptian dialect will not achieve good results with Saudi dialects, which is illustrated in detailed in the following chapter.

Addressing Arabic NLP challenges has attracted a lot of research interest, and this has resulted in a number of scientifically mature tools that attempted to process Arabic NLP. The maturity of these tools is less relevant in dealing with dialectical Arabic because of the complexity of the dialectal Arabic.

2.11. Chapter Summary

The chapter started with overview of sentiment analysis, then presented the importance of social media in Arab countries. The users in social media freely express individual viewpoints on issues that impact their daily lives. One of the main issue in Saudi Arabia is unemployment which is the problem domain used as the case study for this research. After that, the main sentiment analysis approaches were reviewed highlighting their strengths and weaknesses such as lexicon based approach, machine learning approach, the hybrid approach and deep learning approach. The characteristics of the Arabic language were explored in detail with specific focus on the challenges they present to social media sentiment analysis. This chapter illustrate an in-depth exploration of NLP and its application to the Arabic language. Through a comparative review of multiple linguistic traits and variations, these findings have highlighted the need for an alternative ANLP solution capable of reconciling Arabic traits and features during sentiment analysis.

Chapter 3

3 literature review

This chapter reviews studies in the literature that highlights the litruture in sentiment analysis. The majority of the online data is unstructured, high efforts are demanded to extract information from that data to be structured and understandable by utilising different techniques and approaches. Recently, the research community has widely acknowledged the use of sentiment analysis for knowledge representation and understanding people opinion. In this chapter the literature review of sentiment analysis applications and approaches will be illustrated.

3.1. Literature Review of Arabic Natural Language Processing

One of the earliest studies done on Arabic NLP is by Khoja and Garside (1999), which adopts a root-based approach. It utilises morphological analysis to derive the root of a specific Arabic example of vocabulary, in line with derivation patterns and vocalization variation. Khoja and Garside (1999) attempted to locate root Arabic words and based this location on predefined morphological analysis and root lists, thereby creating abstract roots. In a more recent study, Cunningham (2002) developed the General Architecture for Text Engineering (GATE), which is a recognised framework for language engineering applications, supporting text processing as its main function. The GATE tool can handle the Arabic language and it is openly available for public access, developed in Java. It allows for building blocks that can be used to create new modules (plug-ins). This is facilitated via GATE's component-based model and Application Programming Interface (API). It involves a group of re-usable processing resources for everyday NLP tasks. These are collated to form A Nearly New IE System (ANNIE), A, and can be utilised as individual components.

ANNIE is made up of the following main processing resources: semantic tagger, orthomatcher, tokenizer, sentence splitter, POS tagger, Gazetteer, and coreference resolver. The semantic tagger, called a JAPE transducer, is a set of rules written in the JAPE language, describing patterns that can be matched and annotations that can be created. Additionally, GATE has a huge set of plug-ins, including: Stemmer, Chunker, machine learning components, and WordNet. The GATE also supports formats including RTF, HTML, XML, SGML, plain text, and email.

The Stanford POS Tagger was originally developed for the English language and was derived at Stanford University. The tagger is a supervised system mirroring the maximum entropy model. The Arabic version, described in Toutanova (2003), was adapted within the training section of the Arabic Penn Treebank (ATB). It has an accuracy of 96.42% when applied to the Arabic language. ISRI Stemmer, which was developed by Taghva et al. (2005), is an example of light stemming. In the absence of a root dictionary, the ISRI algorithm adopts affix lists and the most common patterns to extract roots. MADA+TOKAN is a collection of tools for POS tagging, stemming and lemmatization, tokenization, morphological disambiguation, and diacritisation provided by Habash et al. (2009). MADA functions by exploring a set of possible analyses for individual words and choosing the best-fit analysis for the current context through support vector machine models that can classify 19 individuals, weighted morphological aspects. TOKAN receives the data gleaned from MADA to generate tokenized output in a vast collection of customisable formats. It achieves an excess of 96% accuracy regarding basic morphological choice and lemmatization. For complete diacritisation, MADA achieves an excess of 86%.

In the Al-Shammari Lemma-based Stemmer, Al-Shammari and Lin (2008) integrated root stemming, light stemming, and dictionary referencing. They utilised a stop list with an excess of 2,200 words containing noun and verb dictionaries as linguistic sources. Additionally, this enhanced success in clustering tasks. The Al-Shammari algorithm was more successful than Khoja and light stemmers regarding over-stemming evaluation (Al-Shammari, 2013). SALMA Tools is a conglomerate of tools, open-source standards, and resources that open the range and ability of Arabic word structure analysis, particularly morphological evaluation, to explore Arabic text corpora within differing formats, genres, and domains in vowelized and nonvowelized samples of text. Sawalha et al. (2013) forward the argument that a more fine-tuned presentation of text is required to consider the complexities of Arabic. The SALMA-Tagger is an example of a fine-grained morphological analyser, dependent upon linguistic data derived from traditional Arabic grammar studies and lexical resources; namely the SALMA–ABC Lexicon. It is a morphological tag-set that adapts a morphological analyser to complement appropriate linguistic data to each particular part or morpheme of the word (suffix, enclitic, proclitic, prefix, and stem). This tool yields a very high accuracy of 98.53%- 100% for the CCA test and 90.11%-100% relating to the Quran test.

Another solution identified as AraNLP by Althobaiti et al. (2014) is a free, Java-based library covering a variety of Arabic preprocessing tools. It is an attempt to harness the majority of widely used Arabic text preprocessing tools into a single library that can be simply utilised by combining or accurately adapting existing tools, hence creating new ones when desired. The library includes a root stemmer, part-of-speech tagger (POS-tagger), word segmenter, sentence detector, tokenizer, light stemmer, normalizer, and a punctuation/diacritic remover. A maximum entropy model has been created and used on Arabic text corpus regarding sentence boundaries. The performance achieved a score of 0.97 precision and .98 recall. For the token boundary detection, a MaxEnt machine learning model was created and trained, achieving a 0.97 precision and recall result. Monroe et al. (2014) developed Word Segmenter. At Stanford University, where a group was working on Arabic NLP, the researchers added to an existing MSA segmenter a simple domain adaptation approach and original features to segment dialectal and informal Arabic text (Monroe et al., 2014). The segmenter itself was based on a sequence classifier (Conditional Random Fields) to produce an Arabic conjunction, preposition, clitic segmentation, and pronoun. The significant advantage of this is that it processes Arabic text at a much speedier rate compared to other systems in existence, enhancing the segmentation F1 score on Egyptian Arabic corpus to 92.09%, compared to 91.60% for a different segmenter designed for the same purpose.

Antoun et al. (2020) designed an Arabic model to boost the recent technological advances in a number of NLU Arabic tasks. They built a bidirectional transformer encoder using the BERT model (Devlin et al., 2018). This model is commonly used as the framework for the most advanced results in various language NLP tasks. In order to achieve the same success as BERT in English, they pre-trained the BERT specifically for the Arabic language. AraBERT's success is comparable to Google's multilingual BERT and other cutting-edge approaches. The results showed that AraBERT has achieved the latest developments in Arabic NLP tasks in most proven applications. In order to democratise the processing of Arabic, Alyafeai et al. (2020) created ARBML with a collection of demonstrations and resources to allow developers, users, and researchers to make use of it easily. They rework and render the NLP pipeline for Arabic.

3.2. Literature Review of Arabic NLP Tools

Throughout the literature in this field, an emergent number of studies have explored the viability of Arabic NLP for MSA. MADAMIRA, for example, was developed by Pasha et al. (2014), and focuses explicitly on MSA. It is not dedicated to MSA in isolation but also takes into account the Egyptian dialect. MADAMIRA, offers a lot of NLP tasks such as phrase chunking, named entity recognition, feature modelling, and tokenization. YAMAMA is a NLP tool for morphological analysis developed by Khalifa et al. (2016). Similar to the work completed by Pasha et al. (2014), Khalifa et al. (2016) concentrated on MSA in relation to the Egyptian dialect, producing a solution that they suggest is around five times faster than MADAMIRA.

Another NLP stemmer tool, Farasa was developed by Abdelali et al. (2016) and adopted a linear SVM solution. To support Farasa, the authors compare it with two additional segmenters: The Stanford Arabic Segmenter (SAS) (Monroe et al., 2014) and MADAMIRA (Pasha et al., 2014). Although the Farassa stemmer achieved better results, it included limited NLP tasks, limiting its overall functionality for more complex analyses. To reduce the complexity of such processes, Abainia et al. (2018) introduced ARLSTem, an Arabic light stemmer solution which eliminates prefixes, infixes, and suffixes. Similar in their approach, Zalmout and Habash (2017) presented a NLP tool based on Recurrent Neural Networks (RNN) which addresses Arabic morphological disambiguation. Their model yielded accurate outcomes; however, as it was designed for MSA, experiments were not conducted with dialectical Arabic

Text. In presenting their findings, Zalmout and Habash (2017) introduced the Long Short-Term Memory Model to demonstrate that their findings had achieved an accurate result.

Drawing upon prior techniques in this field, Shahrour et al. (2015) introduced the CamelParser, a tagger based upon the MADMIRA model originated by Pasha et al. (2014). Using SVMs, the tagger ranks all the possible analyses that resulting from a morphological analyser. It enhances the MADAMIR by utilising syntactic analysis to improve the accuracy of the output. More recently, NUDAR, a Universal Dependence treebank for the Arabic language, was proposed by Taji et al. (2017). This fully automated conversion of the Penn Arabic Treebank was forwarded to the syntactic representation. Subsequently, More et al. (2018) designed their Universal Morphological Lattices for Universal Dependency (UD) Parsing to Arabic.

As researchers strive toward a solution for Arabic NLP, some recent scholars have targeted a variety of orthographic standards, conventions, and rules. For example, Habash et al. (2018) undertook to establish commonly recognised guidelines with adequate specificity to analyse dialectal Arabic. One of the most vital topics were POS tagging. which recently has been attracted the researchers to addressed to process dialectal Arabic. Darwish et al. (2018) also provided a POS tagger, which relied on a sequence mark of Conditional Random Field. This tagger is devoted to address several dialects such as Gulf, Egyptian, Levantine. They manually segmented 350 tweets into each dialect to test their study. Samih et al. (2017) suggested a segmenter that would use neural networks to target 350 annotated tweets. The authors adapted segmentation for the Arabic language by applying sequence grouping based on character models to the outputs (Samih et al., 2017). An alternative approach introduced by Saadane and Habash (2015) demonstrates that Egyptian and Tunisian dialects could be applied to the Algerian dialect, allowing for combinative analysis. Highlighting the importance of dialectical considerations, Zribi et al. (2013) developed a system for adapting a related morphological analyser extracted from the MSA. They used Tunisian dialect lexicons which were developed on the basis of an existing MSA lexicon (Zribi et al., 2013).

For the Gulf dialects, a POS tagger was recently proposed by Alharbi et al. (2018) who noted that SVM derived results were improved by using a Bi-LSTM label.

For Emirati dialectal Arabic, Khalifa et al. (2018) forwarded a manually annotated corpus of on a large-scale morphology. They used about 200,000 Gumar-derived words in Khalifa et al. (2016). Similarly, the Analyser for Dialectal Arabic Morphology (ADAM) was developed by Salloum and Habash (2014). They measured the performance of ADAM in terms of two dialectical Arabic such as Egyptian and Levantine. Another morphological analyser and tagger was studied by Al-Shargi et al. (2016). This analyser was trained upon a morphologically annotated corpus that was manually constructed by the authors and adapted the annotation interface DIWAN (Al-Shargi and Rambow, 2015). Their analyser with a focus on the dialect of Sanaani, Yemeni and Morocco. ADAM is comparable to CALIMA which is designed by Habash et al., 2012, a morphology analyser for Egyptian dialectal that requires great durability and expensive resources to be built.

Continuing to extend these dialectical solutions, an Arabic morphological analyser was introduced by Khalifa et al. (2017) exploring over 2,500 verbs for Gulf dialects. The authors used dual resources including a lexicon of verbs and a collection of root-abstracted paradigms (Khalifa et al., 2017). In contrast, the MADARi interface is an annotation tool developed by Obeid et al. (2018) to assess text derived from only the Egyptian dialect. It is a Web based interface which supports spelling correction and morphological annotation. Zalmout et al. (2018) recently argued for a neural morphological tagging and disambiguation model relating to the Egyptian dialect, with a variety of expansions to cope with the loud and irregular content. The authors adapted the CNN and LSTM models to generate character embedding.

CAMel Software, which includes a group of open-source tools for natural Arabic language processing in Python, has been introduced by Obeid et al. (2020). CAMeL Tools offers pre-processing, dialect recognition, identified object identification, morphological modelling, and sentiment analysis services. In addition, these tools have provided a variety of pre-processing tools popular with Arabic NLP but often re-implemented. A number of tools and packages are not really well-known or exposed over their own use, for certain pre-processing steps. With these utilities in the bundle, the workload of writing of Arabic NLP applications has been reduced and the pre-treatment is consistent between projects. Another NLP tool is ADAWAT, which is developed by Zerrouki et al. (2020) offer a range of applications, verb conjugator, Light stemmer, spell checker, dictionary synonyms, speech system document, Mishkaldiacrtizer, morphological analyser, vocalised texts body, and collocations. The primary approach to constructing rules and data used is rules-based. These tools are built with existing systems such as Hunspell Spell Checkers for millions of users using LibreOffice and Firefox.

3.3. Literature Review of Arabic Stemming

Stemmers have been developed for a wide range of languages, including English, French, Arabic, and Chinese. Across languages, several factors affect stemming. In English, for example, because the usage of affixes is less complex than in Arabic, English-language stemmers mostly focus on the removal of prefixes and suffixes (Al-Omari and Abuata, 2014). In addition, the numbers of words with irregular forms that are not amenable to direct stemming (e.g., 'write' and 'wrote') are limited and can be dealt with explicitly using a root lexicon. The design of stemmers for highly inflected languages such as Arabic (Larkey et al., 2002) requires a deep understanding of these languages' grammar and considerable linguistic expertise (Hammo, 2009).

The Khoja (1999) Arabic stemmer is a fast stemmer that works in two main steps: (1) by removing the longest suffix and prefix present in the input word, and then (2) matching the remaining word with a root library containing lists of known noun and verb patterns. The stemmer considers the inevitable irregularities in the language, with the aim of extracting the correct root from words that do not adhere to the general rules. Notwithstanding numerous tagging errors, the Khoja stemmer has been successfully applied to a variety of natural Arabic-language processing tasks (Larkey et al., 2002). The stemmer does have two drawbacks, however. First, the root dictionary requires maintenance to ensure that newly found words will be stemmed correctly; second, the stemmer does not cover all Arabic patterns and occasionally fails to remove all the affixes attached to words.

While the ISRI stemmer (Taghva et al., 2005) applies the same approach as the Khoja (1999) stemmer to word rooting, it was developed to cope with situations where it is impossible to root a word. In these cases, various normalisation techniques, such as normalising all shapes of the letter ta'a ('ه' - 'ة'), are applied to extract the word's stem. For example, rather than leaving the word unchanged, the ISRI stemmer removes the end patterns or certain determinants. This feature makes the ISRI stemmer capable of stemming rare and new words. It returns a normalised form for non-stemmed words and has more stemming patterns, such as /مفاعل /مفعول / فعلة/ فعيل/ فعول /فعال /فاعل / مفعل/ فعولة, and more than 60 stop words. Bsoul and Mohd (2011) proved that the ISRI stemmer makes excellent improvements to language tasks (for instance, document clustering) compared to non-stemmed approaches. The stemmer is incapable of addressing the irregular plural form, however. Pasha et al. (2014) developed the MADAMIRA tool in 2014, which is a morphological analyser that provides a set of features, including stemming, and is composed of two sub-tools. The first is MADA, created by Habash et al. (2009), and the second is AMIRA, designed by Diab et al. (2009). MADAMIRA also provides light stemming analysis by removing prefixes and suffixes from words. The tool does not explicitly define morphological rules, however. The FARASA stemmer (2016), which falls between heavy and light stemmers, performs an initial grouping of words, which allows for accurately conflating different variants into the same form while limiting over-stemming. The FARASA tool has a tag set of 16 basic tags, although the stemmer is limited in certain cases.

Kanaan et al. (2008) developed a rule-based stemmer in which the input word is compared with a single pre-defined list of Arabic patterns to find matches. If the pattern matches the word, then no changes are made. If the word does not match any patterns, then light stemming is done to remove the prefixes and suffixes. The size of the pattern list cannot be known, however. In their paper, the sample term lists contained only 21 words. Al-Kabi et al. (2015) developed an approach to detect the root using the Khoja (1999) stemmer. As in Khoja (1999), the algorithm in their study begins with the removal of suffixes and prefixes in the input word. The difference between the two algorithms is that the Khoja stemmer depends on matching the extracted stem with patterns in terms of their length. Accordingly, in Al-Kabi et al.'s (2015) approach, the pattern is chosen according two main criteria: the length and the common letters between the stem and the pattern. The results show that, while accuracy is greatly improved over the Khoja (1999) stemmer, the approach fails to extract roots from words shorter than four letters. Recently, Sameer (2016) developed a light stemmer approach to stemming Arabic words. In the proposed algorithm, the pre-defined lists of suffixes and prefixes are removed according to their order in the algorithm. The algorithm was not tested sufficiently, however, as only 14 words were used to test the algorithm. Although the stemmers previously discussed were developed primarily for MSA and cannot be directly applied to Saudi dialectal Arabic, the ISRI stemmer has demonstrated acceptable results when applied to Saudi dialectal Arabic in processing (Abozinadah and Jones, 2016).

Few studies have focussed solely on developing stemmers for Saudi dialectal Arabic. One example presented by Shoukry et al. (2012), implemented a customised stemmer for dialectal Egyptian. The main objective of the stemmer was to reduce the input word to its shortest possible form without compromising its meaning. The researchers tested their implemented stemmer against an available light stemmer and observed that their stemmer produced better results because it addressed dialect-specific issues (Shoukry et al., 2012). Similarly, Al-Gaphari et al. (2012) developed a system for working with the Sana'ani dialectal Arabic and MSA. Their approach is based on morphological rules that assist in the conversion of dialectal Arabic to the corresponding MSA (Al-Gaphari et al., 2012). Because their approach uses dialectal Arabic stemming to translate the Sana'ani dialectal Arabic into MSA, they implemented their method using a simple MSA stemmer.

Sabtan (2018), using a corpus-based approach, implemented a light stem for Arabic. The stemmer groups morphology variants of words into an Arabic corpus using similar characteristics and extracts their prefixes to create their common stem before removing them. Experimental findings show that 86% of the terms in the test set are grouped accurately in a similar reduced state, which is the possible stem. The reduced form is not the legal stem in some situations. The assessment indicates that 72.2% of the test words are reduced to the legitimate stem. Atwan et al. (2019) sought to explain light stemmers efficiency in the restoration of Arabic knowledge. The calculation of the light stemmer is carried out by the use of TFIDF because it considers that the main system is comparable to the primary system, without stemming, as a popular system of weighting compatible with the Linguistic Data Consortium (L DC). In order to achieve the best efficiency, the suggested light stemmer must be used. This study explores the effects of stemming and its effect of improving the text of Arabic. The analysis findings from this study are based on two measurements: accuracy and recalling. Therefore, the author clarifies the effect of the stemmer on improvement of light stemming performance in Arabic documents in this study.

3.4. Literature Review of Lexical Resources for Dialectical Arabic Language Processing

The widespread popularity of social media websites has led to frequent usage of unstructured text throughout the web. This text is often accepted objective, while still reflecting both facts and/or subjective viewpoints that include both sentiments and opinions. Sentiment analysis is a significant research field involving the identification of opinions within a given text and classifying expressions (e.g. positive, negative, neutral). Research has been carried out to develop and analyse a lexicon for the Arabic language. The following sections explore the extant literature in this field, whereby, two methods have been established by researchers to develop lexicon for sentiment analysis approach including manual and automatic.

The manual approach establishes a “sentiment” based on a set list of Arabic vocabulary that is gathered from a set dataset or established domain. The lexicon derived from this approach is generally highly accurate, but a disadvantage is that it is limited in size due the time it takes to collate and annotate. Resources were developed to enhance the value and effect of this approach. One such resource is ArabSenti (Abdul-Mageed et al.,2011); derived from the Arabic Tree Bank (ATB) part 1 V3.0, it includes 3,982 adjectives from 400 texts. Another is SIFFAT (Abdul-Mageed et al., 2012), which contains 3,325 adjectives under the headings neutral, positive or negative. This resource has evidenced clear improvement in accuracy in terms of determining sentiment as a subjectivity analysis tool. Adjectives used in both these resources were manually labelled as neutral, positive and negative by native Arabic speakers and were analysed by a linguistic expert. This area involves MSA and the data is not publicly available.

A further manual resource was created by Abdulla et al. (2013); this resource translated 300 words as a seed set from Sentistrength (Thelwall, 2013) to Arabic. Synonyms, antonyms and emoticons were adapted to expand this seed set. Differing from the previous resources, this resource does not include neutral language. Again, this resource is not publicly available since it is involved with MSA. A recent sentiment lexicon is NileULex (El-Beltagy, 2016; El-Beltagy, 2013). It includes compound phrases, as well as single words, from dialectal Arabic and MSA. Terms and compound phrases were derived from social media automatically, even though they were manually annotated. Stand-alone terms were deliberately unambiguous and compound phrases were adopted to avoid any misleading interpretations. NileULex contains 5,953 annotated terms labelled as positive or negative. This sentiment resource was used by the NileTMRG team (El-Beltagy, 2016) while participating in the SemEval 2016 Task 7 (Kiritchenko, 2016). NileULex was expanded by El-Beltagy (2017) to automatically assign scores or weights to entries and allowed the resulting resource, called "WeightedNileULex", to be available to the public.

Regarding the automatic approaches, it is clear that automatic approaches to sentiment analysis need to be developed. Though it may contain some erroneous data, it is nonetheless cheaper and takes up less time. Traditionally, automatic expansion is achieved through manual intervention and it collects English sentiment lexicon, which is then translated (or finds the nearest equivalent) into Arabic. Mourad and Darwish (2013) for example, revamped the manual ArabSenti to automatically expand the resource with graph reinforcement. ArabSenti was translated into English, and, by using mechanical translation tables (English-Dialectical Arabic, English-MSA), the lexicon was enriched with new terms. MPQA lexicon (Wilson et al., 2005) was also translated from English to Arabic by utilizing the Bing mechanical translation device, combing all lexica and creating an opinion-based classification approach. Once again, the data is not publicly available, and the total number of terms is unknown.

Arabic WordNet (AWN) by Black et al. (2006) and English SentiWordNet (ESWN) by Baccianella et al. (2010) and Farra et al. (2010) were linked by Alhazmi et al. (2013) for utilizing a synset (synonym set) offset approach to data. This was limited in its coverage (approximately 10K lemmas) and no definition was forwarded in terms of adopting the lexicon in more practical applications, bearing in mind the complexity of Arabic morphology. Again, this was not publicly available; hence, no evaluation of sentiment is possible. The availability of ESWN and its success, however, was exploited by Badaro et al. (2014) by developing a lemma-based sentiment based on Arabic lexicon: ArSenL. This was derived by fusing two sub-lexicons. Initially, AWN 2.0 was mapped with ESWN 3.0 via sense map files and utilizing EnglishWordNet, as used by Alhazmi et al. (2013), though one important deviation was to adapt the AWN lemma format to LDC. This approach is vital to allow for integration with various other NLP applications. Secondly, ArSenL-Eng mapped ESWN 3.0 with the Standard Arabic Morphological Analyser (SAMA) (Maamouri et al., 2010) through matching the gloss terms in SAMA with synset terms in ESWN. ArSenL, is available to access with 29K lemmas and their linked POS tags, ESWN sentiment values and EWN synset ID. This approach has certainly improved accuracy and sentiment classification. A recent further extension of ArSenL is ArSEL (Badaro et al., 2018); Arabic Sentiment and Emotion Lexicon with an extra eight emotion values compared to the ArSenL lemmas. Emotion values are derived from a WordNet lexicon application EmoWordNet (Badaro et al., 2018).

A semi-automatic approach for developing sentiment lexicon was devised by Abdulla et al. (2014). 300 words from English to Arabic were manually extracted from SentiStrength and synonym tables were adopted for expansion. Google Translate was utilized as the automatic factor for translation purposes. An annotated corpus was also investigated, allowing sentiment to identify positive and negative words using a term-frequency weight approach. Abdul-Mageed and Diab's (2014) manual sentiment lexicon (SIFAAT) was developed automatically via machine translation and formulaic statistical analysis based on common information to form SANA. This involved 224,564 separate entries covering Egyptian, Levantine dialects and Modern Standard Arabic (MSA). Duplication is evident since a number of sources were used in creating SANA in gloss matching, including Arabic SAMA (Maamouri et al., 2014), THARWA (Diab et al., 2014), English ESWN (Baccianella et al., 2010) and Affect Control Theory (Heise, 2007). SANA automatic influences include Maktoob (Twitter database), Yahoo and a comments database from YouTube (Abdul-Mageed et al., 2011). SANA, predictably, is not publicly available and was not classified as a sentiment task.

Lexicon syntactic rules were adopted by ElSahar and El Beltagy (2014) to automatically derive subjective Arabic phrases. DA rules were applied, concentrating on the Egyptian Cairene dialect. An initial seed set was manually translated into Arabic and a stream of patterns was defined to illustrate subjectivity. Upon deleting slang, Pointwise Mutual Information (PMI) determined the nature of the phrase with annotated tweets. A total of 7.5M cleared tweets produced 633 expressions with an impressive 89% accuracy. Further research by ElSahar and El Beltagy (2015) adopted supervised learning to create sentiment lexicon identifying positive and negative values. A corpus, annotated with 35K sentences, was used to extract bio grams and unigrams, which were SVM sentiment classified. Higher positive extracts were cited as positive sentiment and lower when negative. This is publicly accessible and considers both MSA and dialectal Arabic.

Sentiment lexicons were focused on DA by Al-Twairesh et al. (2016) to enhance the accuracy of opinion-mining approaches in data gained from Twitter. Approximately 2.2 million tweets involving specified positive and negative seed vocabulary were sourced. Coupled with emoticons, the seed words aided annotation in semantic evaluation. Using this dataset, the authors established duel sentiment lexicons: AraSenti-Trans and AraSenti-PMI. Regarding the former, tweets were fed through MADAMIRA (Pasha et al., 2014) (Arabic lemmas) and matched the given English gloss with existent sentiment lexicons Liu Lexicon (Hu and Liu, 2004) and MPQA (Wilson et al., 2005) to label sentiment Arabic sentiment in accordance with established matching rules. 132K Arabic terms were used with 60K being positive and 72K negative. On the other hand, Ara-Senti PMI exploited computing PMI to trace words that appeared 5 times or more in both the positive and negative tweets, ultimately generating a sentiment score. 57K positive terms were traced; 37K were negative. Both lexicons were sentiment evaluated and classified using differing Twitter datasets AraSenti-Tweet (Al-Twairesh et al.,2017), ASTD (Nabil et al., 2015) and RR (Refaee and Rieser, 2014). Average scores were 88.92%, 59.8% and 63.6% respectively. Saudi dialects were targeted by Assiri et al. (2018) through a manual annotation of terms and comments in Saudi social media before merging the results with ArSenL. This updated version removed all diacritics and punctuation from the lemmas. Sadly, it is not publicly available. The resource pool for study is scarce due to the limited access of sentiment models to the public. The author of this work acknowledges this challenging obstacle; however, existing accessible models will be explored to the full.

3.5. Literature Review of Lexicon-based Approaches

Exemplifying the application of a lexicon-based approach to sentiment analysis, Badaro et al. (2014) proposed the ArSenL lexicon, which is similar to the well-known English SentiWordNet2; it is publicly available as a web-based graphical user interface. The ArSenL lexicon uses two lexicons: Arabic WordNet (AWN) and Standard Arabic Morphological Analyser (SAMA) mapped with English SWN. Another Arabic sentiment lexicon (ASL) was used in Mahyoub et al. (2014). In this case, the lexicon was built using a small seed list of positive and negative words and a semi-supervised method was used to assign the polarity scores to 2000 words (600 negative, 800 positive and 600 neutral words). Importantly, these lexicons were built based on the general purpose Arabic WordNet; therefore, in the work by Mahyoub et al. (2014), it is difficult to prove that all the sentiment words that might appear in the reviews were included in the lexicon.

In another study, Abdul-Mageed and Diab (2014) constructed their Arabic sentiment lexicon, SANA, by combining many pre-existing lexicons that contain different approaches, such as automatic machine translation from an existing English lexicon, extensive manual correction and statistical formulation. The SANA included 224,564 entries for MSA and two dialects, Egyptian and Levantine. However, these entries were not distinct, as they contained many duplicates, which limited their resource usability. Similarly, Al-Ayyoub et al. (2015) proposed an unsupervised technique for sentiment analysis of Arabic tweets. The approach to this study was streamlined, as the tweets were first collected and pre-processed, following which a a sentiment lexicon with sentiment scores between 0 and 100 was developed and divided into three sentiment groups. A scalar instrument was used to subdivide these groups, assigning scores from 60 to 100 for positive sentiments, scores between 40 and 60 for neutral sentiments and scores less than 40 for negative sentiments. In this way, the sentiment score of an input sentence was computed by combining the three corresponding scores (Al-Ayyoub et al., 2015). Although the overall accuracy of this technique was reported to be 86.89%, it was limited in its ability to handle different Arabic dialects.

Continuing to extend this lexicon-based approach, Abdulla et al. (2014) developed a four-step solution designed to address dialectical variations in the text. The first step consisted of selecting 300 seed words from the SentiStrength website. Then, in the second step, the synonyms of these selected words were added to the lexicon. In the third step, a term frequency weighting technique was applied to the lexicon in order to identify whether there were still missing words after going through the first two steps. The fourth step added words from different Arabic dialects to the lexicon. Then, based on this lexicon as well as the simple lexicon-based method, the sentiment analysis was performed by computing the polarity of the text without taking into account negation and intensification. The reported results of their study had an accuracy of 70.05% using different lexicon scalability phases, but failed to adequately consider dialectical variations. Extending such research, Al-Twairesh, et al. (2017) presented three different methods for Arabic sentiment analysis; one of them was an enhanced version of a simple lexicon-based method that is capable of handling contextual polarity, such as negation and intensification. By adding these extra features, the authors achieved a better performance (91.75%) than the one obtained in Abdulla et al. (2014), where such contextual polarity was not taken into account.

Although dialectical research is limited in this field of study, a sentiment analysis lexicon approach exploring dialectic Arabic text (mainly Algerian) was proposed by Mataoui et al. (2016). This specific dialect involves a high degree of code switching, particularly between French and Arabic. This example portrays the extreme challenges faced by researchers in this particular field and anticipates new techniques for tackling this problem. Three sentiment lexicons were developed manually. The first lexicon is based on an extant Egyptian dialectical sentiment lexicon; the authors retained only the terms that were commonly used within the Algerian dialect. The second is a list of negative vocabulary frequently present in Algerian dialect, and the final lexicon is a list of intensifiers found within the dialect. The authors then tested differing configurations of the model. The first configuration was satisfactory at the phrase level, presenting similarities of comments that illustrated existent label phrases. The second configuration required word-level analysis subsequent to the application of developed parsers within the Algerian dialect. This was done in order to carry out normalisation and avoid word removal and tokenisation. The next stage was to process the tokens through language detection and a stemming module. This clarifies the root language of the tokens. Stemming was utilised for Arabic tokens, whereas tokens for alternative languages were initially translated into Arabic before being stemmed. The next stage was to match stems with developed sentiment lexicons, which allowed for computation of a text semantic orientation calculation. Mataoui et al. (2016) then amassed their findings and manually annotated the polarity of 7,698 Facebook comments, covering a wide range of topics including Algerian dialect and MSA. When the two configurations were combined, a score of 79/13% was achieved.

AL-Twairesh, et al. (2017) presented three different methods for Arabic sentiment analysis; one of them was an enhanced version of a simple lexical-based method capable of handling contextual polarity, such as negation and intensification. By adding these extra features, the authors achieved 91.75% accuracy. Al-Moslmi et al. (2018) found that the lexicon-based model is normally used in case the data is unlabeled. As for sentiment lexica, they are used to mark the data and to estimate the polarity. Using sentiment lexicon, the sentiment of a text (a review) can be measured using phrases and words in the lexicon. Aloqaily et al. (2020) developed lexicon-based sentiment analysis for Arabic tweets datasets concerning the Syrian civil war and crises. Arabic Tweets, expressed as bag-of-words (BOW), are classified as positive and negative by looking up the mentioned sentiments in an Arabic sentiment lexicon. The registered classification accuracy was 68% and the paper does not report on the analysis of other factors impacting the SA performance such as intensification and negation; dialectical Arabic was not considered by the authors.

3.6. Literature Review of Machine Learning Sentiment Analysis

This section reviews works published in the literature that address the use of machine learning for Arabic sentiment analysis. The SAMAR system for subjective sentiment analysis was initially presented by Abdul-Mageed et al. (2012). The authors used diverse sets of data written in different Arabic dialects and MSA; they considered all social media types in Arab countries, including Wikipedia talk pages, tweets, chats and newswire domain forums. Their proposed approach considered several problem domains, including political, economic, sports and entertainment news. Unfortunately, the SAMAR system performed inefficiently for datasets composed of tweets, producing a positive sentiment F-score equal to 49.41% for the Twitter dataset.

Azmi and Alzanin (2014) assessed a four-level polarity via the mining of remarks from local online newspapers in Saudi Arabia. The set of about 815 comments in Arabic was subdivided into 620 training set comments and 195 testing set comments, resulting in an accuracy of 85%. Also, Itani et al. (2017) researched the applications used for the processing of natural language, such as machine translation, categorization of text and the analysis of sentiment, for which the verification of accuracy and quality needs an annotated corpus. A corpus is basically a group of texts with labels, which are description tags and POS (part of speech) tags from various sources. The corpus of Itani et al. was comprised of 1000 posts gathered from a Facebook page called 'The Voice' and 1000 posts gathered from the 'Al Arabiya' Facebook news page. In this approach by Itani et al., they used Facebook to create the corpus in order to deal with dialectal Arabic. A corpus is also used to predict movie sales; also, it is used in publications to show polarity in sentiment analysis (negative, neutral and positive). POS taggers, tokenisers, vocalisers, and stemmers were used in the processing of natural language to construct the corpus. Manual tagging, IAA (Inter Annotator Agreement) and classifiers like decision trees (DT), support vector machines (SVM), naïve Bayes (NB) and k-nearest neighbours (KNN) were used for the content polarity categorisation. However, Azmi and Alzanin (2014) and Itani et al. (2017) did not handle the negation terms properly in their study, which has a significant effect on sentiment polarity. Also, irrelevant comments were not filtered in the pre-processing stage.

Nabil et al. (2015) forwarded a 4-way sentiment analysis classification that places texts in four distinct categories: objective, subjective negative, subjective positive and subjective mixed. Their dataset was made up of 10,006 Arabic Tweets, annotated manually through the use of the Amazon Mechanical Turk (AMT) service. They took advantage of a number of machine learning algorithms (MBN, BNB SVM, KNN and stochastic gradient descent) on both the balanced and unbalanced datasets. They found that adopting n-grams as unique features in a multi-way classification approach failed to provide encouraging results. They did not apply pre-processing to the set tweets. Al-Obaidi and Samawi (2016) developed a sentiment analysis classification model for dialectical Arabic, Saudi and Jordanian dialects. They enhanced the pre-processing adoption of the system by presenting a bespoke stop-words list for both dialects with a light stemmer specifically forwarded for each. They explored differing classification approaches and strategies along with Bag-of-Words (BOW) and n-gram features. It was found that Maximum Entropy performed best with trigrams. Conclusive findings in these studies indicate negative consensus regarding the ideal experimental approach (length of n-gram or representation) and ultimately that findings are corpus dependent. This is an anticipated result, since the simplicity of these features does not reflect upon the complexity of the exercise.

Alomari et al. (2017) examined a collection of Jordanian tweets and split them into negative and positive ones. The tweets totalled to 1800 written in the Jordanian dialect. A comparison of NB and SVM classifiers was made using two pre-processing strategies and features to analyse Arabic topics on social media written in MSA or in the Jordan dialect using the supervised machine learning sentiment method. Many bigrams, trigrams, n-grams and unigrams were used by Alomari et al. with different weighing methods (term frequency–inverse document frequency (TF-IDF), term frequency (TF)); alternative stemming methods were also applied: light stemmer, no stemmer and stemmer. Through using a SVM with bigrams as well all a TF-IDF using the stemmer, Alomari et al. received the best performance scenario, which gave an 88.72% resolution score and an 88.27% F-score in comparison with the study that used a NB classifier. Al-Rubaiee et al. (2016) studied the application and the layout of Arabic text categorization regarding the thoughts of students at King Abdul-Aziz University. Their approach consisted of five basic steps: data collection, data filtering, data pre-processing, classification and, finally, evaluation. They then prepared the dataset that was collected using a Twitter API; the data comprised of 2000 tweets. By using the RapidMiner program, the light stem, stop word removal and tokenization methods for Arabic NLP were applied. Also, they used the NB and SVM methods for supervised machine learning. However, one of the main challenges in Alomari et al. (2017) and Al-Rubaiee et al. (2016) was the size of the datasets, which is important for training the machine learning algorithms. Thus, the approaches proposed in their studies may improve with larger datasets, also, they experiment few classifiers.

Al-Horaibi and Khan (2016) aimed to improve how emotion is measured in Arabic tweets. This approach for analysing emotion in Arabic tweets consisted of three levels: data collection, data pre-processing, classification and evaluation. The tweets were collected using the Twitter API stream. Moreover, Tweepy library was used in their Python script. While running their scripts, they collected a total of 14,984 tweets. Arabic stop word removal and tokenisation were done using Python language, and a 162-word roster was generated. Then, the light stemmer and the Information Science Research Institute (ISRI) stemmer were applied to get the root of each token within a tweet. Finally, DT and NB classifiers through the Natural Language Toolkit (NLTK) tool were applied. An average accuracy score of 45.60% and a F-measure score of 31.54% were achieved when they used the English SentiWordNet (SWN) classifier in the experiment. However, their results due to the fact that they used English NLP tools on Arabic dialect tweets, which have different characteristics than English. Sghaier et al. (2016) proposed a multi-algorithm based supervised approach for performing Arabic Sentiment Analysis. In particular, KNN, SVM and NB were used in combination with a bag of words to classify the data collected from e-commerce websites. The reported accuracy for SVM and NB were 93.9% and 93.87%, respectively. However, in this study, the corpus contained only 250 documents, which is quite small. Moreover, they did not consider negation in their study.

Baly et al. (2017) explored the complexity of opinion mining Arabic using Twitter, in regard to increased noise through tweets and the multiplicity of dialectical Arabic. The authors carried out a study to assess differing linguistic use in a variety of Arabic areas. They also created a typology of tweets to establish extended comprehension and fuel further investigation. The authors used machine learning on Arabic Twitters with the use of feature engineering and the deep learning approach. They collated datasets via Egyptian tweets (10,006 used by Twitter API, divided into categories: positive (799), negative (1,684), objective (6,691), neutral (832)). They applied both POS tagging and lemmatization utilising MADAMIRA v2.1 to extract features for the SVM classifier. The optimum result achieved with SVM was 55.70%. Baly et al. (2017) forwarded a further study; a comparative evaluation of sentiment analysis methods within Arabic dialect. They presented their analysis of sentiment analysis through a study of different dialects. Overall, they established the Multi-

Dialect Arabic Sentiment Twitter Dataset (MD-ArSenTD), covering 12 Arabic countries. The annotations not only included the topic, but also the overall SA, the target of the sentiment and the means of expression. The authors experimented with two classifiers, LSTM and SVM. It was found that the accuracy of Egyptian tweets was 60.6% in comparison to UAE tweets (51.1%). The variance in the nature of each dialect resulted in a challenging task, hence carried out separately. The author indicated that the SVM classifier performed significantly better upon analysis of the Egyptian tweets. This fact can be justified due to the difficulty in predicting sentiment analysis regarding religious entries, which constitute the majority of UAE social media communication.

Rahab et al. (2017) enhanced sentiment analysis throughout newspaper comments in Algeria, annotating comments made by two Algerian Arabic-speaking native contributors. They held many experiments, to illustrate the impact of a word-weighting strategy, the classification method and light stemming. They also used word-weighting algorithms (Binary Term Occurrence (BTO), Term Frequency, Occurrence and Inverse Document Frequency (TF–IDF). The authors also exploited a significant number of classifiers such as K-Nearest Neighbours (KNN), SVM and NB for which they chose tenfold cross-validation. The optimum results were obtained using TF with an NB classifier, relying on light stemming with an accuracy of 75%. The dataset was constituted of 92 comments, 60 derived from the negative and 32 from the positive category. However, though this research was successful considering the differing weighting scheme, it explored comments differing from tweets, hence the dataset is limited. At the SemEval International Workshop help in 2017, Mulki et al. (2017) provided their valuable contribution. Task 4 entitled "Twitter Sentiment Analysis" was addressed explicitly as a 4A-Arabic subtask. It suggested the use of monitored and un-monitored learning methods to incorporate the two Arabic classifying patterns. Arabic tweets were pre-processed for both models until various bag-of-N-grams schemes were obtained as functions. They studied the classification of sentiments in the Arabic language through two separate learning strategies and classification models. The supervised and lexicon-based models obtained satisfactory results. The supervised learning model has been chosen for the final submission because the highest average recall value and F-score values have been achieved.

However, without any support from stemming, the lexicon model has also produced strong results. It is worth noting that the combined lexica have successfully controlled MSA/multi-dialectal content.

Maghfour et al. (2018) have recently analysed Facebook Arabic comments that were made in Moroccan and MSA. They tried to investigate, before classification, the benefit of classifying the Arabian corpus in its forms (dialect or MSA). Their key concept was to adjust pre-processing text according to each type of language. For example, for dialectical texts the writers used light stemming. For both NB and SVM classificators, they applied their method. It improved its efficiency by taking into account the heterogeneity of MSA and the dialect studied. This two-stage classification was found to be minimising word stemming errors. But in broader multi-dialectical datasets, use of this approach can be difficult. A recent study done by Sayed et al. (2020), they suggested an Arabic-language sentiment research Multidimensional System. They began creating a dataset of 6318 reviews obtained from the hotels reservation website (Booking.com). Both regular and dialect types of the Arabic language were reviewed. They implemented the use of nine machine learning classifiers, which included K-Nearest Neighbour (KNN), Logistic Regression (LR), Multilayer Perceptron (MLP), Ridge Classifier (RC), Decision Tree (DT), Support Vector Machine (SVM), Gradient Boosting (GB), Random Forest (RF), and Naive Bayes (NB) classifiers. Several factors have been explored in the data representations. Moreover, detailed scenarios have been examined for the feature vectors. The designed dataset was used to evaluate the performance of the nine classifiers in different situations. The results of the experiment demonstrated the highest overall performance in recalling, precision, precision, F1 scoring, and training times for Ridge Classifier (RC). In addition, it enhances classification efficiency by applying pre-processing techniques such as stemming and stop-word removal.

3.7. Literature Review of Hybrid Sentiment Analysis

Exemplifying this approach to sentiment analysis, El-Halees (2001) presented a hybrid strategy for sentiment analysis classification of Arabic. The approach commenced with the utilisation of lexicon-based technique to pinpoint either a positive or negative classification to the unannotated work. The resulting data was then fed into the Maximum Entropy (ME) classifier in the format of a training set, to segregate remaining documents, unclassified at the lexicon-based stage. The classified documents, utilising both lexicon-based and the ME approach, were fed into a KNN classifier and treated as a training set in order to classify the finalised set of texts left un-classified. El-Halees (2001) reported an accuracy of 80.29%. However, this study did not consider negation words. More recently, Khalifa and Omar (2014) adapted a hybrid approach for analysing Arabic opinions, inviting answers through the addition of lexicon features to three classifiers: NB, SVM and KNN. Once the data was prepared and differing pre-processing techniques were applied, the authors tested the three classifiers, ignoring lexicon features and subsequently testing the same classifiers, but this time with lexicon features included. In all experiments, they witnessed an enhancement in results through the addition of lexicon features. The experimental results have shown that NB provides the best result, which is better than SVM and KNN.

