Stock Price Prediction based on Stock-Specific and Sub-Industry-Specific News Articles

Yauheniya Shynkevich[1], T.M. McGinnity[1,2], Sonya Coleman[1], Ammar Belatreche[1]
[1]Intelligent Systems Research Centre, University of Ulster, Derry, UK
[2]School of Science and Technology, Nottingham Trent University, Nottingham, UK
shynkevich-y@email.ulster.ac.uk, {tm.mcginnity, sa.coleman, a.belatreche}@ulster.ac.uk

***Abstract*—Accurate forecasting of upcoming trends in the capital markets is extremely important for algorithmic trading and investment management. Before making a trading decision, investors estimate the probability that a certain news item will influence the market based on the available information. Speculation among traders is often caused by the release of a breaking news article and results in price movements. Publications of news articles influence the market state that makes them a powerful source of data in financial forecasting. Recently, researchers have developed trend and price prediction models based on information extracted from news articles. However, to date no previous research that investigates the advantages of using news articles with different levels of relevance to the target stock has been conducted. This research study uses the multiple kernel learning technique to effectively combine information extracted from stock-specific and sub-industry-specific news articles for prediction of an upcoming price movement. News articles are divided into these two categories based on their relevance to a targeted stock and analyzed by separate kernels. The experimental results show that utilizing two categories of news improves the prediction accuracy in comparison with methods based on a single news category.**

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

I. INTRODUCTION

News is a major factor that drives changes in a stock price. Financial investors make decisions based on the available data considering the probability that a new piece of information will have an impact on the market. They analyze recently published news when judging equitable market prices. The information about a company's fundamentals, the activities in which a company is involved and the expectations of other market participants are incorporated into the news articles. The release of major news items often produces speculation among traders which results in price movements [1], [2]. With the development of the internet, real-world trading applications provide a huge amount of textual data with a tremendously increased broadcasting speed [3]. A number of efforts were made to create an automated framework that analyses a large amount of textual data relevant to a particular stock, extracts relevant information and uses it for financial forecasting. A strong relationship between the fluctuation of a stock price and the publication of a related news article has been shown in previous research works [4]. Many researchers have employed or expanded the existing data mining approaches to investigate how news can impact traders' actions and hence affect the stock price [5], [6], [7]. In the reviewed literature, news articles are considered to be relevant to an analyzed stock based on a predefined criterion, then retrieved from a large collection of documents and used for further analyses. Researchers tend to define rules for selecting relevant news items from a whole data set and then analyze every article as having the same potential impact on price changes. So far, no previous work has been done in combining stock-related and industry-related news items with learning the degree of influence for these two categories of news.

This research paper investigates the advantages of using news articles with different levels of relevance to the target stock in financial news based forecasting. In order to achieve this goal, two groups of news articles are formed: stock-specific (SS) and sub-industry-specific (SIS) news items. A number of articles are analyzed and grouped based on their relevance to the target stock. Articles directly relevant to the target stock and articles relevant to the whole sub industry the stock belongs to are used in this study. The Multiple Kernel Learning (MKL) framework is often used for integrating different kinds of information [8], [9], [10], [11], [12]. MKL allows the usage of several kernels each employed for learning from a separate subset of data. The MKL approach applied in this study utilizes from two to six separate kernels each assigned to either a SS or a SIS subset of news articles. The experiments show that an attempt to divide articles into different categories, analyze them separately and then combine the resulting predictions displayed a promisingly improved performance.

The remainder of the paper is organised as follows. Section II presents a literature review of the related research on predictive approaches based on the analysis of financial news articles, outlining the main methods and techniques used. Section III describes raw data and provides details of data pre-processing and machine learning approaches used. Section IV represents and discusses experimental results. Section V concludes the paper and provides incentives for further work.

II. LITERATURE REVIEW

A. Representation of News Articles

An enormous amount of textual data are available to capital market traders in real-world trading systems. For

This research is supported by the companies and organizations involved in the Northern Ireland Capital Markets Engineering Research Initiative (InvestNI, Citi, First Derivatives, Kofax, Fidessa, NYSE Technologies, ILEX, the Queen's University of Belfast and the University of Ulster)

example, financial journals and discussion boards, official reports and analysts' recommendations, news streams from news wire services, news articles about firm's performance, comments and interactive media provide channels of information that are all available to investors. Significant changes in the price of an asset can be caused by reports of an unexpected nature [13].

