Forecasting Movements of Health-Care Stock Prices Based on Different Categories of News Articles using Multiple Kernel Learning

Yauheniya Shynkevich[1,*], T.M. McGinnity[1,2], Sonya Coleman[1], Ammar Belatreche[1]

[1]Intelligent Systems Research Centre, Ulster University, BT48 7JL, Derry, UK

[2]School of Science and Technology, Nottingham Trent University, Nottingham, UK

Abstract —The market state changes when a new piece of information arrives. It affects decisions made by investors and is considered to be an important data source that can be used for financial forecasting. Recently information derived from news articles has become a part of financial predictive systems. The usage of news articles and their forecasting potential have been extensively researched. However, so far no attempts have been made to utilise different categories of news articles simultaneously. This paper studies how the concurrent, and appropriately weighted, usage of news articles, having different degrees of relevance to the target stock, can improve the performance of financial forecasting and support the decision-making process of investors and traders. Stock price movements are predicted using the multiple kernel learning technique which integrates information extracted from multiple news categories while separate kernels are utilised to analyse each category. News articles are partitioned according to their relevance to the target stock, its sub industry, industry, group industry and sector. The experiments are run on stocks from the Health Care sector and show that increasing the number of relevant news categories used as data sources for financial forecasting improves the performance of the predictive system in comparison with approaches based on a lower number of categories.

Keywords—stock price prediction; financial news; text mining; multiple kernel learning; decision support systems

* Corresponding author.
E-mail address: shynkevich-y@email.ulster.ac.uk

Abbreviations: SS (stock-specific), SIS (sub-industry-specific), IS (industry-specific), GIS (group-indusrty-specific), SeS (sector-specific)

1. INTRODUCTION

Investors make investment decisions based on the information available to market participants. News articles bring new information to the market. They contain news about a company, the activities in which it is involved, its fundamentals and what is expected by market participants about its future price changes [1], [2]: stock prices are driven by these publications. With the development of the internet, finance-related websites and applications constantly provide a large amount of textual data containing new information. A system capable of efficiently utilising this new data to predict future changes in prices is required to support the decision making of investors and traders. Researchers have been studying the influence of news articles and developed several automated frameworks that consider large amounts of financial news. These frameworks extract relevant information and employ it to forecast prices and their changes [3]. As has been shown in previous research [4], there is a strong relationship between stock prices fluctuations and publications of relevant news. The effect that news items have on stock prices has been studied using existing data mining techniques [5], [6], [7]. According to the related literature, researchers usually employ a predefined criterion for selecting news articles from a large collection of textual information. Generally, only news articles highly relevant to an analysed stock are selected. After that, equal importance is given to all articles so that every article is treated as impacting the stock price to the same extent. So far no previous studies employ articles that are divided into different news categories and analysed simultaneously yet differently based on their relevance to the analysed stock, which is the focus of this paper.

This paper investigates whether financial news articles that have different degrees of relevance to the target stock can provide an advantage in financial news-based forecasting when used simultaneously and appropriately. Toward this end, the considered stocks are assigned to the corresponding sub industries, industries, group industries and sectors according to the Global Industry Classification Standard (GICS) as in [8]. Then news published about these stocks are allocated to different news categories. We consider five news categories; these are stock-specific (SS), sub-industry-specific (SIS), industry-specific (IS), group-industry-specific (GIS) and sector-specific (SeS) news items. The experiments are performed on stocks from the S&P 500 index belonging the Health Care sector. News

categories are formed from a large database downloaded from the LexisNexis database. News items are allocated to the corresponding categories based on their relevance to the target stock. The SS subset of data includes articles that are only relevant to the target stock. News articles, that are relevant to at least one stock from a list of stocks belonging to the target stock's sub industry, are assigned to the SIS subset of news. Similarly, news articles, relevant to all stocks within the relevant industry, group industry and sector to which the target stock belongs, form the IS, GIS and SeS subsets respectively. A detailed explanation of how the news are allocated to different categories is given in Section 3.2.

Integration of different data types is often performed by the Multiple Kernel Learning (MKL) method [9], [10], [11], [12]. Several kernels are used for learning different data subsets. MKL is applied in this study and it utilises from two to fifteen kernels assigned to either SS, SIS, IS, GIS or SeS subset of articles. The results show that an attempt to allocate news articles into different categories, pre-process them separately, learn from them and integrate their predictions into a single prediction decision improve the prediction performance in comparison with approaches based on a single news subset.

The remainder of the paper is organized as follows. Section 2 gives an overview of the relevant literature. Section 3 discusses the raw dataset, data pre-processing techniques, machine learning approaches and performance metrics utilised for analysis. Section 4 describes the experimental results. Section 5 concludes the research work and outlines directions for future work.

2. RELATED WORK

An extensive review of the research articles published about financial predictions using text mining is presented in [5]. All systems employing text mining for financial prediction have some of the components illustrated in Fig. I. Textual data obtained from online sources and market price data are used as an input to the predictive system, and values predicting the market are outputted from it.
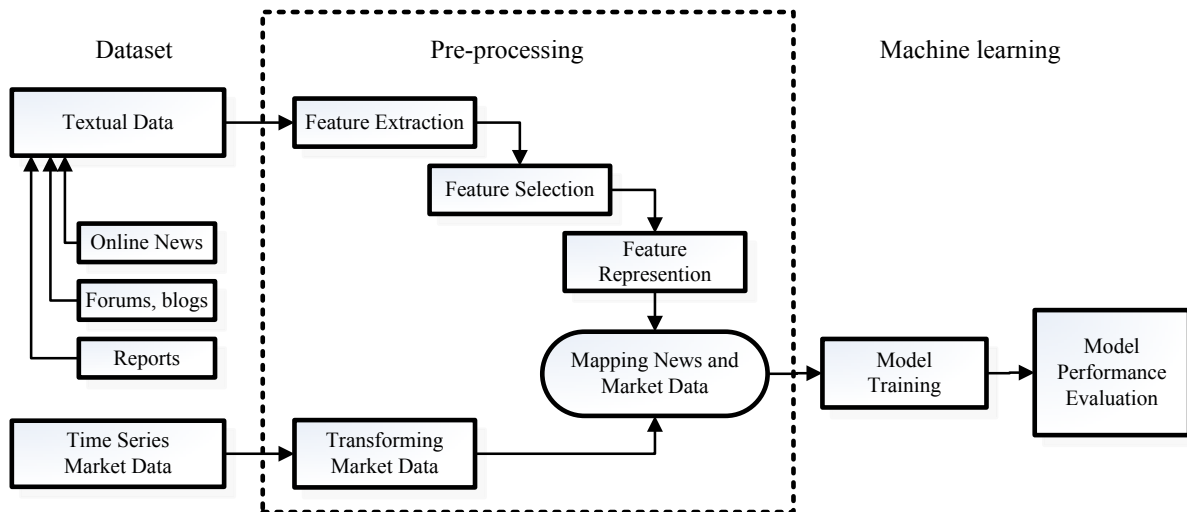
Figure I. Typical components of the news-based financial forecasting system.