Focusing more specifically upon social media, Khasawneh et al. (2015) introduced a hybrid strategy to classify Arabic tweets as positive, negative or neutral. The authors composed their unique dataset, consisting of Arabic tweets and recordings from Twitter. Arabic tweets were collected from three specific domains, news, economics and sport, totalling in 1500 comments and reviews. The Arabic audio was composed of a mere 15 files. From the textual dataset, only 13 lexicons were created manually. The proposed system initially requested the user to choose their review domain; the system would convert the set audio into text. Subsequently, utilising the lexicon-based approach, the text was labelled. The authors adopted duel machine learning classifiers; boosting and bagging. Experimental results illustrated that predictive textual data is an enhanced means of analysis in comparison to predicting audio data, showing an accuracy of 85.95% and 82.95%, respectively whilst adopting the bagging technique, whereas accuracy was 69.25% and 64.52%, respectively using the boosting technique. However, input in this research is audio converted to text, which differs from social media content and input. There were also vital limitations, such as not taking into account sentiments such as laughing or yelling with the audio content.

A further innovative study was done by Alhumoud et al. (2015) which provided a hybrid sentiment analysis approach through the removal of unsentimental words from the training dataset, resulting in the training dataset solely containing instances, sentimental words and labels. The authors utilised two classifiers, KNN and SVM, and compared the assessed approach with machine learning via identical classifiers. The results confirmed that the approach was more accurate when compared with explicit machine learning. However, due to the relatively small dataset and the lack of negation, the results failed to adequately represent the broader and complex domain of dialectical Arabic.

In another study, Abuelenin et al. (2017) utilised the cosine similarity algorithm and the Information Science Research Institute Arabic stemmer (ISRI) to ascertain the most similar word within the lexicon and forge a hybrid approach to enhance the accuracy of Egyptian Arabic. Their proposed hybrid framework increased performance by integrating machine learning techniques with semantic orientation. A further study was presented by Alkubaisi et al. (2018) and focused on the analysis of a Twitter-based dataset. The research proposed a Hybrid Naïve Bayes Classifiers (HNBCs) as the optimum machine learning method for stock market classification. The findings have direct applications for companies, investors and researchers, enabling them to formulate strategies in accordance with underlying public sentiments. Al-Twairesh et al. (2018) suggested the use of a hybrid approach for evaluating sentiment of a particular Saudi dialect in the Arabic language. They used a collection of features designed to be dialect-independent and tested by a functionally reverse-selection process. Then three Saudi Twitter classification models were constructed and contrasted as follows: two approaches (negative and positive), three methods (neutral, negative, and positive) and four methods (neutral) (mixed, negative, positive, and neutral). The authors assessed the effects of all existing classification models of the suggested feature sets. They observed that the AraSenTi lexicon extracted features were there in all the best feature classes of the experimented classification methods.

HILATSA, a hybrid incremental learning approach for ASA was recently launched by Elshakankery et al. (2019). The key idea is to develop a method for sentiment analysis for tweets in Arabic, which can manage the rapid translation and use of terms. The authors have developed some critical lexicons, such as emoticon lexicon, lexicon, idioms lexicon, and special enhanced lexicon terms. Elshakankery et al. tested the usefulness of the Levenshtein Distance Algorithm in order to cope with various word types and misspelling. SVM and LR algorithms and one RNN were used in the experiments that demonstrated a positive outcome in a complex setting with high precision and reliability. Harrag et al. (2020) suggested the use of a hybrid approach that incorporates recommendation system and sentiment analysis to fix issues with data sparsity. This can be done using NLP and text mining models to know the product rating from user feedback. This analysis focuses on a range of Arabic reviews, in which the model is tested with the Arabic Opinion Corpus (OCA) dataset. Their system was effective and almost 85% accurate to predict the rating of reviews.

3.8. Chapter Summary

This chapter has provided a review of the extant literature regarding sentiment analysis and the growing spectrum of applications of these techniques to practical research and enterprise problems. From social media mining to customer feedback comparisons, the advantages of effective and reliable sentiment analysis are significant resources that can be used to predict changes, model opportunities, and improve performance. Whilst a variety of emergent sentiment mining techniques have been developed in recent academic history, the majority are based upon the English language and fail to address more complex or multi-dimensional problems beyond these limited results. Arabic Natural Language Processing was reviewed and the tools that been developed for Arabic language. Stemming is one of the main NLP tool, so, in this chapter literature review of Arabic stemming was presented. There are several sentiment analysis approaches have been reviewed such as Lexicon-Based approach and lexical resources for dialectical Arabic language processing, machine learning sentiment analysis approach, hybrid sentiment analysis.

Chapter 4

4 Developing the Resources and Tools for Lexicon-Based Sentiment Analysis of Dialectical Arabic

With the increasing amounts of dialectal Arabic written text on the web, the efforts of dialectical Arabic natural language processing (NLP), such as morphological analysis and disambiguation, have increased although they are still in the early stages compared to those for Modern Standard Arabic (MSA). According to Habash and Rambow (2006), Jarrar et al. (2014), and Khalifa et al. (2016), the available resources and tools developed for MSA gives limited performance when applied to dialectal Arabic. However, some dialectal Arabic, such as Egyptian Arabic, have received attention from researchers by developing a collection of resources that include annotated datasets and morphological analysers more than other dialectal Arabic, such as Gulf Arabic which is lagging behind. This chapter presents the resources and tools that were developed for the lexicon-based analysis of Saudi dialect. Specifically, a morphologically annotated corpus of the Saudi dialects with all the NLP tasks, such as normalisation has been presented, and novel light stemming methods are outlined. Then, a a novel lexicon construction is introduced that contains sentiment lexicon, negation, emoji, special phrases such as supplication, proverbs and interjection.

4.1. Toward Creating a Morphologically Annotated Corpus of the Saudi Dialects

Through social media, Arabic users communicate with each other and share opinions and ideas utilising unstructured Arabic slang (Abdulla et al., 2013).

Exploiting SA, it is possible to determine aspects of expression, through text polarity, in terms of positive or negative reactions. Despite enhanced interest in SA, limited academic studies have applied this concept to Saudi dialects. This constraint is largely due to limited publicly accessible annotated data (Abdul-Mageed and Diab, 2011). Hence, the current study has undertaken to contribute the first stage of a publicly accessible Saudi domain-specific annotated corpus. This solution could only be achieved by producing a set-procedure regarding manual corpus annotation drawn from a specific data series, in this case, unemployment in Saudi Arabia.

4.1.1 Overview of the Previous Corpus for the Arabic Language

In prior studies targeting an Arabic language corpus for NLP applications exemplifying the linguistic complexities of ANLP, COLABRA is one example of an Arabic corpus created for NLP resources that incorporates four Arabic dialects: Iraqi, Moroccan, Egyptian, and Levantine (introduced by Diab et al., 2010). MAGEAD (Habash and Rambow, 2006) was utilised by the authors, along with the Buckwalter Morphology Analyzer and Generator (BAMA) (Buckwalter, 2004). The Gumar corpus is a further extension of this approach developed by Khalifa et al. (2016) and draws upon a field of Gulf dialects to populate its database of more than 110 million words. The Gulf dialect labels were used to annotate the corpus at the document level; hence, no morphological annotation was evident. In a more recent application of this corpus, approximately 200,000 dialectal Arabic terms from Emirati dialect were selected, with the corpus then being manually annotated to reveal English glosses, lemmas, POS, and tokenization. During the manual annotation period, dialectal Arabic and conventions in spelling were included as factors to consider (Khalifa et al., 2018).

Another, dialect-specific corpus, Curras was developed to account for the Palestinian dialect (Jarrar et al., 2017). Within this targeted solution, 43,000 words were extrapolated from the Palestinian dialect via social networks. The MADAMIRA application was used for conducting the annotation of the corpus and an additional finding during the process was the establishment of a standard form to annotate Levantine dialect through orthographical means (Pasha et el., 2014). This approach is now used as an advancement of Conventional Orthography for Dialect Arabic (CODA), devised by Habash et al. (2012). Initially, the purpose of CODA was to create a unified framework to define dialectal Arabic by exploiting conventional orthography. CODA guidelines were described in great detail for the EGY dialect.

A recent attempt to extend the guidelines covering dialectal Arabic from 25 cities was forwarded by Habesh et al. (2018). The MADAR project is a further investigation of dialectal Arabic, explored by Bouamor et al. (2018). The ultimate aim was to create a single framework unifying annotated guidelines to be used with applications that include Machine Translation (MT) and Dialect Identification (DID). A parallel corpus was devised for the 25 cities by the authors. This was achieved by translating 2000 selected sentences extracted from the Basic Travelling Expression Corpus (BTEC) in MSA, English, and French (Takezawa et al. ,2007). Additionally, a lexicon was created involving 1,045 entries from the specified cities. Regarding the dedicated corpora serving NLP applications in view of the Saudi dialects, great strives have been made to create a corpora reflecting sentiment exploration from Twitter responses (Al-Twairesh et al., 2017; Assiri et al., 2016).

4.2. Collecting the Saudi Twitter Corpus for Sentiment Analysis

Due to the absence of a public Arabic dataset, many Arabic opinion-mining researchers have been obliged to independently collect datasets to advance their research. Emergent research has resulted in limited public datasets for the Arabic language language, such as OCA (The Opinion Corpus) for Arabic movie reviews (Cherif et al., 2015) and the Arabic Book Reviews (LABR) dataset (Aly and Atiya, 2013). Figure 4.1, presents the pipeline model of the methodology applied to the multiple stages of this collection process for a corpus for Arabic sentiment analysis, which includes five main phases (data collection, pre-processing, normalization, light stemming, and the annotation).

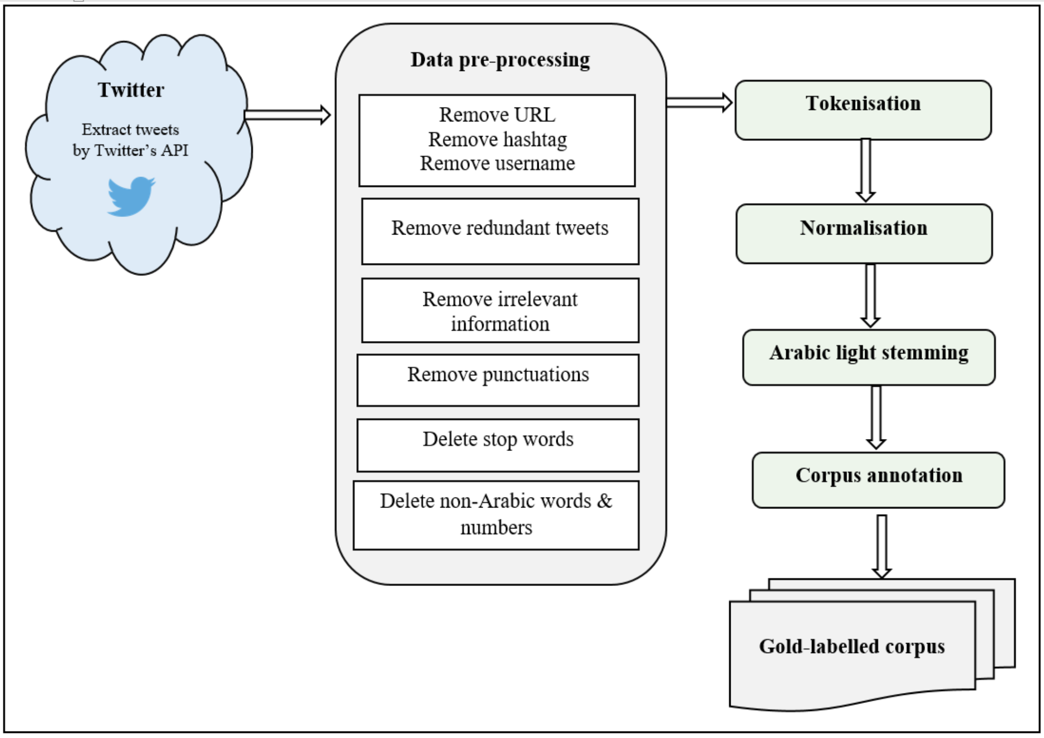


Figure 4.1: Pipeline of collecting the gold-labelled corpus

A Python script was developed to collect tweets in a Saudi dialect using Twitter's API using two methods: first, streaming tweets and, second, searching past tweets. In the first stage of this research, the tweets were collected according to two criteria: first, the geographical location (Saudi Arabia). An effort was made to collect equal tweets for each specific dialect, such as Hejazi and the Najdi using the user location, but some issues were observed, such as the location being locked by users and avoidance of dialectical Arabic in tweets by others. Then, the dataset was collected by selecting specific hashtags that were trending in Saudi Arabia from thousands of tweets. Hashtags are used to discuss social issues, such as السعوديه_للسعوديين (Saudi Arabia for Saudis) and توطين_قطاع_الإتصالات (localisation of the telecommunications sector). Around 23,500 tweets were collected, and, after the removal of redundancies, this number was reduced to around 10,000 tweets. Although comprehensive, this dataset was characterised by significant noise due to irrelevant tweets, such as advertising or tweets related to different topic as shown in Table 4.1. To eliminate this issue, a novel improvement was implemented by focusing on just one of the most well-represented accounts, @JoblessGrads9 (عاطلون بشهادات عليا). Most of the followers of this Twitter account are interested in the domestic unemployment issue, and as a result, the majority of the posts on the account were found to be topically related. From this dataset, approximately 5000 tweets were extracted, resulting in a baseline of just over 3,000 tweets once all redundancies had been removed.

Table 4.1: The number of tweets after removal of redundant tweets

	Hashtag	No. of collected tweets
1	السعوديه_للسعوديين *alsewdyh_lilsaeudiiyn* Saudi Arabia for Saudis	10,000
2	توطين_قطاع_الإتصالات *twtyn_qtae_ali'iitsalat* Localisation of the telecommunications sector	319
3	خريجات_جامعيات_قديمات_عاطلات *khryjat_jamieiat_qidimat_eatulat* Unemployed women graduate for a long time	2,788
4	سعوديات_يطالبون_بالتوظيف *sewdyat_ytalbwn_baltawzif* Saudi women demanding employment	10,000
5	سعوديات_يطالبون_بالتوظيف2 *sewdyat_ytalbwn_baltawzif 2* Saudi women demanding employment 2	358

4.3. NLP Methods for Corpus Pre-Processing:

Prior research and experimentation in this field has highlighted the challenges arising from sentiment analysis in dialectical Arabic. To overcome these challenges, this section presents a practical solution to NLP and pre-processing in order to improve the results of sentiment analysis.

4.3.1. Lexical Normalisation of Raw Tweets

In total, around 40,500 tweets were collected from several hashtags and accounts. The need to lexically normalise the tweets was an initial consideration when applying NLP tools. A lexical normalisation system contains two stages: (1) identification of non-standard words and (2) alternative word selection (Han et al., 2013). To achieve accurate outcomes from a lexical normalisation system, several functions need to be implemented as illustrated in the following pseudo-code model:

The pseudo-code of function: The NLP for Corpus pre-processing

1. **Input:** collected tweets
2. **Output:** the processed tweets
3. **Begin**
4. For each tweet in the dataset
5. Remove URL, remove hashtag, remove username
6. Removal of redundant tweets
7. Removal of irrelevant information
8. Remove punctuations
9. Delete the stop words
10. Delete the non-Arabic words and numbers
11. Apply normalisation
12. Apply Arabic light stemmer
13. Display the result
14. **End**

The first step involves removing redundant tweets (i.e., cleaning the collected tweets). A Python script was developed to remove the redundant tweets, the results of which are shown in Table 4.2.

Table 4.2: The number of tweets after removal of redundant tweets

	Hashtag	No. of collected tweets	After removal of redundant tweets
1	السعوديه_للسعوديين Saudi Arabia for Saudis	10,000	3,552
2	توطين_قطاع_الإتصالات Localisation of the telecommunications sector	319	111
3	خريجات_جامعيات_قديمات_عاطلات Unemployed women graduates for a long time	2,788	936
4	سعوديات_يطالبون_بالتوظيف Saudi women demanding employment	10,000	4,211
5	سعوديات_يطالبون_بالتوظيف2 Saudi women demanding employment 2	358	286

The collected tweets contain a lot of noise as shown in Figure 4.2. Every tweet was processed as follows: User information, URLs, and mentions were removed from the tweets because they included information unrelated to the case study topic. Then, diacritics were removed from words, followed by the removal of redundant letters often used to show emotion. It was important for the SA approach to consider some punctuation like question-marks and exclamation symbols. For example, “The Sauda program is good.”, and “The Sauda program is good !!!!”, the two sentences have the same words, but totally different sentiments. The only information which can help to determine the differences between these statements is the punctuation which shows the actual feeling. In the current dataset, around 3900 tweets (55%) contained question marks or exclamation symbols; therefore, identifying and addressing these punctuation marks was an important part of ensuring accuracy during this process.

Sentiments, which are made up of punctuation also play a role in sentiment analysis such as “(':, O_o, ._., :), :(, -_-, :D, xD” , when processed correctly. However, in this system this type of punctuation is not considered due to the user behaviours that were identified within the dataset. Around 42 tweets of the 7,000, expressed sentiment by using this type of punctuation. In addition, the users usually wrote the Quran quotes, providers, and supplications in brackets preceded by two vertical points which were carefully examined and not considered as indications of sentiment. For examples قال تعالى:(وَاللَّهُ يُحِبُّ الصَّابِرِينَ) , / Allah says:(And Allāh loves the steadfast).

	A	B	C	D	E
1	id	created_at	tweet	favorite_count	retweet_count
2	8.00717449557524E+017	Mon Nov 21 15:08:11 +0000 2016	# https://t.co/RVNPoDnaV0	0	0
3	8.00717456167866E+017	Mon Nov 21 15:08:12 +0000 2016	ق مع الـأمطار فـي منطقـة البـــسفور و التأشيـــرة و الضرائـب كاملـة - شـخص ٤٠٠٠ #	0	0
4	8.00717461221995E+017	Mon Nov 21 15:08:14 +0000 2016	وفي أشدّ أوقات ثباتك.. تهزمك ذكرى.	0	0
5	8.00717463935709E+017	Mon Nov 21 15:08:14 +0000 2016	#بـوب_كورن خخـآص يلا	0	0
6	8.00717466208866E+017	Mon Nov 21 15:08:15 +0000 2016	@t_trok99 ما تـــعطيني شـي منـها ؟	0	0
7	8.00717473376911E+017	Mon Nov 21 15:08:16 +0000 2016	الله اكبر العصر فزت في اسواق المزرعه بهديه والحين ايفون حمدلله https://t.co/l5QUqxKMdL	0	0
8	8.00717481056707E+017	Mon Nov 21 15:08:18 +0000 2016	@albasam20 https://t.co/oF2IzcTrqA	0	0
9	8.00717485347603E+017	Mon Nov 21 15:08:19 +0000 2016	رسميًا: سيرجيو أغويرو و كيفين دي بروين مرشحّين لتشكيلة السنه من دوري ابطال اوروبا. https://t.co/efALJfGRIx	0	0
10	8.00717487037813E+017	Mon Nov 21 15:08:20 +0000 2016	من صـوت الأذان إلا شيـطان فـفي الحديـث اذا نـودي للصـلاة أدبـر الشيـطان ولـه ضـراط#	0	0
11	8.00717495099195E+017	Mon Nov 21 15:08:22 +0000 2016	الاجابـة عليـه @Dahia_Aljazeera @BudaiwiM @HoldenLuminous @alkhobarsea @iraid_0	0	0

Figure 4.2: Screenshot of he collected tweets before pre-processing

Finally, the tweet was tokenised. An example for the cleaning and processing steps of a tweet is shown in Table 4.3.

Table 4.3: Example of pre-processing a tweet

Original tweet	#السعوده مستتتتتتحيل أنا أرفض السعوده الوهمية تماما .. كذا تظلموووووونا !!هذا الشي مرفوض *# alsueuduh musttttttthyl 'ana 'arfid alsueudah alwahmiat tamamaan .. kadha tzlmwwwwwwna !! hdha alshy marfud*
Cleaned tweet	مستحيل انا ارفض السعوده الوهميه تماما . كذا تظلمونا !هذا الشي مرفوض *mustahil 'iinaa arfd alsueudih alwahimiuh tamama. kadha tuzlamuna! hdha alshy marfud*
Tokens	['مستحيل', 'انا', 'ارفض', 'السعوده', 'الوهميه', 'تماما', '.', 'كذا', 'تظلمونا', '!', 'هذا', 'الشي', 'مرفوض'] *'] msthyl' , 'ana' , 'arfd' , 'alsueuduh' , 'alwahmayih' , 'tmama' , '.' , 'kdha' , 'itzalamuna' , '!' , 'hdha' , 'alshy' , ' marfud ['*

As seen in the normalisation example presented in Table 5.3, the process of normalising the word 'تظلموووووونا' to 'تظلمونا' can result in the undesired loss of the intense feeling expressed by the blogger, which is significant to capture the sentiment. One of the most noticeable behaviours of Arab Twitter users is the repetition of one letter to express a strong feeling, whether negative or positive. Nevertheless, this intensity can be preserved if, instead of removing all the repeated letters, two letters are kept. In this case, the words 'تظلموووووونا' and تظلمووووووووووونا will be normalised into 'تظلمووونا'. Although prior investigations have adopted more than one form to justify Arabic normalisation, it is better to just consider some of Arabic vowel letters such as the (ه and ا) to be normalised because of the multiple shapes. Further reconciling these conflicts, the character (ى - alef maqsura) should not be normalised because it is the only vowel changes the semantic meaning of the word, as shown in Table 4.4.

Table 4.4: Examples of Arabic normalization

Letter	**After normalisation**	**Example**
ي *ya'a* ى *alef maqsura*	ي *ya'a*	علي Ali — male name على On top of
ة *ta'a marbuta* ه *ha'a*	ه ha'a	حلوة Beautiful حلوه Beautiful
إ *alef hamza'h* أ *alef hamza 'h* ا *alef wasel*	ا *alef wasel without hamza'h*	افضل Best أفضل Best

4.3.2. Towards an Improved Saudi Dialectal Arabic Stemming (SDS)

Stemming is an important text-processing step for numerous applications, including information extraction, sentiment analysis, and machine translation. The stemming process reduces the size of inflected words to their stem by removing the prefixes and suffixes, but the stem is not necessarily the root of the word (Jaafar et al., 2017). Arabic is a highly inflected language with a rich, complex morphology. Stemming is a critical natural language processing (NLP) task in which words are grouped based on their lexical semantic similarity (El-Beltagy et al., 2016; Albogamy and Ramsay, 2015). For example, the words 'يحب' (he loves), 'يحبون' (they love), 'سيحب' (he will love), and 'أأحببتم' (have you loved?) have similar semantics to 'أحب' (he loved). Hence, stemming allows text-processing applications to manage one target stem instead of five target words. Although most stemmer tools have been designed for Modern Standard Arabic (MSA), dialectal Arabic is more popular than MSA in the Arab world. As people have come to rely more on social networking services to express opinions and to consult others about issues that influence their daily lives, the analysis of social media output has become the subject of increasing interest (Alghamdi et al., 2008).

The stemming of Saudi Dialectal Arabic words has received limited prior attention because of the challenges in Arabic NLP due to the language's complex morphology, which is exacerbated by orthographic mistakes, spelling inconsistencies, the use of abbreviations and slang words, the tendency to repeat letters in writing to convey emotion, and the fact that most posts are written in non-standard dialectal Arabic (Kalwakid et al., 2017). The diversity of dialectal Arabic within Saudi Arabia, including Hejazi and Nejdi, is indicative of the rich variety of dialectal Arabic, even within the same country, which is one of the main challenges for the NLP of Saudi dialectal Arabic (Aldayel and Azmi, 2016). For example, the word 'window' in MSA is نافذه / *nafitha*, in the Hejazi dialect it is طاقه / *taqa*, and in the Nejdi dialect it is شباك / *shobak*. Most of the current stemming techniques focus on dealing with MSA texts; while this delivers a good performance, it falls short of dealing with dialectal Arabic. The current study addresses shortcoming by introducing a new stemming mechanism that can handle Saudi dialectal Arabic. The proposed novel stemming approach integrates the Information Science Research Institute (ISRI) stemmer and a rule-based stemmer purpose-built for this investigation.

4.3.3. Analysing the Effectiveness of Applying MSA Stemmers to Saudi Dialectal Arabic

The primary goal of Improved Saudi dialectal Arabic Stemming is to derive an efficient algorithm for extracting the stem of Saudi dialects words, which are collected from specific trending hashtags in Saudi Arabia. The proposed approach integrates two techniques to address the challenges of stemming Saudi dialects: the ISRI stemmer and a rule-based stemmer. First, attempts were made to retrieve the stem using the ISRI stemmer, as experimental analysis showed that this stemmer performed the best among MSA-based stemmers. Subsequently, the rule-based stemmer was applied to stem any words that ISRI failed to process. This stemming approach comprises three stages:

- 1st Stage: Tweets were collected from trending hashtags in Saudi Arabia. This corpus was pre-processed to achieve good results by considering specific objectives, such as the removal of redundant tweets (duplicates) from the dataset and by tokenisation. The test corpus consisted of 6,000 words extracted from Saudi dialects tweets.
- 2nd Stage: Six Saudi linguists performed manual stemming of the test corpus. Tables 4.5 and 4.6 show the characteristics of the 6,000-word test corpus.

Table 4.5: Characteristics of the test corpus

Word length	Word frequency	Percentage
2	74	1.23 %
3	2,313	38.55 %
4	1,608	31.80 %
5	1,275	21.25 %
6	321	5.35 %
7	409	1.81 %

Table 4.6: Sample characteristics of the test corpus

Word	Word length	Inflection forms	Manual stem
غل / *ghil* Malevolence	2	مغلول / غلهم	غل
حق / *haq* Rights		حقوق / حقوقنا	حق
كرف / *karaf* Boring	3	يكرفوني / مكروفين	كرف
بجح / *bijah* Unashamed		بجحات / بجاحه	بجح
وليف / *walif* Lover	4	وليفي / موالفه	ولف
سنعه / *saneah* Adroit		تسنعنا / سناعه	سنع
يخسون / *yakhsun* Contempt for	5	خسيت / تخسي	خسى
وناسه / *wanasah* Happiness		مونساتنا / ونسناكم	ونس
امينين / *amynin* Trustworthy	6	امانه / امين	امن
هياطهم / *hiatuhum* Haughty		مهايط / هياط	هيط
منفقرين / *munfqirin* Poor	7	فقرانين / فقيرات	فقر
غاثينكم / *ghathynkum* Nausea		غثه / غثيتوني	غثى
تناشبونا / *tanashibuna* Unwelcome person	8	ينشب / مناشبها	نشب

- 3rd Stage: The test-corpus words were stemmed using the MSA stemmers listed in Table 4.7, and then compared against the baseline of manual stems to determine the best-performing stemmer.

Table 4.7: Results of MSA stemmers applied to Saudi dialects words

Stemmer	Accuracy	Correct stem	Incorrect stem
MADAMIRA	25 %	1,500	4,500
Khoja	38 %	2,280	3,720
FARASA	22 %	1,320	4,680
ISRI	64 %	3,840	2,160

The evaluation results indicated that ISRI was the most accurate stemmer, see Table 4.7. Nevertheless, the ISRI stemmer failed to stem some of the words as indicated in Table 4.8, because it was primarily developed for MSA rather than dialectal Arabic. Lexical analysis was then performed on the incorrectly stemmed words to develop a rule-based light stemmer to complement the MSA stemmer.

Table 4.8: Examples of MSA stemming applied to Saudi dialectal Arabic words

Word / Manual stem	MADAMIRA	Khoja	FARASA	ISRI
كرف *karaf* / كرف Boring	رف ×	كرف √	رف ×	كرف √
سنعه *saneah* / سنع Adroit	وعى ×	نعي ×	نع ×	سنع √
يخسون *yakhsun* / خسى Contempt for	خس ×	خسأ √	يخسون ×	خسو ×
منفقرين *munfqirin* / فقر Poor	منفقرين ×	فقر √	منفقرين ×	فقر √
غاثينكم *ghathynkum* / غثى Nausea	غاثينكم ×	غوث ×	غاثينكم ×	غثى √
تناشبونا *tanashibuna* / نشب Unwelcome person	تناشبونا ×	تناشبونا ×	تناشبونا ×	شبو ×

4.3.3.1. New Algorithm for Saudi Dialectal Arabic Stemming

The overall performance of the he ISRI Arabic stemmer algorithm is superior to other MSA stemmers when applied to dialectal Arabic. Based on the lexical analysis of the words that the ISRI approach failed to correctly stem, it was important to subsequently devise a set of rules to extract the stem of the Saudi dialectal Arabic words. This light stemming approach initially processes the smallest Arabic stem (consisting of three letters), since 75% of MSA words have a three-letter root. The main steps of the proposed algorithm are as follows.

1. Check the size of the word (< = 3). This step is performed each time a letter is removed from the word.
2. If the word matches an MSA pattern, then the word is stemmed using the ISRI stemmer.
3. Remove the prefixes commonly used in Saudi dialectal Arabic (بال / هال / بهال / فال / ولل / لل / ال / وال / شهال / شال/ اش / وش).

4. Remove the suffixes commonly used in Saudi dialectal Arabic (ات /الكم /تو /هن / ها /ين /ت /الك /ونهم /هم /ته /ه /ينه /ونه /تني /تكم /تك /ني /تي /وكم /كم /ك /ون / /).
5. Then, following the removal of the prefix, further processing might be required for the first segment of the word, as follows:
   a) If the word starts with a letter or letters from the following set { م/ ن/ ي/ ا/ ت/ وا/ وي/ ون/ وم/ وت/ من}, then remove them.
   b) If the word starts with a letter from the set { ت/ م/ ي/ ن ا/}, and the third letter in the word is (ت), then remove both letters.
6. Check all the vowels:
   a) If one vowel is in the word, then remove it.
   b) If two vowels follow in sequence in the word, then remove one, in the following order: (ا) then (ي) and then (و).
   c) If three vowels follow in sequence in the word, then remove the first and third vowels.

This dialectal Arabic light stemming approach is encoded in the following pseudo-code that was developed explicitly for this study:

Input: MYWORDS a file of the word list for applying stemming on it, MSAP a file of MSA patterns words

Output: WSTEM a file of stemmed words list, FORISRI a file of words list belongs to MSA

Begin

1.While MYWORDS is Not Empty

2.MYW=Read each Word ()

3. While WLength > 3

4. IF MYW have Prefix (بال / هال / بهال / وش / اش / شال/ شهال / وال / ال / لل / ولل / فال) Then

5. Delete Prefix from MYW

6. End If

7. End While

8. If MYW have Suffix (هن / تو / الكم / ات / ون / ك / كم / وكم / تي / ني / تك / تكم / تني / ونه / ينه / ه / ته / هم / ونهم / الك / ت / ين / ها / /) Then

```
        Delete Suffix from MYW
      End If
   If MYW contains Prefix ( ا / ي / ن / م / ت / وا / وي / ون / وم / وت / من )  Then
    Delete Prefix from MYW
 End If
 If first Letter of MYW is ( ا / ي / ن / م / ت )  And third letter of MYW is ( ت ) Then
      Delete Both Letters from MYW
 End If
While not all letters scanned and WLength >3
 Set L to Letter of MYW
   If L is vowel and next letter of L not vowel Then
      Delete L
   End If
   If L is vowel And one next letter of L also a vowel Then
       Delete the vowel with less priority
   End If
   If L is vowel And tow letter after L also vowels Then
       Delete tow vowels with less priority
   End If
End While
End While
End
```

Table 4.9, shows examples from the application of the proposed algorithm to several Saudi words.

Table 4.9: Examples of applying algorithm to Saudi words

Saudi dialects	Manual stem	The rules of algorithm	Stemmer
منظلمات / *munazalamat* Injustice	ظلم	–Remove prefix (من) –Remove suffix (ات)	ظلم
مزيونه / *mazyunah* Beautiful	زين	–Remove prefix (م) –Remove suffix (نه) One vowel (و) → Remove	زين
المبتلش / *almubtalish* In trouble	بلش	–Remove prefix (ال) –1st letter is (م) and 3rd letter is (ت) → Remove both	بلش
مواجع / *mawajie*	وجع	–Remove prefix (م)	وجع

Pain		–1st vowel is (و) and 2nd vowel is (ا) → remove less priority (ا).	
اجاويد / *'ajawid* Generosity اقاويل / *aqawyl* Saying	جود قول	–Remove prefix (ا) –1st vowel is (ا), 2nd vowel is (و), and 3rd vowel is (ي) → remove less priority (ا) then (ي).	جود قول

4.3.3.2. Evaluation of the Saudi Dialectal Arabic Stemmer

The objective of this experimental study was to derive an efficient algorithm for extracting the stem of Saudi dialectal Arabic words by integrating two techniques to address the challenges of stemming: the ISRI stemmer and a rule-based stemmer. The test corpus contains 6,000 words; the correct stemming of these words was determined manually by six Saudi-language linguists. After applying the proposed algorithm to the Saudi dialectal Arabic test corpus, accuracy was found to be 79%. Since no stemmers currently exist for Saudi dialectal Arabic words, comparisons were drawn between the results of four existing stemmers and the results of the proposed stemmer to determine which was the most accurate. The evaluation results (Table 4.10) indicate that the FARASA stemmer, which was developed for MSA, provides the lowest accuracy with 22%, while the experimental stemmer developed for this study is the best for handling Saudi dialectal Arabic words. Figure 4.3. provides a visual representation of the accuracy variations reported across these five stemmer solutions, highlighting the superior accuracy of the experimental stemmer developed for this study.

Table 4.10: Results of experiment applying light stemming algorithm

Stemmer	Accuracy	Correct stem	Incorrect stem
MADAMIRA	25 %	1,500	4,500
Khoja	38 %	2,280	3,720
FARASA	22 %	1,320	4,680
ISRI	64 %	3,840	2,160
SDS Stemmer	79%	4,740	1,260

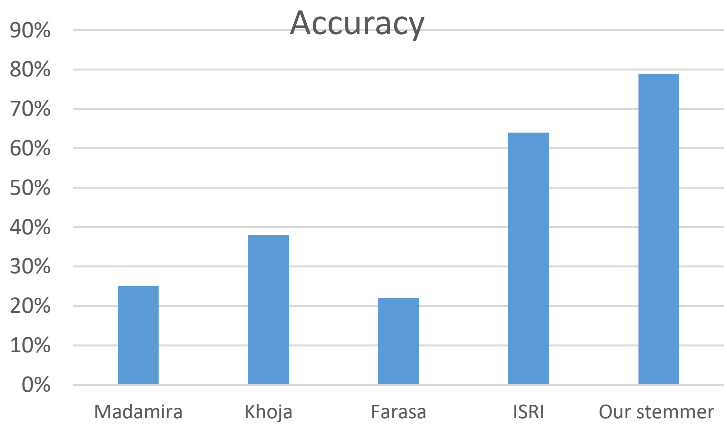


Figure 4.3: Results of experiment of applying light stemming algorithm

This study has confirmed the ISRI stemmer is the most accurate when applied to Saudi dialectal Arabic words. However, the evidence indicates that the ISRI stemmer fails to stem certain words because it was primarily developed for MSA rather than dialectal Arabic. The proposed novel stemming approach integrates the ISRI stemmer and a rule-based stemmer capable of addressing the challenges of Saudi dialectal Arabic stemming. The proposed rule-based algorithm comprises a set of pre-defined rules for extracting the stem of Saudi dialectal Arabic words. This approach can be used in applications where the Saudi dialect is prevalent, such as in social media networks and across varying applications including sentiment analysis, information-retrieval systems, and machine translation. Table 4.11 provides an example of the application of the algorithm to different Saudi dialects.

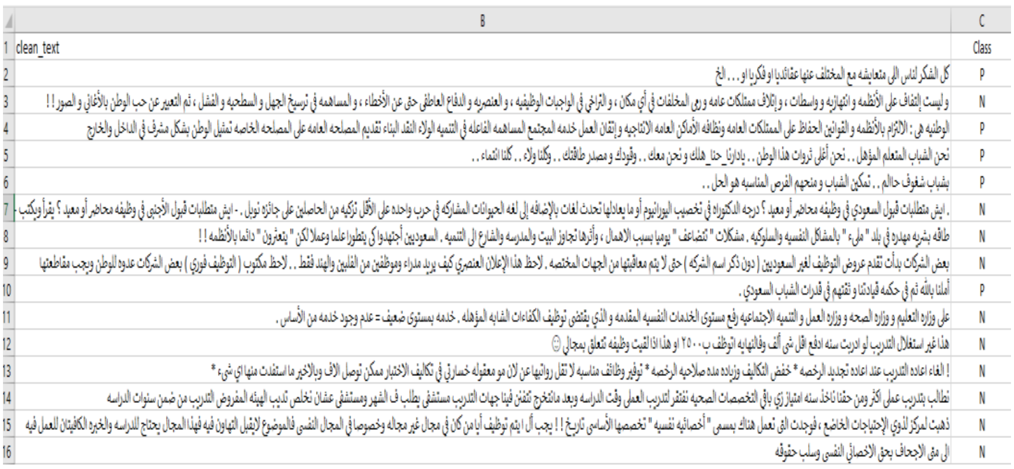
Table 4.11:Example of Algorithm Application to Different Saudi Dialects.

Word / Manual stem	In English	Saudi Dialects	Manual stemmer	SDS stemmer
استهبال *aistihbal*	Idiot	Najdi	هبل	هبال x
مغبون *maghbun*	Defeated		غبن	غبن √
غشيمه *ghshimah*	Stupid		غشيم	غشيمه x
عباطه *eibatah*	Naïve	Hejazi	عبط	عبط √
بكاشين *bakashin*	Liars		بكش	بكش √
يسيبوهم *yusibuhum*	Leave them		سيب	سيب √

قفطوهم *qaftuhum*	Caught them	Sharqawi	قفط	قفط √
تطرهم *tatiruhum*	Sharpened		طر	طر √
عفسونا *eafsuna*	Unorganized		عفس	عفس √
تقرطون *taqratun*	Throw	Shamali	قرط	قرط √
جاعص *jaeis*	Confident		جعص	جعص √
يتنوماسون *yatnawmasun*	Happy		نوماس	تنوماس x
ابثرونا *abthuruna*	Not welcomed	Janubi	بثر	بثر √
مطنوخ *matnukh*	Rich		طنخ	طنخ √
يزرفونا *yazrafuna*	Stole		زرف	زرف √

4.4. The Data Annotation Process

A gold-standard corpus was generated by the manual annotation of 7,000 tweets performed by seven human annotators who labelled the polarity of each tweet with its appropriate sentiment (positive or negative) (See Figure 4.4). Although consideration was given to other variations such as ambiguous or neutral tweets, the evidence suggested that most tweets are either positive or negative in both their construction and intention, increasing the reliability of the annotation exercise. Similar findings presented by Al-Ayyoub et al. (2018) have confirmed that most tweets reflect either a positive or negative sentiment, resulting in a formal procedure termed binary-sentiment analysis (BSA). Other researchers including El-Halees (2011), El-Beltagy et al. (2016), and Biltawi et al. (2017) have employed BSA for similar sentiment analysis purposes.



	B	C
1	clean_text	Class
2	كل الشكر لناس اللى متعايشه مع المختلف عنها عقائديا او فكريا او ... الخ	P
3	و ليست إتفاف على الأنظمه و انتهازيه و واسطات ، و إتلاف ممتلكات عامه ورمى المخلفات في أي مكان ، و التراخي في الواجبات الوظيفيه ، و العنصريه و الدفاع العاطفي حتى عن الأخطاء ، و المساهمه في ترسيخ الجهل و السطحيه و الفشل ، ثم التعبير عن حب الوطن بالأغاني و الصور ! !	N
4	الوطنيه هي : الالتزام بالأنظمه و القوانين الحفاظ على الممتلكات العامه ونظافه الأماكن العامه الانتاجيه وإتقان العمل خدمه المجتمع المساهمه الفاعله في التنميه الولاء النقد البناء تقديم المصلحه العامه على المصلحه الخاصه تمثيل الوطن بشكل مشرف في الداخل والخارج	P
5	نحن الشباب المتعلم المؤهل .. نحن أغلى ثروات هذا الوطن .. ياداربنا_حنا_هلك و نحن معك .. وقودك و مصدر طاقتك .. وكلنا ولاء .. كلنا انتماء ..	P
6	بشباب شغوف حالم .. تمكين الشباب و منحهم الفرص المناسبه هو الحل ..	P
7	. ايش متطلبات قبول السعودي في وظيفه محاضر أو معيد ؟ درجه الدكتوراه في تخصيب اليورانيوم أو ما يعادلها تحدث لغات بالإضافه إلى لغه الحيوانات المشاركه في حرب واحده على الأقل تزكيه من الحاصلين على جائزه نوبل . - ايش متطلبات قبول الأجنبي في وظيفه محاضر أو معيد ؟ يقرأ ويكتب	N
8	طاقه بشريه مهدره في بلد " ملىء " بالمشاكل النفسيه والسلوكيه . مشكلات " تتضاعف " يوميا بسبب الاهمال ، وأثرها تجاوز البيت والمدرسه والشارع الى التنميه . السعوديين أجتهدوا كي يتطوروا علما وعملا لكن " يتعثرون " دائما بالأنظمه ! !	N
9	بعض الشركات بدأت تقدم عروض التوظيف لغير السعوديين (دون ذكر اسم الشركه) حتى لا يتم معاقبتها من الجهات المختصه . لاحظ هذا الإعلان العنصري كيف يريد مدراء وموظفين من الفلبين والهند فقط .. لاحظ مكتوب (التوظيف فوري) بعض الشركات عدوه للوطن ويجب مقاطعتها	N
10	أملنا بالله ثم في حكمه قيادتنا و ثقتهم في قدرات الشباب السعودي .	P
11	على وزاره التعليم و وزاره الصحه و وزاره العمل و التنميه الاجتماعيه رفع مستوى الخدمات النفسيه المقدمه و الذي يقتضي توظيف الكفاءات الشابه المؤهله . خدمه بمستوى ضعيف = عدم وجود خدمه من الأساس .	N
12	هذا غير استغلال التدريب لو ادربت سنه ادفع اقل شي ألف وفالنهايه اتوظف ب٢٥٠٠ او هذا اذا لقيت وظيفه تتعلق بمجالي ☺	N
13	! الغاء اعاده التدريب عند اعاده تجديد الرخصه * خفض التكاليف وزياده مده صلاحيه الرخصه * توفير وظائف مناسبه لا تقل رواتبها عن لان مو معقوله خسارتي في تكاليف الاختبار ممكن توصل الاف وبالاخير ما استفدت منها اي شيء *	N
14	نطالب بتدريب عملي اكثر ومن حقنا ناخذ سنه امتياز زي باقي التخصصات الصحيه نفتقر لتدريب العملي وقت الدراسه وبعد مانتخرج تتفنن فينا جهات التدريب مستشفى يطلب ف الشهر ومستشفى عشان نخلص تديب الهيئه المفروض التدريب من ضمن سنوات الدراسه	N
15	ذهبت لمركز لذوي الإحتياجات الخاصه ، فوجدت التي تعمل هناك بمسمى " أخصائيه نفسيه " تخصصها الأساسي تاريخ ! ! يجب ألا يتم توظيف أيا من كان في مجال غير مجاله وخصوصا في المجال النفسي فالموضوع لايقبل التهاون فيه فهذا المجال يحتاج للدراسه والخبره الكافيتان للعمل فيه	N
16	الى متى الاجحاف بحق الاخصائي النفسي وسلب حقوقه	N

Figure 4.4: Screenshot of annotated data

Annotation represented a principle stage in generating a gold-standard corpus. To streamline this process, all annotators were required to adhere to the following instructions and guidelines:

- Where the chosen tweet is an advertisement, news feed, or fact, where no sentiment is present, labelling was omitted regarding positivity or negativity and replaced with (/), (See Table 4.12).