The first attempt to utilize the textual information available for stock market prediction was performed by Wüthrich et al. [14]. The authors utilized a dictionary of terms provided by a domain expert for assigning weightings, generated probabilistic rules and predicted daily movements of five stock indices. A money making trading strategy was constructed based on the system predictions proving that positive returns can be achieved using financial textual information. Later, Lavrenko et al. [15] designed the Analyst system that created language models, transformed the price time series into trends and classified the incoming news. The authors demonstrated that the developed system was capable of producing profit. Gidofalvi [16] conducted research on financial news articles and real stock market data and created a forecasting system for short-term stock price movements. The system performed alignment and scoring of the articles and then assigned an "up", "down" or "unchanged" label to each news item. The finding shows a strong correlation between the behaviour of the stock and the news article from 20 minutes prior to 20 minutes after the news was published and therefore indicates the ability to predict the direction of a price change. Chan [17] explored the monthly returns using headlines about particular firms and found out that a strong negative drift of the market follows the publication of bad news. Kloptchenko et al. [18] have shown that company generated official reports such as annual and quarterly reports are able to indicate the future performance of the company. For instance, the written style of a financial report changes prior to a dramatic change in company productivity. Fang and Peress [19] examined the relationships between the media coverage of the companies and their average return and pointed out that stocks not featured in the media significantly outperform stocks with high coverage. Tetlock [20] measured the interactions between the stocks and the content of daily articles in the Wall Street Journal. His findings state that highly pessimistic news have a downward effect on market prices and generate high trading volume. Schumaker et al. [6], [8] performed a comparison of several textual analysis techniques. In [21], Schumaker and Chen examined the worth of grouping textual financial news articles by similar industries. The paper is similar to the proposed research in that it uses articles with different levels of relevance but differs in the way in which they are used. Schumaker and Chen compared the performance of the developed prediction system where only one category of news articles was used each time a prediction was made: either articles relevant to the stock itself, its sub industry, industry, group industry or sector, or the whole universal set of articles was utilized. The proposed research creates a predictive system that uses articles from several categories at the same time. To our knowledge, no existing research has involved dividing news articles into different categories, analyzing them separately and then combining multiple predictions made based on these separate news categories into a single prediction.

The pre-processing of textual data is a key part of the text mining processes. When applying to financial forecasting, the aim of the pre-processing is to obtain relevant information that indicates potential stock price changes from a given set of articles [22]. The extracted information should be represented in a computer-readable format. Mittermayer [23] divided pre-processing procedures into three preliminary steps: feature extraction, feature selection and feature representation. The same characterization was applied in later works [24].

In feature extraction, a dictionary of features that describe the documents sufficiently is generated [23]. For this purpose, stop words and symbols such as prepositions, articles and pronouns as well as numbers and punctuation marks are eliminated from the data, all words are processed using word stemming techniques. In the proposed study, the Bag of Words approach is selected as a feature extraction technique. Semantically empty words are removed and the remaining terms are used for representing the text. This technique is popular and typically used in other research papers because of its intuitive meaning and simplicity [25].

In feature selection, features that contain items with the least information are neglected [23]. The literature provides a number of different approaches. Some researchers employ dictionaries containing a list of keyword records selected by domain experts [14]. Others use statistical information extracted from the text, e.g. TF*IDF (term frequency - inverse document frequency) [10], [23], [26], [27]. External market feedback was recently proposed in several research papers. The Chi-square test was used for feature selection by Wang, Liu and Dou [11] for a volatility prediction system. Hagenau et al. [1] explored the usefulness of the Chi-square test and Bi-normal separation methods to evaluate the explanatory power of a word. Both methods utilize the external market feedback and showed promising results. The Chi-square test is employed for feature selection in this study.

The feature representation step represents the whole set of documents in a machine learning friendly manner [24]. For instance, each document is transformed into a feature vector of n elements, where n is the number of features remaining after the feature selection step [23]. Generally, researchers consider the fact that a feature is present in a document as an important factor. Rachlin et al. [28] calculated the membership value for each word and employed binary representation for weightings in the developed financial trading system. Other researchers assign real values to weightings. For example, Luss and d'Aspremont [26] used TF*IDF calculations for obtaining the feature weightings for predicting abnormal returns. In [11], TF*IDF values were used as the weightings for each feature to predict the direction of a change in volatility. Due to its popularity in the literature, TF*IDF is also employed in this study to compute weightings of features for each data point.

After pre-processing is completed, the textual data need to be aligned with the time series data and labelled accordingly. Generally, the impact of the news articles on the movements of the stock prices is classified into two (positive or negative) or three (positive, negative or neutral) categories using exogenous

market feedback defined as the immediate effect of published news on an asset price [10], [23]. Zhai, Hsu and Halgamuge [29] classified the articles into good or bad news. The forecasting system for short-term stock price movements created by Gidofalvi [16] labelled the articles as up, down or unchanged. In order to highlight the degree of influence of a news item Rachlin et al. [28] defined five categories for news classification: Up, Slight-Up, Expected, Slight-Down, Down. In the current study, news items were classified into positive or negative.