2.1 Early works

Wüthrich et al. [13] were the first to try to use textual information for financial forecasting. The authors used knowledge of a domain expert to obtain a dictionary of terms that were later used to assign feature weightings and generate probabilistic rules. Daily price changes were predicted for five stock indices and a trading strategy was formed based on the predictions. The resulting returns were positive and confirmed that profit can be gained with the use of financial news. Lavrenko et al. [14] proposed the Analyst system that employed language models, utilised time series of prices and classified news articles. The authors showed that the designed system is capable of producing profit. Gidofalvi and Elkan [15] developed a system that predicted short term price movements using news articles. Articles were scored using linear regression to the NASDAQ index and assigned with a "down", "unchanged" or "up" label. The authors stated that the behaviour of stock prices is strongly correlated with the information in news articles starting from 20 minutes prior to 20 minutes after its publication. Headlines of news published about companies were examined in [16]. The authors claimed that bad news enforced a strong negative market drift. In [17], official company reports were considered and their ability to indicate future performance of a firm was shown. For instance, a change in written style of documents may indicate a significant change in firm's productivity.

2.2 Key Related Research

Approaches to the financial forecasting that exist in the literature mainly differ in three general aspects: the dataset, the textual pre-processing methods and the machine learning algorithm. Correspondingly, Table I reviews the key related research relevant to the work presented in this paper and provides details about the choices of datasets, textual pre-processing and machine learning techniques made in those papers.

Schumaker and Chen [8] tried to group financial news by similar sectors and industries and studied the predictability of related stock prices based on the news. The authors showed that the ability to predict stock prices varies for different news groups. Schumaker and Chen used only one news group at a time and examined the forecasting performance achieved using articles from the whole dataset of news or relevant to either a stock, its sub industry, industry, group industry or sector. The research proposed in this paper adopts an idea to partition articles by sectors and industries from [8] to create subsets of news articles divided according to their relevance to the target stock. However, these subsets are used simultaneously in order to benefit from news published about the target stock and other stocks across the target stock's industry and sector. The proposed predictive system employs the concurrent use of news articles from all categories. To the best of our knowledge, no existing research has focussed on the simultaneous use of financial news items from different industrial categories and sub categories. Therefore, this paper investigates the importance of including news articles having different stock relevance levels to forecast stock price changes.

Hagenau, Liebmann and Neumann [18] designed a stock price prediction system that uses text mining to automaticaly read corporate announcements and financial news articles and employs market reaction for feature selection process using the Chi-square and bi-normal separation methods, which permit a choice of semantically relevant features. The number of feature extraction methods used in the proposed predictive system and the feedback-based selection of features helped reach a high level of accuracy of 76%. These high results were achieved on several datasets employed in the study, and a simple trading strategy applied to test the system on simulated trading demonstrated its potentially high

profitability. This paper employs the Chi-square method proposed in [18] to select features based on the market reaction to news releases.

Table I. Summary of the most influential works (ordered by relevance to this paper)

Authors	Dataset		Text pre-processing				Machine learning	Forecast type
	Data source	Forecasting target	Feature extraction	Feature selection	Feature re-presentation	Market feedback		
Schumaker and Chen [8]	Financial news	Intraday stock prices	Proper nouns	Minimum occurrence per document	Binary	No	SVR	Price value
Hagenau et al. [1]	Corporate announcement and financial news	Daily stock prices	- Dictionary-based - Bag-of-words - 2-gram - 2-word combination - Noun phrases	- News frequency - Chi-square - Bi-normal-separation	TF-IDF	Yes	SVM	Positive and negative
Luss and D'Aspremont [19]	Press releases from PRNewswire	Intraday stock prices	Bag-of-words	Pre-defined dictionary	TF-IDF	No	MKL	Abnormal and normal returns
Schumaker and Chen [6]	Financial news	Intraday stock prices	- Bag-of-words - Noun phrases - Named entities - Proper nouns	Minimum occurrence per document	Binary	No	SVR	Price value
Mittermayer [20]	Financial news	Intraday stock prices	Bag-of-words	TF-IDF, selecting 1000 terms	TF-IDF	No	SVM	Good news, bad news, no movers
Groth and Muttermann [21]	Adhoc announcement	Daily stock prices	Bag-of-words	Feature scoring using information gain and Chi-square metrics	TF-IDF	No	Naïve Bayes; kNN; ANN; SVM	Positive and negative

Luss and d'Aspremont [19] have studied the predictability of abnormal returns using text and return data. The predictions were made from 10 to 250 minutes after the publication of news articles using intraday data, and news articles published by PRNewswire during an eight year period from 2000 to 2007 were used as textual data for predictions. MKL with several kernels was successfully used to learn from text and price data. The authors highlighted that MKL permits the use of several kernels with different parameters to analyse the same set of data and enhance the prediction performance of the system. In [6], Schumaker and Chen studied the role of financial news using four textual representation methods, bag-of-words, noun phrases, named entities and proper nouns, using the developed AZFinText system. The authors concluded that financial news articles contain useful information valuable for financial forecasting and that the proper nouns technique achieved better textual representation performance than others. Mittermayer [20] developed the NewsCATS (news categorization and trading

system) to predict trends in stock prices immediately after the publication of news releases. The author categorized news articles into three classes: good news, no movers and bad news. Good (bad) news led to at least 3% increase (decrease) at some point during 60 minutes after a news release and had an average price during this period at least 1 % above (below) the price at the moment of a news release. The system was tested on intraday stock price data and the results highlight that it is possible to significantly outperform a random trader by employing predictions made by NewsCATS in trading strategies. The author stated that there is still a lot of room for improvement in the developed system. Groth and Muntermann [21] proposed an intraday risk management approach that makes use of unstructured qualitative data by mining text of adhoc announcements. The approach is designed to forecast market volatility; it classified news items into high volatility-entailing and normal. The authors showed that intraday exposures of market risk can be discovered through text mining and that nowadays technology is able to extract useful information from corporate disclosures and utilise it for risk management purposes.

2.3 Textual Pre-processing

Once news articles are selected, text data pre-processing is required. The target is to extract relevant information from a dataset of news and to prepare it for machine learning. Words and phrases that signal a price change are important and should be extracted. In [20], Mittermayer suggested to divide the textual pre-processing into three major steps: extraction, selection and representation of features. This terminology was then employed in subsequent works [1].

The feature extraction step refers to the process of generating a list of features, which are words or phrases extracted from the documents, that describe the documents sufficiently. According to [5], the bag-of-words approach is the most popular feature extraction method in financial forecasting based on news articles. It is often preferred due to its simplicity and intuitive meaning. In this method, the raw text is cleaned of punctuation marks, pronouns, prepositions and articles. Next, semantically empty terms are removed and the word stemming methods are applied to every word in order to treat different forms of a word as a single feature. The remaining words are used as features that represent the article.

During the feature selection procedure, the most expressive features are chosen from all extracted features, and those containing the least information are eliminated [20]. Some researchers used a dictionary of domain experts selected terms [13]. Others utilise statistical information of term frequencies in news articles, e.g. the Term Frequency - Inverse Document Frequency (TF*IDF) values [10], [19], [20], [22]. Lately, the external market feedback was suggested for use in a number of research papers. In [11], the Chi-square test is chosen to select features for volatility forecasting. Hagenau et al. [1] investigated the effectiveness of the bi-normal separation method and Chi-square test for evaluating the term explanatory ability. Both methods utilised the external market feedback and showed promising results.