Table 4.12: Example of annotation news or facts

Tweet	Type	Label
وزاره الشؤون البلديه والقرويه تعلن عن توفر فرص وظيفيه (للجنسين) لحمله درجه البكالوريوس ، والماجستير ، والدكتوراه ، والتقديم عبر الموقع الالكتروني . . *wazarah alshuwuwn albaladayuh walqarawiuh tuelin ean tuafir furas wazifih (liljinsayn) lihamlih darajah albkalwryus , walmajsitir , waldukturah , waltaqdim eabr almawqie al'iiliktrunii..* The Ministry of Municipal and Rural Affairs announces the availability of career opportunities for both bachelor, master, doctoral and postgraduate degrees.	advertisement	/
الشورى يصوت لصالح استقطاب حمله الشهادات العليا للعمل في القطاع الخاص بما يضمن المزايا الوظيفيه التي تناسب مؤهلاتهم وتلزم جهه التوظيف بها *alshuwraa yusawit lisalih aistiqtab hamlih alshahadat aleulya lileamal fi alqitae alkhasi bma yadman almazaya alwazifiuh alty tanasab muahalatihim watulzim jahh altawzif biha* Shura Council votes in favor of attracting high degree certificates to work in the private sector in order to ensure the functional advantages that suit their qualifications and require the recruitment agency	News	/

- Since viewpoint dictates sentiment positivity and negativity, annotators needed to assign what was said by whom (See Table 4.13).

Table 4.13: Example of annotators description 1

Tweet	Point of view	Label
السعوديه تفرض ضريبه على عائلات المقيمين الأجانب العاملين في القطاع الخاص قد يكون ذلك لتقليل من عددهم *alsaeudiuh tafrid daribah ealaa eayilat almuqimin al'ajanib aleamilin fi alqitae alkhasi qad yakun dhlk litaqlil min eadadihim* Saudi Arabia imposes tax on families of foreign residents working in the private sector may be this happened to reduce them	Positive if the author is Saudi and negative if the author is foreigner	Two annotators label it as positive and one as negative

- Epistemic modality was also considered while assessing the Twitter extracts. This topic considers the judgement of knowledge by the contributor and whether they trust that a statement is true (Palmer, 2001). Root words as hedges could illustrate areas where there is a lack of confidence; for example, “somewhat”, “perhaps” or “maybe”. These are strengthened with such examples as “certainly” and “of course” (Polanyi and Zaenen, 2006). Additionally, epistemic modality can enhance the polarity and subjectivity of a clause within a sentence (See Table 4.14).

Table 4.14: Example of annotators description 2

Tweet	Special word / emoji	Label
القطاع الخاص يضع شروط كثيره للتوظيف *alqitae alkhasu yadae shurut kathirih liltawzif* The private sector asks for many conditions to employment	-	/
للأسف القطاع الخاص يضع شروط كثيره للتوظيف *llasf alqitae alkhasu yadaeu shurut kathirih liltawzif* Unfortunately, the private sector asks for many conditions to employment	للأسف Unfortunately	Negative
القطاع الخاص يضع شروط كثيره للتوظيف □ □ *alqitae alkhasu yadae shurut kathirih liltawzif* 💔 👎 The private sector asks for many conditions to employment 💔 □	□ / □	Negative

- Annotators were strongly instructed not to allow background knowledge or bias to influence their work. Such factors include religious, cultural or social issues. (See Table 4.15).

Table 4.15: Example of annotators description and their point of view

Tweet	Label
بدل الندره يجب أن يتوقف صرفه عن الموظفين *bdl alnadruh yjb 'an yatawaqaf sarfah ean almuazafin* The recompense for scarceness must be stopped given to the employees	Negative for employee and positive for unemployed

Following the annotation process, Cohen's (1960) Kappa, was utilised to examine how reliable the annotations were. This process involves deploying a statistical tool for measuring inter-rater agreement regarding qualitative terms. This tool is recognised as a robust indicator rather than a simplistic percentage calculation, since κ considers agreement by chance. The agreed level of agreement was deemed to be 91.74% and the weighted Kappa came out as κ = 0.816, indicating accurate annotations (Carletta, 1996). The collected tweets were cleaned and then annotated by human annotators. The annotators annotated each tweet with either a positive or negative designation, and the annotator also discarded irrelevant tweets (See Table 4.16).

Table 4.16: Statistics of the tweets in dataset

Dataset	Positive tweets	Negative tweets
Total tweets	2004	4996
Total number of words	16383	33945
Average number of words in each tweet (Tokens)	7.56	10.03
Average number of characters in each tweet	58.89	39.97

4.5. Lexicon Construction for the Saudi Dialect

This experiment has focused on collecting a lexicon for Saudi dialectical Arabic. For Saudi dialects, there are some proposed lexicons, such as the lexicon created by Aldayel and Azmi (2016). This lexicon contains only about 1500 terms. Al-Twairesh et al. (2016) developed AraSenTi-PMI and AraSenTi-Trans which are two large-scale Arabic sentiment lexicons. Another large lexicon contains 14,000 sentiment terms has been built by Assiri (2016), it is based on a pre-created lexicon developed by Badaro et al. (2014) and encoded using the Buckwalter translation. However, all these lexicons are not publicly available with the exception of AraSenTi by Al-Twairesh et al. (2016), which is about multi-domain such as educations, sports, news etc. However, their lexicon was based on extracting lexicon from set of tweets automatically and then review it manually. This method failed to consider the different Saudi dialects such as Hejazi and Najdi.

Due to the lack of freely and publicly available dialectal Arabic sentiment lexicons (either in general or domain specific), a new lexicon construction approach is proposed. The Arabic language consists of MSA and many different regional dialectal Arabic, which are typically used in informal daily communication. In fact, Arabs from different regions or countries usually write their tweets in their own dialects. In particular, Saudi Arabia has six different dialects. In order to address this issue, Saudi dialects attributes for the lexicon are added from different Saudi dialects, such as Hejazi (west region), Najdi (middle region), Shamali (north region), Janubi (south region) and Sharqawi (east region). The lexicon in this study is created both automatically and manually by linguists, with the inclusion of native speakers of and Saudi and Arabic dialects. However, it is important to consider that because of the complicated nature of dialect, most of the efforts have been made to build and enhance the lexicon manually. In addition, the involvement of native speakers of different dialects has been crucial for the development of the lexicon, since, despite its massive popularity, there is a lack of standardisation for colloquial Arabic.

The first phase of lexicon construction involved developing the Arabic sentiment lexicon (MSA and different dialectical Arabic). This sentiment lexicon can be used for any further studies about Arabic sentiment analysis because it contains many terms for various domains and topics, and it may have adopted to other dialects of Arabic. In the second phase, the domain features lexicon was constructed. This lexicon is domain-specific and corresponds to the issue of unemployment in Saudi Arabia. In addition, the lexicon construction process includes all terms that determine the polarity level, such as intensifiers, negations, emojis and special phrases such as supplications, proverbs and interjections.

4.5.1. Building a Dialectical Arabic Sentiment Lexicon

To construct the lexicon, sentiment words and phrases were collected from different resources. In the first place, 1130 sentiment words written in MSA were taken from Azmi and Alzanin (2014). Subsequently, each word was associated with a synonym set and the different Saudi dialects set of the MSA word by eight native speakers. After that, the words were manually classified by annotators into one of four polarity levels: very positive (+1), positive (0.5), negative (-0.5) or very negative (-1), as shown in Table 4.17. In this way, the sentiment lexicon was expanded from 1130 words to 16500 sentiment terms. The following offers an overview of those core stages associated with developing the sentiment lexicon:

1. The sentiment lexicon construction of the specific problem domain involves the collection task of words and phrases. It was collected from different resources. Firstly, 1130 sentiment words written in MSA were taken from Azmi and Alzanin (2014). After that, a list was created from SentiStrength[11] website and then translated it into Arabic using English–Arabic dictionary. Finally, the translated terms were revised manually by Saudi native speakers.
2. Associated the terms with their synonym set in addition to their word forms, each word was associated with a synonym set. Due to the limited coverage of Arabic WordNet, only a few Arabic sentiment words are covered (Abouenour, et.al., 2008). A manual collection of Arabic sentiment lexicon and their synsets is considered for Saudi dialect to have a good coverage of sentiments with their synonym. Several sources were associated that provided words plus their synsets. Then, these were assigned the same polarity as their original word. For each word, 3 synonyms (on average) were added.
3. All of the Saudi dialects, such as Hejazi, Nejdi, Qassmi, Shamali and Janubi were considered in this experiment. There is a degree of difference between Saudi dialects, therefore, the native speakers from different regions of Saudi

[11] sentistrength.wlv.ac.uk/

Arabia manually added the dialects expression on the words list and different words were often used to express the same opinion.

4. The final stage in this study was the assessment of all the sentiment words that were collected and weighted according to the appropriate polarity. In this work, the polarity weighting score of each entry in the sentiment lexicon was determined manually while considering simultaneously. The words were manually classified by three native speaking annotators into one of four polarity levels: very positive (+1), positive (0.5), negative (-0.5) or very negative (-1).

Table 4.17: Example of the lexicon construction

Main word	Synonym set		Dialectal Arabic	
	Word	Polarity	Word	Polarity
جيد *jayid* Good +0.5	حسن *hasan* صالح *salih* جميل *jamil* خير *khayr*	+0.5 +0.5 +1 +1	زين *zyn* حلو *halu* مملوح *mamluh* تمام *tamam*	+0.5 +1 +0.5 +0.5
سيئ *syy* Bad -0.5	فاشل *fashil* طالح *talih* باطل *batil*	-1 -0.5 -1	يفشل *yafshil* خايس *khays* معفن *maefan*	-0.5 -1 -1

4.5.2. Building the Domain Features Lexicon

Unigram provides good coverage and is the simplest of features to extract from the text, enhancing the credibility of the data. In contrast, bigrams and trigrams capture important elements in the text such as sentiment expression patterns or negation. Therefore, the process began with a statistical approach (NB classifier) through extraction of frequent terms, namely the unigrams, bigrams and trigrams in the annotated tweets. Then, for each of these features, the terms were manually checked by domain experts, creating a dictionary for all the candidate features that are relevant to the domain specific with the synonym set, inflections' forms and dialects (as mentioned in detail in a previous section). Although this process determines several domain features, it cannot be relied upon since it does not provide a satisfactory coverage. Therefore, domain experts were responsible for accurately building and enhancing manual lexicon features. The resultant domain feature lexicon is 1987 words, as illustrated in Table 4.18 and Table 4.19.

Table 4.18: Domain categories

Categories	Domain features Lexicon
Saudi cities	42
Countries	25
Saudi government organizations	90
Saudi national programs	7
Nationalities	98
Qualifications	58
Jobs	77

- **Example of the Domain Features Lexicon**

Table 4.19: Example of domain features lexicon

Main word	Category	Synonym set	Dialect	Inflections' forms
سوري *suri* Syrian	Nationalities	شامي	سوارنه سواريه الشوام	سوريه سوريين سوريات
راتب *ratib* Salary	Employee	دخل شهري	معاش مرتب	رواتب معاشات
معلمه *muelimuh* Teacher	Job	مدرسه مربية أجيال أستاذه	أبله	معلم معلمه معلمات معلمين

4.6. The Polarity Level and Intensifiers

Most of the available research in the literature addresses the Arabic sentiment analysis issue as a binary classification problem — that is, a two class (positive or negative sentiment) problem. In this sense, words, phrases or documents that may have different intensities have to be merged into one of these two classes; they have to be classified either as a positive or negative sentiment (Badaro et al., 2019). In the case of word-level sentiment analysis, the two-class approach would lead to the system not being able to recognise the difference between words such as 'nice' and 'beautiful', which both have the same polarity (positive but different intensities). In the case of a phrase-level sentiment analysis, however, these kinds of approaches will not be able to distinguish different sentimental connotations brought about by intensifiers, such as 'very', 'absolutely', 'extremely', etc.

Within the context of these methods, no difference would be noticed between 'nice' and 'extremely nice', as both would be classified as only positive. However, in real-world sentiment analysis, the polarity spectrum of sentiment ranges across a gradient of positives and negatives. In fact, researchers agree that in order to improve the quality of the NLP systems, it is crucial to model intensity at the phrase level, especially in question answering and textual entailment (de Marneffe et al., 2010). Thus, researchers have proposed combining an intensifier (support word), such as 'very', with a polar adjective, such as 'good' or 'bad', as a multiplying effect. This can help to establish different sentiment scores for the phrases 'very good', 'good', 'bad', and 'very bad'.

Throughout the prior research in this field, Saudi sentiment analysis studies have not yet considered the impact of intensifiers on sentiment polarity. Thus, in this study, the intensifiers for Saudi dialects were collected manually by native speakers due to the lack of a pre-existing list of intensifiers in the literature. Around 33 Saudi intensifiers were collected, of which three are presented in Table 4.20.

Table 4.20: Example of some intensifiers

In English	In Arabic
Very	جدا / كثير / واجد / وايد *jiddaan / kthyr / wajid / wayd*
Absolutely	طبعاً / أكيد / من قلب *tbeaan / 'akid / min qalb*
Extremely	مره / حيل *marah / hayl*

4.6.1. Considering Negations

The identification of negations is crucial for the success of sentiment analysis since the presence of such words can alter the whole meaning and orientation of an opinion. Duwairi and Alshboul (2015) proposed an analysis of negation particles for Arabic sentiment analysis that considers two grammar rules: أدوات النصب and أدوات

الجزم. Based on these rules, five important negation particles widely used in Arabic were identified and divided into two different negation representative groups, which are *maa* 'ما', *laa* 'لا', *lam* 'لم', *lan* 'لن', and *laysa* 'ليس'.

Hamouda and El-Taher (2013) proposed a machine learning based sentiment analyser for analysing comments on Arabic Facebook news pages. The authors used different machine learning methods, as well as different features, for training. The machine learning model considered five different negations in order to optimise performance and accuracy. However, it is important to highlight that the proposed model took into account only five MSA negations and did not consider dialects. In addition, Hamouda and El-Taher (2013) used equations to narrow the search process which mean that only the percentage of negation in the post or the comment was taken into account as a feature, which paid no attention to the effect of negation on the phrase. In this study, in order to cope with the complex nature of the Arabic language, particularly the negation issue, it was necessary to use advanced rules that are capable of handling the most relevant and popularly used negation expressions. Therefore, the negation list was collected manually, resulting in a total of 45 negation words used in Saudi dialects, such as مش ، مو ، ماني ، مارح ، محد ، معاد *(msh, mw, mani, marih, mahadun, mueadin*).

4.6.2. Considering Emojis

Emojis are small digital images used in social media to represent moods, thoughts, emotions and feelings (Felbo et al., 2017). In the last few years, the use of emojis on microblogging services and social networks has significantly increased, particularly on Twitter. Without relying on the language or domain, emojis can effectively and quickly convey specific feelings. Therefore, to develop sentiment analysis applications effectively, the detection and classification of the emojis is necessary. There are a few works that consider the use of emojis in Arabic Sentiment Analysis, such as Abdellaoui et al. (2018) and Al-Azani and El-Alfy (2018). Applying this technique to a practical challenge, Al-Azani and El-Alfy (2018) introduced the idea of resorting to new non-verbal features for the sentiment analysis of microblogs as an alternative to using NLP processes The study integrated several machine learning algorithms into a single solution with features extracted from 969 emojis (Al-Azani and El-Alfy, 2018). Experimental results show that the proposed emoji-based features performed well for detecting sentiment polarity.

This study considered non-verbal features within the comprehensive solution to Arabic sentiment analysis. In this way, the intended/overt sentiment of the emojis could be evaluated, as well as their effect on sentiment analysis. The emojis represented four levels of sentiment: very positive (VP), positive (P), negative (N) and very negative (VN). The list of emojis applied to this study originated from Novak et al. (2015) and contains 592 individually distinct characters. The list of the emojis was annotated by human annotators to manually assign polarity class and to score each emoji. The annotators were informed to assign scores of VP= 1.0, P= 0.5, N= -0.5 and VN= -1. The score nearest to the average of the annotators' scores was computed for each emoji. Overall, agreement among the annotators was high at 91.2%, with a Kappa (K) score of 0.85. The partial list of VP, P, N and VN emojis and an example are shown in Tables 4.21 and 4.22.

Table 4.21: Partial list of emojis

Label	Emoji
VP	👌 💪 ❤ 💕
P	💜 👌 💙 👆
N	😐 😟 😶 😟
VN	😣 💔 😓 👎

Table 4.22: Example of a tweet containing some emojis

Tweet	تعرف على برنامج السعوده الرائع واستغل الفرصه لترتقي ببلدك و ☝ فرصه ياشباب الوطن لا تترددد😍 🌸 👌 💙اعر ف حقوقك *taearaf ealaa barnamaj alsueudih alrrayie waistaghala alfirasah litartaqi bibaladik w aer f huquqik farasuh yashbab alwatan la tataradad* 💙 👌 🌸 🌸 ☝ ☝
Translation	Learn about the wonderful program Sauda and take the advantage of the opportunity to improve your country and know your rights 💙 👌 🌸 😍 ☝ it is good opportunity Do not hesitate
Annotation	Positive tweet by all annotators
Emojis	💙 / 👌 / 🌸 / 😍 / ☝

4.6.3. Considering Special Phrases

There are many special phrases in dialectal Arabic that express feelings, such as supplications, proverbs and interjections, and these play an important role in sentiment classification. To improve the developed Arabic sentiment analysis system in this study, it was crucial to be able to analyse these special phrases in a clear and accurate way. To address this issue, a novel, phrase-based method for handling supplications in dialectal Arabic was devised, improving the overall accuracy of the sentiment extraction process.

4.6.3.1. Supplications

Arab people usually use supplications in their daily life, especially the Saudis. This behaviour is also reflected in their social media content. Semantic experts agree that supplication can represent both positive and negative attitudes (Mohammad, 2016). Although supplications are often used in social media to express positive as well as negative feelings, there are only a few studies that address them in the Arabic sentiment analysis context (Ibrahim et al., 2015). In the corpus developed for this study, more than 32% of the tweets contained supplications, whether positive or negative ones, indicating the importance of supplications for determining sentiment. There are several sources of supplication, such as good wishes, as illustrated in Tables 4.23 and 4.24, and bad wishes, as illustrated in Tables 4.25 and 4.26.

Table 4.23: An example of a good wish supplication 1

Tweet	اللهم أسعدنا و وفقنا و بشرنا و سخر لنا عبادك الصالحين الذين يسعون لحل مشكلتنا و نتوظف *allahum 'aseadna w wafiqna w basharna w sakhar lana eibadik alsaalihin aladhin yaseawn lihali mushkilatuna w natawazaf*
Translation	Oh God, give us the happiness, reconcile, tell us good news and let a good people work hard to solve our problems by employing us
Annotation	Positive tweet by two annotators and a negative tweet by one annotator
Source	Personal expression
Special phrases	اللهم أسعدنا / وفقنا / بشرنا / سخر لنا عبادك الصالحين *allahum 'aseadna / wafaqnaa / basharna / sakhar lana eibadik alsaalihin*

Table 4.24: An example of a good wish supplication 2

Tweet	الله يبسط الرزق لمن يشاء و يقدر و هذانا صابرين *allah yabsut alrizq liman yasha' w yuqadar w hidhanana sabirin*
Translation	God simplifies the livelihood for whomever he pleases
Annotation	Positive tweets by two annotators and a negative tweet by one
Source	Qur'an (the holy book)
Special phrases	الله يبسط الرزق *allah yabsut alrizq*

Table 4.25: An example of a bad wish supplication 1

Tweet	الله يحرق قلوبكم مثل ما قلوبنا محترقه و ينتقم منكم .. الى متى و حنا مهمشين؟ *allah yuhariq qulubikum mithl ma qulubuna muhtariquh w yantaqim minkum .. alaa mataa w hanna muhimashina?*
Translation	Oh God, give them the misery feeling like what we feel and take revenge .. until when we are marginalized?
Annotation	Negative tweets by all annotators
Source	Personal expression
Special phrases	الله يحرق قلوبكم *allah yuhriq qulubikum*

Table 4.26: An example of a bad wish supplication 2

Tweet	تم رفضي بدون مقابله حسبي الله و نعم الوكيل *hasbi allah w nem alwakil tama rafdi bidun muqabilih*
Translation	God is enough for me and the best deputy
Annotation	Negative tweets by all annotators
Source	Qur'an (the holy book)
Special phrases	حسبي الله و نعم الوكيل *hasbi allah w nem alwakil*

To address these issues, a seto of common supplications phrases was developed from several resources, such as the Qur'an or common quotes used in everyday speech. The supplications were identified in the tweets if they contained one of these words:

الله or اللهم, see Table 4.27. Most of the supplications phrase were collected from the "ALkalem attayeb" website (Kalemtayeb, 2019). Around 70 supplication phrases were collected that are commonly used in Arabic tweets, in addiction to some other supplications which were added manually.

Table 4.27: An example of a set of common supplications

Positive sentiment supplication	Negative sentiment supplication
الله يوفقك *allah yuafiqik*	حسبي الله و نعم الوكيل *hasbi allah w nem alwakil*
بارك الله فيك *barak allah fik*	أعوذ بالله *'aeudh biallah*
جزاك الله خير *jazak allah khayr*	لا حول و لا قوة الا بالله *la hawl w la quat 'iilaa biallah*
الله يعطيك العافيه *allah yuetik aleafih*	الله المستعان *allah almustaean*

4.6.3.2. Proverbs

Proverbs are short expressions of popular wisdom. A proverbial expression is a type of conventional saying similar to a proverb, which is transmitted by oral tradition. Idiomatic phrases are also similar constructions; it is sometimes difficult to draw a distinction between tideioms and proverbs. For proverbial expressions and idiomatic phrases, the meaning does not immediately follow from the phrase itself. In addition, some experts classify proverbs and proverbial phrases as types of idioms (Ibrahim et al., 2015). In this study, the analysis of the proverbs was included in order to acknowledge expressions of feeling about particular issues. The proverbs were manually collected from a variety of colloquial sources, resulting in a total of 200 proverbs in Saudi dialects. Table 4.28 shows examples of positive and negative proverbs, and Table 4.29 shows an example of a tweet containing a proverb.

Table 4.28: Example of positive and negative proverbs

Positive sentiment proverbs	Negative sentiment proverbs
الصبر مفتاح الفرج *alsabr miftah alfaraj*	بدون حسيب و لا رقيب *bidun husayb w la raqib*
خير خلف لخير سلف *khayr khalf likhayr salaf*	الشق أكبر من الرقعه *alshiqu 'akbar min alraqeih*
ما حرك داواك *ma harak dawak*	و على عينك يا تاجر *w ealaa eaynak ya tajr*

Table 4.29: Example of a tweet containing a proverb

Tweet	تعبنا ونحن نطالب ونستجدي ولا حياه لمن تنادي *taebuna wanahn nutalib wanastajdi wala hiah liman tanadi*
Translate	We are tired of demanding and begging but no one responds
Annotation	Negative tweets by all annotators
Special phrases	ولا حياه لمن تنادي *wala hiah liman tanadi*

4.6.3.3. Interjections

The use of interjections usually expresses a negative feeling (Ortigosa et al., 2014). For instance, expressions like الى متى ، وين قاعدين ، من متى ، وش باقي ، usually come with punctuation marks, such as (؟) and (!). Around 30 interjections were collected manually as shown in Table 4.30.

Table 4.30: Example of a tweet containing an interjection

Tweet	ما العذر في تجاهل ؟ أعلى المؤهلات وخبرات ودورات وأعلى الدرجات في قياس ولا نجد فرص عمل لا في الخدمه المدنيه ولا القطاع الخاص ولا الجامعات . . . ونطالب منذ سنوات . . بلا جدوى . . إلى متى ؟ ؟ *ma aleudhr fi tjahl? 'aelaa almuahalat wakhibrat wadawrat wa'aelaa aldarajat fi. . . wanutalib mundh sanawatin. . bila jadwaa. . 'iilaa mta؟ ؟*
Translation	What excuse is there to ignore us? Higher qualifications, experience, courses and higher grades in Qiass programs but no job opportunities in the civil service or the private sector or the universities. . . We have been demanding them for years. . useless. till when? ?
Annotation	Negative tweets by all annotators
Special phrases	؟ / إلى متى . *'iilaa mta؟*

4.7. Chapter Summary

This chapter has provided an in-depth overview of the resources and collection methods used to develop a corpus for analysing Arabic sentiment across social media channels like Twitter. Due to the lack of open datasets for the Arabic language, this study has created a gold standard corpus for sentiment analysis through the manual annotation of tweets. The dataset was captured from an array of trending Saudi Arabian hashtags and included thousands of tweets that discussed unemployment social issues, such as السعوديه_للسعوديين (Saudi Arabia for Saudis) and توطين_قطاع_الإتصالات (localisation of the telecommunications sector) and one of the most important accounts, @JoblessGrads9 (عاطلون بشهادات عليا). Subsequently, these tweets were by removing the URL, hashtags, redundant tweets and stop words. After that, NLP was applied to the collected tweets, such as tokenisation and normalisation. Regarding the light stemming, this study confirmed that MSA stemming algorithms are not applicable to Arabic dialects, and not many stemmer tools can reconcile the specific dialectic variations. These findings confirm that the ISRI stemmer is the most accurate when applied to Saudi dialectal Arabic words. Still, the ISRI stemmer fails to stem certain words because it was primarily developed for MSA rather than dialectal Arabic. Accordingly, this study has developed a novel stemming approach that integrates the ISRI stemmer with a bespoke rule-based stemmer to address the challenges of Saudi dialectal Arabic stemming. This rule-based algorithm comprises a set of pre-defined rules for extracting the stem of Saudi dialectal Arabic words and was found to provide improved accuracy when compared to other stemming algorithms. Subsequently, a gold-standard corpus comprised of 7,000 manual tweet-based sentiment annotations was developed Finally, a a domain specific lexicon was developed in to assess the polarity, intensifier, negation, emoji, and special phrases related to native Arabic tweets.

Chapter 5

5 Multi-factor Lexicon-Based Sentiment Analysis of Social Media Content in Dialectical Arabic

Due to the structural limitations of social media communication (e.g. character limitations, context, directional), users are frequently challenged to express ideas and arguments with limited linguistic efforts (Albogamy and Ramsay, 2015). For Arabic sentiment analysis, the result of such conditional communication is significant, resulting in a dialectical bias that not only limits the applicability of traditional MSA approaches, but require careful consideration for dialectical influences and biases (El Beltagy et al., 2016). Evidence captured during this study from Twitter revealed a range of abbreviations, acronyms, colloquialisms, emojis, and other lexical limiters that have complicated the application of traditional sentiment analysis tools to high-engagement social media channels like Twitter. This chapter addresses such limitations, drawing upon a multi-factor solution to address the complexities of dialectical Arabic and to increase the accuracy and insightfulness of sentiment analysis outcomes.

Sentiment analysis methods are mainly based on lexical (linguistic) or machine learning (statistical) approaches. In machine learning approaches, the extracted text features are processed using machine learning algorithms, such as a support vector machine (SVM), naïve Bayes (NB) and decision tree that are trained using text that is pre-labelled with the sentiment polarity. In the case of the lexicon-based approach, in the sentiment analysis context, a robust sentiment lexicon with a custom number of terms (each with a known polarity) has to be built. Then, based on this lexicon and the application of statistical-semantical weighing and distribution schemes, the polarities of the unknown words can be established in order to finally determine the polarity of the whole block of text. However, lexicon-based approaches require a significant human effort, since the collection of the opinion words has to be done manually to build a high-quality lexicon (Abdulla et al., 2014). Different Arabic sentiment analysis techniques have been proposed in the literature to analyse Modern Standard Arabic (MSA) or dialectical Arabic (DA). Some studies focus on machine learning approaches using different machine learning algorithms (Duwairi et al., 2016; Hammad et al., 2016; Al-Horaibi et al., 2016). The lexicon-based approach is also considered in the literature (Al- Twairesh et al., 2018; Mataoui et al., 2016; and Abd-Elhamid et al., 2016), and in general, many of these studies focus on sentiment analysis for MSA.

However, little research examines the sentiment analysis of dialects. Although Saudi Arabia has recently been ranked among the countries with the fastest Twitter growth, a major challenge in the sentiment analysis of Saudi Twitter posts is the lack of a gold-labelled corpus and comprehensive sentiment lexicon that covers the different Saudi dialects. Some studies demonstrate an interest in sentiment analysis for the Saudi dialect sentiment analysis (Al-Harbi and Emam, 2015; Assiri et al., 2018; Al-Thubaity et al., 2018; Alahmary et al., 2019); however, the research is still in an early stage. To supplement such emergent theory and experimental models, this study focused on lexicon-based sentiment analysis which incorporates a multi-intensity lexicon-based sentiment analysis algorithm capable of contributing to sentiment analysis in relation to dialectical Arabic.

5.1. Domain Analysis and Feature Extraction

5.1.1 Knowledge Map of the Specific Domain for Arabic Sentiment Analysis

The lexicon-based approach developed for this study fundamentally relies on the comprehensive analysis of the problem domain knowledge. In the context of this work, the overall analysis is critical to the extraction of the semantic features in preparation for their pairing with sentiments. Domain knowledge includes information about a domain's environment, its key concepts, their synonyms and ground facts and the relationship between these items. Domain knowledge in linguistics can be utilised to improve sentiment analysis based on corpus of a data set. The modelling of domain knowledge focuses on capturing relevant information and organising it into concepts connected via relationships. For this case study problem domain of unemployment in Saudi Arabia, the modelled knowledge includes key concepts, such as unemployment, organisation, person, opinion and sentiment; it also includes interrelations, such as interactions with key stakeholders (e.g., citizens and policy makers) and the communication/advice medium (Twitter posts). The concept map is illustrated in Figure 5.1. The concept map was then translated into a formal ontology for use in populating a knowledgebase with semantically tagged information from the Twitter feeds (Khalil and Osman, 2014).

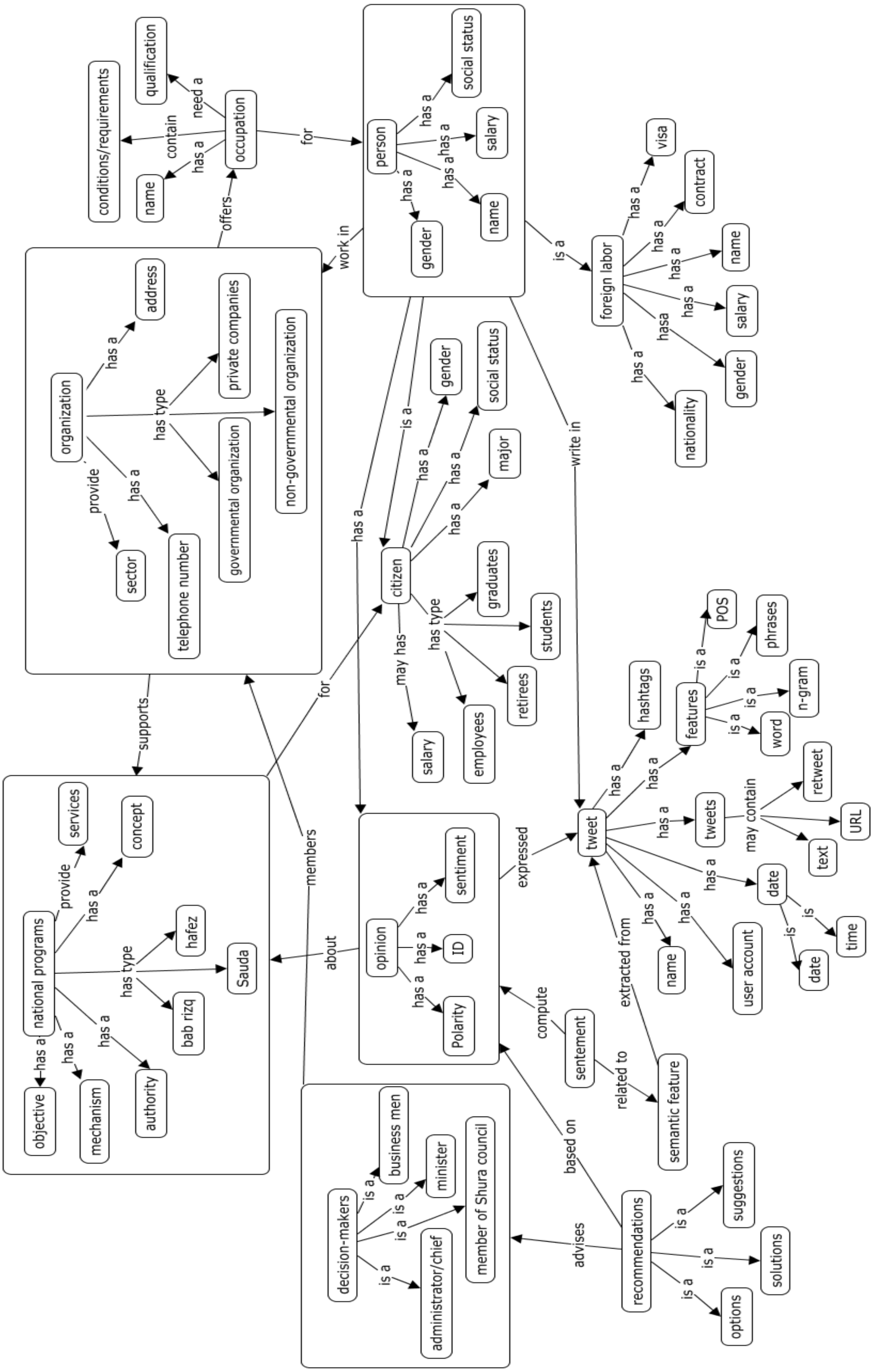


Figure 5.1: Concept diagram

5.1.2 Building an Arabic Sentiment Ontology

Ontology is defined as a set of representational primitives used to model a domain of knowledge (Noy et al., 2001). A JAVA based ontology editor and knowledgebased framework called Protégé [12] was used to develop the targeted ontology. To allow for a good sentiment analysis of tweets, it is necessary to specify which phrases indicate different sentiments, requiring a method capable of assigning which phrases indicate these various sentiments. Therefore, this study involved developing an ontology for the Saudi Arabian dialect that focuses on the semantic relations between the sentiments and their instances. Specifically, there are two polarities in the ontology which are Positive and Negative. In addition, there are different categories of sentiments, for instance, the subtype relation is used to to show that a certain sentiment, e.g. ‘‘فقر” (poor) is a subtype of ‘‘سلبي” (negative) to indicate groupings of sentiments. Along with the sentiment classification, each instance is associated with a polarity (+1) for positive and (-1) for negative.

The primary classes in the formal unemployment ontology are decision-makers, employment offers, national programs, opinions, organizations, people, recommendations, sentiments, Twitter posts and semantic features such as company and city. National programs and organizations are super classes that capture some of the unemployment domain’s key concepts and synonyms. Parts of these key concepts, such as governmental organizations, non-governmental organizations and private companies, are subclasses of the class organization, which represents names of organization as individuals with respect to their roles in the community. The class labelled sentiment contains subclasses such as negative sentiments and positive sentiments, which represents the sentiment orientation of an expressed opinion. The rest of the key concepts, such as foreign labour and citizens are subclasses of the class labelled as people. The class “opinion” contains individuals that characterises the expressed opinions in tweets. All individuals of the created classes were linked together using object relationships, such as “write in”. Figure 5.2 presents a snapshot of the completed unemployment ontology.

[12] http://protege.stanford.edu/

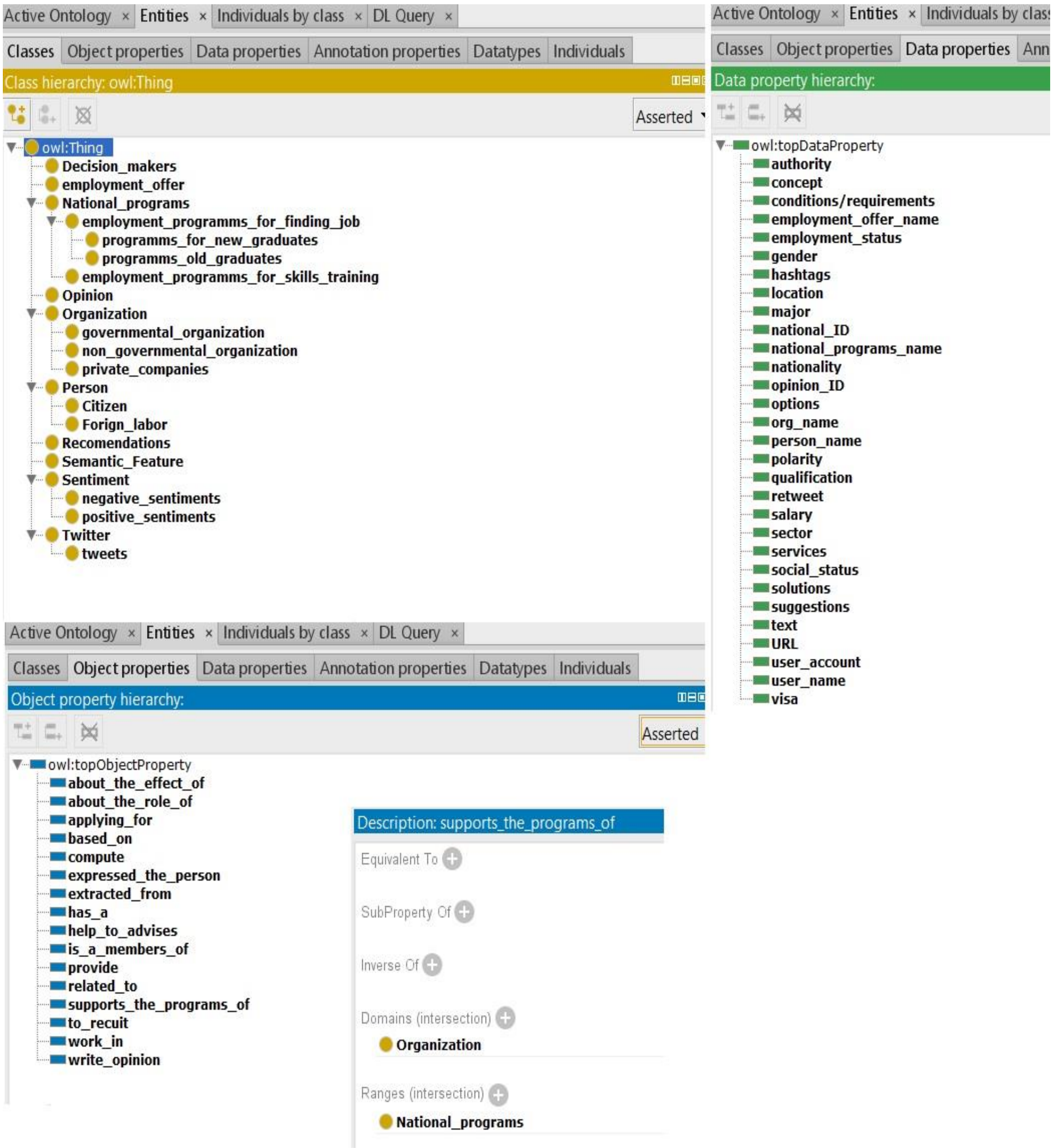


Figure 5.2: Screenshot of the unemployment ontology

5.1.3 Enriching the Gazetteer Lists

GATE[13] is one of the main free software tools currently available that deal with NLP techniques. It was developed as open source software by the University of Sheffield in 1996. Many NLP applications use GATE in multiple languages and media.

[13] https://gate.ac.uk/

GATE automatically manages other standard procedures such as data visualisation, data storage and format analysis. Since the tool is open source, the implementation details, grammar rules and gazetteer lists are available within the tool's source code. These components can be modified to improve the accuracy within the target domain. In this research, each gazetteer list presents a set of names, such as organizations, jobs and cities etc. The gazetteer data was collected from different resources, such as government websites and Wikipedia. Figure 5.3 is a screenshot shows the Arabic gazetteers that were created for this study, exemplifying the positive sentiments written in dialectical Arabic.

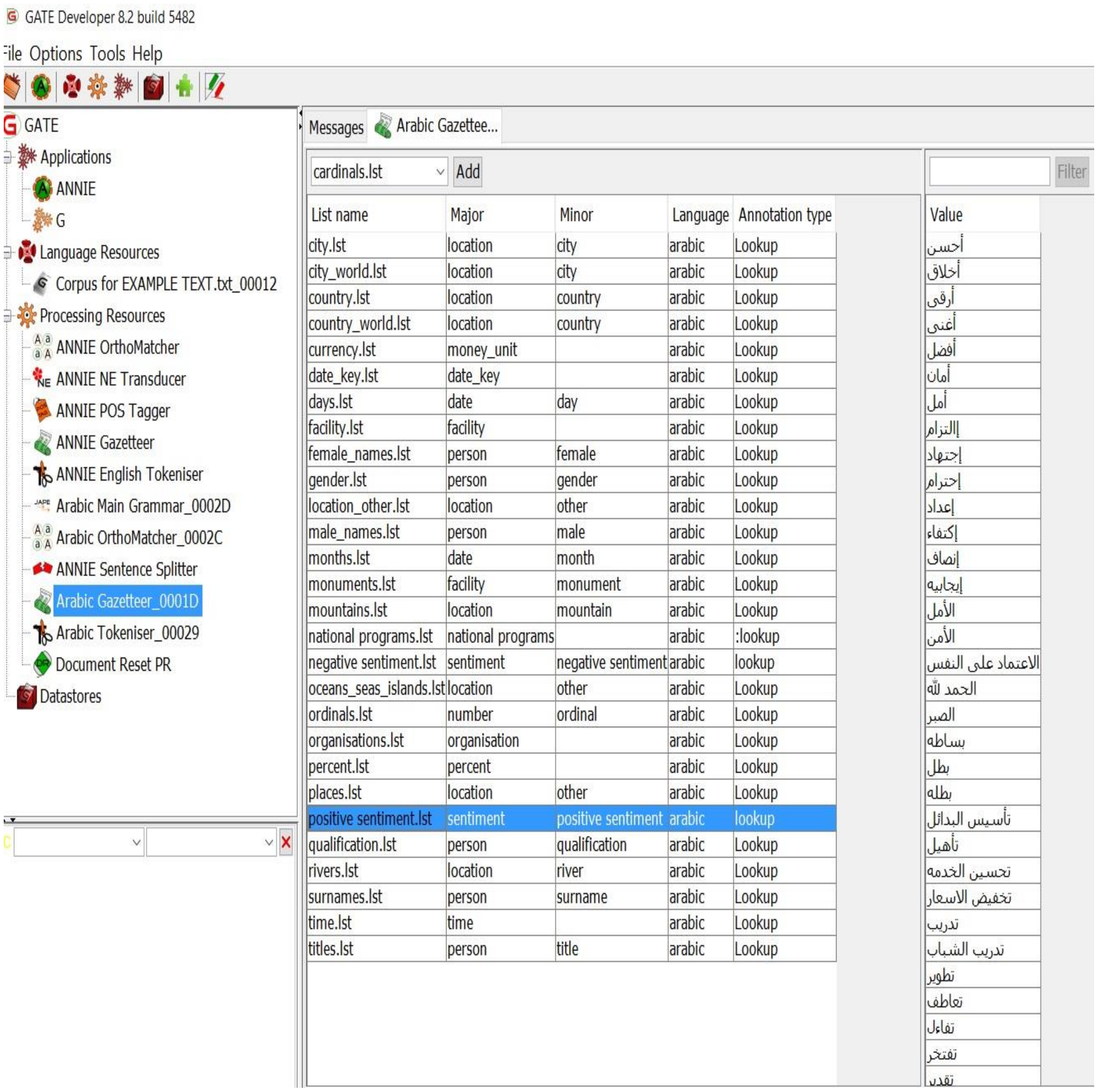


Figure 5.3: Screenshot of the gazetteer lists

5.2. Lexicon-Based Multi-Factor Sentiment Analysis

The lexicon-based sentiment analysis uses two sentiment lexicons (positive and negative) to match the sentiment terms in the tweets. Sentiment terms are counted in the text to calculate the overall polarity of a tweet. The common approach uses some rules to determine the label of the tweet. One rule is that if the number of positive terms in the tweet is larger than the number of the negative terms, then the tweet is labelled as positive, and vice-versa (Pak and Paroubek, 2010). This is a multi-factor process and required several techniques to improve the accuracy of the sentiment analysis.