B. *Predicting from News Articles*

The literature shows many different approaches developed for predicting the market reaction to news articles. In order to extract relevant information from financial textual data, research studies use various artificial intelligence approaches, such as: Artificial Neural Networks (ANN) [30], [31]; Naïve Bayes [16]; Support Vector Machines (SVM) [4], [24], [27] and Genetic Algorithms [32]. For forecasting short-term stock price movements Gidofalvi [16] used a Naïve Bayesian text classifier for training and prediction. Zhao et al. [2] investigated the impact of financial news articles on Chinese stock markets, quantifying their content by utilizing the event study methodology and the cross-sectional analysis of abnormal returns, where Support Vector Regression (SVR) was employed showing that online financial news has a negative impact on the market movements. Groth and Muntermann in [3], [33] designed an approach for supporting market risk and investment management based on textual analysis and machine learning techniques. The authors compared the performance of Naïve Bayes, k-Nearest Neighbour (kNN), ANN and SVM classifiers on forecasting the direction of price movement both 15 and 30 minutes after a news item was published. Considering both computational efficiency and classification results, the paper suggests the use of a SVM for learning. Antweiler and Frank [34] analyzed messages extracted from Raging Bull and Yahoo Finance websites using Naïve Bayes and SVM for classification into bearish, bullish or neutral classes. Their findings state that the posted messages can help to predict the trading volume and stock price volatility. Hagenau et al. [24] employed SVM to classify whether a message had a negative or positive effect on the stock market price. The authors claim that a pilot comparison study found SVM to perform better than ANN and Naïve Bayes. According to the previous findings, SVM is considered to be the most promising available machine learning technique for text classification tasks [24], [35]. Many machine learning techniques employ the Empirical Risk Minimization principle to minimize the training data error that can cause an over-fitting problem. SVM uses the Structural Risk Minimization principle that minimizes the upper limit of the expected risks and build a more robust and precise model to overcome over-fitting.

In addition, the literature describes ensemble methods used for forecasting the financial markets [36], [37]. The performance of base learners can be sufficiently improved using these methods. An ensemble algorithm is a computational intelligence learning approach that combines a set of base learners into an integrated model [37]. In recent papers about financial forecasting, researchers have made use of the MKL technique to select appropriate learning algorithms for different features with the use of financial news articles and/or time series data [8], [9], [10], [11], [12] where linear, polynomial, Sigmoid or Gaussian functions were used as kernels. Li et al. [10] integrated information from stock prices and financial news to improve the accuracy of prediction where the MKL method was applied to combine the information from news articles and past prices. The results reveal that the multi-kernel model outperforms models with simple feature combinations and models based on a single information source. In [11], the MKL framework based on RBF kernels was proposed for prediction of volatility changes. The MKL method showed higher accuracy than single kernel methods. However, the news articles analyzed in both papers [10], [11] are written in Traditional Chinese, and the proposed models were not tested on articles written in English. In [8] Deng et al. presented a stock price prediction system that uses historical time series data with numerical dynamics and semantic analysis of the content of news articles and comments for stock price prediction. In [9] the same authors employed technical analysis of historic price and volume, numerical features of published news articles and communication dynamics of the comments that appeared on the internet. In both papers the model extracts features from all sources of information and forms separate feature sets. The feature sets are integrated and learned by MKL. To the best of our knowledge, no evidence was found that the MKL approach has been ever applied to analyze different categories of news data for financial predictions.

Based on the reviewed literature, the MKL approach was selected to solve the classification problem in the proposed study; SVM and kNN were utilised for comparison purposes. An implementation of the MKL algorithm is proposed by Sonnenburg et al. [38] in the SHOGUN toolbox, which is used in several experimental studies [8], [11], and was selected for use in this proposed system. It is utilized for concurrent estimation of the optimal kernel weights and parameters by repeating the training procedure used for a simple SVM.

III. Research Design

A. *Industry Classification*

Global Industry Classification Standard (GICS) was developed by Morgan Stanley Capital International, an independent provider of global indexes, products and services, and Standard & Poor's (S&P), a financial services company. It aims to support the investment research and asset management. The GICS structure allocates companies to four categories: sector, group industry, industry and sub industry. The GICS was utilised by Schumaker and Chen [21] to study the benefits of using textual analysis of financial news articles grouped by similar industries. In the current study, GICS classification was utilised to retrieve news articles that correspond to the target stock and to all stocks from the sub industry the target stock belongs to. Five stocks from the S&P 500 stock market index were selected for analysis. These stocks are the only stocks from the S&P 500 index that belong to the Managed Health Care sub industry. They were chosen for analysis because a sufficient number of news articles (more than 400 news articles during the period of study) were published about each of them.