Once expressive features are selected, the whole set of news must be represented in a format suitable for applying a machine learning technique. For instance, a vector of n feature elements is constructed for each data point. Usually a feature presence in an article is considered to be an important factor. In the trading system developed in [23], the membership value for each term was computed and then features were represented using the binary format. Other research works utilised real values to assign feature weights. In [19], Luss and d'Aspremont predicted abnormal returns and used TF*IDF to calculate feature weights. In [11], the volatility changes were forecast with TF*IDF values used as weights. After the completion of the text pre-processing steps, the articles are aligned with price time series and subsequently labelled. The documents are often classified into two (negative and positive), e.g. in [1] [21], or three (negative, neutral and positive) categories, e.g. in [20], classes depending on their impact on an asset price. In some papers such as [6] and [8] the stock price value instead of the direction of its change was predicted based on published news.

2.4 Machine learning techniques

When all the preparatory steps are completed, a machine learning approach is usually used to learn from the data and to predict the market reaction. A number of artificial intelligence approaches are generally employed to learn from financial documents, for instance, Support Vector Machines (SVM) [1], [4], [22], Artificial Neural Networks (ANN) [24], k-Nearest Neighbours (kNN) and Naïve Bayes [15]. In [2], Support Vector Regression (SVR) was employed to investigate the impact of financial news

on the Chinese stock market. The authors showed that publications of online financial news items negatively influence the market. In [21], results achieved by the ANN, SVM, Naïve Bayes and kNN classifiers were compared. An approach for supporting risk management and investment decision making was designed using textual analysis and machine learning. Considering both classification results and efficiency of computations, the authors recommended the SVM classifier. In [25], the Naïve Bayes and SVM approaches were applied where messages were classified into bearish, neutral or bullish. Naïve Bayes underperformed in comparison to SVM as measured by the out-of-sample accuracy. In [1], the SVM method classified the effect that a message had on the market price into two classes, positive and negative. The authors mentioned that a pilot comparison of SVM, ANN and Naïve Bayes showed that SVM outperformed the two other techniques. Taking into consideration previous findings, SVM is regarded as a prominent machine learning approach for text mining [1].

Currently, ensemble methods (computational intelligence approaches integrating the results from a set of base learners) are actively employed for forecasting financial markets. The predictions made by the base learners may be enhanced with the help of these methods [26]. The MKL approach combines several kernels and can be used for learning from different kinds of features. Recently researchers have started to employ it for financial forecasting to combine different features, for example extracted from price data and financial news [9], [10], [11], [12]. Luss and d'Aspremont [19] employed the MKL approach with separate kernels assigned to text features and time series of absolute returns. The results were compared to those of MKL utilising textual data only and stock return data only. The majority of kernel weights were assigned to kernels analysing textual data, nevertheless, a combination of both data sources produced higher accuracy and Sharpe ratio than any single data source solely. Therefore, the main finding of the paper is that combining information such as news articles and stock returns for predicting abnormal returns produces promising results and improves the performance in comparison with predictions made based on a single source of data. In [10], these two sources of information were analysed using MKL, and results confirmed that the MKL method outperformed models based on a single information source or a simple feature combination. In [11], MKL with RBF (radial basis function) kernels were proposed to predict movements of volatility and demonstrated higher

performance than methods based on a single kernel. Both papers, [10] and [11], analysed news articles written in traditional Chinese. Therefore, the developed predictive systems were not evaluated on English news. In [9], MKL was used in a stock price prediction system that integrated several sources of information: numerical dynamics of news and corresponding comments such as frequencies of their publications, semantic analysis of their content and time series of prices. The model extracts features and forms separate subsets of features for each source of data; each subset is then analysed by MKL. However, no existing literature provides evidence of employing MKL for analysis of different news categories for financial predictions.

Based on their popularity in the related literature, the bag-of-words approach is employed for feature extraction in this paper, the Chi-square test is applied for feature selection and the TF*IDF values are selected to compute feature weights. This study utilises MKL as the primary machine learning approach to learn from different news categories and employs SVM and kNN that learn from one news category at a time for comparison.

3. THE PROPOSED APPROACH

Details about the designed news-based predictive system are given in this section. We explain how news articles are assigned to different categories, discuss the raw textual data and its pre-processing techniques, and describe the machine learning approaches used and the performance metrics employed for evaluation. Fig. II provides an overview of the proposed predictive system that is discussed in detail in the following subsections. News articles are assigned to different categories based on their relevance to the target stock. Each category is then pre-processed separately and different sets of features are extracted for each of them. Daily prices are employed for selecting the most expressive features and for labelling data points. Then MKL, with separate kernels used for learning from different feature subsets, is applied. The system is validated and then evaluated using performance measures.
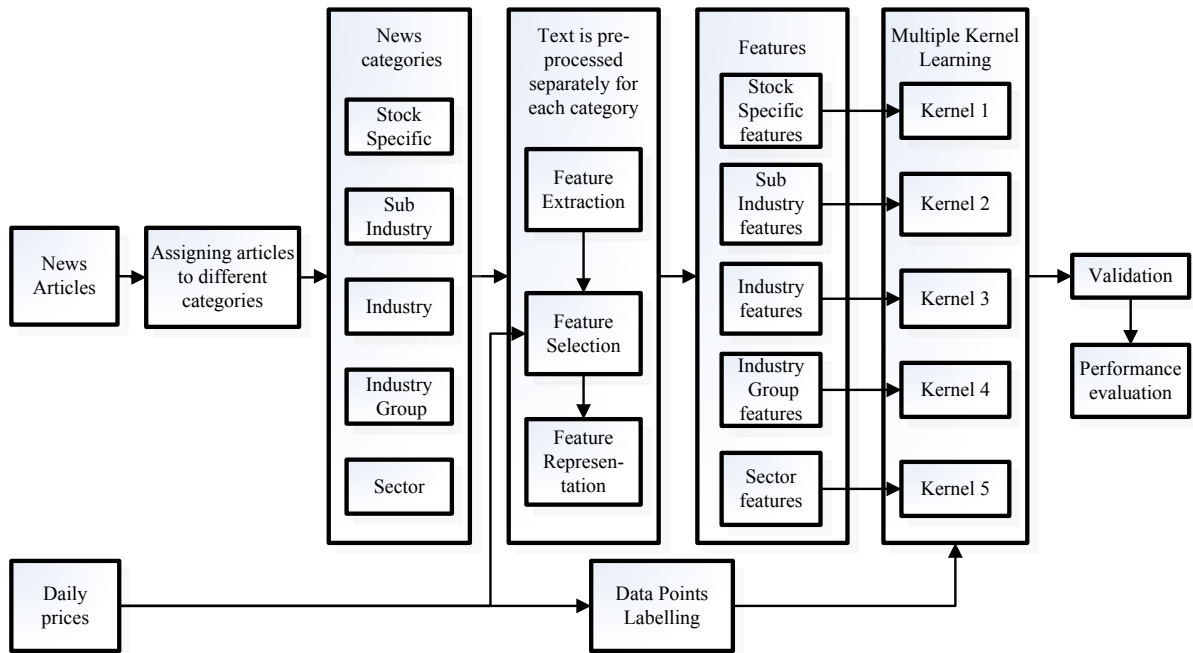
Figure II. An overview of the proposed predictive system

3.1 Industry Classification of News Articles

News articles are grouped by sub industries, industries, group industries and sectors according to the Global Industry Classification Standard (GICS) which was developed by the Standard & Poor's (S&P) and Morgan Stanley Capital International companies to support research and asset management. According to GICS, companies are assigned with a sub industry, industry, group industry and sector to which they belong. In [8], GICS was employed by Schumaker and Chen to explore the benefits of grouping financial news articles by similar sectors and industries before using them for forecasting. In the current study, five news categories are utilised. The categories refer to the target stock and other stocks from the target stock's sub industry, industry, group industry and sector. Here, 28 stocks that are included in the S&P 500 stock market index and belong to the Health Care Equipment and Services group industry are selected as target stocks for forecasting. Only stocks having more than 200 articles released during the period of study are included. Details about the considered stocks and their allocation to sub industry, industry, group industry and sector are given in Table II.