5.2.1. Feature-Sentiment Association

In order to exclude expressed opinions that were irrelevant to the problem domain, an association window (the neighbouring words to the left and right of the target word) was used to determine the sentiments that are located in close proximity to the domain key concepts (features) identified at the knowledge modelling stage. Traditional POS-based referencing techniques (Al-Horaibi and Khan,2016; Ibrahim et al.,2015) cannot be directly used for feature-sentiment association, as dialectal Arabic lacks the grammatical structure of MSA. The proposed feature-sentiment association technique comprises several steps as exemplified in the tweet in Table 6.1:

Step 1. Find positive/negative sentiments (good, excellent, bad) using the sentiment lexicon, as shown in Table 5.1.

Table 5.1: Sentiments in the tweet

Origenal Tweet	الواسطه خربت علينا بجد واضح للمسؤلين نتيجة الفسااااد ..أقول بس يا زين النوم *alwasituh kharabat ealayna bijidin wadih lilmswlin natijat alfsaaaaad ..aqwl bs ya zayn alnuwm*
Translation	“cronyism really ruined us it is clear to the decision makers the corruption .. it is better to sleep”
Light tweet	واسطه خرب على جد واضح مسؤل نتيجه فساد قول بس يا زين نوم

Step 2. If sentiments are found, find semantic domain features (salary, jobs, etc.) around the sentiments using the domain features lexicon, according to a predefined association window (2 words before and after the sentiment) see Table 5.2; this is sufficient for the relatively short sentence length of tweets.

Table 5.2: Sentiments and features in the tweet

Tweet	الواسطه خربت علينا بجد واضح للمسؤلين نتيجة الفسااااااد ..أقول بس يا زين النوم
Light tweet	واسطه خرب على جد واضح مسؤل نتيجه فساد قول بس يا زين نوم

Step 3. Count: consider only sentiments within the window of the feature, as shown in Table 5.3.

Table 5.3: Consideration of sentiments of the domain features

Tweet	الواسطه خربت علينا بجد واضح للمسؤلين نتيجة الفسااااااد..أقول بس يا زين النوم
Light tweet	واسطه خرب على جد واضح مسؤل نتيجه فساد قول بس يا زين نوم

2w 1w

It is clear from the previous example (Table 6.3) that the sentiments خرب and فساد were considered because they surrounded the semantic domain features. The sentiment زين was considered as a non-relevant sentiment and was excluded because it refers to نوم , which is not a domain feature.

Step 4. Associate: associate sentiments with feature, see Table 5.4.

Table 5.4: Sentiment-feature association

واسطه_خرب	Negative
مسؤل_فساد	Negative

5.2.2. Computing Sentiment Score

Using a term-matching technique, a given term is looked up in the lexicon. If there is a match, the score is considered; otherwise, no score is considered for the given term. The score is calculated by the following steps:

pseudocode: Tweet Score Calculation

```
Inputs: A tweet, lexicons
Output: Sentiment score
Set Score ← 0
Words ← Tokenize(Tweet)
FOR EACH word in words DO
    IF word is PositiveLexicon THEN
                        Score ← Score + 0.5
            ELSIF word in VeryPosLexicon THEN
                Score ← Score + 1.0
            ELSEIF word is NegativeLexicon THEN
                        Score ← Score  - 0.5
            ELSIF word is VeryNegLexicon THEN
        Score ← Score - 1.0
Label ←  Classify-Tweet(Score)
RETURN Label
```

The tweet score (TS) is calculated by summing all sentiment scores for all the words of the tweet (WS), as shown in the following equation:

$$TS = \Sigma^{ws}$$

Where TS is tweet score, and WS is word (sentiment) score.

5.3. Techniques to Improve the Basic Sentiment Analysis Process

Experiments were conducted with different techniques to improve the accuracy of the sentiment analysis mechanism, namely light stemming and morphological analysis of Arabic language, negation, intensification words, emojis and special phrases.

5.3.1. Tweet-Score Calculation with Negation

With respect to sentiment analysis, using negation in language reverses the polarity of the sentiment. For example, ‘not happy’ should be considered negative.

Thus, considering negation in sentiment score calculation is important. Negation is represented by a window for terms in tweets. For instance, consider the following tweet: ‘I do not like pizza’,‘انا ما احب البيتزا' , making the window equal to one for the tweet allows us to get the previous word for each word as follows: ‘_ I’, ‘I do’, ‘do not’, ‘not like’ and ‘like pizza’, ‘انا_’,’انا ما ’,’ما احب ’,’احب البيتزا ’. Making the window for dialectical Arabic is exactly the same way in English. The pseudo-code of function tweet-score calculation with negation is as follows.

The pseudo-code of function: tweet-score calculation with negation

```
Inputs: A tweet, lexicons
Output: Sentiment score
Initialize Score ← 0
window_list ← generate-Window(Tweet, 1) // generate a window with size 1
FOR EACH previous_word, word in window_list DO
    IF word is PositiveLexicon AND previous_word in negation_list THEN
                Score ← Score - 0.5
          ELSIF word in VeryPosLexicon AND previous_Word in negation_list THEN
             Score ← Score - 1.0
          ELSEIF word is NegativeLexicon AND previous_word in
          negation_list THEN
                Score ← Score + 0.5
          ELSIF word is VeryNegLexicon AND previous_word in
negation_list THEN
      Score ← Score + 1.0
Label ← Classify-Tweet(Score)
RETURN Label
```

5.3.2. Determining the Sentiments’ Intensity

In this approach, the compiled intensification terms were used to assess the sentiments’ intensity. By considering a window (neighbouring words) for terms in tweets to it is possible to get the previous and next words for each tweet because in some cases, the intensity is not associated with the sentiment as shown in the examples in Table 5.5. This is due to the use of dialectal Arabic on Twitter. The pseudo-code of tweet-score calculation with a consideration of intensification is further described:

Table 5.5: Example of some tweets containing intensification

Tweets	Translation
ماهي أسباب **رفض** أعداد **حيل** كثيره من المتخرجين *mahy 'asbab rafad 'aedad hyl kathirih min almutakharijin*	What are the reasons for **rejecting** **too** **many** graduates
عيال الديره **جدا** سباقين **للخير** *eial aldiyrih jiddaan sabaqin lillkhayr*	Citizens have **very** initiative to do **good** things
لا تكتبوا أمور **حيل** تولد **كراهيه** بيننا *la taktabuu 'umur hyl tulad karahih baynana*	Don't write things that generate **too much** **hatred** among us
السعاده خياراتها **جدا** متعدده .. فقط تأمل حولك *alsaeaduh khiaratuha jiddaan mutaeadiduh .. faqat tamal hawlik*	**Happiness** has **very** different options. Just look around you

The pseudo-code of function: tweet-score calculation with considering intensification

```
Inputs: A tweet, lexicons
Output: Sentiment score
Initialize Score ← 0
window_list ← generate-Window(Tweet, 2) // generate a window with size 2
FOR EACH previous_word, word, next_word in window_list DO
    IF word is PositiveLexicon THEN
                Score ← Score + 0.5
                IF previous_word in intensification_list OR next_word in intensification_list THEN
                        Score ← Score + 0.5
        ELSIF word in VeryPosLexicon THEN
            Score ← Score + 1.0
                IF previous_word in intensification_list OR next_word in intensification_list THEN
                        Score ← Score + 0.5
        ELSEIF word is NegativeLexicon THEN
                Score ← Score  - 0.5
                IF previous_word in intensification_list OR next_word in intensification_list THEN
                        Score ← Score - 0.5
        ELSIF  word is VeryNegLexicon THEN
        Score ← Score - 1.0
        IF previous_word in intensification_list OR next_word in intensification_list THEN
                        Score ← Score - 0.5
Label ←  Classify-Tweet(Score)
RETURN Label
```

5.3.3. Experimental Analysis of a Novel Lexicon-Based Approach

To evaluate the effectiveness of the proposed algorithm, several experiments with multi-intensity lexicon-based sentiment analysis and multi-factor lexicon-based sentiment analysis were conducted, considering emojis, intensifiers, negations and special phrases, such as supplications, proverbs and interjections. To evaluate the different approaches, the traditional measures employed in text classification have been employed including: precision (P), recall (R), accuracy (Acc) and F-measure (F1). However, the F-measure, which is a harmonic mean of recall and precision and the accuracy, is also used to evaluate the performance of the system (Bekkar et al., 2013).

The classification results presented in Table 5.6 and Figure 5.4, display the accuracy and the average F-score between the negative and the positive classes. All tests were applied to the study-derived, gold-labelled dataset. The approach of combining all the factors (Lexicon-based baseline, light stemming, polarity, negations, emojis, intensification words) obtained the best classification results (See Table 6.6) and reached an accuracy score of 89.80% and a F-score of 86.32%, registering an improvement of 5% and 9% respectively over the baseline. However, Table 6.6 also shows a good result for the lexicon-based system baseline with classification accuracy of 84.34% and F-score of 76.47%, which is attributed to two reasons, first, the knowledgebased approach which allows the capture of domain-specific characteristics and the effective lexicon construction of Saudi dialects.

These results also show that emojis exhibit lower accuracy than the baseline system with a score of 82.63% and the F-score of 48.70%, which was in-part due to the sarcastic behaviour of Twitter users. In some cases, for example, they may have used emojis to reflect the opposite feeling, expressing sarcasm, but also affecting the accuracy of the assessment. As shown in Table 6.6, the classification accuracy and the F-score for combining the lexicon-based and polarity was 88.94% and 82.14% respectively. The classification accuracy for combining the lexicon-based and special phrases was 85.39% and the F-score was 76.99%. In terms of light stemming, when applied in conjunction with the lexicon-based method, the classification accuracy score and the F-score were 88.99% and 81.16%, respectively.

Negation is a more complex task which required specific rules to identify all of the negation expressions and avoid misrepresentations due to their inconsistencies in usage. An in-depth linguistic analysis and semantic reconciliation were needed to address the complexities of the Arabic language and the issue of negation. The negation result for the lexicon-based method was poor due to two factors. The first was the use of special characters, such as exception characters in the Arabic language, which are common in tweets (e.g., لن ينجح الا المجتهد/, 'no one successful only the hard working'). The second factor was the free order of words in an Arabic sentence, which led to the wrong match between the negation and the sentiment. The results indicate that the accuracy of the negation with the lexicon-based method was 79.53% and the F-score were 57.70%. As a result, with the exception of negation and emojis, all factors proved individually useful in improving the classification accuracy. Combining the factors resulted in the highest classification accuracy measurement.

Table 5.6: Results of multi-factor lexicon-based sentiment analysis of social media content in dialectical Arabic

Method	**Average Accuracy**	**Average F-score**
Lexicon based	84.34%	76.47%
Lexicon based+ polarity	88.94%	82.14%
Lexicon based+ light stemming	88.99%	81.16%
Lexicon based+ negation	79.53%	57.70%
Lexicon based+ intensification words	86.37%	77.53%
Lexicon based+ emoji	82.63%	48.70%
Lexicon based+ Special phrases	85.39%	76.99%
All enhancement techniques (lexicon based+ light stemming + polarity + negation + emojis + intensification words)	89.80%	86.32%

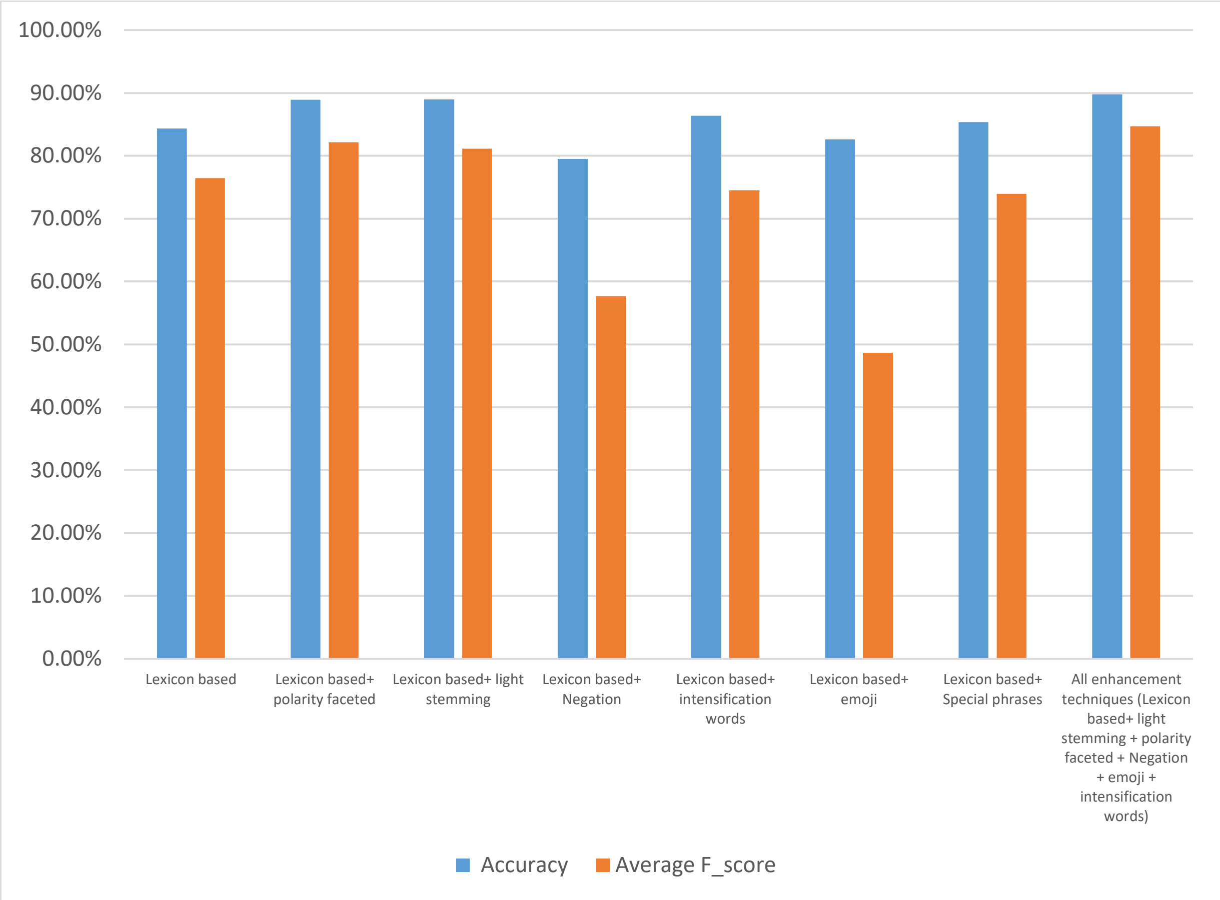


Figure 5.4: Results of multi-factor lexicon-based sentiment analysis of social media content in dialectical Arabic

5.4. Comparison to Similar Work on Dialectal Arabic Sentiment Analysis

The proposed lexicon-based approach fundamentally relies on utilising problem domain knowledge in the sentiment analysis process. Hence, it is useful to evaluate the applicability of this approach to other problem domains. This section compares the performance of the proposed approach against the works of two lexicon-based approaches for Saudi dialects that were published in the Journal of Information Science (JIS).

The first comparative dataset was used by Adayel and Azmi (2016). The authors selected hashtags discussing different social issues in Saudi Arabia, such as #الراتب_مايكفي_الحاجة # *alratb_mayikfy_alhaja* (our salary is not sufficient), #قيادة_26_اكتوبر *#qyadt_26_aktubar* (women driving on 26 October) and #المحتسبون_للديوان_مجدداً # *almhtsbwn_lldywan_mjdda* (Sheikhs went to discuss the women driving with the leader again). Their dataset contained 1103 Arabic annotated tweets. They developed a sentiment analysis system that aims to identify the polarity of the tweets, with two classifications (positive or negative). Regarding Arabic sentiment lexicons, they initially used SentiWordNet to extract some sentiment words after translating the dataset into Arabic and combining the corpus with their own list of essential sentiment-indicating words. The resultant sentiment lexicon consisted of 1500 sentiment words (1000 negative and 500 positive). The overall tweet's polarity was determined according to the cumulative score of the positivity degree of all the sentiment words in that tweet. In their research Adayel and Azmi (2016) considered negation and used regular expression to implement the negation term detector. The algorithm extracted unannotated tweets and returned tweets with sentiment scores. However, some of the returned tweets could not be annotated because the semantic approach depends only on the sentiment words that are found in the lexicon. Therefore, in some instances, the classifier failed to classify a tweet that was devoid of any obvious sentiment word, or the sentiment words were not found in the lexicon. The highest performing results of the Adayel and Azmi (2016) model included a score of 67.60% for the accuracy, F-score of 78.24%, 91.74 for precision and 67.43% for recall.

The second dataset deployed to test the proposed model was collected by Al-Twairesh et al. (2017) in which a total of 14,806 tweets were manually annotated by the recruited annotators. The AraSenti-Tweet corpus is publicly available[14] and the dataset is divided into a training set and test set. The sentiment lexicons 'AraSenti-Trans' were extracted from the datasets of tweets using MADAMIRA tool and contains 131,342 terms. Al-Twairesh et al. (2017) experimented with managing negation in the tweet, compiling a list of negation particles found in the tweets and checking if the tweet contained a negation particle or not. The accuracy assessment revealed that f-score in their work registered 76.31%, the precision was 78.38% and recall was 78.15%.

Table 5.7 and Figure 5.5, compare the performance of the study-proposed sentiment analysis approach against that of Adayel and Azmi (2016) and Al-Twairesh et al. (2017) using their experiments' corpora.

[14] https://github.com/nora-twairesh/AraSenti/tree/AraSenti-Tweet-Corpus

Table 5.7: The results of applying the studylexicon-based approach with Al-Twairesh et al and with Adayel and Azmi corpus

Research	Corpus	Domain	Accuracy	F-score	Precision	Recall
Aldayel and Azmi (2016)	1103 tweets	multi-domain (social issues)	78.22%	76.24%	75.49%	77.02%
Al-Twairesh et al (2017)	4700 tweets	multi-domain	78.61%	62.22%	61.30%	63.17%
Current work	7000 tweets	specific domain (unemployment)	89.80%	86.32%	86%	86.65%

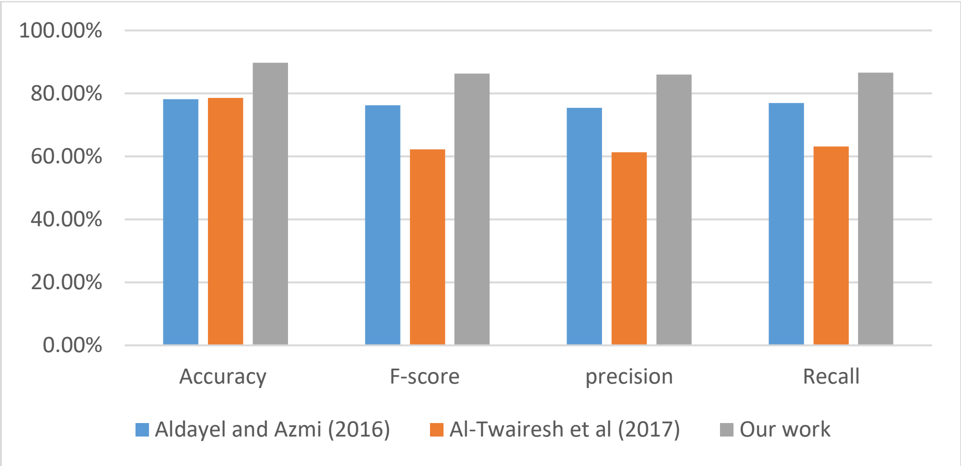


Figure 5.5: The results of applying the study lexicon-based approach with Al-Twairesh et al and with Adayel and Azmi corpus

The experimental results show that the current study's lexicon-based approach clearly outperforms the other across all dimensions including accuracy, F-Score, Precision, and Recall. As the accuracy is not considered in Al-Twairesh et al. (2015), it is assumed that the results indicate an improvement of the f-score around 2% over their work. This is clearly attributed to the fact that the proposed model offers a more comprehensive coverage of the factors that impact lexical analysis including the use of intensifiers, negations, supplication, proverbs and interjections as well as the comprehensive multi-intensity sentiment lexicon for Saudi dialects. These findings confirm that the proposed lexicon-based approach can be re-adapted for other domains, expanding the transferability of these findings to other studies in the future.

5.5. Chapter Summary

This chapter has introduced a novel multi-factor lexicon-based sentiment analysis of social media content in dialectical Arabic social media. This approach integrates the processing of several factors, such as intensification and negation, to improve the classification accuracy. Using unemployment as the target problem domain, the documented research puts forth a sentiment lexicon that is complemented by a comprehensive set of multidialectal sentiment synonyms. Also, this methodology applies an effective light stemming approach which considered knowledge-assisted lexicon-based sentiment analysis and developed the knowledge map of the domain specific for Arabic sentiment analysis. The result was a formal output of an unemployment Arabic Sentiment Ontology which plays a main role in feature-sentiment association in order to exclude expressed opinions that are irrelevant to the problem domain. These results indicate that the proposed combined lexicon approach (light stemming, emojis, intensifiers, negations and special phrases, such as supplications, proverbs and interjections) obtained the best classification result when compared with other prior studies with similar modalities.

Chapter 6

6 Machine Learning Approach for Sentiment Analysis of Social Media Content in Dialectical Arabic

6.1. Introduction

Sentiment Analysis via Machine Learning is a topical area of cutting-edge IT, with classifiers predicting target data based upon a large collection of texts and studies. Machine learning identifies feature of a text through specific applications and approaches. Machine learning searches for patterned straits within texts and exploits them (positive or negative expressions) to classify the most fitting and accurate class. Using machine learning to analyse sentiment from a given body of texts (tweets) has several differing procedures. The initial step is to annotate a corpus for the data field, the Arabic sentiment corpus has been created to achieve this aim and explained, in full, in previous chapters. Subsequently, the text requires converting to a suitable model for machine learning algorithms. This model, otherwise labelled as the vector or feature model, consists of a numerical data matrix. Each individual column within the created matrix signifies an individual word within the corpus. Each individual row signifies the sentence or document from which it is derived, dependant on level classification. The values for both rows and columns illustrate the frequency of the given word within either the sentence or document. In the creation of this matrix, unique features generated or specified to be added. Once the model has been created, the Machine Learning classifier is updated with data, analysed and 'learns' from remaining data to ascertain its own performance.

In linguistic analysis, there are several issues that affect the results. Most of these problems require a robust analytical tool to extrapolate meaning from the evidence. Prior to conducting this study, negation was identified as a critical variable in Arabic text which required careful consideration and analysis. In sentiment analysis, word polarities are affected significantly if negations are ignored, which also affects the text polarities by converting the meaning of the sentence to its opposite. Another issue that was confronted during the lexicon-based sentiment analysis was the need to connect different words together. In Twitter-based social media posts, the length of the tweet is limited, and for this reason, users are likely to connect stop-words with other words that affect the filtering exercise, potentially biasing or amending the outputs. Table 6.1 offers an example of the linguistic problem encountered during this analysis, thereby mandating the application of a machine learning solution to sentiment analysis.

Table 6.1:Example of Linguistic Problem Mandating ML Solution

Word	connecting words
الواسطه Cronyism *alwasituh*	والواسطه/ عالواسطه / الاالواسطه / امساالواسطه / الاالواسطه / بالواسطه / ضدالواسطه / الاالواسطه / اماالواسطه / والاالواسطه/ وبالواسطه / وضدالواسطه / واماالواسطه / وحننجح / وعالواسطه / وبالواسطه / وضدالواسطه *walwasituh / ealwasth / alaalwasth / amaalwasth / alaalwasth / bialwasitih / ddalwasth / alaalwasth / amaalwasituh / walawaastuh / wabialwasitih / wadadalwasitih / wamaaalwasith / wahannjah / waealuasitih / wabialwasitih / wadadaluu*
ننجح We succeed *nanjah*	سننجح / حننجح / فننجح / قدننجح / لاننجح / ولاننجح / وقدننجح / ياننجح *sananjuh / hananjah / fananjih / qadnanjah / lannajah / walannajah / waqadnanajih / yannajah*

The motivation of applying machine learning sentiment analysis is to solve some linguistic analysis issues

6.2. Overview of the Most Common Machine Learning Techniques

This section illustrates the most commonly used machine learning classifiers for sentiment analysis. However, within the Arabic sentiment analysis domain, there is a particular focus on these techniques in the myriad of technologies available for sentiment analysis. For dialectal Arabic, there are a limited number of studies that have been completed to compare with MSA, therefore, there is no prior evidence of the optimal classifier solution.

6.2.1. Naïve Bayes Classifiers

Naïve Bayes classifiers are simplistic and probabilistic based classifiers that apply to Naïve Bayes' theorem. Attributes are independent from each other, resulting in naïve assumptions (Du, 2010). Generally, allowing for independent assumption, class-conditional probability regarding an object (X) (a record or row based within the dataset) is an anticipation of the product of isolated events (Feature Values, X1, X2, X3 …. Xd), conditional on probabilities for the class Y, (d) refer to documents:

$$P(X|Y = y) = \prod_{i=1}^{d} P(Xi|Y = y)$$

Therefore, when predicting class Y:

$$P(Y = y|X) = P(Y = y)(\prod_{i=1}^{d} P(Xi|Y = y)/\ P(X)$$

Given that P(X) is a standard common denominator for class predicted calculations for a record in isolation (X), it has no effect on the class; hence, replacing the previous formula with the following is possible:

$$P(Y = y|X) = P(Y = y)(\prod_{i=1}^{d} P(Xi|Y = y)$$

The main strengths of Naïve Bayes classifiers include nullified values being ignored with irrelevant features uniformly distributed since they fail to have a significant influence on classification. The resultant handling noise data averaged out within the estimated conditional probability. Several Naïve Bayes variations exist, including the Bernoulli Naïve Bayes and the Multinomial Naïve Bayes.

6.2.1.1. Multinomial Naïve Bayes Text Classifiers

Upon adopting the Multinomial Naïve Bayes Text Classifier, the presented probability of a given document (d) being in class (c) is forwarded as (Manning et al., 2010):

$$P(c|d) \propto P(c) \prod_{1 \le k \le n_d} P(t_k|c)$$

P(tk|c) is the conditional probability factor in terms of (tk) throughout the said document regarding terms of class (C). An interpretation of P(tk|c) ascertains that (C) is the appropriate class. P(c) is the previous probability within the document, represented as class (C). Where the document fails to provide evidential basis for one class in relation to another, the case gains a higher prior probability. (t1, t2 tnd) represent tokens in (d) that are a factor of the terms adopted for classification and nd is the sum total of tokens in (d). e.g. (t1, t2 tnd) for a single sentence: “Beijing and Taipei join the WTO” may be (Beijing, Taipei, join, WTO), with nd equalling, when treating the term “and” as a stop word. In text classification, the goal is to ascertain the optimum class for the given transcript. The optimum category in Naïve Bayes classification is the most used; the optimum category could also mean having the Maximum Posteriori (MAP) class cmap:

$$C_{map} = arg_{\ c \in C} max\ \hat{P}(c|d) = arg_{c \in C} \max \hat{P}(c) \prod_{1 \le k \le n_d} \hat{P}(t_k|c)$$

P̂(c) is ascertained by assessing the frequency of class (c), in relation to the overall size of the training data:

$$\hat{P}(C) = \frac{N_c}{N}$$

In which (Nc) represents the overall number of texts in class (c), in which (N) is identified as the mass of documents. P̂(tk|c) is determined by ascertaining the total of occurrences of (t) in relevant documents within class (c), including multi-occurrences of a given term:

$$\hat{P}(t|c) = \frac{T_{ct}}{\sum_{t^1} \in V\ T_{ct^1}}$$

Where (Tct) represents the number of t in training data within class (c), including a multiplex of occurrences of a given example within the given document. Whilst implementing Multinomial Naïve Bayes (MNB), a smoothing addition needs to be added to the conditional probability to avoid negative probability of new terms within the set testing sample unavailable in the training set (Manning et al., 2010).

6.2.1.2. Bernoulli Naïve Bayes Text Classifiers

One alternative to multinomial modelling is the multivariate Bernoulli model. This presents an indicator for each and every term of the text, with 1 establishing the presence of the intended term and 0 indicating negative presence. This estimates $\hat{P}(t|c)$ as the fraction of texts regarding class (c) containing term (t). The Bernoulli model is as complex as the multinomial model (Kim et al., 2006).

6.2.2. Support Vector Machines Classifiers

A Support Vector Machine (SVM) is a non-probabilistic binary classifier based on a linear basis that constructs a set of hyperplanes or a singular hyperplane within an infinite dimensional space and is utilised for classification and regression (Yu and Kim, 2012). The essential underlying concept for SVM regarding Saudi dialectal Arabic classification is to establish a hyper plane, dividing documents or tweets with respect to the sentiment analysis and the marginalising classes as high as possible (Bhuta et al., 2014). For example, here is a training set expressed mathematically:

$$D = \{(x_1, y_1), (x_2, y_2), \dots, (x_m, y_m)\}$$

In which (xi) is represented as an n-dimensional real vector (a document or tweet), with (yi) being either 1 or -1, representing the class to which (xi) is denoted. Initially, the SVM classification F(x) is required to derive positive data points and negative numbers for every point (xi) in (D). Furthermore, F(x) (the hyperplane) requires marginal maximisation. The margin itself is the distance between the hyperplane and the nearest data point (the support vector). This creates the SVM classifier to an optimization constraint dilemma. This can be solved using a language multiplier, such as (Yu and Kim, 2012):

$$F(x) = \sum_i \alpha_i \, y_{i\,x_i} \, . x - b$$

In which the auxiliary non-negative variable (α) is labelled as lagrange multipliers, whereas (b) is the bias, computed by SVM within the training stage. Kuhn–Tucker conditions of optimization theory state that the solution of (α) must satisfy:

$$\propto_i^* \{y_i(w^*.x_i - b) - 1\} = 0 \quad for\ i = 1,2,\dots,m$$

and (α) or corresponding constraints {yi (w.xi – b) -1} must be non-zero. This requirement indicates that when (xi) is a support vector, or when {yi (w.xi – b) -1} = 1, corresponding coefficients (αi) will also be non-zero. Following exploration of theoretical literature regarding SVM, it is not considered as an algorithm, but a mathematical relationship leading to optimal complexity. This complexity requires an optimisation algorithm in order to seek a solution. The algorithm has been termed the SVC () method of “sklearn”.

6.2.2.1. Linear SVC Classifier

Linear SVC, which is related to the Support Vector Machine, determines the optimum linear classification boundary; it attempts to locate hyper line, with the highest margin, derived from the polarity sample from tweets within the dataset. Hence, there is minimum loss in accuracy. It is a popular technique, since it is robust and rarely demands feature selection due to its singular inherent support vectors (Ismail et al., 2016).

6.2.3. Tree Classifier:

6.2.3.1. Random Forest Text Classifier

The Breiman (2001) Random Forest learning method stores and targets classification trees. Tree predictors are formatted so that single trees are dependent on isolated patterned values regarding random vectors. Every tree is distributed uniformly throughout the forest. Since a random forest is a classifier in itself, comprised of tree structured classifiers with an identical distributed random vector, each tree omits a unit vote for the most utilised class. Random forests have been applied to a variety of complex situations in microbiology and genetic epidemiology in recent times. Indeed, random forests have become a major data analysis approach. For example, Ahn et al. (2007) describes their research on Classification on Random Partitions. Classification is a challenge. Based on classifiers, a robust procedure for classification was developed. This predicts random partitions of predictions. A proposed method integrates multiple classifiers in order to achieve enhanced improvement in prediction compared to previous classifiers. This is designed particularly for high dimensional data sets.

6.2.3.2. Decision Tree Classifier

Decision Tree Classifiers imitate decision-making processes found within humans. A tree is a collection of nodes, leaf nodes and links (to children nodes). In the same way, a decision tree also has components with differing interpretations. Each node signifies an attribute. Attaining a child node involves a decision. Subsequently, leaf nodes signify an output. However, it is prone to excessive data since trees have expansive heights. A deep Decision Tree also shows signs of high variance. This algorithm is termed DecisionTreeClassifier, through the method of "sklearn" (Safavian and Landgrebe, 1991).

6.2.4. K-Nearest Neighbor Classifier

K-Nearest Neighbour (KNN) is a simplistic method for classifying text (Tan, 2005). In the proposed approach, with an unannotated document d, the system ascertains that k's nearest neighbour, within the training documents, is classified within the two previous phases. The scored similarity of neighbouring documents, in an attempt to test accuracy, is utilised as the weight of documents. The weighted sum, in kNN classification is

$$score(d, s) = \sum_{d_{j \in knn(d)}} sim(d, d_j)\delta(d_j, s)$$

knn(d) is the set of (k) nearest the neighbours of (d) where (dj) is aligned to sentiment (s), d(dj,s) equals 1, or 0. Document (d) is aligned to the sentiment (s) having the highest rating (Han et al., 2001).

6.2.5. Stochastic Gradient Descent (SGD)

The Stochastic Gradient Descent (SGD) is an algorithm utilised to teach other machines to learn algorithms such as SVM, in which sampling a subset takes place at each stage. It computes gradients from an isolated subset and utilises the gradient to re-evaluate the specified weight vector (w) of SVM classifier. The SGD method calculates the gradient, independent iteration and estimates the overall value of the gradient, recognising randomly chosen examples considered by Bottou (2014).

The stochastic process {w t, t = 1, 2.} is dependent upon randomly chosen examples during each iteration, in which Q (z t, w t) is utilised to limit the risk, as γt is deemed as the learning rate. Convergence of SGD is affected by noisy approximation of the gradient. Where the learning rate decreases, the parameter estimate (w t) also slows down at the same level; however, if the rate decreases too rapidly, the parameter estimate (w t) slowly reaches the optimum point.

This approach is utilised when the extent of training data is large. Due to its computational advantage and simplicity, it is extensively adopted for large-scale machine learning problems (Bottou and Bousquet, 2008).

RT (w∗) , X T t=1 (L(wt , zt) + Ψ(wt)) − X T t=1 (L(w∗ , zt) + Ψ(w∗))

6.3. Developing a Machine Learning Approach for Sentiment Analysis of Dialectical Arabic Social Media Content

Working with the same problem domain of unemployment in Saudi Arabia, this study has investigated the application of a machine learning approach for sentiment analysis in multi-dialect Saudi tweets. The task involved building aclassifier that classifies tweets into positive or negative labels. There were several stages involved in this procedure: First, the dataset was pre-processed; then, features were extracted from the corpus, including different n-grams features and weighting schemes. After that, training was applied to different machine learning algorithms. In the last stage, the performance of different combinations of features were evaluated including the weighting schemes and machine learning algorithms. The performance metric that was adopted for this study is based upon the traditional measures employed in text classification: precision, recall, accuracy and F-measure, as discussed in previous chapters.

Figure 6.1 illustrates the workflow of the machine learning approach. This includes the NLP pre-processing stage, feature extraction, and classification as detailed in the following sections.

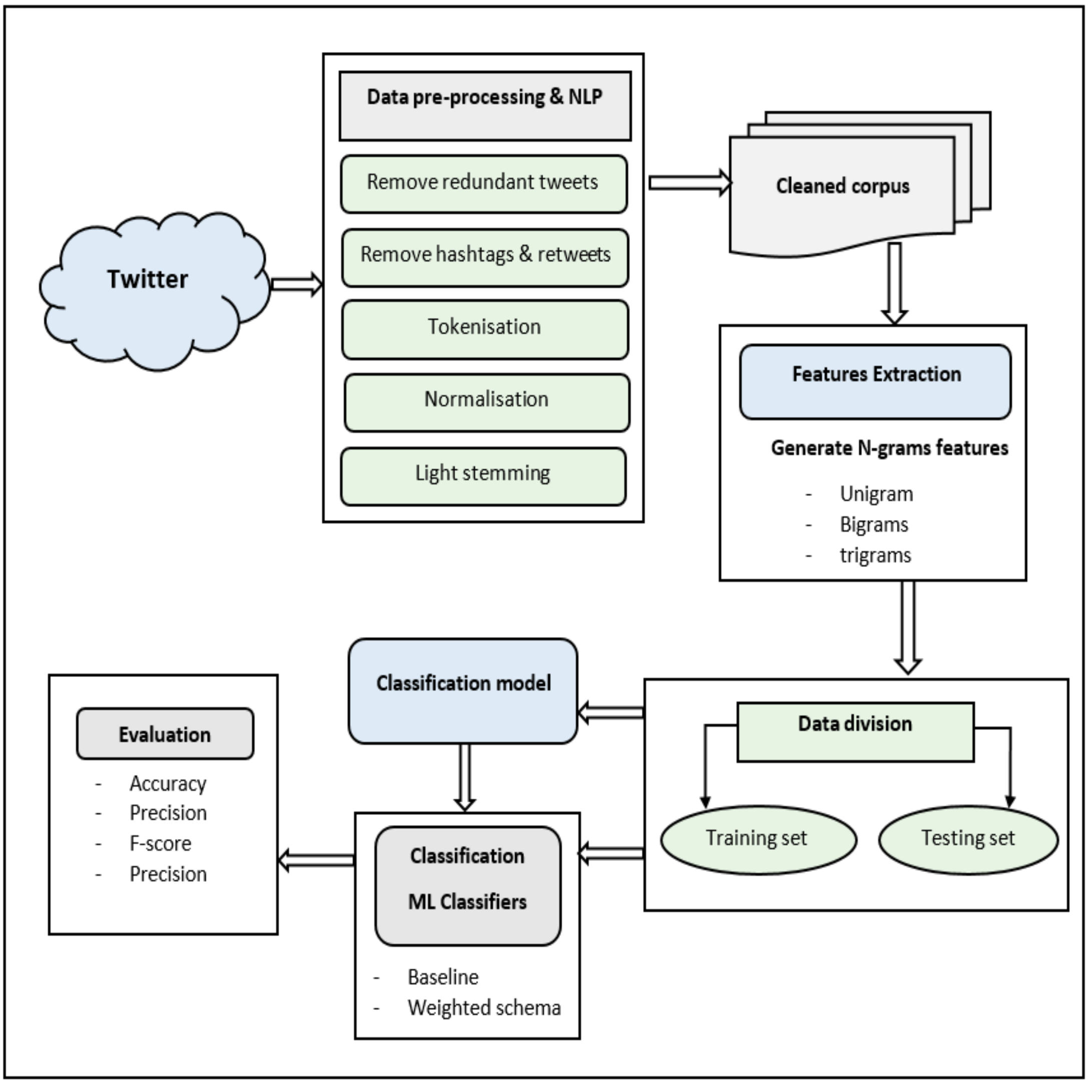


Figure 6.1: Frame work of the system Machine Learning for Sentiment Analysis of dialect Arabic

6.3.1. Pre-processing of the Tweets

In this step, tweets are cleaned by removing links, hashtags and special twitter characters such as (RT), which is shorthand for “retweet”. Then, the text is normalized by removing diacritics or redundant letters (more than two). This step is important because it reduces the number of variations of word features. Also, light stemming was applied to the tweets, as outlined in the pre-processing stage.

6.3.2. Feature Extraction for Machine Learning Sentiment Analysis

The fundamental objective of feature-selection was to ascertain the most indicative features for classification by deleting irrelevant, redundant and noisy data (Liu and Zhang, 2012). Feature selection also included a secondary aim, in reducing both special featural dimensionality and processing longevity. A plethora of text features require consideration for sentiment analysis (Pang et al., 2002), including POS and n-grams models. The former is utilised to locate opinionated adjectives. The N-grams models consist of a continuous sequence of n terms within a set text. When n = 1, features are labelled as unigrams or Bag of Words (BOW), considering the text as unstructured non-contextual vocabulary. Where n = 2, features are labelled as bigrams (dual grams), extracted from the text within a sequence of two words. This application retains textual context. In a similar way, trigrams (triple grams) are extracted in an identical way. N-grams may also be combined to illustrate text in a contextual basis.

Larger and more significant N-grams are classified by n’s value and retain vocabulary with the highest score achieved in accordance with a pre-defined, accepted threshold (a pre-determined indication of the word’s importance). In the feature extraction stage, text is translated into vector representation. Within this model, the feature (weight) of the text is assessed in accordance with the document wherein the word exists. Several weighting approaches, such as the TF-IDF, TF, Inverse Document Frequency (IDF) and Binary (Boolean), offer efficient approaches and schemes. TF-IDF is a numerical statistic reflecting the significance of lexis within the entire document. Scikit-learn creates vectorisers that interpret input documents into featural vectors. The library function TF-IDF Vectoriser can be utilized, in which parameters for the desired features are maintained through reference to the minimum acceptable frequency features.

S1 = "This film is bad"

S2 = "This film is good"

The TF-IDF weighting approach is highly relevant here; it is a standard and popular tool regarding document classification. The formula is:

$$tf(w,d) = f_d(w): frequency\ of\ w\ in\ document\ d$$

$$idf(w,D) = log\frac{1+|D|}{1+df(d,w)}$$

The value of TF (w, d) is the repetition of certain words (w) appearing within a set document (d).

$$tfidf(w,d,D) = tf(w,d) \times idf(w,D)$$

The aggregated value of TF-IDF equals the complete number of documents within the corpus divided by the number of times w is repeated within the set corpus for the IDF (w, D). A more concise labelling for frequency, IDF is a numbered statistical identifier to illustrate the importance of a word within a document in a collated piece of work or corpus. The TF-IDF value expands proportionally and is dependent upon the repeated times the given lexis appears within the document itself; offset by the overall number of documents found within the corpus that contain this specific word, this is an aid in adjusting for the situation that words used more frequently as a general rule (Mohammad et al., 2016).

The most common features used in machine learning sentiment analysis are surface features that generally include n-grams and syntactic features. The Syntactic Features are utilised to reflect the structural nature of the text and comprehend how words combine and function as a process of conveying meaning. Since Arabic is both a rich and morphologically complex language, incorporating morphological and syntactic evaluation is of vital importance when considering sentiment analysis. In the research literature, a very early grammatical approach forwarded the notion to simplify both nominal and verbal phrases into a single distinct format based on “actions” and “actors”, subsequently training SVM to use the following features: adjectives, nouns, actors, actions, syntactic sentence structure, word sentiment polarity and conjunctions relating to previous sentences (Farra et al., 2010).

The recent advanced Arabic NLP resources and tools allowed for the automatic emergence of morphological and syntactic features, utilised in mitigating the impact of complex SA. Such resources include SAMA (Ibrahim et al., 2015), ATB (Maamouri, 2004) and MADAMIRA (Pasha et al., 2014). For example, complementing word-level inflectional morphological features (number, voice and gender) to standard features improves the enhancement of sentiment analysis classification regarding MSA data (Abdul-Mageed et al., 2011), whereas performance lapsed when applied to Twitter which usually written in dialectical Arabic (Refaee and Rieser, 2014).