B. News Articles Data

A period of five years from 1 Sep 2009 to 1 Sep 2014 was considered for analysis. To collect financial news articles published about the stocks of interest during this time period, LexisNexis database was explored. LexisNexis database contains news articles published in major newspapers, and has been used in other research studies, e.g. Fang and Peress investigated the cross sectional relationships between expected stock returns and media coverage of firms in news [39]. An important feature of the LexisNexis database is that a list of relevant companies and their relevance scores supplement each article. A relevance score corresponds to a particular company and specifies a degree of relevance of the article to the company. It is expressed as a percentage in a range (0%,100%). Business Wire, PR Newswire and McClatchy-Tribune Business News sources were selected as providers of news stories because they showed sufficient press coverage of stocks constituting the S&P 500 market index. The total number of articles downloaded from LexisNexis is 8264. For each news article, the following information was retrieved: the date, month, year, heading and body as well as lists of relevant companies, relevant tickers and corresponding relevance scores. The combination of a date, month and year defines a date of a publication for every article. Words from the heading and body of the message were combined and used as the textual data for further information extraction. The heading is often repeated in the body of the message, though it emphasizes the main idea of the publication and its addition may intensify the significance of the words used. Lists of relevant tickers and corresponding relevance scores were employed to identify the degree of relevance of an article to the companies. When forming a set of articles relevant to the target company, an automated procedure examined every article whether it was related to the target company or not. If the target company's ticker was among relevant tickers of the article and a corresponding relevance score was higher than or equal to 85%, the article was selected for further analysis. When a set of articles relevant to the sub industry is formed, every article is checked whether at least one ticker of a company from the sub industry of interest is present among the article's tickers, and then whether its corresponding relevance score is higher than or equal to 85%. If these conditions are satisfied, the article is selected for further analysis. As a result, two subsets corresponding to SS and SIS news items are formed in this example.

Articles published on the same date are combined together so that no information published about stocks is lost. The system searched for articles published on the same date and concatenated them. At the same time it checked whether an article is already included because some articles could be downloaded more than once or republished by another news source. If the same article is already processed, it is ignored and not considered for further analysis. This is necessary for eliminating the repetition. The described procedure is carried separately through SS and SIS textual data subsets. In the developed system, predictions are made only for dates when at least one article was published about the target stock. Here a number of data points is defined. A single data point corresponds to a date when a news article specific to a target stock was published. A list of those dates is utilized when collecting articles relevant to the sub industry where only articles published on those dates are considered.

TABLE I. DESCRIPTION OF ANALYSED STOCKS AND THEIR DATASETS

Ticker	Company Name	# data points	Fraction of '1' labelled data points, %	Fraction of '-1' labelled data points, %
AET	Cigna Corp	836	52.51	47.49
CI	Aetna Inc	806	56.08	43.92
HUM	Humana Inc	482	54.56	45.44
UNH	United Health Group Inc	583	52.49	47.51
WLP	WellPoint Inc	392	52.55	47.45

C. Historical prices data

Historical prices data were downloaded from the publicly available website Yahoo! Finance [40]. Prices were used for feature selection and data points labelling. The feature selection was based on the market feedback depending on a price move on the next trading day following the day of publication. A price move was computed as the difference between open and close stock price values on the next trading day. The problem considered in this study is a two class classification problem. Labels '1' or '-1' are assigned to each data point and correspond to an increase or decrease in the target stock price, respectively. Daily prices were selected for the analysis due to the absence of intraday data. However, it should be noted that daily data were used in a number of research papers concerning financial predictions from textual data extraction [3], [14], [34] and [41] that showed that market reaction to new information is slow enough to be captured and studied using daily price observations. Table I provides information about the stocks, the total number of data points obtained and the fractions of '1' or '-1' labelled data points.

D. Textual data pre-processing

The textual data pre-processing is essential and particularly important when creating a predictive system based on financial news articles. It usually contains three preliminary steps: feature extraction, selection and representation. As mentioned previously, the Bag-of-Words approach is used for feature extraction in this approach where all articles are processed in the following way. First, emails, hyperlinks, websites' addresses, numbers, punctuation and symbols other than letters are removed from the text. Next, words containing capital letters are converted to lowercase, words consisting of two or less characters are filtered out and stop words are deleted. Finally, Porter's stemming algorithm is applied to each word to extract stems in order that different forms of the same word are treated equally [42].

After the completion of these steps, a list of unique features extracted from the dataset is formed. Then features occurred in two or less articles are filtered out, and every article is labelled as positive or negative based on changes in the target stock prices. In feature selection, the Chi-square method is used to calculate the scores for each unique feature

based on the market feedback as the sum of normalized deviations [1] using the following formula:

$$\chi^2 = \sum_{j=1}^{4} \frac{\left(O_{ij} - E_{ij}\right)^2}{E_{ij}} \quad (1)$$

where O_{il} is the observed frequency of the feature *i* within the set of positive messages, E_{il} is the expected frequency, and *j* indicates four possible outcomes when a feature occurred within positive messages (*j=1*); occurred within negative messages *(j=2)*; did not occur within positive messages *(j=3)*; did not occur within negative messages *(j=4)*. Thereby, the observed frequency of the feature *i* in positive messages is calculated as a fraction of positive messages containing the feature *i* among all positive messages. The observed frequency of the feature *i* occurring in negative messages and not occurring in positive or negative messages is computed analogously. If a feature has neutral meaning and does not imply any positive or negative context, it should occur uniformly among positive and negative messages. Thus, the expected frequency of the feature *i* occurring in positive and negative messages is the overall frequency of the feature occurring in the whole set of articles. Similarly, the expected frequency of the feature *i* not occurring in all messages is the overall frequency of the feature not occurring within all documents. Hence, a feature has a p-value close to zero when it occurs uniformly among positive and negative documents. If a feature appearance in positive articles significantly exceeds that in negative articles or vice versa, its p-value is noticeably higher than zero.