Table II. Description of Analysed Stocks and Datasets

Company Name	# data points	'Up' labelled data points, %	'Down' labelled data points, %	Stock	Sub Industry	Industry	Group Industry	Sector
Medtronic plc	715	53.93	46.07	MDT	Health Care Equipment	Health Care Equipment & Supplies	Health Care Equipment & Services	Health Care
Agilent Technologies Inc	691	55.81	44.19	A				
Abbott Laboratories	569	49.30	50.70	ABT				
Boston Scientific Corp.	542	57.78	42.22	BSX				
Johnson & Johnson	508	49.61	50.39	JNJ				
Baxter International Inc	463	45.22	54.78	BAX				
PerkinElmer Inc	451	52.68	47.32	PKI				
Becton, Dickinson and Co.	359	57.10	42.90	BDX				
Thermo Fisher Scientific, Inc.	337	46.88	53.12	TMO				
Varian Medical Systems, Inc.	325	56.31	43.69	VAR				
CR Bard Inc.	235	53.62	46.38	BCR				
CareFusion Corp.	229	54.59	45.41	CFN				
Hospira Inc.	201	50.25	49.75	HSP				
Covidien plc	440	51.82	48.18	COV				
St. Jude Medical Inc.	416	47.12	52.88	STJ				
Bristol-Myers Squibb Co.	647	54.66	45.34	BMY	Health Care Distributors	Health Care Providers & Services		
Express Scripts Holding Co.	373	51.21	48.79	ESRX				
Cardinal Health, Inc.	263	49.05	50.95	CAH				
McKesson Corp.	520	53.85	46.15	MCK				
Quest Diagnostics Inc.	320	47.81	52.19	DGX	Health Care Facilities			
DaVita HealthCare Partners Inc.	291	52.92	47.08	DVA				
Lab. Corp. of America Holdings	205	56.10	43.90	LH				
Tenet Healthcare Corp.	287	50.17	49.83	THC				
Aetna Inc.	844	52.61	47.39	AET	Managed Health Care			
Cigna Corp.	812	57.14	42.86	CI				
UnitedHealth Group Inc.	598	51.68	48.32	UNH				
Humana Inc.	486	56.20	43.80	HUM				
WellPoint, Inc.	480	55.00	45.00	WLP				

3.2 News Articles Data

A five-year period, which started on September 1, 2009, and finished on September 1, 2014 was selected to study the importance of including news articles having different relevance. News articles that mention stocks of interest and are released during this period were obtained from the LexisNexis database. This database contains news published by major newspapers and was used in previous studies, e.g. in [27] Fang and Peress studied the relationship between a firm's media coverage and their average returns using news articles downloaded from LexisNexis. Three providers that showed sufficient media coverage of the considered stocks were selected: PR Newswire, McClatchy-Tribune Business News and Business Wire. An important feature of the LexisNexis database is that additional information such as

relevant companies and their relevance scores supplement its news articles. A relevance score is expressed as a percentage that represents the degree of relevance of a news article to a given company. The dataset of news articles was downloaded from the LexisNexis database on October 30, 2014. On that day, 53 stocks of the S&P 500 index were allocated to the Health Care sector according to the GICS. In order to analyse the importance of including news articles relevant to the whole sector, all news published during the analysed period by the considered news providers and relevant to at least one of the 53 stocks were downloaded from the LexisNexis database. As a result, a large dataset of news was retrieved where the total number of news articles was equal to 51,435. Table III gives details about the number of articles retrieved per news provider. The following information is saved for every article: heading, body, month, day and year, lists of relevant companies, their tickers and corresponding relevance scores. The date of publication is made up of the day, month and year values. The heading and body are concatenated into a pool of words and used as the raw text for information extraction.

Table III. Number of Articles per News Providers

News providers	# news articles	Percentage of news articles
PR Newswire	18,767	36.5%
McClatchy-Tribune Business News	6,603	12.8%
Business Wire	26,065	50.7%
Total	51,435	100.0%

A subset of articles relevant to the target stock is formed in the following way. To define how relevant an article is to a company, its tickers and relevance scores are checked. Every article is examined to consider whether the target company's ticker is included in a list of relevant companies' tickers linked to that article. If the target ticker is present among relevant tickers of the article and its corresponding relevance score is more than or equal to 85%, then the article is selected and included in the SS subset. In [27], only articles having a relevance score equal to or higher than 90% are analysed. In this paper, a slightly lower threshold of 85% was selected in order to include a bigger number of articles in the analysis.

To form the SIS subset for the target stock, the following steps are taken. First, a list of companies belonging to the same sub industry where the target stock belongs to is identified. For example, when predictions are made for the Aetna stock (ticker AET) which belongs to the Managed Health Care sub industry, a list of companies from this sub industry includes companies with tickers AET, CI, UNH, HUM and WLP (see Table II). Second, the whole dataset of 51,435 news articles is examined so that every article is checked whether its list of relevant tickers contains either AET, CI, UNH, HUM or WLP. If this condition is satisfied, then the relevance score of the found ticker is checked and, if it is equal to or higher than 85%, then the article is included in the SIS subset of news. Once each article from the original dataset is examined, the SIS data subset is formed. A similar procedure is followed when forming the IS, GIS and SeS subsets for the target stock: every article from the original dataset is checked to determine that at least one company belonging to the target stock's industry, group industry and sector respectively, is present among article's companies, and then that its relevance score is more than or equal to 85%. If both conditions are satisfied, then the article is added to the corresponding data subset.

After all news articles are assigned to the corresponding SS, SIS, IS, GIS and/or SeS subsets, the following procedure is carried out separately for each textual data subset. The articles released on the same day are checked for uniqueness. This step is necessary to remove articles downloaded several times or republished by several news sources. Then all unique news articles released on the same day are concatenated and treated as a single document. After that, only the dates for which there is at least one article published about the target stock are kept. Thus, price movements are predicted only for days following publications of target stock related articles. All news articles published on other days are neglected. The number of data instances for every stock is equal to the number of dates when a relevant publication is released.

3.3 Historical Prices Data

Time series of a stock price are used in feature selection and data labelling. Yahoo! Finance, a publicly available website, is chosen as a provider of historical daily prices as in [28]. The most expressive features are selected based on the market reaction to the publication of a news item. The reaction is derived from a movement of a stock price defined as the difference between the open and close prices

on the next trading day following the day of publication. Data instances are classified into two classes in this paper. Labels 'Up' or 'Down' that correspond to an increase or a decrease in a price of the target stock, respectively, are given to each data point. Daily prices are used in the analysis to compute the amplitude of a price movement. Previous studies of financial forecasting from news articles used daily price observations [13], [25] and showed that the market adapts to new information slowly and its reaction can be explored and studied using daily data. Details about the stocks used, the number of data instances for each stock and fractions of each class are given in Table II.