The reason for the performance lapse when applying Arabic NLP to dialectical Arabic text such as tweets is predominantly due to the fact that the majority of Arabic NLP tools are designed for MSA texts. Studies by Abdul-Mageed (2017), Abuaiadah et al. (2017), Al-Harbi (2017) and Cherif et al, (2016) proved that considering N-grams features show better performance than POS tags features in dialectal Arabic text. Which mean that considering POS feature fails to provide enhanced improvement for sentiment analysis of dialectal Arabic. In this study, the syntax features, which are dependent on NLP and grammar, are not useful, as proven in the literature. This is due to the lack of specific grammar in the dialectical Arabic text to allow the use of NLP tools, such as extracting POS. So, in this work, unigrams, bigrams and trigram features were extracted from the corpus as shown in Figures 6.2, 6.3 and 6.4. Analysis was stopped at trigrams due to the nature of tweets (short messages) and to avoid the potential for noise.

sample of selected features:

['الفاسد', 'الوزراء', 'الأجنبيه', 'مكافحه', 'مختلف', 'الاقارب', 'الشهادات', 'مناسبه', 'الكفاءات', 'التمييز', 'حجج', 'مفتاح'']

[alfasda', 'alwzra'a', 'al'ajnabyha', 'mkafhha', 'mkhtalfa', 'alaqarba', 'alshahadata', 'mnasbha', 'alkafa'ata', 'altamyiza', 'hjaj', 'mfatah'']

Translate into English:

[‘Corrupt ',' Ministers', 'Foreigners',' Elimination ',' Various', 'Relatives',' Certificates', 'Appropriateness',' Competencies', 'Discrimination',’ Arguments’, ‘Key’]

Figure 6.2: Screenshot of a sample of unigram features

sample of selected features:

['ليه الظلم', 'للمطالبه بحقوقهم', 'سوق العمل', 'بنات الوطن', 'الاختيار الأفضل', 'شهادات جامعيه', 'الكبر والتعالي', 'حسبي الله', 'تستقطب كفاءات', 'المعاناه الماديه']

['binnat alwtn', 'laikhtiar alafdl' , 'ishhadat jamieih', 'alkabir waltaealy',,' suq aleml', 'llmatalibih bihuquqihim', 'lyh alzlm' 'almeanah almadih' , 'hsbi allh', 'tsataqatib kfa'at']

Translate into English:

['Saudi girls’, ‘Best choice', 'university degrees', 'Arrogance and Ego', 'The labor market', 'to demand their rights', 'Why this injustices' 'financial issue', 'God with me-supplication', 'attracts competencies']

Figure 6.3: Screenshot of a sample of bigrams features

sample of selected features:

['مشكله العقليات القديمه', 'تستقطب كافة الكفاءات', 'خريجات الجامعيات القديمات', 'تدني مستوى المخرجات', 'الرؤيه الوطنيه الواعده', 'العاطلون حمله الماجستير', 'الاقارب ذوي الواسطات']

[' tadaniy mustawaa almukharijat', 'khryjat aljamieiat alqdimat', 'tsataqtib kaft alkfa'at', 'mshklh aleaqliat alqdymh', 'alaqarb dhwyi alwastat', 'aleatilun hamlah almajstyr', 'alrwiyih alwataniuh alwaeiduh']

Translate into English:

['Low level of outputs', 'Old female university graduates', 'Attracting all competencies', 'The old mental problem', 'Relatives who have Cronyism ', 'Unemployed holders of a master's degree', 'A promising national vision']

Figure 6.4: Screenshot of a sample of trigrams features

N-grams that have frequency lower than a predefined certain threshold are discarded (5 in this experiment). This value is also called a “cut off” in the literature. The threshold is a filter that removes features that have a probability (value) less than a certain threshold. The goal was to reduce n-gram features to avoid noise. The value of the threshold was determined in these experiments by trial and error (try multiple values and set it to the value that gives the best result). In Figure 6.5, min_df= 5 parameter means discard n-gram features that occurred in 5 documents or less, max_df= 0.95 parameter means discard n-gram features that occurred in 95% of the documents.

Threshold = min frequency / number of all features

min frequency (trial and error): 1, 3, 5, 7, ……..etc

used value is 5

```
pipeline = Pipeline([
    ('vect', TfidfVectorizer(min_df=5, max_df=0.95,
                             analyzer='word', lowercase=False,
                             ngram_range=(1, n))),
    ('clf', my_classifier),
])
```

Figure 6.5: Determined the threshold in experiments

Word features (unigrams), bigrams and tri-grams were weighted using the TF-IDF weighting scheme, which defines the importance of a feature based on the term frequency/inverse term frequency. Thus, features and their weights form a so-called document-term matrix. This paradigm is called Vector Space Model or VSM. Another weighting scheme is called Binary weighting (0/1 indicates the absence/presence respectively of n-grams feature in the tweet) is also commonly used in text classification. This study has not formally addressed the TF; this is due to the difference of TF and TF/IDF, which is based on whether the corpus-frequencies of words are used or not. The TF/IDF is by far a better choice, independent of classifier. Thus, some common words e.g. articles received a large weight even if they contribute no real information. In TF/IDF the more frequent a word appears in the dataset, the lower the weight it received. Accordingly, the common words such as articles received small weights, however, the rare words, which are assumed to carry more information, received larger weights.

Also, it should be noted that negation is considered due to the use of bigrams and trigram features. Scikit-learn (sklearn), a machine learning library in Python, was adopted to implement these experiments. scikit-learn (sklearn) has different machine learning algorithms and preprocessing operators. In this work, two files were fed to the classifier, one file contains the positive tweets and a second file contains negative tweets. Each line in the both files represents a tweet. The classifiers have a set of parameters that are useful when dealing with text.

6.3.3. Application of Machine Learning for Sentiment Classification

In this step, different machine learning algorithms are applied to different combinations of n-gram features and weighting schemes. Before that, the dataset is prepared by splitting the corpus into two parts: training (80%) and test (20%). The training part is used to train the machine learning algorithm, and the test part is used to evaluate the performance of the machine learning models. The experiments in this section used different machine learning classifiers to carry out a targeted, novel approach to the analysis of dialectal Arabic sentiment analysis. The classifiers are Naïve Bayes Variant (BernoulliNB, MultinomialNB), Support vector machine (SVC, LinearSVC), Trees (DecisionTree, RandomForest), KNeighbors and SGD. Most of the machine learning sentiment analysis approaches for MSA language used two classifiers, which are Naïve Bayes (NB) and Support Vector Machine (SVM), because in NLP and MSA sentiment analysis field these machine learning classifiers are the state-of-the-art that are usually used (Abbasi et al., 2008; Abdul-Mageed et al., 2011; Pang and Lee, 2008).

On the other hand, for dialectal Arabic, there is limited number of studies completed when compared with the scope of MSA in prior research... Some of those studies that were completed achieved the best result by applying SVM, such as Al-Rubaiee et al. (2016) and Boudad et al. (2017); however, El-Masri et al. (2017) and Mahmoud and Elghazaly (2018) demonstrated that NB outperforms the other classifiers. On the other hand, Nuseir et al. (2017) and Ismail et al. (2018) found that KNN provides the best result, while Altawaier and Tiun (2016) and Abo et al. (2018) selected DT to be the best classifier. SGD provides the best results in some research for dialectical Arabic, such as Rizkallah et al. (2018) and Gamal et al. (2019).

Pseudocode: Applying Machine Learning Algorithms

1. Inputs: Load dataset
2. Output: Sentiment classification
3. Extract n-gram features and filter data
4. Compute TF-IDF features
5. Apply Machine Learning Algorithms on TF-IDF features
6. Use the trained model to get the class label for test data

6.4. Experimental Evaluation of Utilising Machine Learning for Sentiment Analysis of Dialectical Arabic

To evaluate the effectiveness of the machine learning capabilities, three separate, but similar experiments were conducted. The first stage was to conduct the machine learning baseline experiment. A baseline refers to the measurements of key conditions (indicators) prior to the commencement of experiments. Secondly, the machine learning approach was implemented with the TF-IDF weighting scheme, and finally with the binary weighting scheme. The weighting scheme is intended to illustrate how vital words are to a document within a corpus (O'Keefe and Koprinska, 2009). The main aim of the weighting of sentiment analysis is to assign an accurate weight for individual words to reflect its relative significance within the feature, in turn, allowing for accurate prediction of sentiment polarity.

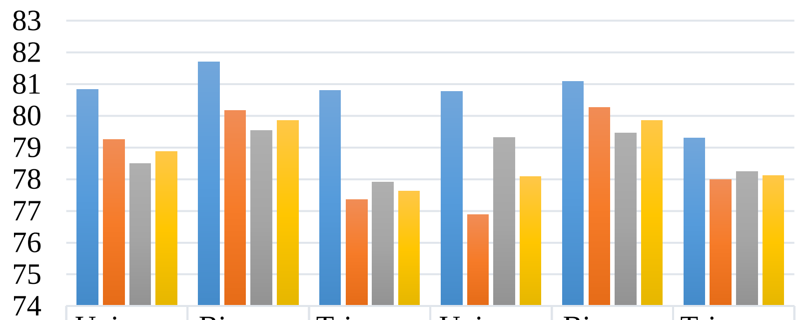
6.4.1. Experimental Evaluation of Machine Learning Baseline

It is essential to initially establish a standard baseline experiment in order to compare results. This allows for a method to present a comparison between the performances of differing classifiers and relevant feature sets. It is difficult to estimate the optimal baseline experiment results, since it will vary, depending upon the precise nature of the experiment itself. In this situation, varied n-gram feature model was chosen as a baseline for the applied pre-processing and the light stemming tool for Saudi dialectal Arabic, since they provide a point of reference to judge alternative feature set experiments for each classifier explored. This baseline is functionally adequate, since it retains basic knowledge of the text classification quandary, which yields the primary underlying challenge in Sentiment Analysis (SA). To begin, an experiment was conducted to find out the effectiveness of pre-processing and the Saudi light stemming (see appendix A). The results indicated that applying pre-processing improved the performance of machine learning around 8% against the results without light stemming. In the experiment of machine learning baseline, differing n-gram models were created and explore their effect on machine learning classifiers. A secondary objective of this experiment was to ascertain which N-Gram model was most successful in regard to the text classification for Arabic dialectal text and also to examine its effect of pre-processing and dialectal Arabic light stemmimg on differing machine learning classifiers.

The results of three N-Gram experiments are shown in Figures 6.6, 6.7 and 6.8 which display the varying classification results (baseline experiments with Saudi light stemming) in terms of the accuracy, recall, precision and F-score. It can be noted from the results that the BernoulliNB classifier obtained the best classification result with bigrams features; it reached an accuracy score of 81.71% and a F-score of 79.85%. However, LinearSVC also shows a good result with bigrams; the classification actuary had a score of 81.20% and a F-score of 81.77, with a small difference from BernoulliNB around 0.51%.

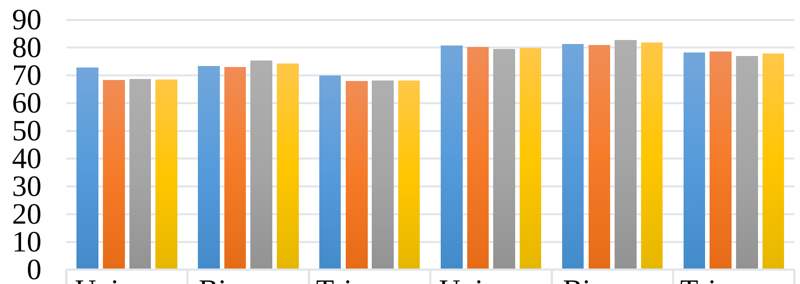
also It was also evidence in the review of both classifiers that the highest results were achieved with the bigrams features. This is attributed to the common use of genitive Arabic phrases in Saudi dialect that consists of two words (bigram) such as "السعوده الوهميه" (unreal Saudisation), foreign labour, localisation of sectors (السعوده الوهميه, العماله الاجنبيه توطين القطاعات).—The results also show that trigrams have the worst results in all classifiers experiments, which can be attributed to the noise captured in the additional gram that suggests false relationships between the words. The result of the trigram experiment confirmed the original hypothesis that that the result will decline in accuracy if more grams are considered.

Figure 7.8 demonstrates that the tree classifiers (DecisionTree and RandomForest) have poor performance in all evaluation measures (precision, recall, accuracy and F-measure). The best precision and recall (80.84% and 82.73%) was achieved by LinearSVC with bigrams features The classification results of the KNeighbors and SGD (see Appendix B) show that it achieved good performance in dialectal Arabic sentiment analysis, especially SGD with Unigram. In this case, the classification accuracy score was 81.12% and the F-score was 80.71%. However, BernoulliNB and LinearSVC achieved the best classification measurement with bigrams features.



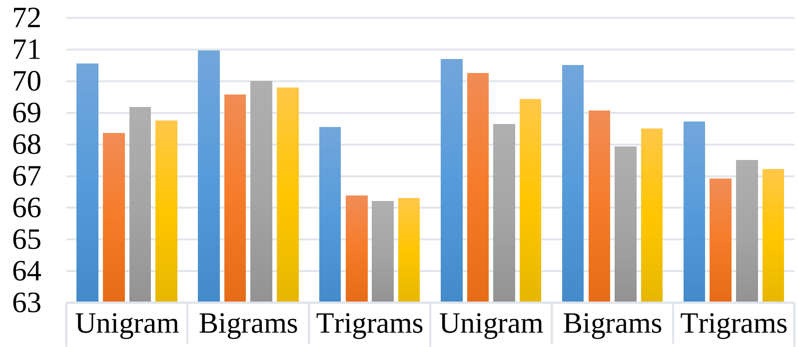
	BernoulliNB			MultinomialNB		
	Unigram	Bigrams	Trigrams	Unigram	Bigrams	Trigrams
Accuracy	80.84	81.71	80.81	80.77	81.09	79.3
Recall	79.25	80.17	77.36	76.89	80.27	78
precision	78.5	79.54	77.92	79.31	79.46	78.24
F-score	78.87	79.85	77.64	78.08	79.86	78.12

Figure 6.6: The results of Machine Learning Baseline - Naive Bayes classifiers



	SVC			LinearSVC		
	Unigram	Bigrams	Trigrams	Unigram	Bigrams	Trigrams
Accuracy	72.74	73.21	69.73	80.65	81.2	78.07
Recall	68.19	72.97	67.98	80.18	80.84	78.48
precision	68.65	75.3	68.02	79.37	82.73	76.91
F-score	68.42	74.12	68.00	79.77	81.77	77.69

Figure 6.7: The results of Machine Learning Baseline - SVM classifiers



	DecisionTree			RandomForest		
	Unigram	Bigrams	Trigrams	Unigram	Bigrams	Trigrams
Accuracy	70.55	70.97	68.54	70.69	70.51	68.72
Recall	68.35	69.57	66.38	70.25	69.07	66.93
precision	69.17	70	66.21	68.63	67.94	67.5
F-score	68.76	69.78	66.29	69.43	68.50	67.21

Figure 6.8: The results of Machine Learning Baseline - Trees classifiers

6.4.2. Experimental Evaluation of Machine Learning Techniques with weighted schemes features

This experimental process has incorporated two commonly used weighting schemes in sentiment classification, TF-IDF and Binary. According to Oussous et al. (2019), the binary model has been effectively validated and employed in a varied range of prior studies. However, Naz et al. (2018) used three different weighting schemes (TF, TF-IDF and Binary) to understand the impact of weighting on classifier accuracy. They observed that the TF-IDF weighting scheme works best.

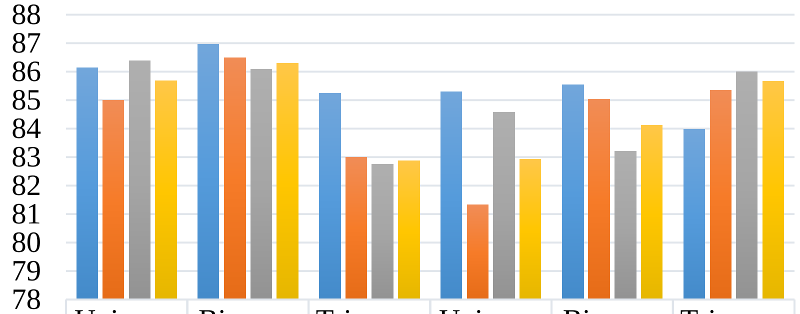
This study has investigated sentiment analysis by using the machine learning approach for dialectal Arabic f to analyse the results and study the impact of different weighting schemes on classification in the dialectical Arabic text. The N-Grams features were weighted using the TF-IDF weighting scheme, which defines the importance of a feature based on the term frequency/inverse term frequency. This paradigm is called Vector Space Model or VSM. Another weighting scheme called Binary weighting (0/1 indicates the absence/presence, respectively, of N-Grams features in the tweet) and is also commonly used in text classification. This study has applied several machine learning classifiers and the results show all the evaluation measures.

6.4.2.1. Machine Learning Techniques with TF-IDF Features Experiment

This section presents the experimental evaluation of machine learning with TF-IDF features and provides results in Figures 6.9, 6.10 and 6.11. The initial example demonstrates how the experiment works with a sample of tweets.

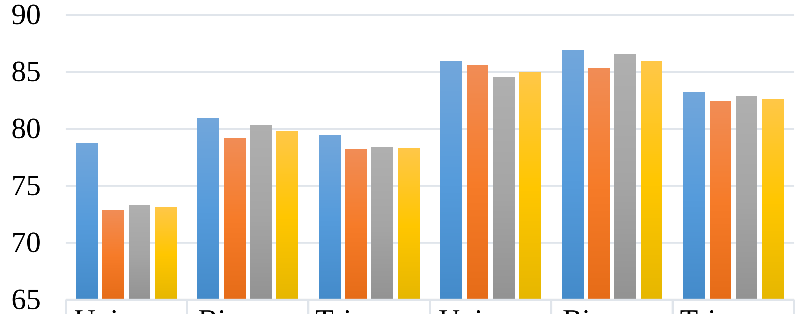
```
from sklearn.feature_extraction.text import Tfidf Vectorizer

corpus = [
    'وظفونا يا جماعة وظفونا',
    'يجب توطين قطاع الاتصالات',
    'وظفونا حسبي الله ونعم الوكيل',
    'العمالة الاجنبية مسيطرة على قطاع الاتصالات',
]

features = TfidfVectorizer()
X = features.fit_transform(corpus)
print(features.get_feature_names())
print(X.shape)
for word in features.get_feature_names():
    print(word, features.vocabulary_.get(word), end='\t')
print(X)
print(X.toarray())

--------------------------------------------------------------------------------
SKLearn features
features: ['الاتصالات', 'الاجنبية', 'العمالة', 'الله', 'الوكيل', 'توطين', 'جماعة', 'حسبي', 'على', 'قطاع', 'مسيطرة', 'وظفونا', 'ونعم', 'يا', 'يجب']
dim: (4, 15)
0 الاتصالات    1 الاجنبية    2 العمالة    3 الله    4 الوكيل    5 توطين    6 جماعة    7 حسبي    8 على    9    قطاع
    10 مسيطرة 11 وظفونا 12 ونعم    13 يا    14 يجب    (0, 11)    0.7444497035180324

  (0, 13)    0.47212002654617047
  (0, 6)     0.47212002654617047
  (1, 14)    0.5552826649411127
  (1, 5)     0.5552826649411127
  (1, 9)     0.43779123108611473
  (1, 0)     0.43779123108611473
  (2, 11)    0.3667390112974172
  (2, 7)     0.4651619335222394
  (2, 3)     0.4651619335222394
  (2, 12)    0.4651619335222394
  (2, 4)     0.4651619335222394
  (3, 9)     0.3443145201184689
  (3, 0)     0.3443145201184689
  (3, 2)     0.43671930987511215
  (3, 1)     0.43671930987511215
  (3, 10)    0.43671930987511215
  (3, 8)     0.43671930987511215
[[0.         0.         0.         0.         0.         0.
  0.47212003 0.         0.         0.         0.         0.7444497
  0.         0.47212003 0.        ]
 [0.43779123 0.         0.         0.         0.         0.55528266
  0.         0.         0.         0.43779123 0.         0.
  0.         0.         0.55528266]
 [0.         0.         0.         0.46516193 0.46516193 0.
  0.         0.46516193 0.         0.         0.         0.36673901
  0.46516193 0.         0.        ]
 [0.34431452 0.43671931 0.43671931 0.         0.         0.
  0.         0.         0.43671931 0.34431452 0.43671931 0.
  0.         0.         0.        ]]
```

The results in Figures 6.9-6.11 represent the outputs of three classifiers: Naïve Bayes Variant (BernoulliNB, MultinomialNB), Support vector machine (SVC, LinearSVC) and SGD.



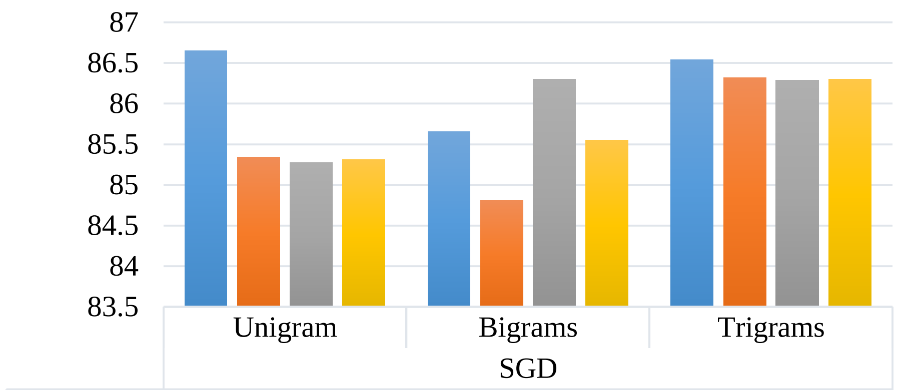
	BernoulliNB			MultinomialNB		
	Unigram	Bigrams	Trigrams	Unigram	Bigrams	Trigrams
Accuracy	86.14	86.97	85.25	85.31	85.54	83.98
Recall	84.99	86.5	83	81.33	85.04	85.35
precision	86.38	86.08	82.74	84.57	83.2	85.99
F-score	85.68	86.29	82.87	82.92	84.11	85.67

Figure 6.9: The results of Machine Learning with TF-IDF features - Naive Bayes classifiers



	SVC			LinearSVC		
	Unigram	Bigrams	Trigrams	Unigram	Bigrams	Trigrams
Accuracy	78.74	80.93	79.45	85.88	86.85	83.19
Recall	72.85	79.19	78.16	85.52	85.27	82.37
precision	73.32	80.33	78.34	84.5	86.54	82.86
F-score	73.08	79.76	78.25	85.01	85.90	82.61

Figure 6.10: The results of Machine Learning with TF-IDF features - SVM classifiers



	Unigram	Bigrams	Trigrams
Accuracy	86.65	85.66	86.54
Recall	85.34	84.81	86.32
Persian	85.28	86.3	86.29
F-score	85.31	85.55	86.30

Figure 6.11: The results of Machine Learning with TF-IDF features – SGD classifiers

6.4.2.2. Machine Learning Techniques with Binary Features Experiment

This section illustrates the experiment for Machine Learning binary features, and presents the associated results. Figure 6.12 and 6.13 show an example of most informative binary (Boolean) feature. Figure 6.14 shows a sample of frequencies features. The included example shows the function of this experimental approach.

d: w0 w1
p(pos|w0) 0.7
p(neg|w0) 0.3

p(pos|w1) 0.4
p(neg|w1) 0.6

p(pos| w0, w1) = 0.61
p(neg| w0, w1) = 0.39

```
Most Informative Features
          has(اهل البلد) = True
      has(التغيير للأفضل) = True
```

Figure 6.12: Example of machine learning classifier with binary features

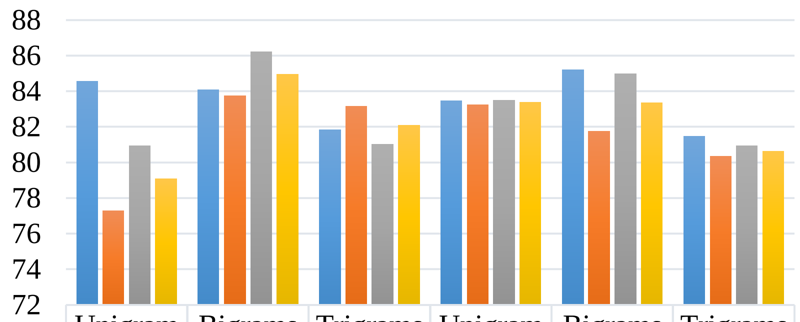
```
Most Informative Features

        has(❤) = True          pos : neg    =    15.6 : 1.0
     has(للأفضل) = True          pos : neg    =    12.6 : 1.0
      has(الصبر) = True          pos : neg    =    12.1 : 1.0
        has(فرج) = True          pos : neg    =    12.1 : 1.0
      has(مكرمه) = True          pos : neg    =    12.1 : 1.0
     has(الطيبه) = True          pos : neg    =    12.1 : 1.0
     has(الشروط) = True          neg : pos    =    10.1 : 1.0
      has(مفتاح) = True          pos : neg    =     9.7 : 1.0
        has(👌) = True          pos : neg    =     9.7 : 1.0
      has(توصيه) = True          pos : neg    =     9.7 : 1.0
       has(نقدر) = True          pos : neg    =     9.7 : 1.0
        has(♡) = True          pos : neg    =     9.7 : 1.0
```

Figure 6.13: Screenshot about a sample of most informative features for binary

sample frequencies
[('العماله الاجنبيه', 835), ('الكفاءات الوطنيه', 634), ('القطاع الخاص', 925), ('امتداد التخصص', 641)]

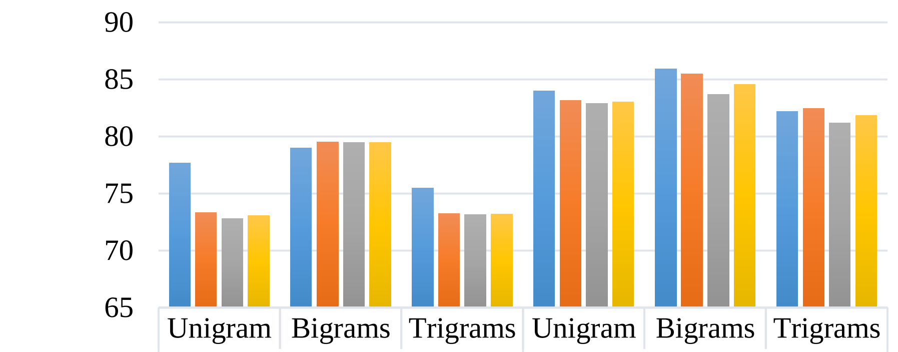
Figure 6.14: Sample of frequencies features

The results in Figures 6.15-6.17 reflect three classifiers: Naïve Bayes Variant (BernoulliNB, MultinomialNB), Support vector machine (SVC, LinearSVC) and SGD.



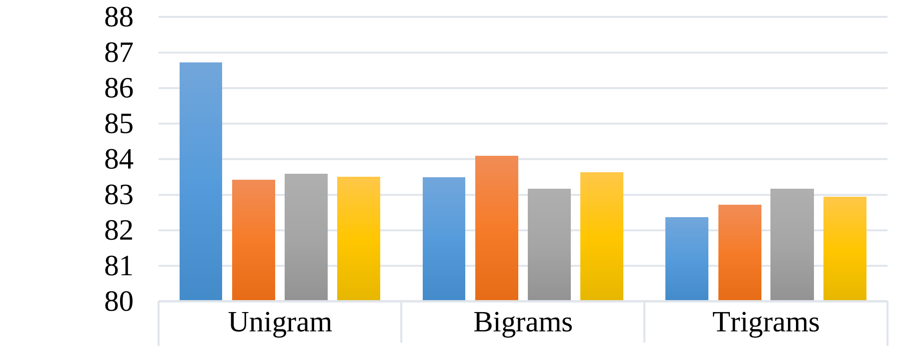
	BernoulliNB			MultinomialNB		
	Unigram	Bigrams	Trigrams	Unigram	Bigrams	Trigrams
Accuracy	84.56	84.08	81.85	83.47	85.21	81.47
Recall	77.29	83.75	83.16	83.24	81.76	80.36
precision	80.93	86.21	81.01	83.5	85	80.92
F-score	79.07	84.96	82.07	83.37	83.35	80.64

Figure 6.15: The results of Machine Learning with binary features - Naive Bayes classifiers



	Unigram	Bigrams	Trigrams	Unigram	Bigrams	Trigrams
	SVC			LinearSVC		
Accuracy	77.7	79.03	75.5	84	85.94	82.2
Recall	73.34	79.53	73.27	83.14	85.51	82.48
precision	72.8	79.46	73.15	82.92	83.69	81.22
F-score	73.07	79.49	73.21	83.03	84.59	81.85

Figure 6.16: The results of Machine Learning with binary features – SVM classifiers



	Unigram	Bigrams	Trigrams
		SGD	
Accuracy	86.71	83.49	82.37
Recall	83.42	84.09	82.72
Precision	83.58	83.17	83.16
F-score	83.50	83.63	82.94

Figure 6.17: The results of Machine Learning with binary features – SGD classifiers

6.4.3. Results and Discussion of Machine Learning Techniques with Weighted Schemes Features

It is clear from the results that the accuracy of machine learning techniques with TF-IDF features is better than machine learning techniques with binary features. The best result achieved was by the BernoulliNB classifier with TF-IDF for bigrams; the accuracy was 86.97% and the F-score was 86.29%. This is higher than the BernoulliNB classifier with binary for bigrams, which achieved an accuracy score of

84.08%. Also, this result is higher by around 5% of baseline results, which is a significant improvement over prior models. This improved outcome was due to the pre-processing stage of the tweets and to the negation consideration, which has been previously proven in relation to MSA experimentation. These results, thereby extend such research, confirming that the pre-processing stage has improved the classifier's performance in dialectal Arabic. Figure 6.18 compares the results of BernoulliNB classifier with Baseline, TF-IDF and Binary features.

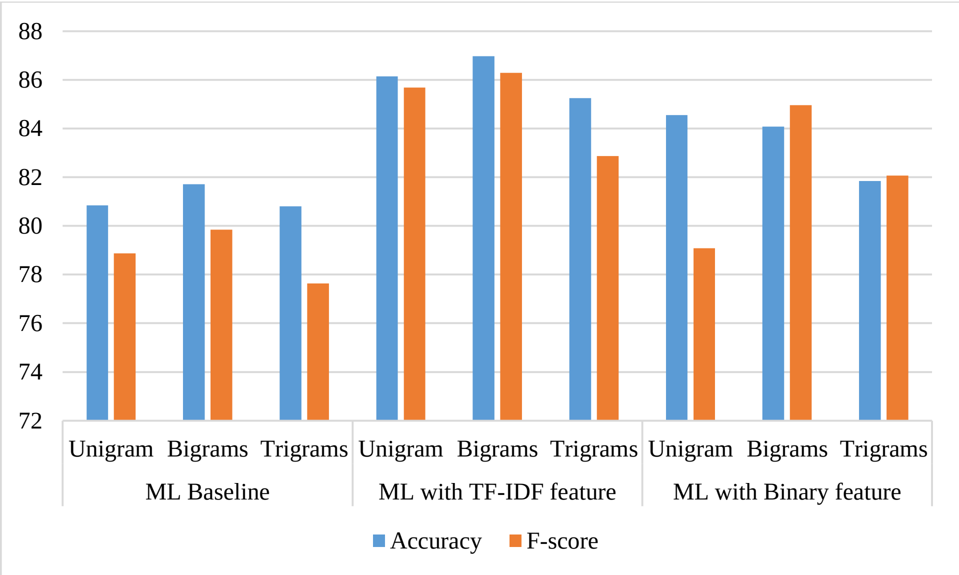


Figure 6.18: Comparison of the results of BernoulliNB classifier with Baseline, TF-IDF and Binary features

From these findings, the highest performing outcome with binary features for unigram was achieved by the SGD classifier with 86.71% of accuracy; the F-score was 83.50% and around 83% for recall and precision. At the same time, the SGD classifier with TF-IDF for trigram achieved good results; the classification accuracy score was 86.54%. The SGD classifier provides good results with most N-Grams and with both weighting schemes. Moreover, the KNeighbors classifier also demonstrates good performance with all N-Grams features and with both weighted schemes; the accuracy and F-score of machine learning with TF-IDF of unigram was 84.58% and the accuracy and F-score of machine learning with binary of bigrams were 82.29 and 83.32%, respectively. These findings yield positive results that suggest that these solutions can process large datasets and yield computationally efficient results.

The findings from these experiments also illustrate that tree classifiers were subject to the lowest performing results within all experimental models that integrated machine learning with TF-IDF and binary. The lowest accuracy achieved was by the RandomForest classifier with trigrams, with an accuracy and f-score reported at 73.12%. The most common problem observed in relation to the DT classifier is its inability to incorporate variations in data, including noise, when trees increased and deepened. This inadequacy is commonly termed as overfitting. Additionally, the structure of the tree would inevitably be altered due to the addition of data (see Appendix B for KNeighbors and trees classifiers results). LinearSVC results show a good performance of this classifier with all evaluation measures. It shows the highest precision for machine learning with TF-IDF of bigrams feature, which was 86.54%. On the other hand, the best recall is by the BernoulliNB classifier, which reached a score of 86.50% with bigrams, as well as for machine learning with TF-IDF. Based upon these findings, it can be concluded that Naïve Bayes (BernoulliNB), Support vector machine (LinearSVC) and SGD classifiers provide the best results for sentiment analysis of machine learning approach with TF-IDF of dialectal Arabic.

6.5. Comparing The Machine Learning Approach for Sentiment Analysis of Saudi Tweets Against Similar Works

This section compares the experimental results captured during this multi-stage process with the results of two other works related to machine learning approaches for Saudi dialects. The current study has adopted a machine learning approach for sentiment analysis in tweets related to a specific domain, and the tweets were a mixture of MSA and Saudi dialects. Due to the lack of a specific domain, this comparative assessment determines whether these findings could be transferrable to other specific domain problems or experimental conditions. These additional experiments were used to study how the machine learning approach could perform on two different corpora that were used in two other works related to the analysis of machine learning approaches for tweets in Saudi dialects.

The first dataset that experimented with was collected by Adayel and Azmi (2015). In their study, they selected hashtags discussing different social issues in Saudi

Arabia (multi-domain), such as #الراتب_مايكفي_الحاجة (our salary is not sufficient), #قيادة_26_اكتوبر and #المحتسبون_للديوان_مجدد (these tweet regards women driving). Their dataset contained 1103 Arabic annotated tweets and was provided for research purposes. They developed a sentiment analysis system that was designed to identify the polarity of the tweets using two classifications (positive or negative). For the machine learning classifier, they used Support Vector Machine (SVM). All three N-Grams models (unigrams, bigrams and trigrams) were applied on the annotated tweets and they used TF-IDF weighting scheme on the N-Grams and all the features that have frequencies greater than a certain threshold were selected. Their results were a score of 78% for the accuracy, 79% for precision and around 77% for recall and F-score.

The second dataset was collected by Al-Twairesh et al. (2017). Their study was published in the Procedia Computer Science. Initially, they collected around 6.3 million Arabic tweets in three months, and applied cleaning and pre-processing to the output, resulting in a remainder of 2.2 million tweets. Then, they decreased the number of tweets to 13,226 and add 6,090 newly collected tweets. Also, they collected 2090 tweets from three trending Saudi hashtags in 2016. The resulting number of tweets after cleaning was 1,580. A total of 14,806 tweets were manually annotated by the recruited annotators. The AraSenti-Tweet corpus is publicly available[15] . The dataset is divided into a training set and test set. Al-Twairesh et al. (2017) conducted several experiments for multi-way sentiment classification. In the current study, only the two-way classification (positive and negative tweets) was considered). For classification, they used SVM with a linear kernel, and, for the term feature, they tested the TF-IDF features. They only show the F-score in their work, which is 60.05%.

6.5.1. Evaluating the Machine Learning Approach (Experiment 1)

The results of the comparative machine learning experiment are shown in Table 6.2 (The best result shown in green colour and the worst result in red colour). These findings applied the Adayel and Azmi (2015) corpus to the range of classifier

[15] https://github.com/nora-twairesh/AraSenti/tree/AraSenti-Tweet-Corpus

schemes developed for this study, extrapolating accuracy and performance across three levels of N-Grams.

Table 6.2: The results of applying machine learning approach with Adayel and Azmi corpus

Classifier	N-grams	Accuracy	Precision	Recall	F-score
BernoulliNB	Unigram	72.33%	72.05%	72.71%	72.38%
	Bigrams	75.86%	74.21%	71.50%	72.83%
	Trigrams	71.41%	75.85%	72.74%	74.26%
SVC	Unigram	72.22%	69.87%	71.34%	70.60%
	Bigrams	75.43%	74.55%	73.83%	74.19%
	Trigrams	68.50%	67.74%	66.77%	67.25%
RandomForest	Unigram	55.29%	59.57%	55.00%	57.19%
	Bigrams	58.86%	63.56%	58.84%	61.11%
	Trigrams	58.03%	64.83%	58.15%	61.31%
LinearSVC	Unigram	69.07%	69.31%	69.57%	69.44%
	Bigrams	69.46%	69.63%	69.19%	69.41%
	Trigrams	69.88%	69.20%	69.88%	69.54%
MultinomialNB	Unigram	75.53%	73.79%	73.06%	73.42%
	Bigrams	79.84%	77.98%	76.41%	77.19%
	Trigrams	73.19%	73.22%	73.54%	73.38%
KNeighbors	Unigram	56.66%	61.45%	56.69%	58.97%
	Bigrams	60.21%	62.00%	60.40%	61.19%
	Trigrams	56.74%	57.63%	56.18%	56.90%
SGD	Unigram	71.91%	71.54%	71.43%	71.48%
	Bigrams	68.89%	68.07%	68.73%	68.40%
	Trigrams	69.43%	69.43%	69.43%	69.43%
DecisionTree	Unigram	50.36%	39.18%	50.53%	44.14%
	Bigrams	50.97%	40.50%	50.85%	45.09%
	Trigrams	50.23%	39.18%	50.39%	44.08%

6.5.2. Evaluating The Machine Learning Approach (Experiment 2)

The results of the comparative machine learning experiment are shown in Table 6.3 (The best result shown in green colour and the worst result in red colour). These findings applied the Al-Twairesh et al. (2017) corpus to the range of classifier schemes developed for this study, extrapolating accuracy and performance across three levels of N-Grams

Table 6.3: The results of applying the machine learning approach with Al-Twairesh et al corpus

Classifier	N-grams	Accuracy	Precision	Recall	F-score
BernoulliNB	Unigram	67.23%	64.86%	64.91%	64.88%
	Bigrams	67.08%	64.42%	64.93%	64.67%
	Trigrams	65.56%	64.58%	64.58%	64.58%
SVC	Unigram	61.25%	73.15%	58.92%	65.27%
	Bigrams	61.33%	73.24%	58.60%	65.11%
	Trigrams	59.32%	73.77%	55.45%	63.31%
KNeighbors	Unigram	53.93%	49.87%	50.08%	49.97%
	Bigrams	54.14%	58.25%	51.82%	54.85%
	Trigrams	54.90%	58.45%	51.63%	54.83%
LinearSVC	Unigram	64.05%	64.13%	63.13%	63.63%
	Bigrams	66.28%	64.41%	63.52%	63.96%
	Trigrams	62.66%	63.07%	63.44%	63.25%
MultinomialNB	Unigram	65.86%	65.58%	63.82%	64.69%
	Bigrams	64.57%	66.14%	63.42%	64.75%
	Trigrams	61.34%	65.72%	63.92%	64.81%
RandomForest	Unigram	47.50%	23.75%	50.48%	32.30%
	Bigrams	47.19%	23.82%	50.79%	32.43%
	Trigrams	47.04%	23.85%	50.17%	32.33%
SGD	Unigram	66.37%	66.86%	66.60%	66.73%
	Bigrams	69.74%	66.64%	66.22%	66.43%
	Trigrams	63.87%	65.35%	64.16%	64.75%
DecisionTree	Unigram	47.62%	23.87%	50.32%	32.38%
	Bigrams	47.81%	23.61%	50.49%	32.17%
	Trigrams	47.63%	23.55%	50.09%	32.04%

The evidence captured from these two experimental outputs reveals that the machine learning approach attained positive, high-performing results with both of these distinct, dialectical datasets (See Figure 6.19). Overall, the Adayel and Azmi corpus shows better performance than the Al-Twairesh et al. corpus. The MultinomialNB classifier achieved the best result for bigrams features in the Adayel and Azmi corpus; the accuracy was 79.84%, which is around their original result of 78%. The accuracy in the proposed approach is around 2% higher than their findings yielded. On the other hand, the best result achieved with the Al-Twairesh et al. corpus was the SGD classifier, which had 69.74% accuracy and 66.43% F-score. The F-score in the proposed approach is 5% higher than their findings yielded. Indicative of a transferrable finding from this study to future modelling, the evidence confirmed the poor performance of trees with both corpuses. The accuracy of DecisionTree with the trigrams of the Adayel and Azmi corpus was 50.23%, while the accuracy of RandomForest with trigrams of the Al-Twairesh et al. corpus was 47.04%. Even though the proposed approach to machine learning analysis is domain knowledge informed, it is applicable with any domain specific problem, indicating that it could be re-adapted to other domains in future studies.

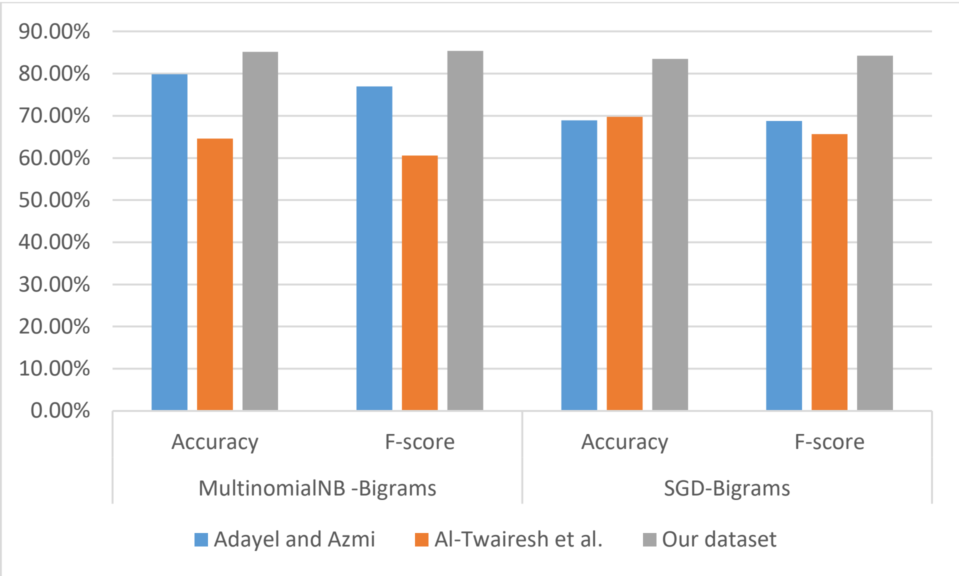


Figure 6.19: The results of applying our approach with two other domain specific datasets

To summarise our experiences, in these study we experiment two approaches for sentiment analysis and we run our system in different datasets as shown in table 6.4.

Table 6.4:Result of Comparative Analysis of Lexicon-Based Approach and ML Variations

		Aldayel and Azmi (2016)		Al-Twairesh et al (2017)		Our work	
		Accuracy	F-score	Accuracy	F-score	Accuracy	F-score
Lexicon		78.22 %	76.24 %	78.61 %	62.22 %	89.80 %	86.32 %

ML With TF-IDF weighted schemes features	Multinomial NB	Bigram	79.84 %	77.19 %	64.57 %	64.75 %	85.54 %	84.11 %
	SGD	Bigrams	68.89 %	68.40 %	69.74 %	66.43 %	85.66 %	85.22 %
	DecisionTree	Trigrams	50.23 %	44.08 %	47.63 %	32.97 %	68.54 %	68.55 %
	RandomForest	Trigrams	58.03 %	61.31 %	47.04 %	32.33 %	68.72 %	68.34 %

The first approach is the Lexicon-Based approach. The experimental results presented in Table 7.4 show that the lexicon-based approach outperforms the prior models developed in this field. For example, there was an improvement in accuracy of around 10% over the Adayel and Azmi corpus. As the accuracy is not considered in Al-Twairesh et al, however, the results indicate an improvement in the f-score of around 2% over the prior outputs of that study. This performance improvement is attributed to the more comprehensive coverage of the factors that impact lexical analysis including the use of intensifiers, negations, supplication, proverbs and interjections as well as the comprehensive multi-intensity sentiment lexicon for Saudi dialects.