Consequently, a list of unique features is sorted based on the computed Chi-square values and 500 features with the highest scores were selected as input features. 500 is the number of features considered to be sufficient to represent news articles. This number is close to the 567 features selected by Hagenau et al. [1] for the Bag-of-words approach. Next, the whole set of documents is required to be represented in a machine learning friendly manner [24]. In this research study each document is transformed into a feature vector of 500 elements [23] and features are represented using TF*IDF values. Since the TF*IDF value is equal to zero when the feature is not present in the article, the result is a sparse matrix of size [number of articles]*[number of features = 500] where each value is equal to the corresponding TF*IDF value. It is worth noticing that SS and SIS subsets of news are processed separately through the above described procedure of textual pre-processing, and different lists of unique features and therefore different feature matrices are constructed for these subsets. After textual pre-processing is completed, labels are assigned to data points. Each data point has a label '1' or '-1', 500 values for SS features and 500 values for SIS features.

E. Machine learning techniques

In MKL, the resulting kernel is combined from several sub-kernels according to the formula below where $K_{comb}(x,y)$ is the combined kernel, K is the number of kernels and the weights β_j are learnt from the data for each sub-kernel $K_j(x,y)$:

$$K_{comb}(x,y) = \sum_{j=1}^{K} \beta_j K_j(x,y), \text{ where } \beta_j \geq 0, \sum_{j=1}^{K} \beta_j = 1. \quad (2)$$

Multiple kernel learning (MKL) allows the assignment of a separate kernel to each category of news articles. There is a number of different types of kernels available, and in this study the MKL technique with different combinations of linear, polynomial and Gaussian kernels was examined. Two news categories, SS and SIS, were considered with a need to utilize at least two separate kernels. At the same time, in order to treat both news categories equally, kernels of different types were taken in pairs where the first kernel in a pair was assigned to the SS news and the second kernel was assigned to the SIS news. Weights learnt for every kernel indicate its usefulness and contribution to making the final prediction decision. Table II gives details about kernel functions applied to SS and SIS subsets of data. The first column corresponds to the order of an independent experimental setting. The second column indicates the machine learning technique applied. The third and the fourth columns specify whether SS and SIS data were used and which kernels were applied to them. Six kinds of experimental settings were used for MKL to find the combination of kernels that would give the highest results. For comparison purposes, Support Vector Machines (SVM) with different types of kernels and kNN were applied to SS and SIS datasets in turn.

For each stock the dataset was divided into training, validation and testing datasets in a chronological order. First 50% of the data points were used for training the MKL model. The subsequent 25% of the data points were used for validation, a phase required to tune model parameters. The C parameter is used within MKL and SVM and is defined as a penalty parameter for classification error [43]. Some data points are allowed to be misclassified while being penalized at the rate C [26]. This parameter is required to be tuned during validation. It was selected from a range of reasonable values

TABLE II. A DESCRIPTION OF EXPERIMENTAL SETTINGS LISTING THE MACHINE LEARNING METHODS AND CORRESPONDING NUMBER AND TYPES OF KERNELS USED FOR EACH DATA SUBSET

#	ML method	Kernels	
		Stock-specific data	***Sub-industry-specific data***
1	MKL	One linear	One linear
2		One polynomial	One polynomial
3		One Gaussian	One Gaussian
4		One linear and one polynomial	One linear and one polynomial
5		One Gaussian and one polynomial	One Gaussian and one polynomial
6		One linear, one polynomial, and one Gaussian	One linear, one polynomial, and one Gaussian
7	SVM	One linear	not used
8		One polynomial	not used
9		One Gaussian	not used
10		not used	One linear
11		not used	One polynomial
12		not used	One Gaussian
13	kNN	applied directly to SS data	not used
14		not used	applied directly to SS data

{1, 5, 10, 50, 100, 1000, 10000, 100000}. Furthermore, when Gaussian and polynomial kernels were utilised, the Gaussian kernel width and the degree of a polynomial were also tuned during the validation [43]. For kNN, an optimal number of neighbours was found during the validation. The remaining 25% of the data instances were utilised to test the developed predictive system. The prediction accuracy was applied to measure the model performance amid different settings of parameters during the validation and testing phases. The above described procedure was followed separately for every stock.