3.4 Textual Data Pre-processing

Textual data pre-processing is an essential part of text mining, and is particularly important for developing news-based predictive models. As mentioned in Section 2, the bag-of-words approach is employed for feature extraction. In every article, symbols other than letters from the English alphabet as well as hyperlinks, emails and website addresses are filtered out. Uppercase letters are transformed to lowercase. Words having only one or two characters and semantically empty words are removed. Then each word is stemmed using the Porter's stemming algorithm [29]. Word stems extracted from the data subset are examined and a list of unique features is formed, where each feature corresponds to a unique single word stem. Finally, features that appeared in less than three articles are eliminated.

In order to select features that carry the most important information, Chi-square values are computed for each unique feature based on the market reaction as a sum of normalized deviations of observed term frequency from its expected value[1]:

$$\chi_i^2 = \sum_{j=1}^{4} \left(O_{ij} - E_{ij} \right)^2 \Big/ E_{ij}, \tag{1}$$

where i is the order of a feature, O_{ij} and E_{ij} are its observed and expected frequencies of occurence in the news dataset respectively, and j refers to four possible outcomes: the feature appeared among positive news, $j=1$; it appeared among negative news, $j=2$; it did not appear among positive news, $j=3$; it did not appear among negative news, $j=4$. News articles were considered to be positive or negative depending on whether the stock price increased or decreased on the next trading day after the news publication. The observed frequency of appearing in positive news is computed as a fraction of positive

15

articles where the feature occurred. The observed frequencies of appearing among negative news and not appearing among positive or negative news are computed in a similar way. When a feature does not carry any positive or negative meaning, it is likely to occur uniformly among all documents. Thus, the expected frequency of appearing in positive or negative articles is the overall frequency of appearing in all documents. Similarly, the expected frequency of not appearing in positive or negative articles is the overall frequency of not appearing within the whole dataset of news. Consequently, a feature that appears uniformly in positive and negative articles has a zero Chi-square value. On the opposite side, a feature that appears more often in either positive or negative articles has a Chi-square value significantly higher than zero.

After the Chi-square values for each feature are calculated, unique features are sorted in descending order according to their corresponding scores. 500 terms that have the highest Chi-square scores are chosen and used as an input into the machine learning technique. This is consistent with the approach in [1], where Hagenau et al. selected 567 features using bag-of-words.

The final preliminary step is to convert subsets of articles to a format suitable for applying a machine learning technique. In this paper each news article is represented as a vector of 500 TF*IDF values each of which corresponds to a feature. If a feature is not present in an article, then it has a zero TF*IDF value. Therefore, a sparse matrix of size [number of data points]*500 is constructed. It is important to note that the above described procedure is applied separately to the SS, SIS, IS, GIS and SeS subsets of documents. Lists of unique features extracted for each subset differ from each other. Therefore, feature matrices formed for each subset are also different. When pre-processing is completed, each data instance is assigned an 'Up' or 'Down' label. As a result, each instance has 500 feature values for each of the five subsets and a label.

3.5 Machine Learning Techniques

The MKL applied to the prepared dataset is based on a linear combination of sub-kernels:

$$K_{comb}(x,y) = \sum_{j=1}^{K} \beta_j K_j(x,y), \qquad (2)$$

where $\beta_j \geq 0$ and $\sum_{j=1}^{K} \beta_j = 1$, $K_{comb}(x,y)$ is a kernel combined from K sub-kernels $K_j(x,y)$ using weights β_j learnt during a training process. A separate kernel or several kernels can be assigned to each news category. In this work we employ MKL with various combinations of linear, Gaussian and polynomial kernels. Five news categories, SS, SIS, IS, GIS and SeS, are considered and separate kernels are utilised to learn from them. To determine which combination of categories achieves the highest performance, several combinations are examined. When a single subset of news is utilised independently from others to forecast movements of stock price, SVM with either a linear, Gaussian or polynomial kernel or kNN is employed for learning. In this case, subsets are used as an input one by one. After that a combination of the SIS and SS subsets is fed into a MKL algorithm that uses different kernel types. Next, subsets of categories that included a broader range of news, IS, GIS and SeS, are added successively. All categories are treated in the same way. For this purpose, when a certain kernel type is used, separate kernels of this type are applied for learning from each news category. The most complex combination analyses five subsets with three kernel types assigned to each subset. Kernel weights that are learnt during the training procedure reflect the contribution of each individual kernel to the combined kernel. Algorithms implemented in the Shogun toolbox [30] for the MKL, SVM and kNN methods are utilised in this study. This toolbox was also used in previous studies [9], [11]. When training the MKL, its parameters and optimal weights are estimated concurrently by repeating the procedure employed for a simple SVM.

The training, validation and testing are performed separately for each stock whose dataset is split into training, validation and testing in a chronological order. Training of the predictive system is based on the first 50% of the instances. A validation phase is required to tune system's parameters and is conducted using the subsequent 25% of the instances. Tuning of the parameter C, which is a penalty rate for data misclassification, is required for both MKL and SVM. Additionally, the width of the Gaussian kernel and the polynomial degree are tuned during the validation phase. Optimal parameter values are determined using a grid search. C and gamma (γ) values are chosen from exponentially growing sequences $C=\{2^{-3}, 2^{-1}, \ldots, 2^{19}\}$ and $\gamma=\{2^{-15}, 2^{-13}, \ldots, 2^{-1}\}$, as suggested in [31]. The grid search is also used for finding an optimal number of neighbours, k_{opt}, for the kNN approach. A range of k

values is chosen according to an empirical rule of thumb suggested in [32] where k is set approximately equal to the square root of the total number of training instances. For the considered stocks, the number of training points varies from 101 to 422, and a slightly broader range of $k=\{5,6,\ldots,30\}$ is used. During the validation, the performance of the model with different parameter settings is measured by classification accuracy. For testing the developed predictive system on out-of-sample data, the remaining 25% of the instances are employed.

3.6 Performance Metrics

The forecasting accuracy and return from simulated trades are employed to evaluate the predictive performance of the employed techniques for each of the selected 28 stocks. Forecasting accuracy is used to measure classification performance of each machine learning technique. The prediction accuracy achieved by a single stock is computed using (3):

$$Accuracy = (TrueUp + TrueDown)/N$$

(3)

where N is the total number of classified data instances during the testing phase, $TrueDown$ and $TrueUp$ are correctly classified down and up movements respectively.