The second approach is the ML Approach, in addition to the lexicon performance improvements, Table 6.4 also demonstrates improved performance within the range of ML solutions for the current study when compared with the two prior experiments. From the corpus-based proposition, the Adayel and Azmi corpus indicates superior performance over the Al-Twairesh et al. corpus. The MultinomialNB classifier achieved the best result for bigrams features in the Adayel and Azmi corpus; the accuracy was 79.84%, which is around their result of 78%. The accuracy in the current approach was around 2% higher than the prior experimental results in Adayel and Azmi (2016). The best results were achieved with the Al-Twairesh et al. corpus by applying the SGD classifier, which had 69.74% accuracy and 66.43% F-score. The F-score in the current approach is 5% higher than the performance in the Al-Twairesh et al. (2017) experiment. Even though the current approach was domain knowledge informed, it is applicable with any domain specific problem features.

After find out the performance of two different sentiment analysis approaches, we will develop a hybrid lexicon based-machine learning approach for sentiment analysis of social media content in dialectical Arabic, this hybrid approach will benefit from the advantages of both approaches.

6.6. Chapter Summary

This chapter has introduced and analysed a machine learning approach for sentiment analysis of social media content in dialectical Arabic. This approach has been used to investigate the sentiment analysis performance of the proposed model with the Arabic language, and in particular the dialectical Arabic written on Twitter. To solve the problem related to a lack of specific grammar in the feature selection stage, this experimental approach has focused on N-Grams features (unigrams, bigrams and trigram). The experimental results indicated positive outpcomes which favoured the BernoulliNB classifier with TF-IDF for bigrams that suggested a 5% improvement through pre-processing and N-Gram weighting. Finally, the effectiveness of the proposed approach was evaluated by applying the model to two other researcher corpuses, resulting in positive, higher performing accuracy results across both of the experiments. Even though the proposed approach is domain knowledge informed, these findings confirm that it does yield adequate results for any domain specific problem features.

Chapter 7

7 Linguistic-Machine Learning Hybrid Approach for Dialectical Arabic

7.1. Introduction

A central objective of the hybrid approach in sentiment analysis is to establish a more content-rich analysis for classifiers and to carry out sentiment analysis to a more precise degree, with the objective of attaining accurate results for sentiment classification. This study has explored a hybrid approach for sentiment analysis of dialectical Arabic tweets by adopting two different, but interrelated methods. The first method involved a hybrid lexicon-based machine learning approach. The lexicon-based approach considers central linguistic features and ensures the transparency of the classification criteria and useful treatment of the syntax. Machine-based sentiment analysis calculates the polarity values through statistical estimation and enables the creation and adaptation of the trained data set. The second method was a hybrid semantic knowledgebase machine learning approach. This approach adopted a semantic knowledgebase approach to analyse a collection of tweets at the domain feature level and produce a set of structured information that associates the expressed sentiments with domain specific features.

The motivation of developing a hybrid approach is to combine two different methodologies or systems to create a new and better model. According to Gupta et al. (2019) the hybrid approach of sentiment analysis exploits both statistical methods and knowledge-based methods for polarity detection. It inherits high accuracy from the machine learning (statistical methods) and stability from the lexicon-based approach.

7.2. Proposed Hybrid Approach

Drawing upon the prior methods and models, the current study has proposed a linguistic-machine learning hybrid approach for sentiment analysis of social media content in dialectical Arabic. Two possible methods have been explored, combining linguistic and machine learning approaches for sentiment analysis based upon domain knowledge. The first method adopts domain-specific features and sentiment lexicon modelled in the semantic ontology as the token training features for the opinion classification. In the second method, the sentiment score resulting from the lexicon-based sentiment analysis was included in the training feature-set. The utilisation of domain knowledge in a second method was implicit as it was used to associate the sentiment mention with the domain features. Finally, the performance of these features has been assessed and recommendations regarding the best performing solution have been made.

7.2.1. Hybrid Semantic Knowledgebase-Machine Learning Approach

In spite of a growing compendium of research in this field, there are limited studies that have undertaken to combine machine learning with semantic features. The proposed methodology combines the advantages of both approaches: machine learning approach and ontologies and semantic knowledge to enhance the performance of sentiment analysis.

The technique developed employed three N-Grams features and the TF-IDF weighting scheme, and added semantic features from the ontology. For example, the semantic domain features are: (unemployment-البطاله – *albitaluh /* Saudah program-السعوده – *Saudah /* foreign labour- العماله الاجنبيه - *aleamalat al'ajnabia)*. Each tweet was inspected and the semantic features were detected in order to extract the corresponding features for this tweet and map these features into a higher concept. The domain ontology is used to map the tweets as demonstrated in the following examples:

Positive Tweet:

"برنامج قوى مفيد مره اشكر الوزير على اهتمامه بنا كعاطلين"

"*barnamaj quaa mufid marih 'ashkura alwazir ealaa aihtimamah bina kaeatilin"*

Translate: "Quaa program is very useful, I thank the minister for his interest in us as unemployed"

Mapping by domain ontology: "'National Program' is 'Intensifiers' 'Positive Sentiment', I 'Positive Sentiment' the 'Decision Makers' for his 'Positive Sentiment' in us as 'Domain Feature'"

Negative Tweet:

" السعودة المزيفة هي سبب الفشل في الحد من البطالة ... لقد أصبحنا عبئًا على مجتمعنا .. لقد تولى المصريون والهنود وظائفنا"

"*alsueudat almuziafat hi sbb alfashal fi alhadi min albitala ... laqad 'asbahna ebyana ealaa mujtamaeina .. laqad tawalaa almisriuwn walhunud wazayifna"*

Translate: "Fake Saudization is the cause of failure to reduce unemployment ... we become burden on our society .. Egyptian and Indians took our jobs"

Mapping by domain ontology: "'Negative Sentiment' 'National Program' is the cause of 'Negative Sentiment' to reduce 'Domain Feature' ... we become 'Negative Sentiment' on 'Domain Feature' .. 'Foreign Labor' and 'Foreign Labor' took 'Domain Feature'"

Different machine learning algorithms were tested with these features and the pseudocode of combining features in hybrid semantic knowledgebase-machine learning approach has been presented. Figure 7.1 shows the framework of the hybrid semantic knowledgebase-machine learning approach.

Pseudocode: Combining Semantic Features with Text Features

1. Inputs: Load dataset
2. Output: Class label of sentiment classification
3. FOR every tweet DO
4. Extract semantic features
5. Extract n-gram features
6. Combine semantics features with n-gram features
7. Train the machine learning algorithm on the hybrid features
8. Use the trained model to get the class label for test data

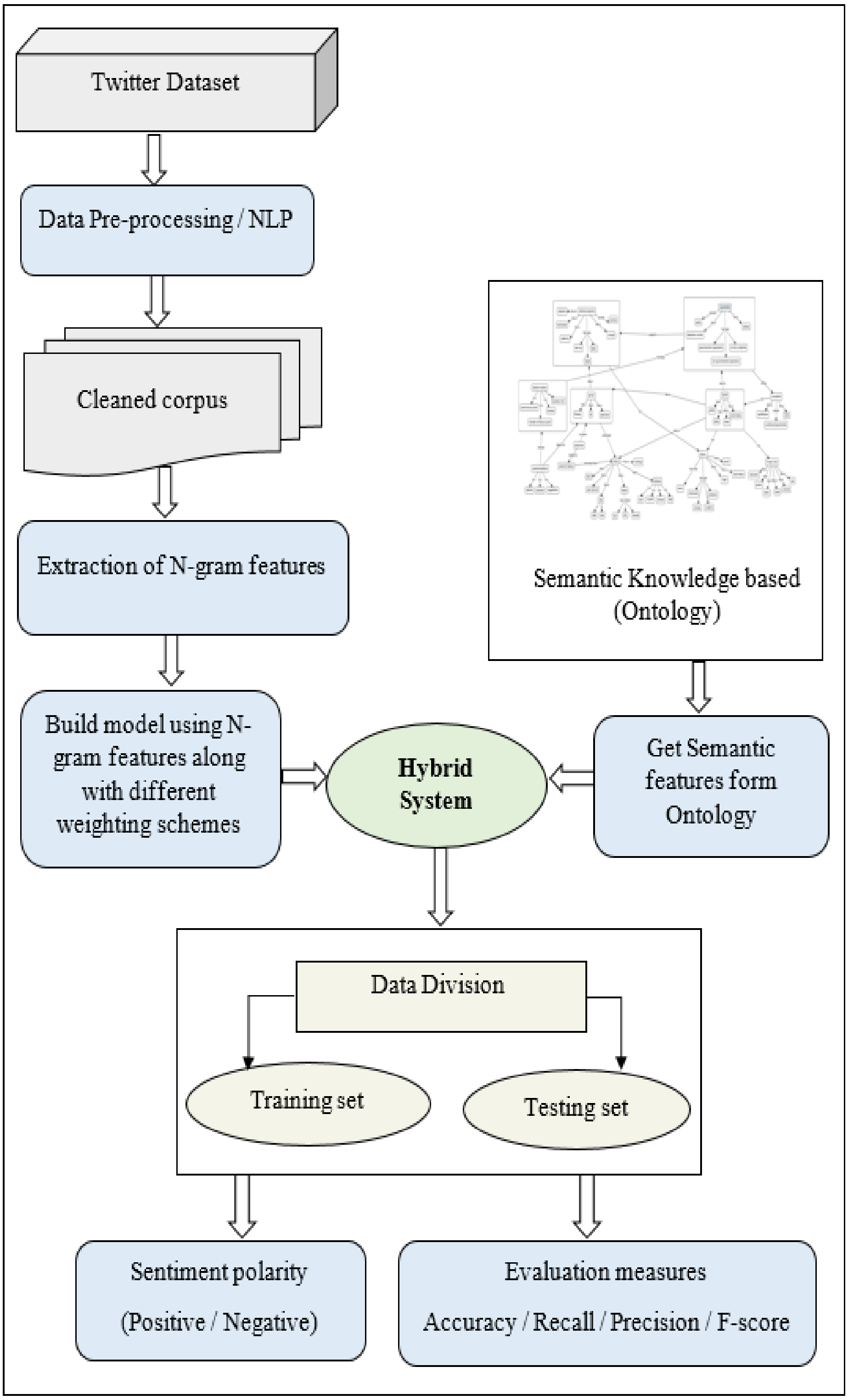


Figure 7.1: Frame work of the hybrid approach algorithm of Incorporating TF-IDF features with semantic features (domain features)

7.2.2. Hybrid Lexicon Based-Machine Learning Approach

In this hybrid approach, the output of the linguistic sentiment analysis (the lexicon-based approach) was used to enhance the training of the machine learning approach. Using three N-Grams features and the TF-IDF weighting scheme, a sentiment score was then extracted from the lexicon and added to the construct. Each tweet and its lexicon was inspected to extract the corresponding total sentiment score for each individual tweet. This final, resultant score was added into the machine learning approach and considered several underlying factors including feature-sentiment association, light stemming, emojis, intensifiers, negations and special phrases, such as supplications, proverbs and interjections. The final scores were then aggregated with 3-Gram features with the TF-IDF weighting scheme, and the machine learning classifier was then trained to classify tweets into positive or negative. Several different machine learning algorithms for these features were investigated and the pseudocode of combining sentiment score features with TF-IDF features has been presented. Figure 7.2 considers the framework of the hybrid approach algorithm of Incorporating sentiment score features.

Pseudocode: Combining Sentiment Score Features with TF-IDF Features

1. Inputs: Load dataset
2. Output: Class label of sentiment classification
3. FOR every tweet DO
4. Get the sentiment score using the lexicon-based method
5. Extract n-gram features
6. Combine n-gram features with the sentiment score
7. Train the machine learning algorithm on the hybrid features
8. Use the trained model to get the class label for test data

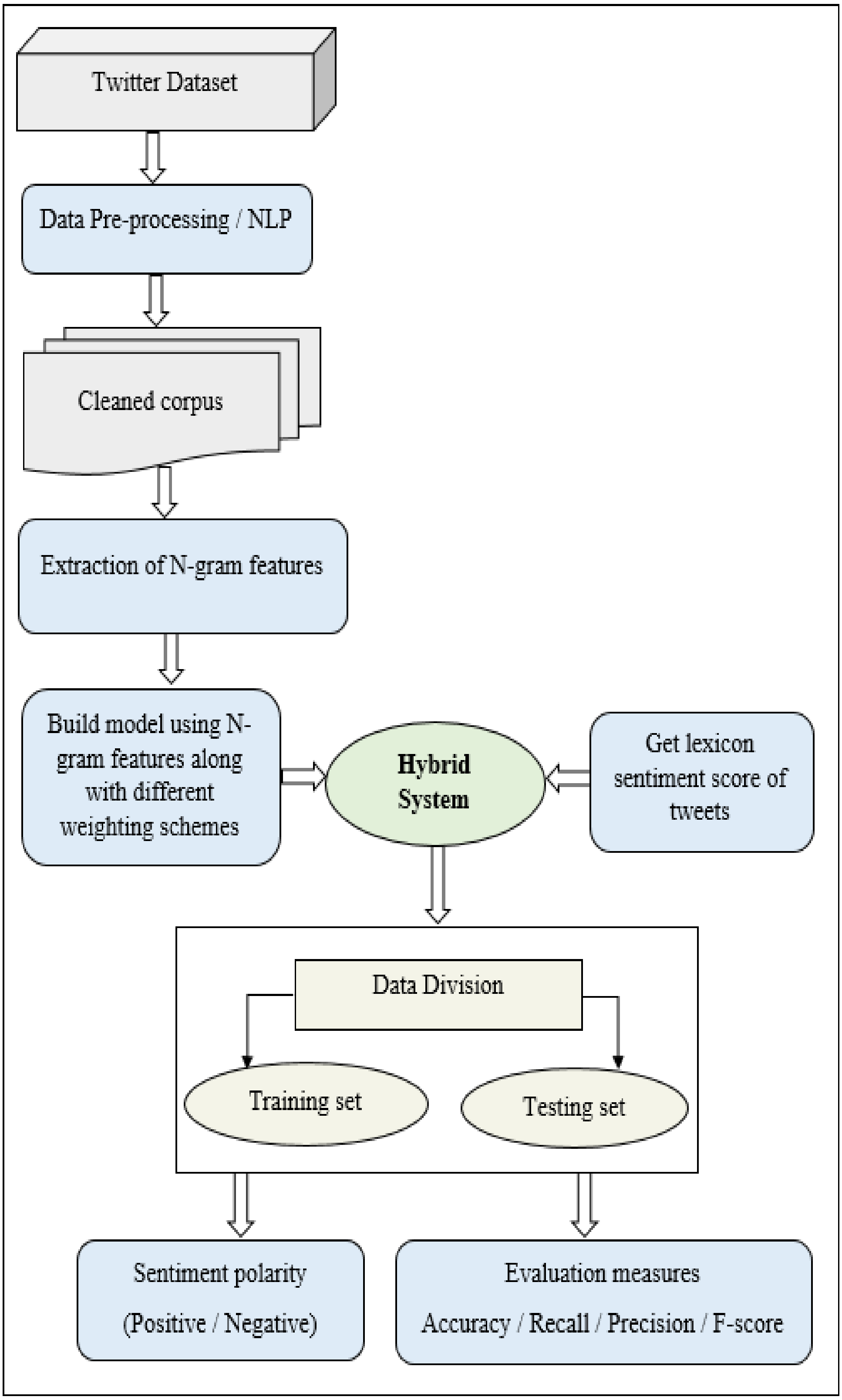


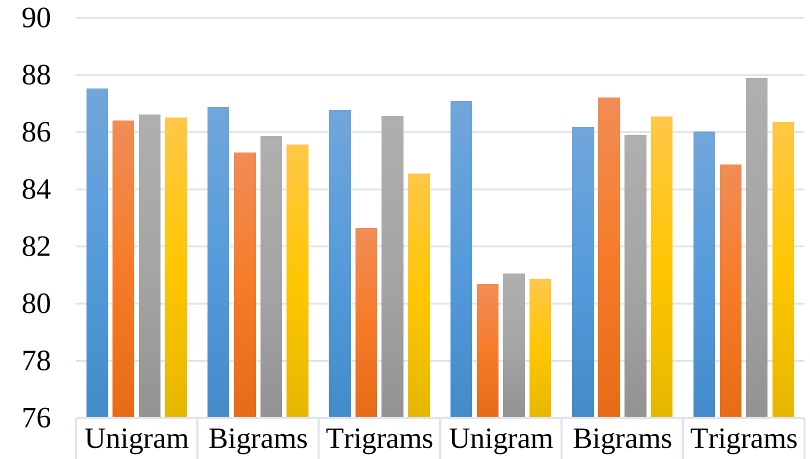
Figure 7.2: Frame work of the hybrid approach algorithm of Incorporating TF-IDF features with sentiment score features

7.3. Results and Discussion of the Proposed Hybrid Approach

This section presents the results of the linguistic-machine learning hybrid approach for sentiment analysis of social media content in dialectical Arabic.

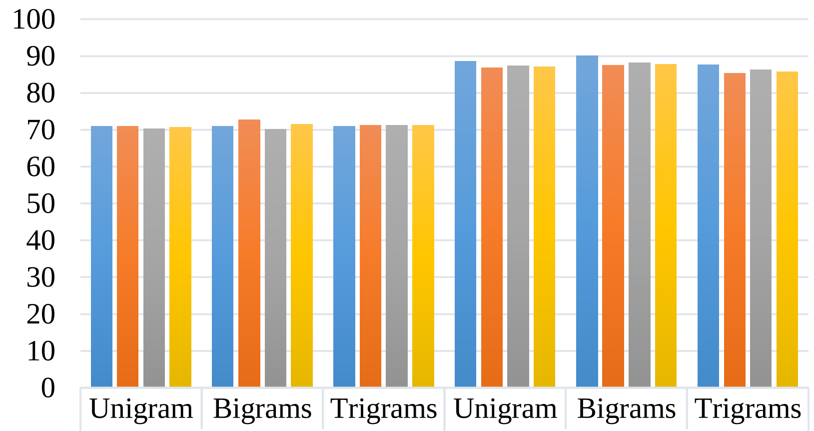
7.3.1. Hybrid Semantic Knowledgebase-Machine Learning Approach

The result of the hybrid approach of the machine learning (TF-IDF features) and lexicon-based approaches (Semantic Knowledgebase) is illustrated in Figures 7.3, 7.4 and 7.5. It is clear from the results that the LinearSVC classifier with bigrams shows the best performance, with 90.07% accuracy and an 87.82% F-score. This classifier shows a good result with unigram and trigrams; the accuracy was 88.59% and 87.72%, respectively. Another interesting observation from Figure 7.3 is that the NB Variant (BernoulliNB and MultinomialNB) shows a good performance with accuracy between 88% and 86%. However, the trees classifiers (DecisionTree and RandomForest) show the worst results across all the evaluation measures. The accuracy of RandomForest with all n-grams features was 70.77%. According to recall, the best result was for the KNeighbors classifier with trigrams, which was 87.61%; another observation is that this result is higher by around 8.50% than bigrams with same classifier. The SGD classifier shows good results, which was around 86% in all the evaluation measures (see Appendix C for KNeighbors and SGD classifiers results). These results confirm a 4% improvement from the semantic knowledge included from the ontology to enhance the accuracy of the hybrid approach.



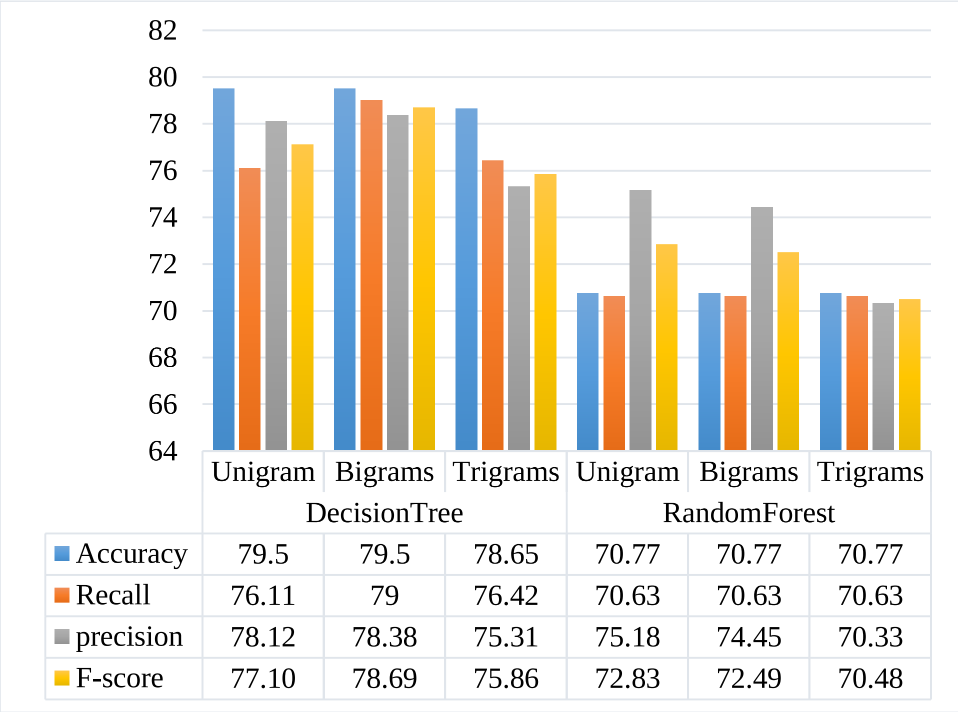
	BernoulliNB Unigram	BernoulliNB Bigrams	BernoulliNB Trigrams	MultinomialNB Unigram	MultinomialNB Bigrams	MultinomialNB Trigrams
Accuracy	87.51	86.87	86.77	87.09	86.17	86.01
Recall	86.4	85.27	82.63	80.67	87.21	84.86
precision	86.62	85.86	86.56	81.04	85.89	87.9
F-score	86.51	85.56	84.55	80.85	86.54	86.35

Figure 7.3: The results of hybrid approach (TF-IDF features + semantic features) - NB classifiers



	SVC Unigram	SVC Bigrams	SVC Trigrams	LinearSVC Unigram	LinearSVC Bigrams	LinearSVC Trigrams
Accuracy	70.98	70.9	70.9	88.59	90.07	87.72
Recall	71.02	72.77	71.21	86.87	87.5	85.38
precision	70.33	70.14	71.31	87.33	88.14	86.21
F-score	70.67	71.43	71.26	87.10	87.82	85.79

Figure 7.4: The results of hybrid approach (TF-IDF features + semantic features) - SVM classifiers



	DecisionTree			RandomForest		
	Unigram	Bigrams	Trigrams	Unigram	Bigrams	Trigrams
Accuracy	79.5	79.5	78.65	70.77	70.77	70.77
Recall	76.11	79	76.42	70.63	70.63	70.63
precision	78.12	78.38	75.31	75.18	74.45	70.33
F-score	77.10	78.69	75.86	72.83	72.49	70.48

Figure 7.5: The results of hybrid approach (TF-IDF features + semantic features) - Trees classifiers

7.3.2. Hybrid Lexicon Based-Machine Learning Approach

The results of the hybrid lexicon based-machine learning approach are illustrated in Table 7.1 and Figure 7.6. Three experiments that combined the three highest results of the machine learning approach with the highest result of the multi-factor Lexicon-based approach were conducted and included all enhancement techniques (light stemming, polarity, negation, emojis and intensification words). Based upon these findings, this particular method achieved the highest recall (88.34%) and the best precision (89.55%). The accuracy of the highest performing method, the linear SVC classifier of bigrams was 93.45% and the F-score was 89.55%. Across all methods, improved accuracy was reported, such as the accuracy of the BernoulliNB and SGD of bigrams feature with Lexicon method (All levels), which was 91.85% and 92.33%, respectively. Based upon these findings, it is confirmed that the hybrid approach of machine learning (TF-IDF features) and the lexicon-based approaches provide the best results and confirms the originating hypothesis about the improvement that the hybrid approach brings to sentiment analysis

It is clear that this hybrid approach between the machine learning approach and the lexicon-based approach (the final lexicon sentiment score of each tweet as feature) provided a significant improvement in accuracy compared with the result of the lexicon-based approach alone which is around 4%. This improvement confirmed that there is a benefit from the final score from the sentiment lexicon to enhance the hybrid approach. In general, the core observation from these experiments is that the three machine learning classifiers performance is quite similar to what is achieved via the machine learning approach, which provided good performance. What makes the difference is the final lexicon sentiment score of each tweet's features, which was added to the classifiers. This improvement was due to the multi-factor lexicon-based sentiment analysis, in which the tweet went through several stages, such as pre-processed, NLP and lexicon feature extraction process, until the final score was issued. It was processed with deep linguistic analysis of social media content in dialectal Arabic. This confirms the hypothesis that the hybrid approach improves the sentiment classification to a more precise level and attains accurate results.

Table 7.1: The results of hybrid approach (TF-IDF features + lexicon sentiment score feature)

Method	Accuracy	Recall	Precision	F- score
linear SVC + bigrams feature + Lexicon method (All levels)	93.45%	88.34%	90.80%	89.55%
BernoulliNB + bigrams feature + Lexicon method (All levels)	91.85%	83.11%	89.97%	86.40%
SGD + Unigram feature + Lexicon method (All levels)	92.33%	87.50%	90.00%	88.73%

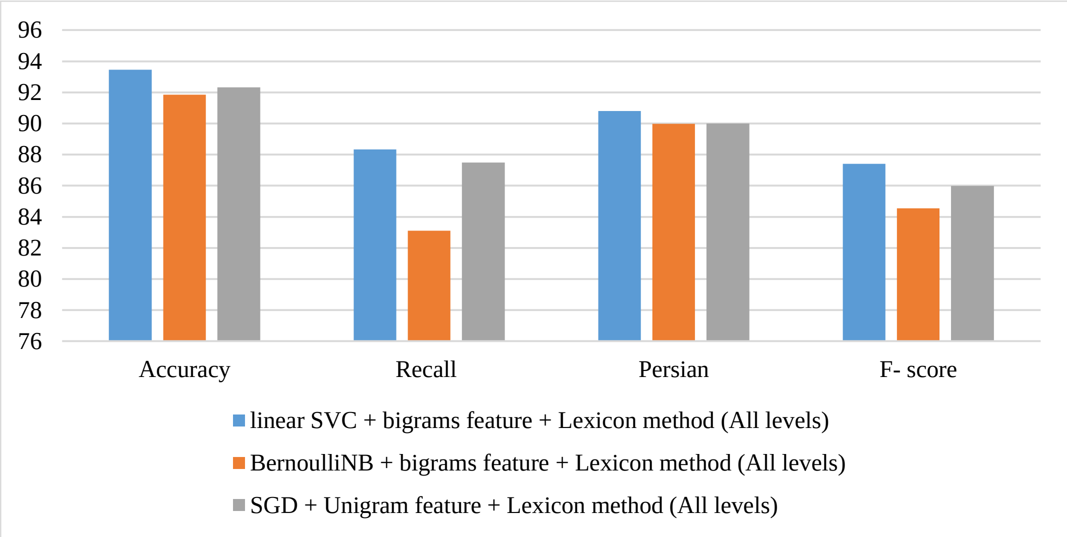


Figure 7.6: The results of hybrid approach (TF-IDF features + lexicon sentiment score feature)

7.4. The Usability of Sentiment analysis to Aid Government and Decision Makers

Using opinions expressed through social media can be seen as inexpensive and may be more accurate than surveys, since the latter may not reflect an honest opinion. Assessing and exploring social networks allows researchers to interpret the the generalised expressions of users, reflecting their immediate and personal reactions to an event or topic. Sentiment analysis (SA) of user opinions is vital for analysing matters concerning the provision of public services, government policy and emerging political policy. An overall aim of this study was to support policy makers in their decisions, with the ultimate objective to improve and enhance the day-to-day life of the local and national community. more effective SA models, government policy makers will have the opportunity to explore, identify, arrange and evaluate public opinion and reaction to policy decisions in both a qualitative and quantitative manner, with sentiment analysis highlighting the positive and negative reactions.

Generally, this study has determined that most of the information expressed in social networks can help to predict future strategies regarding political and social issues. The proposed novel dataset regarding opinions and issues of concern will inevitably expand on a massive scale, as millions of users regularly express their views through social networks. As a vehicle of purposive communication, Twitter is now a prime source of information and can be exploited by sentiment analysis for decision makers. Sentiment analysis, from this perspective, allows for improved decision-making due to the evaluation of opinions expressed and products and services reviewed. Additionally, it explores and categorises all collated data available and translates it into reliable and useable information, allowing decision makers to make informed choices.

Sentiment analysis has the ability to facilitate a close bond between public decision makers and the general public. From this position, these social comments are capable of communicating strong opinions, whilst SA functionality helps policy makers to implement, or avoid, decisions and ineffective strategy. Sentiment analysis conducted throughout social networks can be exploited by public bodies to learn about and satisfy user demands/expectations. It also enhances the confidence of the general public, since their opinions are being heard and acted upon to influence policy, change existing service delivery and update existing services. Based on previously published analysis (Corallo et al., 2015), and the findings collected over the course of this research, it is evident that utilising sentiment scores (searching for both positive and negative aspects) shapes and influences decision-making. There are, however, significant obstacles that have been identified in using sentiment analysis for decision-making in light of the complexity of the content and usage of tweets. One vital issue, is that opinionated lexicon may be considered positive in one scenario, but negative in another. A further challenge discovered over this research is that opinions are not always expressed in the same manner, taking into consideration the differences in style between texts and tweets. Individual opinions can be contradictory, and sentiment analysis exploration regarding Twitter data and alternative micro-blogs face challenges regarding interpretation, primarily due to the limited length of the entry and the irregular structure of the expression.

In this sentiment analysis exploration study, alternative techniques and approaches may enhance the decision-making procedure regarding a particular controversial social issue, such as unemployment issues in Saudi Arabia and the evaluation of positive and negative reactions via tweets. To investigate such phenomena, this research has proposed hybrid sentiment analysis to pinpoint and classify sentiment expressed via electronic text using Twitter, where posts express opinions, feelings and reactions concerning particular social concerns. The overall intention is to improve decision-making by policy makers, local councils and journalists. These bodies can take advantage of these information insights and, thereby achieve an informed position to improve their decision-making process and better reflect public opinion. This strategy is termed as the 'machine learning approach' and is adapted through knowledge-based language to exploit all approaches to sentiment classification and to maximize accuracy. The hybrid approach has been introduced, experimentally confirmed, and is thereby proposed, with n-grams features derived from machine learning for the Arabic dialect to fulfil this informational gap in policymaking and public service. Initially, n-grams were fused with semantic elements (semantic features) from semantic ontology, followed by sentiment collaborative language elements (score features) from the multi-factor lexicon-based sentiment analysis of social media content in dialectal Arabic. The results gained from this hybrid approach resulted in an accuracy of 93.45 and an F-score of 89.55, illustrating a clear indication that individual feelings were understood based on Twitter entries throughout Saudi Arabia regarding unemployment.

Since sentiment analysis is considered a state-of-the-art approach regarding classification, this study sanctioned an optimized evaluation approach when analysing tweets related to specified social issues, such as unemployment figures. This ensures that it can cope with potential obstacles when utilising sentiment analysis in decision-making and can be used to the benefit of government policy makers. The ability to determine attitudes from social media and to classify such sentiments, adapting differing sentiment analysis approaches, ensures that the proposed approach is a vital supporting tool for decision makers. Thus, a central objective was to present an optimized and reliable sentiment analysis approach concerning the general public's opinion about institutions; in particular, this work helps determine the efficiency of delivery and infrastructure and the degree of public satisfaction revolving around social issues. Table 7.2 shows some examples of the usability and informative information from the dataset. The model demonstrates the simplest way to practically exploit the sentiment analysis approach which involves extracting the positive and negative sentiments about a particular phrase/keyword identified by the policy maker. Subsequently, opinion polarity can be calculated about the topic, whilst also providing the top 100 tweets (in terms of scores) expressing either negative or positive opinions.

Table 7.2: examples of the usability and informative information from the dataset

Keyword	Current Dataset	Positive tweet	Negative tweet
السعوده *Alsueuduh* (Saudisation) Program of the ministry of labour	# of positive tweets: 125 # of negative tweets: 1013 12.33% positive 87.66% negative	بكل مايقدرون وزاره العمل شغالة بطريقه رائعه بسعوده الوظاءف جهودهم مشكوره	الصحيح انه السعوديه وخيرها لاجانب يكنون الكره و البغيضه لنا . والوزاره تدعم الكلام هذا بدليل حملات السعوده الوهميه
الواسطه *Alwasituh* Cronyism	# of positive tweets: 0 # of negative tweets: 2254 0.00% positive 100.00% negative	/	العنصريه هي الممارسات الاستفزازيه الجائره التي تمارسها جهات التوظيف ضد المواطن في وطنه بتفضيل الوافدين أو اصحاب الواسطه الأقل كفاءه
طاقات *Taqat* Program of the ministry of labour	# of positive tweets: 1072 # of negative tweets: 325 69.68% positive 30.31% negative	جميل القرار الجديد و ينفع المواطنين صاحب العمل قبل يسمح له استخراج فيزه لازم يعلن عن الوظيفه بطاقات تابعوا موقعهم	نظام العمل يجبر المؤسسات على التوطين بالاعلان عن طريق طاقات لفتره محدده واذا ماتوفر سعوديين يسمح لهم بتوظيف أجنبي فيحطون مواصفات تعجيزه
العماله الاجنبيه *aleumaluh* alajnibih Foreign labour	# of positive tweets: 192 # of negative tweets: 870 22.06% positive 77.93% negative	براي القرارات الاخيره مفيده ومشجعه لشريحه كبيره من العماله الاجنبيه اللي بيستثمرون اموالهم في بلدنا	كفانا ظلم و قهر سعوديون ينتفضون ضد العماله الاجنبيه بعباره احنا اولي بفلوسنا
كفاءات *kafa'at* Program of the ministry of labour	# of positive tweets: 2909 # of negative tweets: 765 73.70% positive 26.29% negative	اتمنى ان يطبق برنامج) نطاقات (على كل مؤسسه وجامعه بحيث كل ماقلت النسبه كلما تمت عدم موافقه على التعاقد البرنامج نجح جدا مع القطاع الخاص	نطاقات تم ايجاده لحل مؤقت لمشكلة التوظيف الكبيره ما هو مفيد احنا نبغى حل جذري للمشكله
البطاله *Albitaluh* Unemployment	# of positive tweets: 53 # of negative tweets: 2141 2.47% positive 97.52% negative	أنظمه الدوله في مصلحه المواطن فهي لا تسمح بالتعاقد لأكثر من عشر سنوات لإتاحه الفرص للمواطنين والحد من البطاله	فوق غبن البطاله و المعاناه اليوميه من آثارها الاقتصاديه و النفسيه والاجتماعيه يزيدوننا غبنا باتهامنا بالتكاسل و التعالي و التشكيك في جدارتنا
بدل الندره *bdl alnadruh* amount of money added to the salary for scarcity	# of positive tweets: 75 # of negative tweets: 483 15.52% positive 84.47% negative	الوطن يحتاج للتنميه وقبلها يحتاج إلى قلوب تنبض بحبه لا تركزوا على بدل الندره لانه ليست المشكله	الجريمه الكبرى هي أن يتعاون أبناء البلد مع الأجانب لإقصاء السعودي من أجل الحصول على بدل الندره

7.5. Chapter Summary

This chapter presented the lingustic-machine learning hybrid approach for sentiment analysis of social media content in dialectical Arabic. This approach has utilised the advantages of several methods when integrating different features together and two techniques were used to determine the best hybrid approach. The first method adopted the domain-specific features and sentiment lexicon modelled in the semantic ontology as the token training features for the opinion classification. In the second method, the sentiment score resulting from the lexicon-based sentiment analysis was included in the training feature-set. The utilisation of domain knowledge in a second method was implicit as it was used to associate the sentiment mention with the domain features. In the first method, the domain-specific features and sentiment lexicon were modelled in the semantic ontology as the token training features for the opinion classification. In the second method, the sentiment score resulting from the lexicon-based sentiment analysis were included in the training feature-set. The utilisation of domain knowledge in a second method is implicit as it was used to associate the sentiment mention with the domain features. The best result was achieved by the hybrid approach that incorporates TF-IDF with lexicon sentiment score features. It combined the linear SVC classifier with bigrams feature and the best multi-factor lexicon method (all levels); the accuracy was 93.45% and the F-score was 89.55%. It was observed that both of the hybrid approach methods provided higher results than applying approaches separately (lexicon-based approach alone and the machine learning approach alone). This improvement confirms the initial hypothesis which suggested that the hybrid approach improved the sentiment classification of dialectical Arabic text.

Chapter 8

8 Conclusion and Suggestions for Future Work

This chapter is an overview of the findings explored within this study along with the overall outcome, contributions, PhD anticipated research limitations and recommended proposals for further investigation.

8.1 Overview

The expansion of social media allows posting opinions on a variety of subjects such as eating out and politics. Posted opinions contain personal viewpoints to commercial and institutional organizations, since information gleaned can steer and direct marketing policy and aid in ascertaining the general publics' opinion and mood towards events such as general elections or product campaigns. However, the expanse and seemingly chaotic nature of online data means that the evaluation and classification of text sentiment is a challenge.

Twitter is deemed a valuable Sentiment Analysis (SA) resource, since individuals turn to media outlets to express their personal views on a variety of subjects. It is one of the most popular social media apps in Saudi Arabia. Platforms such as Twitter, allow for a lucrative capture of the sentiments of the general public, particularly in terms of social issues and politics. This study has focused on unemployment in Saudi Arabia as a case study, analysing documents and an in-depth exploration of challenges identified in capturing sentiments posted by Arabic users of the net. It was ultimately concluded that Arabic NLP can be utilised for dialectic expressions, but this requires further research.

This study introduced a novel hybrid strategy that integrated lexicon-based sentiment analysis and machine learning approaches to extract and evaluate opinions from dialectal Arabic tweets. This approach followed several stages, with each developed to meet sentiment analysis challenges yet consider viable alternatives.

In the initial stages, for a corpus sentiment analysis of a specific domain (unemployment) was created by collating tweets posted in Saudi dialect via Twitter's API. W The data was amassed through hashtags trending in Saudi Arabia and attracted a huge volume of tweets. Approximately 23,500 tweets were collected, and once the redundancies had been deleted, approximately 10,000 tweets were assessed. A significant contribution to the collected tweets was from the popular account, @JoblessGrads9 (عاطلون بشهادات عليا), which specialises in unemployment issues. From this account alone, approximately 5000 tweets were extracted and ultimately reduced to 3000 after deleting redundant entries. The final dataset involved 7000 tweets. Lexical normalization of tweets was deemed a vital factor in the application of NLP tools.

As MSA stemming algorithms do not apply to Arabic dialects and only limited available stemmer tools compute dialectic vocabulary. Some of NLP tools for MSA have been experimented with in past studies, yielding poor results. For this reason, the current investigation has developed a novel stemming strategy that marries the Information Science Research Institute (ISRI) with a rule-based stemmer, to meet the obstacle posed by Saudi dialectical Arabic. The ISRI Arabic stemmer algorithm outperforms other MSA stemmers when applied to dialectal Arabic. Based on the lexical analysis of the words that the ISRI approach failed to correctly stem, a set of rules were devised to extract the stem of the Saudi dialectal Arabic words. The algorithm improved the stemming accuracy of dialectical Arabic in comparison to alternative stemming algorithms. This strategy can be utilised in applications where Saudi dialect is used, within social media, information-retrieval applications and machine translation.

In the second stage of this research, since public dialectal Arabic sentiment resources are rare, Saudi dialectal resources were developed. First, a sentiment lexicon was collected to analyse the idiosyncratic elements in online posts. Several stages followed to collate 16,500 sentiment terms. Initially, the early work by Azmi and

Alzanin (2014) and their lexicon of 1,130 sentiment items of vocabulary, presented in MSA was used to direct this exercise. Then, eight native speakers worked on linking each word to synonym sets, considering all the Saudi Arabian different dialects such as Hejazi and Najdi. Finally, the collected lexicon was manually classified by 3 annotators to confirm sentiment. Additional lexicons were developed including intensifiers, negation, emojis, supplication, proverbs, and interjections. Initially, statistical algorithms were constructed to locate frequently used terms, i.e. unigrams, bigrams, and trigrams. For each of these features, information was manually checked, constructing a dictionary for all relevant candidate features of the domain, with reference to, inflection forms, dialect and synonym sets.

In the third stage, a semantic knowledgebase of detailed knowledge within the domain was devised. Amassing this knowledgebase commenced with modelling the domain representing data collected from opinions and reviews. The model was converted into a formal ontology representing the schemata to populate the domain knowledgebase with data. To extract features, the semantic domain knowledgebase identified synonyms and vital issues from pre-viewed tweets. Identifying key concepts was carried out by linking root words in pre-assessed tweets with terms in the semantic knowledgebase via GATE's onto Root Gazetteer. From this baseline, a domain feature sentiment process was compiled to select domain features with corresponding features. Sentiment lexicon was the utilised to detect sentiment vocabulary from tweets and the elevant sentiment was considered by calculating the associated scores.

In the fourth stage, a multi-factor lexicon-based sentiment within dialectical Arabic social media was created. A novel multi-intensity lexicon-based sentiment analysis algorithm was developed that took into account several factors that will improve classification accuracy by considering emojis, negations, intensifiers and special phrases. This sentiment analysis approach also integrated a light stemming mechanism that matched sentiments to the corresponding root word in the multidialectal sentiment lexicon. Experimental analysis was performed to evaluate the accuracy of this multi-intensity lexicon-based strategy. The evaluation evidenced that, combined with light-stemming, the consideration of multiple factors (as opposed to single factors) contributed to the enhancement of the performance of the algorithm in relation to tweet sentiment classification. The accuracy of the implemented algorithms improved by approximately 6%, which compared positively with two prior research projects that had enacted a lexicon-based approach for the sentiment analysis of Saudi dialects.

The penultimate stage of this research involved a machine learning approach to classify the overall sentiment in dialectical Arabic posts. Differing machine learning algorithms were used to combine weighting schemes and n-gram features. The training section was employed to teach the machine learning algorithm, whereas the test was adapted to evaluate performance of machine learning models. The resultant experiments adopted differing machine learning classifiers with the intention to process the strategy with dialectal Arabic sentiment analysis. The accuracy of machine learning techniques, in comparison with TF-IDF features was superior to machine learning techniques with binary features. Optimum results were achieved via with TF-IDF for bigrams and BernoulliNB classifiers, which created a solid coverage of phrases, with an accuracy score of 86.97%.

Finally, this study has proposed a linguistic-machine learning hybrid approach for sentiment analysis of social media content in dialectical Arabic. The main objective of the proposed hybrid approach was to improve the sentiment classification to a more precise level and attain more accurate results. The proposed hybrid approach uses two methods to integrate machine learning and computational linguistics. The first method uses the domain semantic knowledgebase, and the second, the lexicon-based sentiment classification. Regarding the hybrid semantic knowledgebase-machine learning approach, this study combined the machine learning approach with semantic features extracted from the ontology. Three n-grams features and the TF-IDF weighting scheme were used, in addition to semantic features from the ontology. t. In the hybrid lexicon based-machine learning approach, the final score of the multi-factor lexicon-based approach was extracted and added into the machine learning approach. Three N-Grams features and the TF-IDF weighting scheme were used, and several factors such as feature-sentiment association, light stemming, emojis, intensifiers, negations and special phrases, such as supplications, proverbs and interjections were also considered.