IV. Experimental Results

Table III represents the experimental results obtained using the MKL approach with different combinations of kernels. The first column displays the stock ticker. The second column specifies which value is shown in a row where *w_PolySS*, *w_PolySIS*, *w_LinSS*, *w_LinSIS*, *w_GausSS* and *w_GausSIS* are the notations of kernel weightings corresponding to a polynomial kernel assigned to the SS data, a polynomial kernel assigned to the SIS data, a linear kernel assigned to the SS data, a linear kernel assigned to the SIS data, a Gaussian kernel assigned to the SS data and a Gaussian kernel assigned to the SIS data respectively. The highest accuracy obtained for each stock is highlighted in bold. Columns 3-8 contain weights percentages assigned to kernels while training MKL, C parameter values and prediction accuracies for each stock, and correspond to the cases where only two polynomial, only two linear, only two Gaussian, a combination of two polynomial and two linear, a combination of two polynomial and two Gaussian, and finally a combination of two polynomial, two linear and two Gaussian kernels were employed respectively. Additionally, a sign '-' is used to indicate that a corresponding kernel was not used to form the MKL. The weight '0.00' illustrates that zero weighting was assigned to the corresponding kernel during learning phase.

Let's begin with the analysis of results in columns 3-5 where only two kernels of the same type were employed to learn from SS and SIS data. In the majority of cases both kinds of information were considered important and were used to make the final prediction decision. Kernel weights assigned to both kernels were of comparable values. For linear and Gaussian kernels larger weights were given to the kernel learning from SS data when for polynomial kernels larger weights were assigned to the kernel based on SIS news. This behaviour is reproducible for all analyzed stocks. Moreover, when comparing the prediction accuracies of the MKL approach based on linear, Gaussian or polynomial kernels, the polynomial kernels produced higher accuracies than others for every stock. This behaviour may be caused by the fact that polynomial kernels are able to learn more information from SIS news articles than other kernels and can take advantage of utilizing this information for predictions. It is worth noting that the optimal value of the parameter C differs for various types of kernels. It varies in the range 1-10 for MKL consisting of polynomial kernels, the range 5-100 for Gaussian kernels and the range 1000-10000 for linear kernels.

Columns 6 and 7 relate to the MKL with four kernels: two polynomial and two linear, and two polynomial and two Gaussian, respectively. In column 6, for every stock zero weightings were assigned to linear kernels, and values of polynomial kernel weightings, C parameters and accuracies are equal to those of column 3 which implies that linear kernels do not add any value to the overall prediction performance and decisions are made solely based on the results obtained using polynomial kernels. In column 7, the same behaviour is observed for the AET stock. For UNH, Gaussian kernel weightings are zeros but values of polynomial kernel weightings, C parameters and accuracies slightly differ from that of column 3. For CI, HUM and WLP stocks weightings assigned to the Gaussian kernel used for SS data are between 40-66% which indicates that information learnt with the use of a Gaussian kernel adds significant value to the final prediction. However, the prediction accuracy is only higher for the WLP stock than that made using polynomial kernels whereas predictions for all other stocks showed lower performance.

TABLE III. Experimental Results for the MKL Approach

Stock	Parameter	Kernel Types used in MKL					
		Poly	***Lin***	***Gauss***	***Poly& Lin***	***Poly& Gauss***	***Poly& Lin& Gauss***
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
AET	w_PolySS, %	44.77	-	-	44.77	44.77	44.77
	w_PolySIS, %	55.23	-	-	55.23	55.23	55.23
	w_LinSS, %	-	64.92	-	0.00	-	0.00
	w_LinSIS, %	-	35.08	-	0.00	-	0.00
	w_GausSS, %	-	-	64.54	-	0.00	0.00
	w_GausSIS, %	-	-	35.46	-	0.00	0.00
	C	10	10000	100	10	10	10
	Accuracy, %	**77.03**	73.21	66.51	**77.03**	72.25	70.33
CI	w_PolySS, %	50.37	-	-	50.37	23.52	49.59
	w_PolySIS, %	49.63	-	-	49.63	36.29	50.41
	w_LinSS, %	-	65.98	-	0.00	-	0.00
	w_LinSIS, %	-	34.02	-	0.00	-	0.00
	w_GausSS, %	-	-	70.50	-	40.19	0.00
	w_GausSIS, %	-	-	29.50	-	0.00	0.00
	C	10	10000	100	10	5	1
	Accuracy, %	73.13	68.66	67.66	73.13	71.64	**74.13**
HUM	w_PolySS, %	45.90	-	-	45.90	10.08	41.13
	w_PolySIS, %	54.10	-	-	54.10	24.73	58.87
	w_LinSS, %	-	100.00	-	0.00	-	0.00
	w_LinSIS, %	-	0.00	-	0.00	-	0.00
	w_GausSS, %	-	-	88.96	-	65.19	0.00
	w_GausSIS, %	-	-	11.04	-	0.00	0.00
	C	5	1000	100	5	5	100
	Accuracy, %	**80.83**	68.33	66.67	**80.83**	74.17	75.83
UNH	w_PolySS, %	32.86	-	-	32.86	39.65	39.65
	w_PolySIS, %	67.14	-	-	67.14	60.35	60.35
	w_LinSS, %	-	67.57	-	0.00	-	0.00
	w_LinSIS, %	-	32.43	-	0.00	-	0.00
	w_GausSS, %	-	-	74.28	-	0.00	0.00
	w_GausSIS, %	-	-	25.72	-	0.00	0.00
	C	1	10000	5	1	100	100
	Accuracy, %	**76.55**	71.03	58.62	**76.55**	69.66	62.76
WLP	w_PolySS, %	21.63	-	-	21.63	0.00	0.00
	w_PolySIS, %	78.37	-	-	78.37	47.48	53.71
	w_LinSS, %	-	57.27	-	0.00	-	0.00
	w_LinSIS, %	-	42.73	-	0.00	-	0.00
	w_GausSS, %	-	-	100.00	-	52.52	46.29
	w_GausSIS, %	-	-	0.00	-	0.00	0.00
	C	5	10000	5	5	100	100
	Accuracy	78.57	73.47	78.57	78.57	80.61	**81.63**