Determining the price direction is important when making predictions, however, identifying large price changes is significantly more important than identifying small changes. Incorrectly classified movements with returns close to zero have little effect on the total return from the trading system. Averaged return from simulated trades describes the performance of a predictive system from a trading point of view and hence the trades are simulated using the following procedure. When the system predicts an increase in a stock price (an 'Up' movement), it is treated as a signal to buy so that an amount X is invested in the stock of interest at the opening price on the next trading day. The acquired stocks are sold at the end of the day. The return per single trade is calculated as:

$$R_t = (C_t - O_t)/O_t$$

(4)

where O_t and C_t are the open and close stock prices on the trading day that followed the day of news publication respectively. When the system predicts a 'Down' price movement, it is regarded as a signal

to sell. In this case, assuming that an amount of money X is currently invested in the considered stock, the stocks are short sold at the opening price on the following trading day and bought back at the closing price of that day. Therefore, the return per single trade is calculated as:

$$R_t = (O_t - C_t)/O_t \qquad (5)$$

Returns obtained from single trades are averaged over the whole testing period for each stock and then the returns are averaged over 28 stocks to compare different techniques. In order to get a better understanding of the returns obtained using different techniques, the highest possible return is computed. The highest possible return would be achieved if all predictions made regarding the direction of the price movement were correct. The highest possible return is averaged over 28 stocks, its value is equal to 0.81% per trade with a standard deviation of 0.16%.

4. EXPERIMENTAL RESULTS

This section discusses the results produced by the designed news-based prediction system. Both forecasting accuracy and return shown in the tables of this section are averaged over 28 analysed stocks. Standard deviations are also reported for each metric preceded by the '±' sign. The value of the parameter C displayed in Tables IV and V is the most common value of this parameter during the validation process for these 28 stocks. In Tables IV and V, 'Acc.', 'R.' and 'C' correspond to the accuracy, return and parameter C used in MKL and SVM, respectively.

4.1 The SVM and kNN Approaches

News subsets created for each level of the GISC classification are employed for prediction independently from each other in order to investigate their usefulness before combining them together. The SVM and kNN approaches are employed for learning in this case. The experimental settings are similar to [8] where the predictions were made separately from news relevant to each GICS classification level, however, [8] utilised a universal set of news that combined all available articles. A universal dataset of news is not considered in this study. Table IV outlines the prediction results achieved by the SVM, with different kernel types, and kNN machine learning approaches applied to either SS, SIS, IS, GIS or SeS data subsets. The highest forecasting accuracy and return reached for

every subset are highlighted in bold. SVM performs better than kNN for all data subsets in terms of both performance measures. When comparing results achieved by different kernel types, the SVM method with a polynomial kernel performs on average slightly better than that with Gaussian and linear kernels. Nevertheless, all three kernel types performed comparatively well. It is worth noting that the forecasting accuracy increases with a broader range of articles.

TABLE IV. Experimental results obtained for the SVM and KNN approaches

Machine Learning Technique	Data subset														
	Stock-specific data			Sub-industry-specific data			Industry-specific data			Group-industry-specific data			Sector-specific data		
	Acc., %	R., %	C	Acc., %	R., %	C	Acc., %	R., %	C	Acc., %	R., %	C	Acc., %	R., %	C
SVM, Gaussian	66.11 ±6.34	**0.29** ±0.11	8	72.12 ±6.40	0.37 ±0.18	32	75.92 ±5.23	0.43 ±0.17	128	77.29 ±6.90	0.45 ±0.16	32	76.65 ±4.90	**0.43** ±0.14	512
SVM, Linear	**66.97** ±5.12	**0.29** ±0.11	512	72.57 ±5.81	**0.38** ±0.19	32768	75.68 ±5.99	0.43 ±0.18	32768	<u>**77.66**</u> ±7.84	<u>**0.47**</u> ±0.19	131072	75.92 ±5.74	0.42 ±0.15	131072
SVM, Polynomial	66.44 ±5.84	0.27 ±0.13	2	**73.02** ±6.53	0.37 ±0.16	2	**77.42** ±5.02	**0.45** ±0.15	2	76.86 ±6.74	0.44 ±0.15	2	**77.55** ±5.64	**0.43** ±0.15	2
kNN	55.73 ±6.10	0.11 ±0.11	-	58.02 ±4.55	0.14 ±0.11	-	58.55 ±6.77	0.16 ±0.15	-	56.96 ±8.41	0.14 ±0.14	-	56.93 ±6.40	0.12 ±0.12	-

In Table IV, the highest performance measures obtained among all subsets of data are underlined. The highest performance corresponds to the group industry subset of news articles. These results might be due to the following reasons. Firstly, news articles relevant to the group industry may contain some additional information that is useful for forecasting stock price changes but is missed in news articles relevant to the stock or its industry only. Secondly, news relevant to the whole sector may include too many articles containing little relevant information thus causing the prediction performance to deteriorate. Similar behaviour was observed in [8], where the highest prediction performance was achieved for the sector-based system and steadily decreased when more specific or more general (universal) news were added. The way in which experiments are conducted in this paper and in [8], for instance, the usage of different datasets or daily vs intraday data, are likely to cause the slight differences in the experimental results. In [8] the authors did not attempt to combine news articles from different categories. This paper aims to improve the forecasting performance achieved by SVM and kNN by

considering all news categories simultaneously. The following subsection presents the proposed approach.

4.2 The Proposed MKL Approach

Table V displays the experimental results obtained using the MKL approach with different combinations of kernels and data subsets. The highest forecasting accuracy and return values achieved for each data subset are marked in bold. The first column of Table V shows the results produced when the SS and SIS data subsets are combined using MKL. For the purpose of treating both subsets equally, different kernel types are taken in pairs. Thus, the same set of kernels is used to analyse each data subset. A number of kernel combinations are considered: two linear, two polynomial, two Gaussian, a combination of two linear and two polynomial, a combination of two linear and two Gaussian, a combination of two polynomial and two Gaussian, and finally a combination of two linear, two polynomial and two Gaussian kernels. The highest forecasting accuracy (74.95%) is reached when all Gaussian and polynomial kernel types are utilised. It is higher than the accuracies achieved by SVM and kNN for both SS and SIS subsets. Several kernel combinations produced a return of 0.42%, which is higher than those obtained using either SVM or kNN with SS or SIS data subsets. This is consistent with [33] and confirms that the simultaneous usage of the SS and SIS subsets and the employment of the MKL method for their analysis achieves better prediction performance than SVM and kNN based on a single data subset at a time.

Results of the concurrent employment of the SS, SIS and IS subsets are presented in the second column of Table V. Kernel combinations are formed as in the first column with three kernels of each type taken instead of two. Polynomial kernels produced the highest forecasting accuracy (78.77%) and return per trade (0.47%). A combination of linear and polynomial kernels showed the same values of accuracy and return, but linear kernels received zero weights for all 28 stocks. This indicates that the contribution from linear kernels to the resulting combined kernel is minimal and insignificant. As discussed in [33], the most likely reason why zero weights are assigned to the linear kernels when they are combined with polynomial and/or Gaussian kernels is the optimal value selected for the parameter C in MKL. For polynomial and Gaussian kernels this value typically lies in a range $[2:2^9]$. However,

for linear kernels it usually lies in a range [2^9:2^{17}]. Taking into account that, as shown in Table IV, generally polynomial kernels achieve higher prediction performance than linear kernels, the MKL approach selects a value of the parameter C that is more favourable for polynomial rather than linear kernels. This difference in optimal parameter values is likely to be the main factor for linear kernels having zero weights during the learning procedure when they are combined with polynomial and/or Gaussian kernels. Performance measures achieved using the MKL that uses three data subsets are greater than those for the MKL analysing two subsets, and than those of SVM and kNN learning from either SS, SIS or IS subset. These results confirm that adding industry related news in an appropriately weighted manner enhances the news-based prediction system.