The results of these experiments confirm that the hybrid approach, utilising machine learning approach and lexicon- based approach, improves performance and accuracy of the sentiment analysis of dialectal Arabic. The best sentiment classification results were achieved by the hybrid approach, incorporating sentiment scores with TF-IDF, combined favoured lexicon methods (multi-factor) and linear SVC classifier with bigram features. The accuracy was 93.45%. Importantly, this result was enhanced by approximately 4% compared with optimum results when using the lexicon-based approach, attaining better results than machine learning by approximately 6.5%. Ultimately, both of the hybrid approach experiments provided higher results than applying the individual approaches separately (lexicon-based approach alone and the machine learning approach alone). This improvement confirms that the core hypothesis, that the hybrid approach improved the sentiment classification of dialectical Arabic text.

8.2. Thesis Contributions

The main aim of this research was to propose a technique for achieving high sentiment analysis accuracy for tweets written in non-standard dialectical Arabic extracted from social media (Twitter). It has been achieved by addressing research and development challenges of a novel hybrid sentiment analysis combining two approaches: machine learning and lexicon-based. The subsequent sections will discuss how the findings answered each of the two primary research questions identified at the onset of this study and each of the 6 sub-questions that were also answered through this multi-stage research.

Primary Research Question 1: Can hybrid approach combining domain Semantic Knowledgebase features with machine learning improve the performance of sentiment analysis?

In the hybrid semantic knowledgebase-machine learning approach, this study evaluated the combination of a machine learning approach with domain features extracted from an ontology. Through this approach, the central advantages of both approaches were used to improve the accuracy of the analysis: Three N-Grams features were employed and the TF-IDF weighting scheme was adopted, with specific domain features adopted from the ontology. It is clear from the results achieved in this study that the LinearSVC classifier with bigrams shows the best performance and also achieved positive results with both unigram and bigrams. Further, when compared with singular machine learning approaches the results demonstrated a significant improvement of more than 4%, suggesting a positive benefit from the semantic knowledge extracted from the ontology to enhance the hybrid approach.

Primary Research Question 2: Can a hybrid approach combining multi-factor lexicon-based sentiment analysis score with machine learning improve the performance of sentiment analysis?

In the hybrid lexicon based-machine learning approach, the output of the linguistic sentiment analysis (the lexicon-based approach) was used to enhance the training of the machine learning approach. Three n-grams features and the TF-IDF weighting scheme were adopted and a sentiment score was then extracted from the sentiment lexicon. By inspecting each tweet and detecting the sentiments lexicon, the corresponding total sentiment score for an individual tweet was extracted. The final score was then added into the machine learning approach and considered several factors such as feature-sentiment association. The final scores were aggregated with 3-Gram features with the TF-IDF weighting scheme, and the machine learning classifier was subsequently trained to classify tweets into positive or negative groupings. These results confirmed the effectiveness of this hybrid approach; achieving the highest performing results with the combination of the linear SVC classifier of bigrams feature with the lexicon method (all levels). Both of the hybrid approach experiments provided higher results than the lexicon-based approach and the machine learning approach.

In addition to answers to these two primary research questions, the following series of 6 sub-questions were also answered over the course of this study:

RQ1. What are the main challenges in utilising the methods and tools designed for MSA in the NLP of dialectal Arabic?

In social media, many individuals write in dialectal Arabic, writing as they speak; e.g., the emotional Arabic tweeters and frequent habits of repeating letters to exaggerate, (helooooo)’. Appling NLP tools to dialectic vocabulary is challenging due to its complex specific features such as, slang, ironic sentences, unstructured language, contractions, colloquial expressions, idiomatic expressions, abbreviations, spelling errors, use of conjunctives and lack of punctuation. In literature dealing with MSA, scholars apply speech patterns and syntactic dependency regarding Arabic text with the view to extract sentiment and featured, utilising guidelines such as nouns and adjective order, extracting both grammatical terms. This approach allows for satisfactory results because of the grammatical structure of the MSA. However, the dialectal Arabic is considered as a spoken language rather than written language. There is no itemised structure that generally concurs with written standard for dialect, hence, applying NLP tasks in terms of dialectical language is not viable. With clearer comprehension of the characteristics of dialectical Arabic, with its challenges regarding structure and meaning, this study has demonstrated that the NLP tools analysing MSA Arabic is not an efficient solution when presented with the dialectical Arabic texts.

This research has concluded that NLP tasks such as stemming and normalisation have a clear impact on results. The most challenging NLP task in this research is stemming Saudi dialect, primarily due to obstacles in Arabic NLP because of language complexity and morphology. Diversity of dialects within Saudi Arabia, e.g. Hejazi and Nejdi, provide typical examples of the varied dialects within the Arabic language, which presents a challenge to NLP analysis of Saudi dialect. Most of the current stemming techniques focus on dealing with MSA texts and some specific dialects such as Egyptian. Following experiments involving existing NLP tools on Saudi dialect text, it was confirmed that the Information Science Research Institute (ISRI) stemming tool delivers a solid performance, though it fails to deal with dialect. In this research, this problem was addressed by enabling a novel stemming mechanism for Saudi dialect. The proposed stemming approach couples a rule-based stemmer developed in-house with the ISRI stemmer, thereby enhancing the accuracy of Saudi dialectical Arabic stemming.

RQ2. Can a domain specific framework support a knowledge-based approach to dialectical Arabic sentiment analysis?

Regarding the corpus, there are very limited open-source datasets for MSA and there is no annotated based morphological corpus for Saudi dialect of specific domain. Hence, this investigation created the gold-standard dataset for domain-specific research (e.g. unemployment). Tweets were collated by searching through streaming and previously posted tweets. As a result of in-depth research and discrimination, the gold-standard corpus for sentiment analysis was created via manual annotation of tweets by labelled polarity with their sentiment: positive or negative, representing the absolute value. The number of annotated tweets was 7,000.

At the core of this approach, lexicon is one of the main resources for sentiment analysis and it has a very important role in sentiment classification. This study revealed that Saudi Arabia has six different dialects, leading to lexical challenges due to a lack of adequate sources. In order to address this issue, dialect attributes for the lexicon add Saudi dialects, Hejazi (west region), Najdi (middle region), Shamali (north region), Janubi (south region) and Sharqawi (east region). The lexicon in this study was created both automatically and manually by experienced linguists who are native speakers of Saudi dialects. To build the lexicon, sentiment vocabulary and phrases were collated from various resources. Initially, 1130 sentiment words, created in MSA, were gleaned from Azmi and Alzanin (2014). Each word was linked with a synonym set and applied to a unique Saudi dialect set of MSA by 8 native speakers. Subsequently, vocabulary was manually classified by 3 annotators with polarity levels of very positive (+1), positive (0.5), negative (-0.5) or very negative (-1). The sentiment lexicon was hence increased from 1130 words to 16,500. Finally, several lexicons were created such as, the domain features lexicon (1987 words), intensifiers (33 words), negation (45 words), emojis (969 emojis), supplication (70 phrase), proverbs (200 phrase), interjections (30 words).

RQ 3. Which linguistic features of the Arabic language can impact the lexicon-based sentiment analysis, and how can these features be collectively considered to improve the accuracy of the analysis?

This study has investigated several linguistic features of the Arabic language that enhance accuracy of the lexicon-based sentiment analysis classification. Different techniques have been adopted over the course of this resarch to determine the final classification of tweets, such as light stemming that integrates two strategies for dialect-specific words: the ISRI Arabic stemmer and an in-house rule-based stemmer. Also, morphological analysis was conducted, drawing upon the effects of intensifying vocabulary and emojis. To cope with complex Arabic constructs such as negation, this study has adopted advanced means of exploiting the most frequently used negative expressions. In addition, intensifiers were combined to increase the accuracy of the analysis (very nice). Finally, a window technique was adopted for tweets terms (the neighbouring words to the left and right of the target word) to resolve the free word order characteristics of the Arabic language. During this investigation, it was determined that colloquial expressions and interjections were likely to influence the polarity of the textual sentiment, requiring a phrase-based modelling method capable of manging supplications in dialectical Arabic and increase the accuracy of tweet sentiment analysis.

Over the course of this process, a multi-dimensional lexicon-based sentiment analysis algorithm was developed for Saudi dialects that synthesised all of these specific, targeted features into a singular, high-performing analytical resource. Experimentation and model validation was subsequently conducted in order to confirm the accuracy of this multi-factor approach and to interpret the potential advantages of the proposed model over prior research in this field. This study has confirmed that with light stemming and multi-factor consideration (e.g. emojis, supplications, proverbs), the accuracy of the sentiment analysis was significantly improved. As a result, future applications of this model to Arabic tweet analysis and sentiment modelling are predicted to yield valuable, accurate, and domain-specific results for a variety of applications.

RQ4. Can the Semantic knowledgebase improve the accuracy of the feature extraction task? How can the semantic modelling of the domain knowledge further contribute to improving lexicon-based sentiment analysis?

The lexicon-based approach developed for this study fundamentally relies on the comprehensive analysis of domain-specific knowledge. Due to topical, domain-specific variations, effective analysis is critical to the extraction of the domain features in preparation for their pairing with sentiments. Domain knowledge includes information about a domain's environment, its key concepts, their synonyms and the relationships between these items. Domain knowledge in linguistics can be utilised to improve sentiment analysis based on rigorous dataset. The modelling of domain knowledge captures relevant information and organises it into concepts connected via relationships. For example, the specific domain problem adopted for this study, unemployment in Saudi Arabia, drew from a range of key concepts, such as unemployment, organisation, person, opinion and sentiment; it also includes interrelations, such as interactions with key stakeholders (e.g., citizens and policy makers) and the communication/advice medium (Twitter posts). The concept map was then translated into a formal ontology for use in populating a knowledgebase with semantically tagged information from the Twitter feeds.

One of the primary strategies for improving the accuracy of the feature extraction task was the feature-sentiment association which relied on the Semantic knowledgebase. After a sentiment was identified, the approach evaluated specific semantic domain features (salary, jobs, etc.) related to the sentiment using the domain features lexicon. According to a predefined association window (the neighbouring words/ two to the left and two to right of the target word) this approach was sufficient for the relatively short sentence length of tweets. Subsequently, the sentiments were associated with these particular features, and only domain-specific opinions were thereby accepted. To ensure that these sentiments were proximally linked to the domain concepts, an association window was employed, connecting the knowledge modelling output with the domain-specific indicators.

RQ5. What is the impact of Arabic language light stemming on the performance of machine learning sentiment classification?

For dialectal Arabic, a limited number of studies have been completed, restricting the comparable evidence regarding classifier performance and optimality. Throughout the literature in this field, it was evident that the most common features used in machine learning sentiment analysis are surface features that generally include n-grams and syntactic features such as POS. The Syntactic Features are utilised to reflect the structural nature of the text in order to understand how words combine and function as a process of conveying meaning. Since Arabic is both a rich and morphologically complex language, incorporating morphological and syntactic evaluation is of vital importance when considering sentiment analysis. However, the difficulty in applying Arabic NLP to dialectical Arabic text such as tweets is predominantly due to the fact that the majority of Arabic NLP tools are related to MSA data and designed for MSA texts. In this study, the syntax features, which are dependent on NLP and grammar, are not useful, as proven in the literature. This is due to the fact that dialectical Arabic text does not have a specific grammar to allow the use of NLP tools, such as extracting POS. So, in this study, unigrams, bigrams and trigram features were extracted from the corpus.

RQ6. What is the impact of applying pre-processing and the impact of adding weighting schemes on the performance of machine learning?

This study investigated the impact of pre-processing on the effectiveness of the Saudi light stemming tool that was created for the machine learning sentiment analysis. The corpus was cleaned by removing links, hashtags and special twitter characters such as retweets. Then, the text was normalized by removing diacritics or redundant letters (more than two). This step was important because it reduced the number of variations in the associated word features. Also, the Saudi light stemming was applied to this analysis. Through experimentation, it was observed that pre-processing improved the performance of machine learning around 8% against the results without light stemming.

Supplementing this pre-processing technique, this study has also investigated sentiment analysis by using a machine learning approach to analyse the results and studying the impact of different weighting schemes on classification in the dialectical Arabic text. The N-Grams features were weighted using the TF-IDF weighting scheme, which defined the importance of a feature based on the term frequency/inverse term frequency. In addition, several machine learning classifiers were applied to the output, resulting in the superior performance of machine learning with TF-IDF features than machine learning with binary features. These findings indicate that future studies using domain-specific sentiment analysis could appropriate the machine learning with TF-IDF features techniques developed for this study to significantly improve the accuracy and reliability of the sentiment analysis outputs.

8.3. Future Work

Having completed this study and critically assessed literature and explorative =findings and results, there are certain fields of interest and relevance requiring further exploration regarding sentiment analysis and its application to dialectical Arabic. The following provides several possible research directions that could be adopted in the future in order to expand the depth and accuracy of this high-performing solution:

Investigating the feasibility of applying the hybrid lexicon based- machine learning approach to text reviews.

Corporations could adopt sentiment-based analytical tools to analyse web-based customer reviews in order to trace the strengths and weaknesses of their products or services. Organisations invest time and money into resources collating and analysing online materials to pinpoint customer habits and expectations regarding consumer sentiment. They utilize this data to enhance the quality and effect of their products, along with services, production targets and marketing direction.

However, merely posting comments on a website allows for complete subjective opinions to be expressed without limits or specifications. Tese variations present a challenge when compared with other social media channels such as Twitter since sentiment analysis is designed to track positive or negative word use across limited comments and feedback. The general reviews on the internet are likely to vary by scope and specificity, drawing upon a variety of issues and complaints without a specific topical domain. This issue will require the adoption of a domain-oriented solution, requiring expert input on effect and consequence in order to increase the accuracy of the sentiment analysis. It is predicted that the proposed hybrid lexicon based-machine learning approach could be broadened to explore general reviews on the net rather than just tweets.

Improve lexicon construction (sentiment lexicon - domain features lexicon - the polarity level and intensifier - negations - emojis - special phrases) by adding a numerical score.

Throughout this study, a novel lexicon approach was proposed, assessing five different dialects within Saudi Arabia. For future work, this lexicon could be expanded and positioned for use by the general public. By applying a numerical value to each lexical item in order to indicate a positive or negative sentiment, the accuracy and depth of the proposed model could be expanded to improve transferability and applicability across a range of other analytical problems.

Investigating the development of a recommendation system based on data analysis - Application on tweets SA.

Despite many previous works proposing novel methods of analysing English text, limited studies pinpoint Arabic work. Future research will explore various approaches to the analysis of dialectical Arabic based on SA, taking advantage of Spark, an online application framework offering extensive data, to extract the plethora of information available on the internet and social networks. In detecting general opinions concerning a specific domain covering multiple tweets, the SPARQL query is the chosen recommender function, retrieving the average quality of predicted rates within all tweets analysed. Complicated SPARQL queries may be adopted to analyse usage of domain knowledgebase for complex interrogative evaluation of opinions regarding the recommender functions (Nabil et al., 2018). KnowledgebaseIn future work, an application exploiting a semantic knowledgebase will be developed in order to to offer and satisfy demand for the recommender system user.

Develop deep learning models that can be used to analyse variations in word embeddings and varying class sizes

This study adopted six Deep learning models with the use of different word embeddings such as Arabic Online Commentary Word Embeddings (AOC), Twitter, Twitter-City Word Embeddings, FastText Arabic Wiki Word Embeddings, Mazajak Word Embeddings. The results yielded interesting and relevant findings, and, in most cases, they were superior to the most frequent class baseline. When compared with ML approach it was found that classical ML models do better than deep learning models. This finding was expected considering the number of classes and the size of the dataset. Deep learning models work better with much more data. In future work, improvements to the current dataset are proposed in order to make expand it in scale and then apply a deep learning approach to the modified result.

9 References

Abainia, K., Ouamour, S. and Sayoud, H., 2017. A novel robust Arabic light stemmer. *Journal of Experimental & Theoretical Artificial Intelligence*, *29*(3), pp.557-573.

Abbasi, A., Chen, H. and Salem, A., 2008. Sentiment analysis in multiple languages: Feature selection for opinion classification in web forums. *ACM Transactions on Information Systems (TOIS)*, *26*(3), p.12.

Abdelali, A., Cowie, J. and Soliman, H.S., 2004, April. Arabic information retrieval perspectives. In *Proceedings of the 11th Conference on Natural Language Processing, Journes d'Etude sur la Parole-Traitement Automatique des Langues Naturelles (JEP-TALN)* (pp. 391-400).

Abdelali, A., Darwish, K., Durrani, N. and Mubarak, H., 2016, June. Farasa: A fast and furious segmenter for arabic. In *Proceedings of the 2016 conference of the North American chapter of the association for computational linguistics: Demonstrations* (pp. 11-16).

Abd-Elhamid, L., Elzanfaly, D. and Eldin, A.S., 2016, December. Feature-based sentiment analysis in online Arabic reviews. In *2016 11th International Conference on Computer Engineering & Systems (ICCES)* (pp. 260-265). IEEE.

Abdellaoui, H. and Zrigui, M., 2018. Using tweets and emojis to build TEAD: an Arabic dataset for sentiment analysis. *Computación y Sistemas*, *22*(3).

Abdulla, N., Mohammed, S., Al-Ayyoub, M. and Al-Kabi, M., 2014, August. Automatic lexicon construction for Arabic sentiment analysis. In *2014 International Conference on Future Internet of Things and Cloud* (pp. 547-552). IEEE.

Abdulla, N.A., Ahmed, N.A., Shehab, M.A. and Al-Ayyoub, M., 2013, December. Arabic sentiment analysis: Lexicon-based and corpus-based. In *2013 IEEE Jordan conference on applied electrical engineering and computing technologies (AEECT)* (pp. 1-6). IEEE.

Abdul-Mageed, M. and Diab, M., 2012. Toward building a large-scale Arabic sentiment lexicon. In *Proceedings of the 6th international global WordNet conference* (pp. 18-22).

Abdul-Mageed, M. and Diab, M.T., 2011, June. Subjectivity and sentiment annotation of modern standard arabic newswire. In *Proceedings of the 5th linguistic annotation workshop* (pp. 110-118). Association for Computational Linguistics.

Abdul-Mageed, M. and Diab, M.T., 2014, May. Sana: A large scale multi-genre, multi-dialect lexicon for arabic subjectivity and sentiment analysis. In *LREC* (pp. 1162-1169).

Abdul-Mageed, M., 2017. Modeling Arabic subjectivity and sentiment in lexical space. *Information Processing & Management*.

Abdul-Mageed, M., Diab, M.T. and Korayem, M., 2011, June. Subjectivity and sentiment analysis of modern standard Arabic. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: short papers-Volume 2* (pp. 587-591). Association for Computational Linguistics.

Abdul-Mageed, M., Kuebler, S. and Diab, M., 2012. SAMAR: A System for Subjectivity and Sentiment Analysis of Sicial Media Arabic. In *Proc. of the Third Workshop on Computational Approaches to Subjectivity and Sentiment Analysis (WASSA'12), ICC Jeju, Republic of Korea, 2012*.

Abo, M.E.M., Shah, N.A.K., Balakrishnan, V. and Abdelaziz, A., 2018, August. Sentiment analysis algorithms: evaluation performance of the Arabic and English language. In *2018 International Conference on Computer, Control, Electrical, and Electronics Engineering (ICCCEEE)* (pp. 1-5). IEEE.

Abouenour, L., Bouzoubaa, K. and Rosso, P., 2008, December. Improving Q/A using Arabic WordNet. In *Proc. The 2008 International Arab Conference on Information Technology (ACIT'2008), Tunisia, December*.

Abozinadah, E.A. and Jones, J.H., 2016. Improved micro-blog classification for detecting abusive Arabic Twitter accounts. *International Journal of Data Mining & Knowledge Management Process (IJDKP)*, *6*(6), pp.17-28.

Abu Shquier, M. and Abu Shqeer, O. (2014). Words ordering and corresponding verb-subject agreements in English-Arabic machine translation: Hybrid-based approach. *International Journal of Soft Computing And Software Engineering (JSCSE)*, [online] Vol.2(2251-7545). Available at: http://www.jscse.com/papers/vol2.no8/vol2.no8.5.pdf [Accessed 21 Jun. 2017].

Abuaiadah, D., Rajendran, D. and Jarrar, M., 2017, October. Clustering Arabic tweets for sentiment analysis. In *2017 IEEE/ACS 14th International Conference on Computer Systems and Applications (AICCSA)* (pp. 449-456). IEEE.

Abuelenin, S., Elmougy, S. and Naguib, E., 2017, September. Twitter sentiment analysis for arabic tweets. In *International Conference on Advanced Intelligent Systems and Informatics* (pp. 467-476). Springer, Cham.

Adawood, M. and Mohammed, M. (2008). The Ambiguity of Gender in English - Arabic Translation. *ADAB AL-RAFIDAYN*, [online] 52(1). Available at: http://www.iasj.net/iasj?func=fulltext&aId=32936 [Accessed 12 May. 2017].

Ahmad, K., Cheng, D. and Almas, Y., 2007, May. Multi-lingual sentiment analysis of financial news streams. In *1st International Workshop on Grid Technology for Financial Modeling and Simulation* (Vol. 26, p. 001). SISSA Medialab.

Ahmed, S., Pasquier, M. and Qadah, G. (2013). Key Issues in Conducting Sentiment Analysis on Arabic Social Media Text. [online] Available at: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6544396.

Ahn, H., Moon, H., Fazzari, M.J., Lim, N., Chen, J.J. and Kodell, R.L., 2007. Classification by ensembles from random partitions of high-dimensional data. *Computational Statistics & Data Analysis*, *51*(12), pp.6166-6179.

Al Aqad, M. (2013). Syntactic analysis of Arabic adverb's between Arabic and English: X bar theory. *International Journal of Language and Linguistics*, [online] 1(3), pp.70--74. Available at: http://article.sciencepublishinggroup.com/pdf/10.11648.j.ijll.20130103.11.pdf [Accessed 22 May. 2017].

Alahmary, R.M., Al-Dossari, H.Z. and Emam, A.Z., 2019, January. Sentiment Analysis of Saudi Dialect Using Deep Learning Techniques. In *2019 International Conference on Electronics, Information, and Communication (ICEIC)* (pp. 1-6). IEEE.

Al-Ayyoub, M., Essa, S.B. and Alsmadi, I., 2015. Lexicon-based sentiment analysis of Arabic tweets. *IJSNM*, *2*(2), pp.101-114.

Al-Ayyoub, M., Nuseir, A., Alsmearat, K., Jararweh, Y. and Gupta, B., 2018. Deep learning for Arabic NLP: A survey. *Journal of Computational Science*, *26*, pp.522-531.

Al-Azani, S. and El-Alfy, E.S.M., 2018, April. Combining emojis with Arabic textual features for sentiment classification. In *2018 9th International Conference on Information and Communication Systems (ICICS)* (pp. 139-144). IEEE.

Albogamy, F. and Ramsay, A., 2015, September. POS tagging for Arabic tweets. In *Proceedings of the International Conference Recent Advances in Natural Language Processing* (pp. 1-8).

Aldayel, H.K. and Azmi, A.M., 2016. Arabic tweets sentiment analysis–a hybrid scheme. *Journal of Information Science*, *42*(6), pp.782-797.

Alduais, A. (2012). Simple Sentence Structure of Standard Arabic Language and Standard English Language: A Contrastive Study. *International Journal of Linguistics*, [online] 4(4), p.--500. Available at: http://www.macrothink.org/journal/index.php/ijl/article/viewFile/2621/pdf [Accessed 3 Jun. 2017].

Al-Gaphari, G.H. and Al-Yadoumi, M., 2012. A method to convert Sana'ani accent to Modern Standard Arabic. *International Journal of Information Science and Management (IJISM)*, *8*(1), pp.39-49.

Alghamdi, M., Alhargan, F., Alkanhal, M., Alkhairy, A., Eldesouki, M. and Alenazi, A., 2008. Saudi accented Arabic voice bank. *Journal of King Saud University-Computer and Information Sciences*, *20*, pp.45-64.

Alhajjar, A., Hajjar, M. and Zreik, K., 2009, April. Classification of Arabic information extraction methods. In *2nd International conference on Arabic language resources and tools*.

Alhanjouri, M.A., 2017. Pre Processing Techniques for Arabic Documents Clustering. *Pre Processing Techniques for Arabic Documents Clustering*, *7*(2).

Al-Harbi, O., 2017. Using objective words in the reviews to improve the colloquial arabic sentiment analysis. *arXiv preprint arXiv:1709.08521*.

Alharbi, R., Magdy, W., Darwish, K., Abdelali, A. and Mubarak, H., 2018, May. Part-of-speech tagging for arabic gulf dialect using bi-lstm. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*.

Al-Harbi, W.A. and Emam, A., 2015. Effect of Saudi dialect preprocessing on Arabic sentiment analysis. *International Journal of Advanced Computer Technology*, *4*(6), pp.91-99.

Alhazmi, S., Black, W. and McNaught, J., 2013. Arabic SentiWordNet in relation to SentiWordNet 3.0. *2180*, *1266*(4), p.1.

Alhindi, W.A., Talha, M. and Sulong, G.B., 2012. The role of modern technology in Arab Spring. *Archives Des Sciences*, *65*(8), pp.101-112.

Al-Horaibi, L. and Khan, M.B., 2016, July. Sentiment analysis of Arabic tweets using text mining techniques. In *First International Workshop on Pattern Recognition* (Vol. 10011, p. 100111F). International Society for Optics and Photonics.

Alhumoud, S., Albuhairi, T. and Altuwaijri, M., 2015, November. Arabic sentiment analysis using WEKA a hybrid learning approach. In *2015 7th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K)* (Vol. 1, pp. 402-408). IEEE.

Alison, 2018. *Top 10 Languages Used on the Internet*. [online] Accredited Language Services. Available at: https://www.accreditedlanguage.com/2016/09/13/top-10-languages-used-on-the-internet [Accessed 3 Sep. 2018].

Al-Kabi, M.N., Kazakzeh, S.A., Ata, B.M.A., Al-Rababah, S.A. and Alsmadi, I.M., 2015. A novel root based Arabic stemmer. *Journal of King Saud University-Computer and Information Sciences*, *27*(2), pp.94-103.

Alkubaisi, G.A.A.J., Kamaruddin, S.S. and Husni, H., 2018. Stock Market Classification Model Using Sentiment Analysis on Twitter Based on Hybrid Naive Bayes Classifiers. *Computer and Information Science*, *11*(1), pp.52-64.

Almas, Y. and Ahmad, K., 2007, July. A note on extracting 'sentiments' in financial news in English, Arabic & Urdu. In *The Second Workshop on Computation, al Approaches to Arabic Script-based Languages* (Vol. 21, pp. 21-22).

Al-Moslmi, T., Albared, M., Al-Shabi, A., Omar, N. and Abdullah, S., 2018. Arabic senti-lexicon: Constructing publicly available language resources for Arabic sentiment analysis. *Journal of information science*, *44*(3), pp.345-362.

Al-Obaidi, A.Y. and Samawi, V.W., 2016. Opinion mining: Analysis of comments written in Arabic colloquial. In *Proceedings of the World Congress on Engineering and Computer Science* (Vol. 1).

Al-Omari, A. and Abuata, B., 2014. Arabic light stemmer (ARS). *Journal of Engineering Science and Technology*, *9*(6), pp.702-717.

Alomari, K.M., ElSherif, H.M. and Shaalan, K., 2017, June. Arabic tweets sentimental analysis using machine learning. In *International Conference on Industrial,* *Engineering and Other Applications of Applied Intelligent Systems* (pp. 602-610). Springer, Cham.

Aloqaily, A., Al-Hassan, M.A.L.A.K., Salah, K., Elshqeirat, B. & Almashagbah, M. 2020. Sentiment analysis for Arabic tweets datasets: Lexicon-based and machine learning approaches. *Journal of Theoretical and Applied Information Technology*, *98*(4): 1-14.
Aloqaily, A., Al-Hassan, M.A.L.A.K., Salah, K., Elshqeirat, B. & Almashagbah, M. 2020. Sentiment analysis for Arabic tweets datasets: Lexicon-based and machine learning approaches. *Journal of Theoretical and Applied Information Technology*, *98*(4): 1-14.

Alrefai, M.A., Faris, H. and Aljarah, I., 2018. Sentiment analysis for Arabic language: A brief survey of approaches and techniques. *arXiv preprint arXiv:1809.02782*.

Al-Rubaiee, H., Qiu, R. and Li, D., 2016, March. Identifying Mubasher software products through sentiment analysis of Arabic tweets. In *2016 International Conference on Industrial Informatics and Computer Systems (CIICS)* (pp. 1-6). IEEE.

Al-Shalabi, R. and Obeidat, R., 2008, March. Improving KNN Arabic text classification with n-grams based document indexing. In *Proceedings of the Sixth International Conference on Informatics and Systems, Cairo, Egypt* (pp. 108-112).

Al-Shammari, E. and Lin, J., 2008, July. A novel Arabic lemmatization algorithm. In *Proceedings of the second workshop on Analytics for noisy unstructured text data* (pp. 113-118). ACM.

Al-Shammari, E., 2013. *Lemmatizing, stemming, and query expansion method and system*. U.S. Patent 8,473,279.

Al-Shargi, F. and Rambow, O., 2015, July. DIWAN: A dialectal word annotation tool for Arabic. In *Proceedings of the Second Workshop on Arabic Natural Language Processing* (pp. 49-58).

Al-Shargi, F., Kaplan, A., Eskander, R., Habash, N. and Rambow, O., 2016, May. Morphologically annotated corpora and morphological analyzers for Moroccan and Sanaani Yemeni Arabic.

Al-smayel, S. (2016). *7000 عاطل في المملكة من أصحاب الشهادات العليا*. [online] Watanksa. Available at: https://www.alwatan.com.sa/article/312692/7000--عاطل-في-المملكة-من-أصحاب-الشهادات-العليا [Accessed 5 Sep. 2016].

Altawaier, M.M. and Tiun, S., 2016. Comparison of machine learning approaches on arabic twitter sentiment analysis. *International Journal on Advanced Science, Engineering and Information Technology*, *6*(6), pp.1067-1073.

Althobaiti, M., Kruschwitz, U. and Poesio, M., 2014. Aranlp: A java-based library for the processing of arabic text.

Al-Thubaity, A., Alharbi, M., Alqahtani, S. and Aljandal, A., 2018, April. A Saudi dialect Twitter Corpus for sentiment and emotion analysis. In *2018 21st Saudi Computer Society National Computer Conference (NCC)* (pp. 1-6). IEEE.

Al-Twairesh, N., Al-Khalifa, H. and AlSalman, A., 2016. AraSenTi: large-scale twitter-specific Arabic sentiment lexicons. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (Vol. 1, pp. 697-705).

Al-Twairesh, N., Al-Khalifa, H., Al-Salman, A. and Al-Ohali, Y., 2017. Arasenti-tweet: A corpus for Arabic sentiment analysis of Saudi tweets. *Procedia Computer Science*, *117*, pp.63-72.

Al-Twairesh, N., Al-Khalifa, H., Alsalman, A. and Al-Ohali, Y., 2018. Sentiment analysis of Arabic tweets: Feature engineering and a hybrid approach. *arXiv preprint arXiv:1805.08533*.

Alwakid, G., Osman, T. and Hughes-Roberts, T., 2017. Challenges in sentiment analysis for Arabic social networks. *Procedia Computer Science*, *117*, pp.89-100.

Alwakid, G., Osman, T. and Hughes-Roberts, T., 2019, April. Towards Improved Saudi Dialectal Arabic Stemming. In *2019 International Conference on Computer and Information Sciences (ICCIS)* (pp. 1-5). IEEE.

Aly, M. and Atiya, A., 2013. Labr: A large scale Arabic book reviews dataset. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)* (Vol. 2, pp. 494-498).

Alyafeai, Z. and Al-Shaibani, M., 2020, November. ARBML: Democritizing Arabic Natural Language Processing Tools. In *Proceedings of Second Workshop for NLP Open Source Software (NLP-OSS)* (pp. 8-13).

Antoun, W., Baly, F. and Hajj, H., 2020. AraBERT: Transformer-based model for Arabic language understanding. *arXiv preprint arXiv:2003.00104*.

Arts, T., Belinkov, Y., Habash, N., Kilgarriff, A. and Suchomel, V., 2014. arTenTen: Arabic corpus and word sketches. *Journal of King Saud University-Computer and Information Sciences*, *26*(4), pp.357-371.

Asghar, M.Z., Kundi, F.M., Ahmad, S., Khan, A., and Khan, F. 2017. T-SAF: Twitter sentiment analysis framework using a hybrid classification scheme. *Expert Systems*, 35, 1-19.

Assiri, A., Emam, A. and Al-Dossari, H., 2016. Saudi twitter corpus for sentiment analysis. *World Academy of Science, Engineering and Technology, International Journal of Computer, Electrical, Automation, Control and Information Engineering*, *10*(2), pp.272-275.

Assiri, A., Emam, A. and Al-Dossari, H., 2018. Towards enhancement of a lexicon-based approach for Saudi dialect sentiment analysis. *Journal of Information Science*, *44*(2), pp.184-202.

Atwan, J., Wedyan, M. and Al-Zoubi, H., 2019. Arabic text light stemmer. *Int. J. Comput. Acad. Res*, *8*(2), pp.17-23.

Azmi, A.M. and Alzanin, S.M., 2014. Aara'–a system for mining the polarity of Saudi public opinion through e-newspaper comments. *Journal of Information Science*, *40*(3), pp.398-410.

Baccianella, S., Esuli, A. and Sebastiani, F., 2010, May. Sentiwordnet 3.0: an enhanced lexical resource for sentiment analysis and opinion mining. In *Lrec* (Vol. 10, No. 2010, pp. 2200-2204).

Badaro, G., Baly, R., Hajj, H., El-Hajj, W., Shaban, K.B., Habash, N., Al-Sallab, A. and Hamdi, A., 2019. A survey of opinion mining in Arabic: A comprehensive system perspective covering challenges and advances in tools, resources, models, applications, and visualizations. *ACM Transactions on Asian and Low-Resource Language Information Processing (TALLIP)*, *18*(3), p.27.

Badaro, G., Baly, R., Hajj, H., Habash, N. and El-Hajj, W., 2014, October. A large scale Arabic sentiment lexicon for Arabic opinion mining. In *Proceedings of the EMNLP 2014 workshop on Arabic natural language processing (ANLP)* (pp. 165-173).

Badaro, G., El Jundi, O., Khaddaj, A., Maarouf, A., Kain, R., Hajj, H. and El-Hajj, W., 2018, June. Ema at semeval-2018 task 1: Emotion mining for Arabic. In *Proceedings of The 12th International Workshop on Semantic Evaluation* (pp. 236-244).

Badaro, G., Jundi, H., Hajj, H. and El-Hajj, W., 2018, June. EmoWordNet: Automatic expansion of emotion lexicon using English WordNet. In *Proceedings of the Seventh Joint Conference on Lexical and Computational Semantics* (pp. 86-93).

Badr, I., Zbib, R. and Glass, J. (2009). Syntactic phrase reordering for English-to-Arabic statistical machine translation. [online] pp.86--93. Available at: https://groups.csail.mit.edu/sls/publications/2009/Badr-EACL09.pdf [Accessed 28 May. 2014].

Baly, R., El-Khoury, G., Moukalled, R., Aoun, R., Hajj, H., Shaban, K.B. and El-Hajj, W., 2017. Comparative evaluation of sentiment analysis methods across Arabic dialects. *Procedia Computer Science*, *117*, pp.266-273.

Baly, R., Hajj, H., Habash, N., Shaban, K.B. and El-Hajj, W., 2017. A sentiment treebank and morphologically enriched recursive deep models for effective sentiment analysis in arabic. *ACM Transactions on Asian and Low-Resource Language Information Processing (TALLIP)*, *16*(4), p.23.

Bekkar, M., Djemaa, H.K. and Alitouche, T.A., 2013. Evaluation measures for models assessment over imbalanced data sets. *J Inf Eng Appl*, *3*(10).

Bhuta, S., Doshi, A., Doshi, U. and Narvekar, M., 2014, February. A review of techniques for sentiment analysis Of Twitter data. In *2014 International conference on issues and challenges in intelligent computing techniques (ICICT)* (pp. 583-591). IEEE.

Biltawi, M., Al-Naymat, G. and Tedmori, S., 2017, October. Arabic sentiment classification: A hybrid approach. In *2017 International Conference on New Trends in Computing Sciences (ICTCS)* (pp. 104-108). IEEE.

Black, W., Elkateb, S., Rodriguez, H., Alkhalifa, M., Vossen, P., Pease, A. and Fellbaum, C., 2006, January. Introducing the Arabic WordNet project. In *Proceedings of the third international WordNet conference* (pp. 295-300).

Bottou, L. and Bousquet, O., 2008. The tradeoffs of large scale learning. In *Advances in neural information processing systems* (pp. 161-168).

Bottou, L., 2014. From machine learning to machine reasoning. *Machine learning*, *94*(2), pp.133-149.

Bouamor, H., Habash, N., Salameh, M., Zaghouani, W., Rambow, O., Abdulrahim, D., Obeid, O., Khalifa, S., Eryani, F., Erdmann, A. and Oflazer, K., 2018, May. The MADAR

Arabic dialect corpus and lexicon. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*.

Boudad, N., Faizi, R., Thami, R.O.H. and Chiheb, R., 2017. Sentiment Classification of Arabic Tweets: A Supervised Approach. *J. Mobile Multimedia*, *13*(3&4), pp.233-243.

Boudad, N., Faizi, R., Thami, R.O.H. and Chiheb, R., 2018. Sentiment analysis in Arabic: A review of the literature. *Ain Shams Engineering Journal*, *9*(4), pp.2479-2490.

Breiman, L., 2001. Random Forests. Vol. 45. *Mach Learn*, *Kluwer Academic Publishers,* [online]. Available at: https://doi.org/10.1023/A:1010933404324 [Accessed 29 Jan. 2018].

Bsoul, Q.W. and Mohd, M., 2011, December. Effect of ISRI stemming on similarity measure for Arabic document clustering. In *Asia Information Retrieval Symposium* (pp. 584-593). Springer, Berlin, Heidelberg.

Buccoliero, L., Bellio, E., Crestini, G. and Arkoudas, A., 2020. Twitter and politics: Evidence from the US presidential elections 2016. *Journal of Marketing Communications*, *26*(1), pp.88-114.

Buckwalter, T., 2004. *Buckwalter arabic morphological analyzer version 2.0. linguistic data consortium, University of Pennsylvania, 2002. ldc cat alog no*. Ldc2004l02. Technical report.

Carletta, J., 1996. Assessing agreement on classification tasks: the kappa statistic. *Computational Linguistics*, *22*(2), pp.249-254.

Chapelle, O., Scholkopf, B. and Zien, A., 2009. Semi-supervised learning (Chapelle, o. et al., eds.; 2006) [book reviews]. *IEEE Transactions on Neural Networks, 20*(3), pp.542-542.

Chau, M. and Xu, J., 2007. Mining communities and their relationships in blogs: A study of online hate groups. *International Journal of Human-Computer Studies*, *65*(1), pp.57-70.

Chen, Z., Mukherjee, A. and Liu, B., 2014, June. Aspect extraction with automated prior knowledge learning. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (pp. 347-358).

Cherif, W., Madani, A. and Kissi, M., 2015, March. A new modeling approach for Arabic opinion mining recognition. In *2015 Intelligent Systems and Computer Vision (ISCV)* (pp. 1-6). IEEE.

Cherif, W., Madani, A. and Kissi, M., 2016, October. A combination of low-level light stemming and support vector machines for the classification of Arabic opinions. In *2016 11th International Conference on Intelligent Systems: Theories and Applications (SITA)* (pp. 1-5). IEEE.

Cohen, J., 1960. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, *20*(1), pp.37-46.

Corallo, A., Fortunato, L., Matera, M., Alessi, M., Camillò, A., Chetta, V., Giangreco, E. and Storelli, D., 2015, July. Sentiment Analysis for Government: an optimized approach. In *International Workshop on Machine Learning and Data Mining in Pattern Recognition* (pp. 98-112). Springer, Cham.

Cunningham, H., 2002. GATE: A framework and graphical development environment for robust NLP tools and applications. In *Proc. 40th annual meeting of the association for computational linguistics (ACL 2002)* (pp. 168-175).

Darwish, K., Magdy, W. and Mourad, A., 2012, October. Language processing for arabic microblog retrieval. In *Proceedings of the 21st ACM international conference on Information and knowledge management* (pp. 2427-2430). ACM.

Darwish, K., Mubarak, H., Abdelali, A., Eldesouki, M., Samih, Y., Alharbi, R., Attia, M., Magdy, W. and Kallmeyer, L., 2018, May. Multi-dialect Arabic pos tagging: a CRF approach. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018).*

Dashtipour, K., Poria, S., Hussain, A., Cambria, E., Hawalah, A.Y., Gelbukh, A. and Zhou, Q., 2016. Multilingual sentiment analysis: state of the art and independent comparison of techniques. *Cognitive computation*, *8*(4), pp.757-771.

de Marneffe, M.C., Manning, C.D. and Potts, C., 2010, July. Was it good? It was provocative. Learning the meaning of scalar adjectives. In *Proceedings of the 48th annual meeting of the association for computational linguistics* (pp. 167-176). Association for Computational Linguistics.

Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805.

Diab, M., 2009. Second generation AMIRA tools for Arabic processing: Fast and robust tokenization, POS tagging, and base phrase chunking. In *2nd International Conference on Arabic Language Resources and Tools* (Vol. 110, p. 198).

Diab, M., Habash, N., Rambow, O., Altantawy, M. and Benajiba, Y., 2010, May. COLABA: Arabic dialect annotation and processing. In *Lrec workshop on Semitic language processing* (pp. 66-74).

Diab, M.T., Al-Badrashiny, M., Aminian, M., Attia, M., Elfardy, H., Habash, N., Hawwari, A., Salloum, W., Dasigi, P. and Eskander, R., 2014, May. Tharwa: A Large Scale Dialectal Arabic-Standard Arabic-English Lexicon. In *LREC* (pp. 3782-3789).

Diab, M.T., Levin, L., Mitamura, T., Rambow, O., Prabhakaran, V. and Guo, W., 2009, August. Committed belief annotation and tagging. In *Proceedings of the Third Linguistic Annotation Workshop* (pp. 68-73). Association for Computational Linguistics.

Du, H., 2010. *Data mining techniques and applications: An introduction* (pp. 113-147). Boston: Cengage Learning.

Duwairi, R.M. and Alshboul, M.A., 2015, August. Negation-aware framework for sentiment analysis in Arabic reviews. In *2015 3rd International Conference on Future Internet of Things and Cloud* (pp. 731-735). IEEE.

Duwairi, R.M. and Qarqaz, I., 2016. A framework for Arabic sentiment analysis using supervised classification. *International Journal of Data Mining, Modelling and Management*, *8*(4), pp.369-381.

El-Beltagy, S.R. and Ali, A., 2013, March. Open issues in the sentiment analysis of Arabic social media: A case study. In *2013 9th International Conference on Innovations in Information Technology (IIT)* (pp. 215-220). IEEE.

El-Beltagy, S.R., 2016, May. NileULex: a phrase and word level sentiment lexicon for Egyptian and modern standard Arabic. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)* (pp. 2900-2905).

El-Beltagy, S.R., 2017. WeightedNileULex: A Scored Arabic Sentiment Lexicon for Improved Sentiment Analysis. *Language Processing, Pattern Recognition and Intelligent Systems. Special Issue on Computational Linguistics, Speech& Image Processing for Arabic Language. World Scientific Publishing Co*.

El-Beltagy, S.R., Khalil, T., Halaby, A. and Hammad, M., 2016, April. Combining lexical features and a supervised learning approach for Arabic sentiment analysis. In *International Conference on Intelligent Text Processing and Computational Linguistics* (pp. 307-319). Springer, Cham.