The last column (8) corresponds to the case when pairs of polynomial, linear and Gaussian kernels are used together. For AET, CI, HUM and UNH only polynomial kernels were assigned with values noticeably higher than zeros. But the prediction accuracy is lower than in column 3 for three out of those four stocks. The reason may lie in the fact that parameters were tuned differently in cases when two or six kernels were used, especially when the same parameter C is used for all kernels in MKL whereas, as it was mentioned earlier, the optimal values of parameter C differ for different kernels. However, for CI and WLP the combination of six different kernels resulted in higher prediction accuracy than that of any other combination. This demonstrates that the use of polynomial kernels only generally gives higher prediction performance than that of other kernels and kernel combinations, however adding more kernels and properly tuning their parameters may provide higher performance for particular stocks.

Table IV describes the prediction results obtained using SVM and kNN machine learning techniques applied to either SS or SIS datasets. The second column in the table indicates which subset of data is used for predictions. SVM outperforms kNN for all stocks and datasets. Comparing the use of different kernels, SVM based on a polynomial kernel performs slightly better than that based on Gaussian and significantly outperform SVM with linear kernels. It is important to highlight that on average accuracy of predictions made with the use of SIS news is higher than that of SS news. This indicates that more important information can be extracted from the news articles relevant to the whole sub industry rather than from news published about the stock only. This conclusion agrees with findings in [21]. When comparing the prediction performance in Tables III and IV, the highest values of accuracies obtained using MKL are higher than those of SVM for each stock apart from UNH for which these values are equal. The finding illustrates that integrated usage of SS and SIS news articles helps to gather more information about the stock and the market, to extract relevant indicators and to make more informed predictions about future movements of a stock price. It is important to distinguish between different categories of news, to treat them differently when pre-processing and extracting the information, and to integrate the extracted and learnt information at the later stage instead of combining everything in the early stage and throwing all this information to the machine learning method.

TABLE IV. EXPERIMENTAL RESULTS OBTAINED FOR THE SVM AND KNN APPROACHES

Stock	Data Type	Parameters	SVM			kNN
			Polynomial kernel	*Linear kernel*	*Gaussian kernel*	
AET	SS	C	1	1000	5	-
		Accuracy, %	71.29	70.33	66.03	53.59
	SIS	C	10	10000	100	-
		Accuracy, %	69.38	**72.25**	**72.25**	56.94
CI	SS	C	100	10000	100	-
		Accuracy, %	66.17	63.68	58.71	53.73
	SIS	C	5	10000	100	-
		Accuracy, %	68.16	68.66	**70.65**	56.72
HUM	SS	C	5	10000	100	-
		Accuracy, %	63.33	67.50	68.33	57.50
	SIS	C	1	1000	100	-
		Accuracy, %	**77.50**	70.34	70.83	61.67
UNH	SS	C	5	1000	0.0055	-
		Accuracy, %	67.59	70.34	68.97	62.07
	SIS	C	1	10000	100	-
		Accuracy, %	**76.55**	71.72	70.34	57.24
WLP	SS	C	5	1000	10	-
		Accuracy, %	76.53	71.43	69.39	59.18
	SIS	C	1	10000	100	-
		Accuracy, %	**79.59**	77.55	**79.59**	64.29

V. CONCLUSION AND FUTURE WORK

This research paper studies whether the simultaneous usage of different categories of news articles can improve the prediction performance of a news based financial prediction system. Two categories of news were examined: news articles related to a target stock and news articles relevant to its sub industry. News articles from each category were pre-processed separately and two different subsets of data were formed. The MKL approach was applied so that separate kernels were used for learning from different news subsets. A number of kernels and kernel combinations were applied. The results have shown that for the majority of stocks the highest prediction accuracy was achieved for MKL when polynomial kernels were applied to both categories of news items. For other stocks utilizing pairs of Gaussian, linear and polynomial kernels together in MKL leads to higher prediction performance. Both SVM and kNN methods showed worse performance than MKL. These results indicate that classifying news items based on their relevance to the target stock and using separate kernels for learning from different categories of news allows the system to learn and utilize more information about the future price behaviour. This gives an advantage for more accurate predictions.