TABLE V.   Experimental results for the MKL approach

| Data subset | | | | | | | | | | | | | | | |
| SS and SIS data | | | | SS, SIS and IS data | | | | SS, SIS, IS and GIS data | | | | SS, SIS, IS, GIS and SeS data | | | |
Kernels	Acc., %	R., %	C	Kernels	Acc., %	R., %	C	Kernels	Acc., %	R., %	C	Kernels	Acc., %	R., %	C
2 Gaussian	65.63 ±7.01	0.26 ±0.14	8192	3 Gaussian	69.29 ±8.39	0.35 ±0.18	8	4 Gaussian	68.84 ±8.08	0.32 ±0.17	0.5	5 Gaussian	65.14 ±8.96	0.27 ±0.19	0.5
2 linear	70.44 ±4.39	0.33 ±0.09	8192	3 linear	70.64 ±5.01	0.34 ±0.10	8192	4 linear	70.58 ±5.43	0.34 ±0.12	2048	5 linear	71.32 ±5.30	0.35 ±0.13	512
2 polynomial	74.24 ±5.46	0.39 ±0.16	2	3 polynomial	**78.77** ±4.72	**0.47** ±0.14	2	4 polynomial	**80.37** ±6.66	**0.51** ±0.16	8	5 polynomial	81.25 ±6.10	0.52 ±0.16	2
2 Gaussian & 2 linear	66.17 ±6.65	0.26 ±0.14	2	3 Gaussian & 3 linear	69.12 ±8.62	0.34 ±0.19	8	4 Gaussian & 4 linear	68.62 ±8.01	0.32 ±0.17	0.5	5 Gaussian & 5 linear	65.82 ±8.76	0.28 ±0.18	0.5
2 Gaussian & 2 polynomial	**74.95** ±5.88	**0.42** ±0.16	2	3 Gaussian & 3 polynomial	76.25 ±5.79	0.42 ±0.12	2	4 Gaussian & 4 polynomial	79.91 ±6.23	0.50 ±0.17	2	5 Gaussian & 5 polynomial	**81.31** ±6.39	**0.53** ±0.17	2
2 linear & 2 polynomial	74.33 ±5.29	0.39 ±0.16	2	3 linear & 3 polynomial	**78.77** ±4.72	**0.47** ±0.14	2	4 linear & 4 polynomial	**80.37** ±6.66	**0.51** ±0.16	8	5 linear & 5 polynomial	81.28 ±6.09	0.52 ±0.16	2
2 Gaussian, 2 linear & 2 polynomial	74.88 ±5.77	**0.42** ±0.16	2	3 Gaussian, 3 linear & 3 polynomial	76.14 ±5.80	0.42 ±0.12	2	4 Gaussian, 4 linear & 4 polynomial	79.91 ±6.23	0.50 ±0.17	2	5 Gaussian, 5 linear & 5 polynomial	**81.31** ±6.39	**0.53** ±0.17	2

Results obtained when four news categories are employed for learning are shown in the third column of Table V. The highest forecasting accuracy (80.37%) and return (0.51%) are again achieved when only polynomial kernels are used. As in the second column, the same performance is also observed for a combination of four linear and four polynomial kernels, but the linear kernels received zero weights. The highest accuracy and return in the third column are greater than those produced by MKL analysing two or three subsets, and those for the SVM and kNN approaches learning from any single subset. These

results demonstrate that adding news, relevant to the target stock's group industry, to the prediction system improves the forecasting results and therefore brings new useful information.

The fourth column of Table V presents the results obtained when all five news categories are employed for predicting price movements. The highest forecasting accuracy (81.31%) and return (0.53%) are reached using a combination of polynomial and Gaussian kernels, and a combination of kernels that included all three kernel types with zero weights learnt for the linear kernels. Comparing with the MKL results obtained when fewer subsets are considered, it is important to note that the simultaneous analysis of five news categories shows the best performance among all considered combinations. This prediction performance is also higher than that obtained by SVM or kNN based on any single subset. These results highlight the importance of including different categories of news with different relevance to the target stock for financial forecasting where separate sub-learners are explicitly specified for each category. The information in Tables IV and V emphasises the usefulness of appropriately distinguishing between different news categories, processing them separately and integrating the resultant knowledge at the final stage. Machine learning methods produce worse results when news articles relevant to industries and sectors are mixed up and not separately pre-processed. The proposed approach for predicting stock price movements that uses MKL for learning from multiple news categories demonstrates the improved performance in comparison to approaches that use a lower number of news categories.

Table VI provides additional information regarding the weights assigned to kernels during the training of MKL. It presents an example of weights that are learnt for five polynomial and five Gaussian kernels when the highest forecasting performance is reached. These weights are averaged over 28 analysed stocks. In general, polynomial kernels have higher weights than Gaussian kernels. However, non-zero weights are obtained for every subset of news and both types of kernels. It means that every news category contains some unique information useful for forecasting and contributes to the resulting prediction decision.

TABLE VI.   Weights assigned to different kernels and data subsets when the highest performance is achieved for MKL with ten kernels where one polynomial and one Gaussian kernels are assigned to each category of news

Data subset	Kernel Type		Total
	Gaussian	Polynomial	
Sector-specific	3.21%	26.52%	29.73%
Group-industry-specific	5.50%	14.92%	20.42%
Industry-specific	6.37%	11.46%	17.83%
Sub-industry-specific	5.19%	8.13%	13.32%
Stock-specific	9.74%	8.95%	18.69%
Total	30.01%	69.99%	100.00%

5. CONCLUSION AND FUTURE WORK

It is impossible for a single person to read all the news articles published about stocks and sectors of interest, therefore the investors and traders can potentially benefit from employing an automated system that can gauge information from multiple sources and accurately forecast the changes in the market prices. This research paper explores how the simultaneous analysis of different categories of financial news can lead to improved algorithms for financial forecasting. The proposed approach of using multiple news categories with the MKL technique has resulted in the improved performance when compared with SVM and kNN based on a single news category; methods based on a lower number of news categories demonstrated reduced performance. In this paper, news articles were allocated to five categories based on their relevance to a target stock and its sub industry, industry, group industry and sector. Every news category was pre-processed independently from others and five subsets of data were formed. MKL was used to learn from these news subsets so that independent kernels were utilised for each subset. Different kernel types and several kernel combinations were employed. Stocks from the Health Care sector were selected to test the proposed approach. The obtained results showed that the highest forecasting accuracy and trading return were reached for MKL with five news categories utilised and two kernels, polynomial and Gaussian, used for each category. Polynomial kernels received the highest kernel weights during the learning process meaning that polynomial kernels contributed the most to the final decision. SVM and kNN based on a single news category, either SS, SIS, IS, GIS or

SeS, showed worse performance than MKL. This indicated that allocating news articles to different categories according to their relevance to the target stock and employing separate kernels to model these categories helps the predictive system to capture more information and make more accurate predictions about the future price. A possible direction of future work is to include additional data sources such as historical prices and make predictions based on both textual and time series data. Additional kernels can be employed for different data sources. This may potentially improve the forecasting performance. Enhancing the textual pre-processing techniques utilised in the developed predictive system is another direction for future work.