El-Halees, A.M., 2011. Arabic opinion mining using combined classification approach. *Arabic opinion mining using combined classification approach*.

El-Masri, M., Altrabsheh, N., Mansour, H. and Ramsay, A., 2017. A web-based tool for Arabic sentiment analysis. *Procedia Computer Science*, *117*, pp.38-45.

ElSahar, H. and El-Beltagy, S.R., 2014, April. A fully automated approach for arabic slang lexicon extraction from microblogs. In *International conference on intelligent text processing and computational linguistics* (pp. 79-91). Springer, Berlin, Heidelberg.

ElSahar, H. and El-Beltagy, S.R., 2015, April. Building large Arabic multi-domain resources for sentiment analysis. In *International Conference on Intelligent Text Processing and Computational Linguistics* (pp. 23-34). Springer, Cham.

Elshakankery, K. and Ahmed, M.F., 2019. HILATSA: A hybrid incremental learning approach for Arabic tweets sentiment analysis. *Egyptian Informatics Journal*, *20*(3), pp.163-171.

Farra, N., Challita, E., Assi, R.A. and Hajj, H., 2010, December. Sentence-level and document-level sentiment mining for Arabic texts. In *2010 IEEE international conference on data mining workshops* (pp. 1114-1119). IEEE.

Felbo, B., Mislove, A., Søgaard, A., Rahwan, I. and Lehmann, S., 2017. Using millions of emoji occurrences to learn any-domain representations for detecting sentiment, emotion and sarcasm. *arXiv preprint arXiv:1708.00524*.

Froud, H., Benslimane, R., Lachkar, A. and Ouatik, S.A., 2010, September. Stemming and similarity measures for Arabic Documents Clustering. In *2010 5th International Symposium On I/V Communications and Mobile Network* (pp. 1-4). IEEE.

Gamal, D., Alfonse, M., El-Horbaty, E.S.M. and Salem, A.B.M., 2019. Implementation of Machine Learning Algorithms in Arabic Sentiment Analysis Using N-Gram Features. *Procedia Computer Science*, *154*, pp.332-340.

General Authority for Statistics. (2016). *General Authority for Statistics*. [online] Available at: https://www.stats.gov.sa/en/node [Accessed 11 Apr. 2017].

Ghadeer, A.S., Aljarah, I. and Alsawalqah, H., 2017. Enhancing the Arabic sentiment analysis using different preprocessing operators. *New Trends Inf. Technol*, *113*, pp.113-117.

Ghosh, S., 2009. Application of Natural Language Processing (NLP) Techniques in E–Governance. In *E-Government Development and Diffusion: Inhibitors and Facilitators of Digital Democracy* (pp. 122-132). IGI Global.

Gmi_blogger, 2019. *Saudi Arabia Social Media Statistics 2019 (Infographics) - GMI Blog*. [online] Globalmediainsight.com. Available at: https://www.globalmediainsight.com/blog/saudi-arabia-social-media-statistics/ [Accessed 8 Nov. 2019].

Guellil, I., Saâdane, H., Azouaou, F., Gueni, B. and Nouvel, D., 2019. Arabic natural language processing: An overview. *Journal of King Saud University-Computer and Information Sciences*.

Gupta, I. and Joshi, N., 2019. Enhanced Twitter Sentiment Analysis Using Hybrid Approach and by Accounting Local Contextual Semantic. *Journal of Intelligent Systems*, *29*(1), pp.1611-1625.

Habash, N. and Rambow, O., 2005, June. Arabic tokenization, part-of-speech tagging and morphological disambiguation in one fell swoop. In *Proceedings of the 43rd annual meeting of the association for computational linguistics (ACL'05)* (pp. 573-580).

Habash, N. and Rambow, O., 2006, July. MAGEAD: a morphological analyzer and generator for the Arabic dialects. In *Proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the Association for Computational Linguistics* (pp. 681-688).

Habash, N. and Roth, R.M., 2009, August. Catib: The columbia arabic treebank. In *Proceedings of the ACL-IJCNLP 2009 conference short papers* (pp. 221-224). Association for Computational Linguistics.

Habash, N., Diab, M.T. and Rambow, O., 2012, May. Conventional Orthography for Dialectal Arabic. In *LREC* (pp. 711-718).

Habash, N., Eryani, F., Khalifa, S., Rambow, O., Abdulrahim, D., Erdmann, A., Faraj, R., Zaghouani, W., Bouamor, H., Zalmout, N. and Hassan, S., 2018, May. Unified guidelines and resources for Arabic dialect orthography. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018).*

Habash, N., Eryani, F., Khalifa, S., Rambow, O., Abdulrahim, D., Erdmann, A., Faraj, R., Zaghouani, W., Bouamor, H., Zalmout, N. and Hassan, S., 2018, May. Unified guidelines and resources for Arabic dialect orthography. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*.

Habash, N., Eskander, R. and Hawwari, A., 2012, June. A morphological analyzer for Egyptian Arabic. In *Proceedings of the twelfth meeting of the special interest group on computational morphology and phonology* (pp. 1-9). Association for Computational Linguistics.

Habash, N., Rambow, O. and Roth, R., 2009, April. MADA+ TOKAN: A toolkit for Arabic tokenization, diacritization, morphological disambiguation, POS tagging, stemming and lemmatization. In *Proceedings of the 2nd international conference on Arabic language resources and tools (MEDAR), Cairo, Egypt* (Vol. 41, p. 62).

Habash, N.Y., 2010. Introduction to Arabic natural language processing. *Synthesis Lectures on Human Language Technologies*, *3*(1), pp.1-187.

Hajmohammadi, M.S., Ibrahim, R., Selamat, A. and Fujita, H., 2015. Combination of active learning and self-training for cross-lingual sentiment classification with density analysis of unlabelled samples. *Information sciences*, *317*, pp.67-77.

Hammad, M. and Al-awadi, M., 2016. Sentiment analysis for Arabic reviews in social networks using machine learning. In *Information Technology: New Generations* (pp. 131-139). Springer, Cham.

Hammo, B.H., 2009. Towards enhancing retrieval effectiveness of search engines for diacritisized Arabic documents. *Information retrieval*, *12*(3), pp.300-323.

Hamouda, A.E.D.A. and El-taher, F.E.Z., 2013. Sentiment analyzer for Arabic comments system. *Int. J. Adv. Comput. Sci. Appl*, *4*(3).

Han, B., Cook, P. and Baldwin, T., 2013. Lexical normalization for social media text. *ACM Transactions on Intelligent Systems and Technology (TIST)*, *4*(1), p.5.

Han, E.H.S., Karypis, G. and Kumar, V., 2001, April. Text categorization using weight adjusted k-nearest neighbor classification. In *Pacific-asia conference on knowledge discovery and data mining* (pp. 53-65). Springer, Berlin, Heidelberg.

Harrag, F., Al-Salman, A.S. and Alquahtani, A., 2020. Arabic Opinion Mining Using a Hybrid Recommender System Approach. *arXiv preprint arXiv:2009.07397*.

Heise, D.R., 2007. *Expressive order: Confirming sentiments in social actions*. Springer Science & Business Media.

Holes, C. (2004). *Modern Arabic*. 1st ed. Washington, D.C.: Georgetown University Press.

Hu, M. and Liu, B., 2004, August. Mining and summarizing customer reviews. In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 168-177). ACM.

Ibrahim, H.S., Abdou, S.M. and Gheith, M., 2015. Idioms-proverbs lexicon for modern standard Arabic and colloquial sentiment analysis. *arXiv preprint arXiv:1506.01906*.

Ibrahim, H.S., Abdou, S.M. and Gheith, M., 2015. Sentiment analysis for modern standard Arabic and colloquial. *arXiv preprint arXiv:1505.03105*.

Ismail, H., Harous, S. and Belkhouche, B., 2016. A Comparative Analysis of Machine Learning Classifiers for Twitter Sentiment Analysis. *Research in Computing Science*, *110*, pp.71-83.

Ismail, R., Omer, M., Tabir, M., Mahadi, N. and Amin, I., 2018, August. Sentiment Analysis for Arabic Dialect Using Supervised Learning. In *2018 International Conference on Computer, Control, Electrical, and Electronics Engineering (ICCCEEE)* (pp. 1-6). IEEE.

Itani, M., Roast, C. and Al-Khayatt, S., 2017, April. Corpora for sentiment analysis of Arabic text in social media. In *2017 8th international conference on information and communication systems (ICICS)* (pp. 64-69). IEEE.

Jaafar, Y., Namly, D., Bouzoubaa, K. and Yousfi, A., 2017. Enhancing Arabic stemming process using resources and benchmarking tools. *Journal of King Saud University-Computer and Information Sciences*, *29*(2), pp.164-170.

Jarrar, M., Habash, N., Akra, D.F. and Zalmout, N., 2014. Building a corpus for Palestinian Arabic: a preliminary study.

Jarrar, M., Habash, N., Alrimawi, F., Akra, D. and Zalmout, N., 2017. Curras: an annotated corpus for the Palestinian Arabic dialect. *Language Resources and Evaluation*, *51*(3), pp.745-775.

Kalemtayeb. (2019). *الكلم الطيب - موسوعة الفوائد والحكم والأدعية والأذكار والأقوال المأثورة*. [online] Available at: https://kalemtayeb.com/, (Accessed 13 Jul. 2019).

Kanaan, G., Al-Shalabi, R., Ababneh, M. and Al-Nobani, A., 2008, December. Building an effective rule-based light stemmer for Arabic language to improve search effectiveness. In *2008 International Conference on Innovations in Information Technology* (pp. 312-316). IEEE.

Kang, H., Yoo, S.J. and Han, D., 2012. Senti-lexicon and improved Naïve Bayes algorithms for sentiment analysis of restaurant reviews. Expert Systems with Applications, 39(5), pp.6000-6010.

Keshtkar, F. and Inkpen, D., 2013. A bootstrapping method for extracting paraphrases of emotion expressions from texts. *Computational Intelligence*, *29*(3), pp.417-435.

Khalifa, K. and Omar, N., 2014. A Hybrid method using Lexicon-based Approach and Naive Bayes Classifier for Arabic Opinion Question Answering. *JCS*, *10*(10), pp.1961-1968.

Khalifa, S., Habash, N., Abdulrahim, D. and Hassan, S., 2016. A large scale corpus of Gulf Arabic. *arXiv preprint arXiv:1609.02960*.

Khalifa, S., Habash, N., Eryani, F., Obeid, O., Abdulrahim, D. and Al Kaabi, M., 2018, May. A morphologically annotated corpus of Emirati Arabic. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*.

Khalifa, S., Hassan, S. and Habash, N., 2017, April. A morphological analyzer for Gulf Arabic verbs. In *Proceedings of the Third Arabic Natural Language Processing Workshop* (pp. 35-45).

Khalifa, S., Zalmout, N. and Habash, N., 2016, December. Yamama: Yet another multi-dialect Arabic morphological analyzer. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: System Demonstrations* (pp. 223-227).

Khalife, L. (2019). *Fix Your CV: Saudi Arabia to create over half a million jobs by 2030*. [online] StepFeed. Available at: https://stepfeed.com/fix-your-cv-saudi-arabia-to-create-over-half-a-million-jobs-by-2030-3425 [Accessed 16 Jun. 2019].

Khalil, H. and Osman, T., 2014, March. Challenges in information retrieval from unstructured Arabic data. In *2014 UKSim-AMSS 16th International Conference on Computer Modelling and Simulation* (pp. 456-461). IEEE.

Khasawneh, R.T., Wahsheh, H.A., Alsmadi, I.M. and AI-Kabi, M.N., 2015, April. Arabic sentiment polarity identification using a hybrid approach. In *2015 6th International Conference on Information and Communication Systems (ICICS)* (pp. 148-153). IEEE.

Khoja, S. and Garside, R., 1999. Stemming arabic text. *Lancaster, UK, Computing Department, Lancaster University*.

Kim, S.B., Han, K.S., Rim, H.C. and Myaeng, S.H., 2006. Some effective techniques for naive bayes text classification. *IEEE transactions on knowledge and data engineering*, *18*(11), pp.1457-1466.

Kiritchenko, S., Mohammad, S. and Salameh, M., 2016. Semeval-2016 task 7: Determining sentiment intensity of English and Arabic phrases. In *Proceedings of the 10th international workshop on semantic evaluation (SEMEVAL-2016)* (pp. 42-51).

Larkey, L.S., Ballesteros, L. and Connell, M.E., 2002, August. Improving stemming for Arabic information retrieval: light stemming and co-occurrence analysis. In *Proceedings of the 25th annual international ACM SIGIR conference on Research and development in information retrieval* (pp. 275-282). ACM.

Liu, B. and Zhang, L., 2012. A survey of opinion mining and sentiment analysis. In *Mining text data* (pp. 415-463). Springer, Boston, MA.

Maamouri, M., Bies, A., Buckwalter, T. and Mekki, W., 2004, September. The penn arabic treebank: Building a large-scale annotated arabic corpus. In *NEMLAR conference on Arabic language resources and tools* (Vol. 27, pp. 466-467).

Maamouri, M., Bies, A., Kulick, S., Ciul, M., Habash, N. and Eskander, R., 2014, January. Developing an Egyptian Arabic Treebank: Impact of Dialectal Morphology on Annotation and Tool Development. In *LREC* (pp. 2348-2354).

Maamouri, M., Graff, D., Bouziri, B., Krouna, S., Bies, A. and Kulick, S., 2010. Standard Arabic morphological analyzer (SAMA) version 3.1. *Linguistic Data Consortium, Catalog No.: LDC2010L01*.

Maghfour, M. and Elouardighi, A., 2018, October. Standard and dialectal Arabic text classification for sentiment analysis. In *International Conference on Model and Data Engineering* (pp. 282-291). Springer, Cham.

Magumba, M.A., Nabende, P., and Mwebaze, E. 2018. Ontology boosted deep learning for disease name extraction from Twitter messages. *Journal of Big Data*, 5(31), 1-19.

Mahmoud, A. and Elghazaly, T., 2018. Using Twitter to monitor political sentiment for Arabic slang. In *Intelligent Natural Language Processing: Trends and Applications* (pp. 53-66). Springer, Cham.

Mahyoub, F.H., Siddiqui, M.A. and Dahab, M.Y., 2014. Building an Arabic sentiment lexicon using semi-supervised learning. *Journal of King Saud University-Computer and Information Sciences*, *26*(4), pp.417-424.

Mala, P.R. and Devi, S.S., 2017, June. Product response analytics in Facebook. In *2017 International Conference on Intelligent Computing and Control Systems (ICICCS)* (pp. 1265-1269). IEEE.

Manning, C., Raghavan, P. and Schütze, H., 2010. Introduction to information retrieval. *Natural Language Engineering*, *16*(1), pp.100-103.

Mataoui, M.H., Zelmati, O. and Boumechache, M., 2016. A proposed lexicon-based sentiment analysis approach for the vernacular Algerian Arabic. *Res. Comput. Sci*, *110*, pp.55-70.

McCarus, E.N., 2008. Modern Standard Arabic. *Encyclopedia of Arabic Language and Linguistics III, Leiden–Boston–Cologne: Brill*, pp.238-262.

MCIT - Media Center , 2020. *Over 18 Million Users Of Social Media Programs And Applications In Saudi Arabia*. [online] Mcit.gov.sa. Available at: <https://www.mcit.gov.sa/en/media-center/news/89698> [Accessed 3 November 2020].

Mohammad, A.H., Alwada'n, T. and Al-Momani, O., 2016. Arabic text categorization using support vector machine, Naïve Bayes and neural network. *GSTF Journal on Computing (JoC)*, *5*(1), p.108.

Mohammad, S., 2016, June. A practical guide to sentiment annotation: Challenges and solutions. In *Proceedings of the 7th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis* (pp. 174-179).

Mohammed, T., 2016. Review of Sentiment Analysis for Classification Arabic Tweets. International Journal of Emerging Technology and Advanced Engineering ISSN 2250-2459, ISO 9001:2008 Certified Journal. [online] 6(3). Available at: http://www.ijetae.com/files/Volume6Issue3/IJETAE_0316_06.pdf. [

Monroe, W., Green, S. and Manning, C.D., 2014, June. Word segmentation of informal Arabic with domain adaptation. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)* (pp. 206-211).

More, A., Çetinoğlu, Ö., Çöltekin, Ç., Habash, N., Sagot, B., Seddah, D., Taji, D. and Tsarfaty, R., 2018, May. CoNLL-UL: Universal morphological lattices for Universal Dependency parsing. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018).*

Mourad, A. and Darwish, K., 2013, June. Subjectivity and sentiment analysis of modern standard Arabic and Arabic microblogs. In *Proceedings of the 4th workshop on computational approaches to subjectivity, sentiment and social media analysis* (pp. 55-64).

Mridula, A. and Kavitha, C.R., 2018, April. Opinion mining and sentiment study of tweets polarity using machine learning. In *2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT)* (pp. 621-626). IEEE.

Mulki, H., Haddad, H., Gridach, M. and Babaoğlu, I., 2017, August. Tw-star at semeval-2017 task 4: Sentiment classification of arabic tweets. In *Proceedings of the 11th international workshop on semantic evaluation (SEMEVAL-2017)* (pp. 664-669).

Mumtaz, D. and Ahuja, B., 2016, July. Sentiment analysis of movie review data using Senti-lexicon algorithm. In *2016 2nd International Conference on Applied and Theoretical Computing and Communication Technology (iCATccT)* (pp. 592-597). IEEE.

Nabil, M., Aly, M. and Atiya, A., 2015, September. Astd: Arabic sentiment tweets dataset. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing* (pp. 2515-2519).

Nabil, M., Aly, M. and Atiya, A., 2015. Astd: Arabic Sentiment Tweets Dataset. In Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, Lisbon, Portugal, September 2015. Association for Computational Linguistics, pp. 2515–2519. Available at: http://www.aclweb.org/anthology/D15-1299.

Napoleoncat, 2020. *Facebook Users In Egypt - January 2020*. [online] Napoleoncat.com. Available at: <https://napoleoncat.com/stats/facebook-users-in-egypt/2020/01> [Accessed 3 November 2020].

Naz, S., Sharan, A. and Malik, N., 2018, December. sentiment classification on twitter data using support vector machine. In *2018 IEEE/WIC/ACM International Conference on Web Intelligence (WI)* (pp. 676-679). IEEE.

Nogueira, E. and Tsunoda, D.F., 2018. A Proposed Model for Consumer-Based Brand Equity Analysis on Social Media Using Data Mining and Social Network Analysis. *Journal of Relationship Marketing*, *17*(2), pp.95-117.

Novak, P.K., Smailović, J., Sluban, B. and Mozetič, I., 2015. Sentiment of emojis. *PloS one*, *10*(12), p.e0144296.

Noy, N.F. and McGuinness, D.L., 2001. Ontology development 101: A guide to creating your first ontology.

Nuseir, A., Al-Ayyoub, M., Al-Kabi, M., Kanaan, G. and Al-Shalabi, R., 2017. Improved Hierarchical Classifiers for Multi-Way Sentiment Analysis. *International Arab Journal of Information Technology (IAJIT), 14.*

O'Keefe, T. and Koprinska, I., 2009, December. Feature selection and weighting methods in sentiment analysis. In *Proceedings of the 14th Australasian document computing symposium, Sydney* (pp. 67-74).

Obeid, O., Khalifa, S., Habash, N., Bouamor, H., Zaghouani, W. and Oflazer, K., 2018. Madari: A web interface for joint Arabic morphological annotation and spelling correction. *arXiv preprint arXiv:1808.08392*.

Obeid, O., Zalmout, N., Khalifa, S., Taji, D., Oudah, M., Alhafni, B., Inoue, G., Eryani, F., Erdmann, A. and Habash, N., 2020, May. Camel tools: An open source python toolkit for Arabic natural language processing. In *Proceedings of The 12th Language Resources and Evaluation Conference* (pp. 7022-7032).

Olson, D.L., Delen, D. and Meng, Y., 2012. Comparative analysis of data mining methods for bankruptcy prediction. *Decision Support Systems*, *52*(2), pp.464-473.

Ortigosa, A., Martín, J.M. and Carro, R.M., 2014. Sentiment analysis in Facebook and its application to e-learning. *Computers in Human Behavior*, *31*, pp.527-541.

Oussous, A., Benjelloun, F.Z., Lahcen, A.A. and Belfkih, S., 2019. ASA: A framework for Arabic sentiment analysis. *Journal of Information Science*, p.0165551519849516.

Pak, A. and Paroubek, P., 2010, May. Twitter as a corpus for sentiment analysis and opinion mining. In *LREc* (Vol. 10, No. 2010, pp. 1320-1326).

Palmer, F.R., 2001. *Mood and modality*. Cambridge: Cambridge University Press. PLANK, FRANS. *The modals story retold*. Studies in Language 8:305-64

Pang, B. and Lee, L., 2008. Opinion mining and sentiment analysis. *Foundations and Trends® in Information Retrieval*, *2*(1–2), pp.1-135.

Pang, B., Lee, L. and Vaithyanathan, S., 2002, July. Thumbs up?: sentiment classification using machine learning techniques. In *Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10* (pp. 79-86). Association for Computational Linguistics.

Pasha, A., Al-Badrashiny, M., Diab, M.T., El Kholy, A., Eskander, R., Habash, N., Pooleery, M., Rambow, O. and Roth, R., 2014, May. Madamira: A fast, comprehensive tool for morphological analysis and disambiguation of arabic. In *LREC* (Vol. 14, pp. 1094-1101).

Polanyi, L. and Zaenen, A., 2006. Contextual valence shifters. In *Computing attitude and affect in text: Theory and applications* (pp. 1-10). Springer, Dordrecht.

Qassim, A., 2011. Translation grammatically viewed. *University of Baghdad*. [online] Available at: http://www.uobaghdad.edu.iq/uploads/PDFs/book.pdf [Accessed 15 May. 2017].

Rahab, H., Zitouni, A. and Djoudi, M., 2017, September. SIAAC: Sentiment polarity identification on Arabic Algerian newspaper comments. In *Proceedings of the Computational Methods in Systems and Software* (pp. 139-149). Springer, Cham.

Refaee, E. and Rieser, V., 2014, May. An Arabic twitter corpus for subjectivity and sentiment analysis. In *LREC* (pp. 2268-2273).

Refaee, E. and Rieser, V., 2014, May. Subjectivity and sentiment analysis of arabic twitter feeds with limited resources. In *Workshop on Free/Open-Source Arabic Corpora and Corpora Processing Tools Workshop Programme* (p. 16).

Rizkallah, S., Atiya, A., Mahgoub, H.E. and Heragy, M., 2018, February. Dialect versus MSA sentiment analysis. In *International Conference on Advanced Machine Learning Technologies and Applications* (pp. 605-613). Springer, Cham.

Rubin, A. and Babbie, E.R., 2016. *Empowerment series: Research methods for social work*. Cengage Learning.

Ryding, K. (2005). *A reference grammar of modern standard Arabic*. 1st ed. Cambridge: Cambridge University Press.

Saadane, H. and Habash, N., 2015. A conventional orthography for Algerian Arabic. In *Proceedings of the Second Workshop on Arabic Natural Language Processing* (pp. 69-79).

Sabtan, Y.M.N., 2018. Towards Corpus-Based Stemming for Arabic Texts. *International Journal of Linguistics, Literature and Translation*, *1*(4), pp.119-129.

Sadegh, M. and Othman, R.I.Z.A., 2012. Combining lexicon-based and learning-based methods for twitter sentiment analysis. *International Journal of Computers & Technology*, *2*(3), pp.171-178.

Safavian, S.R. and Landgrebe, D., 1991. A survey of decision tree classifier methodology. *IEEE transactions on systems, man, and cybernetics*, *21*(3), pp.660-674.

Sakaki, T., Okazaki, M. and Matsuo, Y., 2010, April. Earthquake shakes Twitter users: real-time event detection by social sensors. In *Proceedings of the 19th international conference on World wide web* (pp. 851-860). ACM.

Salem, Y., 2009. A generic framework for Arabic to English machine translation of simplex sentences using the Role and Reference Grammar linguistic model. [online] Available at: http://linguistics.buffalo.edu/people/faculty/vanvalin/rrg/Yasser_Salem_MSc_thesis.pdf [Accessed 25 Jun. 2017].

Salloum, W. and Habash, N., 2014. ADAM: Analyzer for dialectal Arabic morphology. *Journal of King Saud University-Computer and Information Sciences*, *26*(4), pp.372-378.

Sameer, R.A., 2016. Modified light stemming algorithm for Arabic language. *Iraqi Journal of Science*, *57*(1B), pp.507-513.

Samih, Y., Attia, M., Eldesouki, M., Abdelali, A., Mubarak, H., Kallmeyer, L. and Darwish, K., 2017, April. A neural architecture for dialectal Arabic segmentation. In *Proceedings of the Third Arabic Natural Language Processing Workshop* (pp. 46-54).

Sawalha, M., Atwell, E. and Abushariah, M.A., 2013, February. SALMA: standard Arabic language morphological analysis. In *2013 1st International Conference on Communications, Signal Processing, and their Applications (ICCSPA)* (pp. 1-6). IEEE.

Sayed, A.A., Elgeldawi, E., Zaki, A.M. and Galal, A.R., 2020, February. Sentiment Analysis for Arabic Reviews using Machine Learning Classification Algorithms. In *2020 International Conference on Innovative Trends in Communication and Computer Engineering (ITCE)* (pp. 56-63). IEEE.

Sghaier, M.A. and Zrigui, M., 2016, September. Sentiment Analysis for Arabic e-commerce websites. In *2016 International Conference on Engineering & MIS (ICEMIS)* (pp. 1-7). IEEE.

Shahrour, A., Khalifa, S. and Habash, N., 2015, September. Improving arabic diacritization through syntactic analysis. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing* (pp. 1309-1315)..

Shahrour, A., Khalifa, S., Taji, D. and Habash, N., 2016, December. Camelparser: A system for arabic syntactic analysis and morphological disambiguation. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: System Demonstrations* (pp. 228-232).

Sharath, T.S., Tandon, S. 2017. Topic based sentiment analysis using deep learning. *arXiv*, 1-8.

Shearlaw, M., 2016. Egypt Five Years On: Was It Ever a 'Social Media Revolution'?. *The Guardian*. *The Guardian*. [online] Available at: https://www.theguardian.com/world/2016/jan/25/egypt-5-years-on-was-it-ever-a-social-media-revolution [Accessed 15 Jun. 2019]

Shoukry, A. and Rafea, A., 2012, November. Preprocessing Egyptian dialect tweets for sentiment mining. In *The Fourth Workshop on Computational Approaches to Arabic Script-based Languages* (p. 47).

Shoukry, A. and Rafea, A., 2015, April. A hybrid approach for sentiment classification of Egyptian Dialect Tweets. In *2015 First International Conference on Arabic Computational Linguistics (ACLing)* (pp. 78-85). IEEE.

Somprasertsri, G. and Lalitrojwong, P., 2010. Mining Feature-Opinion in Online Customer Reviews for Opinion Summarization. *J. UCS*, *16*(6), pp.938-955.

Statista. 2020. Most common languages on the internet. Statista, Available at: https://www.statista.com/statistics/262946/share-of-the-most-common-languages-on-the-internet/.

Stojanovski, D., Strezoski, G., Madjarov, G., Dimitrovski, I., Chorbev, I. 2018. Deep neural network architecture for sentiment analysis and emotion identification of Twitter messages. *Multimedia Tools Applications*, 77, 32213-32242.

Suchdev, R., Kotkar, P., Ravindran, R. and Swamy, S., 2014. Twitter Sentiment Analysis using Machine Learning and Knowledge-based Approach. *International Journal of Computer Applications*, *103*(4).

Taghva, K., Elkhoury, R. and Coombs, J., 2005, April. Arabic stemming without a root dictionary. In *International Conference on Information Technology: Coding and Computing (ITCC'05)-Volume II* (Vol. 1, pp. 152-157). IEEE.

Taji, D., Habash, N. and Zeman, D., 2017, April. Universal dependencies for Arabic. In *Proceedings of the Third Arabic Natural Language Processing Workshop* (pp. 166-176).

Takezawa, T., Kikui, G., Mizushima, M. and Sumita, E., 2007, September. Multilingual spoken language corpus development for communication research. In *International Journal of Computational Linguistics & Chinese Language Processing, Volume 12, Number 3, September 2007: Special Issue on Invited Papers from ISCSLP 2006* (pp. 303-324).

Tan, S., 2005. Neighbor-weighted k-nearest neighbor for unbalanced text corpus. *Expert Systems with Applications*, *28*(4), pp.667-671.

Thelwall, M.,2013, Heart and soul: Sentiment strength detection in the social web with sentistrength (summary book chapter). *Cyberemotions: Collective emotions in cyberspace. Berlin, Germany: Springer*, pp.119-134.

Toutanova, K., Klein, D., Manning, C.D. and Singer, Y., 2003, May. Feature-rich part-of-speech tagging with a cyclic dependency network. In *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology-Volume 1* (pp. 173-180). Association for computational Linguistics.

Tsytsarau, M. and Palpanas, T., 2012. Survey on mining subjective data on the web. *Data Mining and Knowledge Discovery*, *24*(3), pp.478-514.

Turban, E., Sharda, R. and Delen, D., 2014. *Business intelligence and analytics: systems for decision support*. Pearson Higher Ed.

Turney, P. D. (2002). Thumbs up or Thumbs down?: Semantic Orientation Applied to Unsupervised Classification of Reviews. In Proceedings of the 40th annual meeting on association for computational linguistics, pages 417–424. Association for Computational Linguistics

Wiebe, J. M.; Bruce, R. F. and O'Hara, T. P. (1999). Development and Use of a GoldStandard Data Set for Subjectivity Classifications. In: Proceedings of the Association for Computational Linguistics (AC L-1999).

Wightwick, J. and Gaafar, M., 2005. *Easy Arabic Grammar*. New York: McGraw-Hill.

Wilson, T., Wiebe, J. and Hoffmann, P., 2005. Recognizing contextual polarity in phrase-level sentiment analysis. In *Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing*.

Wonder, J. (2019). *The State of Social Media 2019 | Crowd Analyzer*. [online] Digital Marketing Community. Available at: https://www.digitalmarketingcommunity.com/researches/the-state-of-social-media-in-middle-east-2019-crowd-analyzer/ [Accessed 12 Dec. 2019].

Yi, J., Nasukawa, T., Bunescu, R. and Niblack, W., 2003, November. Sentiment analyzer: Extracting sentiments about a given topic using natural language processing techniques. In *Third IEEE international conference on data mining* (pp. 427-434). IEEE.

Yu, H. and Kim, S., 2012. SVM tutorial—classification, regression and ranking. *Handbook of Natural computing*, pp.479-506.

Zalmout, N. and Habash, N., 2017, September. Don't throw those morphological analyzers away just yet: Neural morphological disambiguation for Arabic. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing* (pp. 704-713).

Zalmout, N., Erdmann, A. and Habash, N., 2018, June. Noise-robust morphological disambiguation for dialectal Arabic. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)* (pp. 953-964).

Zerrouki, T., 2020. Towards An Open Platform For Arabic Language Processing.

Zhang, X. and Yu, Q., 2017, September. Hotel reviews sentiment analysis based on word vector clustering. In *2017 2nd IEEE International Conference on Computational Intelligence and Applications (ICCIA)* (pp. 260-264). IEEE.

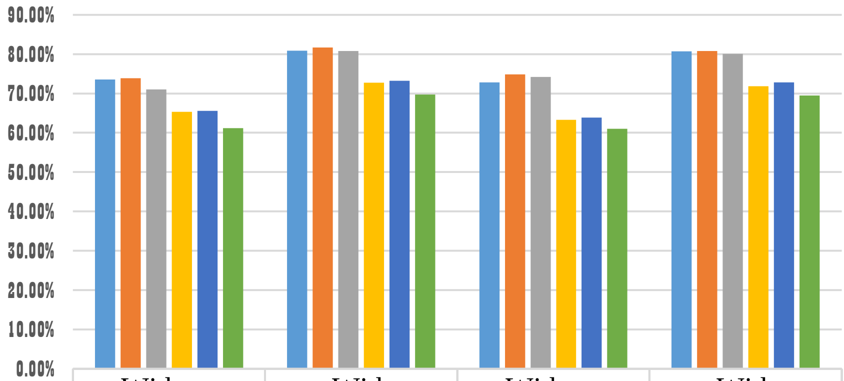
Zhou, S., Chen, Q. and Wang, X., 2014. Fuzzy deep belief networks for semi-supervised sentiment classification. *Neurocomputing*, *131*, pp.312-322.

Zribi, I., Khemakhem, M.E. and Belguith, L.H., 2013, October. Morphological analysis of Tunisian dialect. In *Proceedings of the Sixth International Joint Conference on Natural Language Processing* (pp. 992-996).

10 Appendix

Appendix A

The Impact of Pre-processing on Machine Learning Classifiers for Dialectical Arabic Content

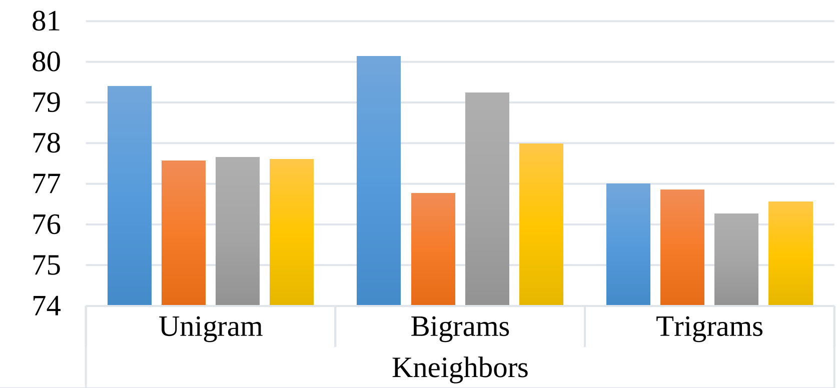


	Accuracy: Without Preprocessing	Accuracy: With Preprocessing	F-Score: Without Preprocessing	F-Score: With Preprocessing
BernouliNB Unigram	73.51%	80.84%	72.85%	80.67%
BernouliNB Bigrams	73.89%	81.71%	74.85%	80.81%
BernouliNB Trigrams	71.05%	80.81%	74.18%	80.02%
SVC Unigram	65.35%	72.74%	63.25%	71.84%
SVC Bigrams	65.58%	73.21%	63.83%	72.79%
SVC Trigrams	61.14%	69.73%	61.02%	69.49%

Figure 10.1: The Impact of Pre-processing on Machine Learning Classifiers

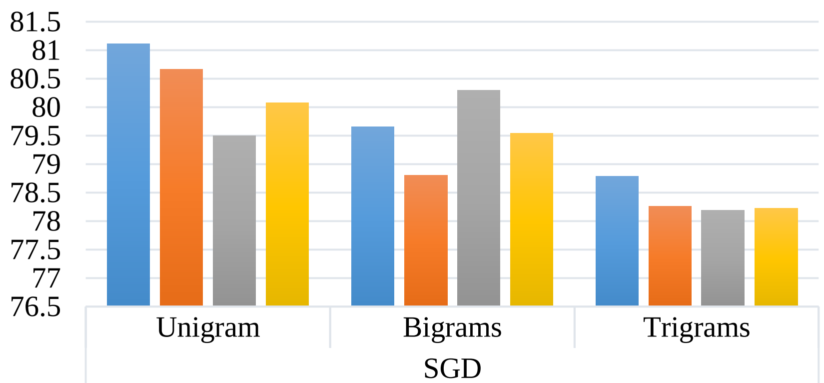
Appendix B

1. **The results regarding KNeighbors and SGD for Machine Learning Baseline Experiments**



	Unigram	Bigrams	Trigrams
Accuracy	79.4	80.13	77
Recall	77.57	76.77	76.86
precision	77.65	79.24	76.26
F-score	77.61	77.99	76.56

Figure 10.2: The results of Machine Learning Baseline - KNeighbors classifiers

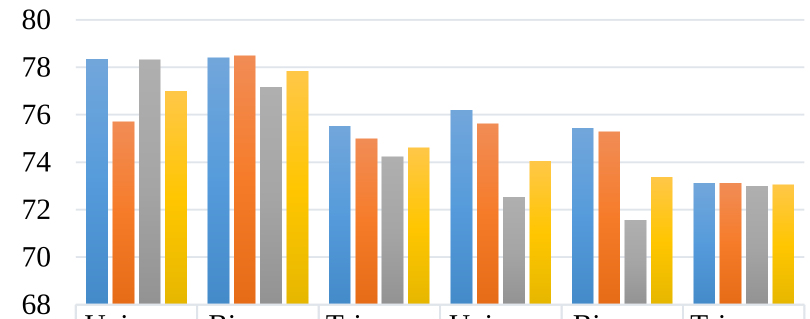


	Unigram	Bigrams	Trigrams
Accuracy	81.12	79.66	78.79
Recall	80.67	78.81	78.26
precision	79.5	80.3	78.19
F-score	80.08	79.55	78.22

Figure 10.3: The results of Machine Learning Baseline - SGD classifiers

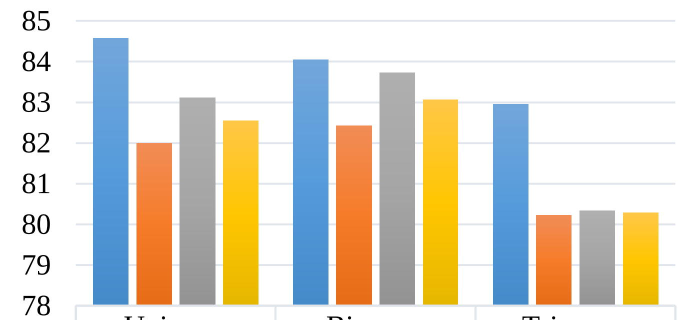
2. **The result and discussion of Machine Learning Techniques with weighted schemes features for Trees and KNeighbors classifiers results**

2.1. Machine Learning Techniques with TF-IDF features



	DecisionTree			RandomForest		
	Unigram	Bigrams	Trigrams	Unigram	Bigrams	Trigrams
Accuracy	78.33	78.4	75.5	76.18	75.43	73.12
Recall	75.69	78.49	74.98	75.62	75.27	73.11
precision	78.32	77.16	74.24	72.53	71.55	72.98
F-score	76.98	77.82	74.61	74.04	73.36	73.04

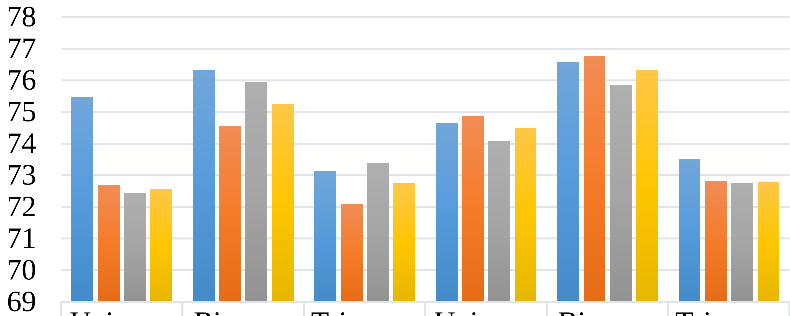
Figure 10.4: The results of Machine Learning with TF-IDF features - Tree classifiers



	Kneighbors		
	Unigram	Bigrams	Trigrams
Accuracy	84.58	84.05	82.96
Recall	81.99	82.42	80.22
precision	83.11	83.73	80.34
F-score	82.55	83.07	80.28

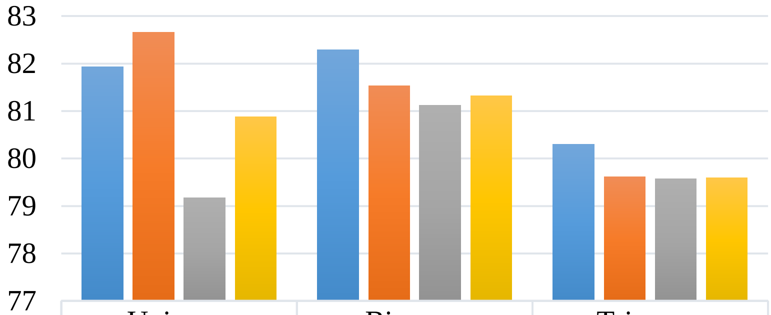
Figure 10.5: The results of Machine Learning with TF-IDF features - KNeighbors classifiers

2.2.Machine Learning Techniques with binary features



	DecisionTree			RandomForest		
	Unigram	Bigrams	Trigrams	Unigram	Bigrams	Trigrams
Accuracy	75.47	76.32	73.14	74.65	76.58	73.49
Recall	72.67	74.56	72.08	74.88	76.77	72.81
precision	72.43	75.95	73.39	74.07	75.86	72.73
F-score	72.55	75.25	72.73	74.47	76.31	72.77

Figure 10.6: The results of Machine Learning with binary features – Tree classifiers

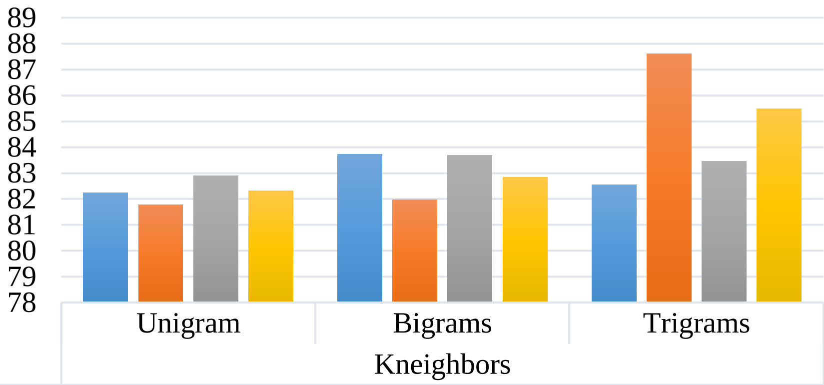


	Kneighbors		
	Unigram	Bigrams	Trigrams
Accuracy	81.93	82.29	80.3
Recall	82.66	81.53	79.62
precision	79.17	81.12	79.57
F-score	80.88	81.32	79.59

Figure 10.7: The results of Machine Learning with binary features – KNeighbors classifiers

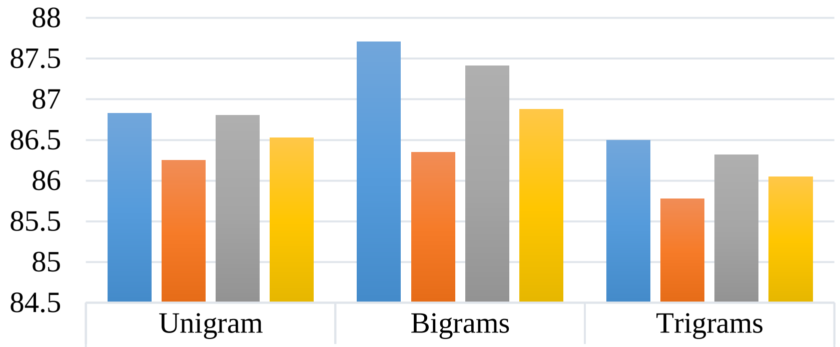
Appendix C

1. **Hybrid approach for TF-IDF features + semantic features (domain features) for KNeighbors and SGD classifiers results.**



	Unigram	Bigrams	Trigrams
Accuracy	82.24	83.74	82.55
Recall	81.77	81.98	87.61
precision	82.89	83.7	83.45
F-score	82.33	82.83	85.48

Figure 10.8: The results of Hybrid approach (TF-IDF features + semantic features) - KNeighbors classifiers



SGD	Unigram	Bigrams	Trigrams
Accuracy	86.83	87.71	86.5
Recall	86.25	86.35	85.78
precision	86.8	87.41	86.32
F-score	86.52	86.88	86.05

Figure 10.9: The results of Hybrid approach (TF-IDF features + semantic features) - SGD classifiers