The results achieved to date are promising and hence the directions of further work are the enhancement of the developed prediction system to include news items relevant to the stock's industry, group industry and sector. This will help to investigate the importance of news items with different levels of relevance to the target stock. Later, additional sources of data such as historical prices will be employed for financial forecasting together with news articles.

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, 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] S. S. Groth and J. Muntermann, "Supporting Investment Management Processes with Machine Learning Technique," in *Business Services: Konzepte, Technologien, Anwendungen - 9, Internationale Tagung Wirtschaftsinformatik*, 2009.

[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] G. Fung, J. Yu, and W. Lam, "News sensitive stock trend prediction," in *Proceedings of the 6th Pacific-Asia Conference on Advances in Knowledge Discovery and Data Mining*, 2002, vol. 2336, pp. 481–493.

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

[7] M. Mittermayer and G. Knolmayer, "Text mining systems for market response to news: A survey," vol. 41, no. 184. University of Bern, 2006.

[8] 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.

[9] S. Deng, T. Mitsubuchi, K. Shioda, T. Shimada, and A. Sakurai, "Multiple Kernel Learning on Time Series Data and Social Networks for Stock Price Prediction," in *Proceedings of the 10th International Conference on Machine Learning and Applications and Workshops*, 2011, pp. 228–234.

[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, 2011.

[13] R. P. Schumaker and H. Chen, "Textual Analysis of Stock Market Prediction Using Financial News Articles," in *Proceedings of the 12th Americas Conference on Information Systems (AMCIS)*, 2006, pp. 1–20.

[14] B. Wiithrich, 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 minin*, 1998.

[15] 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.

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

[17] 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, 2003.

[18] 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.

[19] 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.

[20] P. C. Tetlock, "Giving Content to Investor Sentiment: The Role of Media in the Stock Market," *Journal of Finance*, vol. 62, no. 3, pp. 1139–1168, 2007.

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

[22] D. Wu, G. Fung, J. Yu, and Z. Liu, "Integrating multiple data sources for stock prediction," in *Proceedings of the 9th international conference on Web Information Systems Engineering*, 2008, pp. 77–89.

[23] 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.

[24] 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.

[25] T. Loughran and B. McDonald, "When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks," *The Journal of Finance*, vol. 66, no. 1, pp. 35–65, 2011.

[26] R. Luss and A. D'Aspremont, "Predicting abnormal returns from news using text classification," *Quantitative Finance*, 2012.

[27] Y. Zhai, A. Hsu, and S. Halgamuge, "Combining news and technical indicators in daily stock price trends prediction," *Advances in Neural Networks*, vol. 4493, pp. 1087–1096, 2007.

[28] 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.

[29] F. Zhai, X. Lin, Z. Yang, and Y. Song, "Financial time series prediction based on Echo State Network," in *Proceedings of the Sixth International Conference on Natural Computation*, 2010, pp. 3983–3987.

[30] L. Bucur and A. Florea, "Techniques for prediction in chaos – a comparative study on financial data," *U.P.B. Scientific bulletin, series C*, vol. 73, no. 3, pp. 17–32, 2011.

[31] 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, 2012.

[32] D. Kato and T. Nagao, "Stock Prediction Using Multiple Time Series of Stock Prices and News Articles," in *Proceedings of IEEE Symposium on Computers & Informatics*, 2012, pp. 11–16.

[33] 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, 2011.

[34] 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.

[35] T. Joachims, "Text Categorization with Support Vector Machines : Learning with Many Relevant Features," in *Proceedings of the 10th European Conference on Machine Learning*, 1998, pp. 137–142.

[36] Y. Kwon and B. Moon, "Evolutionary ensemble for stock prediction," in *Proceedings of Genetic and Evolutionary Computation Conference, Part II*, 2004, pp. 1102–1113.

[37] 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.

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

[39] L. Fang and J. Peress, "Media Coverage and the Cross-section of Stock Returns," *The Journal of Finance*, vol. 64, no. 5, pp. 2023–2052, 2009.

[40] 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 and Economics*, 2014, pp. 341–348.

[41] F. Li, "The Information Content of Forward-Looking Statements in Corporate Filings-A Naïve Bayesian Machine Learning Approach," *Journal of Accounting Research*, vol. 48, no. 5, pp. 1049–1102, 2010.

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

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