6. ACKNOWLEDGEMENTS

REFERENCES

[1]   M. Hagenau, M. Liebmann, and D. Neumann, "Automated news reading: Stock price prediction based on financial news using context-capturing features," *Decision Support Systems*, vol. 55, no. 3, pp. 685–697, Jun. 2013.

[2]   X. Zhao, J. Yang, L. Zhao, and Q. Li, "The impact of news on stock market: Quantifying the content of internet-based financial news," in *Proceedings of the 11th International DSI and 16th APDSI Joint meeting*, 2011, pp. 12–16.

[3]   G. Mitra and L. Mitra, *The handbook of news analytics in finance*. Wiley-Finance, 2011.

[4]   G. Fung, J. Yu, and H. Lu, "The predicting power of textual information on financial markets," *IEEE Intelligent Informatics Bulletin*, vol. 5, no. 1, 2005.

[5]   A. K. Nassirtoussi, T. Y. Wah, S. R. Aghabozorgi, and D. N. C. Ling, "Text Mining for Market Prediction: A Systematic Review," *Expert Systems with Applications*, vol. 41, no. 16, pp. 7653–7670, Jun. 2014.

[6]   R. P. Schumaker and H. Chen, "Textual analysis of stock market prediction using breaking financial news: The AZFin text system," *ACM Transactions on Information Systems*, vol. 27, no. 2, pp. 1–19, Feb. 2009.

[7]   M. Mittermayer and G. F. Knolmayer, "Text mining systems for market response to news: A survey." Institute of Informations System, University of Bern, 2006.

[8]   R. Schumaker and H. Chen, "A quantitative stock prediction system based on financial news," *Information Processing & Management*, vol. 45, no. 5, pp. 571–583, Sep. 2009.

[9]   S. Deng, T. Mitsubuchi, K. Shioda, T. Shimada, and A. Sakurai, "Combining Technical Analysis with Sentiment Analysis for Stock Price Prediction," in *Proceedings of the IEEE Ninth International Conference on Dependable, Autonomic and Secure Computing*, 2011, pp. 800–807.

[10]   X. Li, C. Wang, J. Dong, and F. Wang, "Improving stock market prediction by integrating both market news and stock prices," *Database and Expert Systems Applications, Lecture Notes in Computer Science*, vol. 6861, pp. 279–293, 2011.

[11]   F. Wang, L. Liu, and C. Dou, "Stock Market Volatility Prediction: A Service-Oriented Multi-kernel Learning Approach," in *Proceedings of IEEE Ninth International Conference on Services Computing*, 2012, vol. d, pp. 49–56.

[12]   C.-Y. Yeh, C.-W. Huang, and S.-J. Lee, "A multiple-kernel support vector regression approach for stock market price forecasting," *Expert Systems with Applications*, vol. 38, no. 3, pp. 2177–2186, Mar. 2011.

[13]   B. Wuthrich, D. Permunetilleke, and S. Leung, "Daily prediction of major stock indices from textual www data," in *Proceedings of the 4th international conference on knowledge discovery and data mining*, 1998.

[14]   V. Lavrenko, M. Schmill, and D. Lawrie, "Mining of concurrent text and time series," in *Proceedings of the 6h ACM SIGKDD Int'l Conference on Knowledge Discovery and Data Mining*, 2000.

[15]   G. Gidófalvi and C. Elkan, "Using news articles to predict stock price movements," *Department of Computer Science and Engineering, University of California*. 2001.

[16]   W. S. Chan, "Stock price reaction to news and no-news: drift and reversal after headlines," *Journal of Financial Economics*, vol. 70, no. 2, pp. 223–260, Nov. 2003.

[17]   A. Kloptchenko, T. Eklund, B. Back, J. Karlsson, H. Vanharanta, and A. Visa, "Combining data and text mining techniques for analysing financial reports," *International Journal of Intelligent Systems in Accounting and Finance Management*, vol. 12, no. 1, pp. 29 – 41, 2004.

[18]   M. Hagenau, M. Liebmann, M. Hedwig, and D. Neumann, "Automated News Reading: Stock Price Prediction Based on Financial News Using Context-Specific Features," in *Proceedings of the 45th Hawaii International Conference on System Sciences*, 2012, pp. 1040–1049.

[19]   R. Luss and A. D'Aspremont, "Predicting abnormal returns from news using text classification," *Quantitative Finance*, vol. 15, no. 6, pp. 999–1012, 2015.

[20]   M. Mittermayer, "Forecasting intraday stock price trends with text mining techniques," in *Proceedings of the 37th Annual Hawaii International Conference on System Sciences*, 2004, pp. 1–10.

[21]   S. S. Groth and J. Muntermann, "An intraday market risk management approach based on textual analysis," *Decision Support Systems*, vol. 50, no. 4, pp. 680–691, Mar. 2011.

[22]   Y. Zhai, A. Hsu, and S. K. Halgamuge, "Combining news and technical indicators in daily stock price trends prediction," in *Proceedings of the 4th International Symposium on Neural Networks: Advances in Neural Networks*, 2007, pp. 1087–1096.

[23]   G. Rachlin, M. Last, D. Alberg, and A. Kandel, "Admiral: A data mining based financial trading system," in *Proceedings of the IEEE Symposium on Computational Intelligence and Data Mining*, 2007, no. Cidm, pp. 720–725.

[24]   S. Simon and A. Raoot, "Accuracy Driven Artificial Neural Networks in Stock Market Prediction," *International Journal on Soft Computing*, vol. 3, no. 2, pp. 35–44, May 2012.

[25]   W. Antweiler and M. Frank, "Is all that talk just noise? The information content of internet stock message boards," *The Journal of Finance*, vol. 59, no. 3, pp. 1259–1294, 2004.

[26]   C. Cheng, W. Xu, and J. Wang, "A Comparison of Ensemble Methods in Financial Market Prediction," in *Proceedings of the Fifth International Joint Conference on Computational Sciences and Optimization*, 2012, pp. 755–759.

[27]   L. Fang and J. Peress, "Media Coverage and the Cross-section of Stock Returns," *The Journal of Finance*, vol. LXIV, no. 5, pp. 2023–2052, 2009.

[28]   Y. Shynkevich, T. M. McGinnity, S. Coleman, Y. Li, and A. Belatreche, "Forecasting stock price directional movements using technical indicators: investigating window size effects on one-step-ahead forecasting," in *Proceedings of the IEEE Conference on Computational Intelligence for Financial Engineering & Economics*, 2014, pp. 341–348.

[29]   M. Porter, *An algorithm for suffix stripping*. 1980, pp. 313–316.

[30]   S. Sonnenburg and G. Rätsch, "The SHOGUN machine learning toolbox," *Journal of Machine Learning Research*, vol. 11, pp. 1799–1802, 2010.

[31]   C. Hsu, C. Chang, and C. Lin, "A practical guide to support vector classification," *Department of Computer Science, National Taiwan University*, 2010.

[32]   R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern classification*, 2nd Editio. Wiley-Interscience, 2000.

[33]   Y. Shynkevich, T. M. McGinnity, S. Coleman, and A. Belatreche, "Stock Price Prediction based on Stock-Specific and Sub-Industry-Specific News Articles," in *Proceedings of the IEEE International Joint Conference on Neural Networks*, 2015, pp. 1 − 8.