Journal of Soft Computing and Information Technology
Vol. 3, No. 4, Winter 2015

An Intelligent News Based Decision Support System for Trading Stocks in Tehran Stock Exchange

Nima Shayanfar[1], Vali Derhami[2]
[1]Electrical and Computer Engineering Department, Yazd University, nshayanfar@stu.yazd.ac.ir
[2] Electrical and Computer Engineering Department, Yazd University, vderhami@yazd.ac.ir

Abstract: *Stock market is known as a stochastic, nonlinear, and uncertain environment. Hence, decision making in its trading is a challenging task. Indeed, stock markets are influenced by many parameters such as stock market indicators, micro-economic and macro-economic parameters and news articles. The latter is known as one of the effective parameters, attracting analysts' attention in recent years. If news could be successfully analyzed, predicting the stock market's reaction to the news would be achievable. This paper proposes an intelligent decision support system for trading stocks, based on news in Tehran Stock Exchange. Here, instead of predicting stock prices, we predicted the trend of stock prices. To this aim, text mining was used to extract several types of features from the news. Afterwards, the most effective features were selected and then provided for Support Vector Machine (SVM) in order to be classified as a positive or negative trend. Based on comparison of several types of features we concluded that the word combination is the best solution for Farsi. The simulation results depicted that stock trend prediction is successful and profitable in many cases. Moreover, due to linguistic similarities between Farsi, Arabic and Urdu, having some powerful stock markets of the world, our proposed intelligent algorithm could pave the way for data analysis in those languages.*

Key-Words: *Text mining, Stock price, Intelligent trend prediction, News, Farsi*

1. Introduction

Stock market is a stochastic process with high dimensionality and fast dynamics. Hence, suitable decision making in this market is a complex task. Decision making is considered as a mapping of some inputs to predicted prices (or price rates). In one hand, it has been historically assumed that these inputs are in form of market history or fundamental analysis data [1]. Market history is represented by the time series of a symbol while fundamental analysis data is considered as macro and micro economic factors. On the other hand, the news is another factor that greatly influences the stock price. Recent advances in data mining have enabled researchers to utilize such data and turn them into quantitative ones.

In the earlier works [2–6], one of the key elements that was used as input data was market history e.g. closing price, volume and various market indicators. These factors have been widely used to predict the outcome of the market. Using time series are completely acceptable for long term investments' analyses since the fact that after a short period of uncertainty a market reaches equilibrium, and, consequently the classifier can predict long-term events. But in short term and intraday tradings, one can benefit from using the news to trade before anybody else has traded on the basis of that particular news. In other words, for a short period of time post news publication the market is inefficient and has not reached an equilibrium [7]. The classifier's prediction can be in two forms, as an estimate of price level and as a price trend or price direction. Leung et. al. showed that predicting the price trend is less prone to errors and in simulations lead to better profits [8]. Despite the fact that there have been numerous studies on using the time series and their relative stock market indicators, using the textual data is relatively a new field of study. For example, Mittermayer [9] studied an end to end system, named NEWSCATS, to categorize news into three categories namely positive, negative and neutral, and then ran a simulation to trade on the basis of this classifier. He concluded that even though the use of news provides additional information in comparison with classical inputs, a trade strategy is of utmost importance. In another study, Hagenau et. al. [10] presented some novel approaches of feature selection for extracting useful features. They concluded that it is possible and profitable to successfully use news to predict Germany's stock market. Nizer and Nievola [11] used Spanish news in Brazil and successfully simulated trading in Brazilian market.

In spite of the fact that humans can accurately judge the effect of news articles on financial markets, their capabilities are frequently hampered by their speed limitations, impartiality in their judgments and psychological factors [12]. An automated engine which utilizes a classifier to predict the effect of news on the market enables us to trade shares according to defined strategies, promoting the speed and effectiveness of the sale. Indeed, an automated engine would access published news earlier than human counterparts, and, therefore it makes faster decisions. Undoubtedly, such engines are the final goal of researchers. As it has been stated above, there are a few studies on the use of news for stock market prediction. Indeed, to the best of our knowledge there is not any paper on applying news in the Iranian market. This paper presents a news based decision support system for trading stocks in Tehran Stock Exchange (TSE).

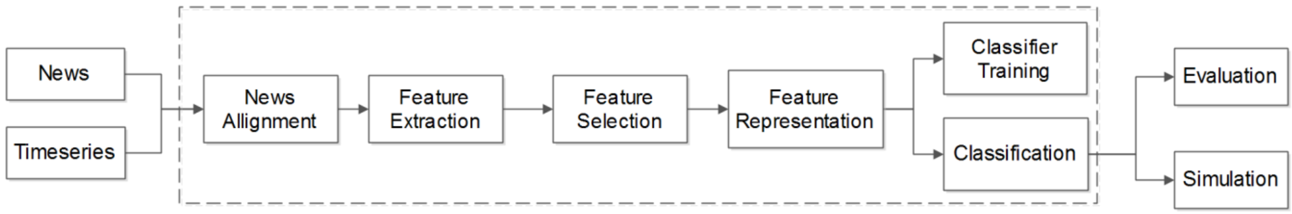

Fig. 1: Overall View of the processes involved in classifying the news

The remainder of this paper is structured as follows: in section 2 basic concepts and our proposed model are discussed. The evaluation and simulation results are given is section 3, and section 4 further discusses the implications of our work. Finally section 5 concludes the paper.

2. Materials and Methods

2.1. Basic Concepts

This section is dedicated to the presentation of basic concepts. Fig. 1 depicts the framework of methods, used for classifying the news articles. The aim is to classify the news to "negative" and "positive" classes. Text classification is a complex task and, subsequently necessitates the use of sophisticated solutions. Here, we have described the steps in Fig. 1. Our inputs are symbol's time series and related news. In order to label the training data we used exogenous feedback process, published by Hagenau et.al [10] . In this process, the label would be the reaction of the time series to published news. After news alignment, in the feature extraction phase articles were stemmed and their stop words, non-Farsi characters and numbers were removed. In contrast, the rest of features including words or a combination of them were extracted. Because of the sheer number of extracted features, it is important to reduce their numerosity to achieve better results. In the feature selection section, this would be achieved by utilizing some statistical tools such as IG[13], BNS[13] and Chi Square [13] statistic. Finally each article was represented by a TF·IDF vector which was considered as the input of classifier.

2.1.1. News Alignment

News alignment is the process of determining each article's label. In order to do this, an interval was extracted and the price movements in this interval was evaluated. This interval ranged from the immediate data point after news publication to an arbitrary data point in the future. We evaluated the classification on various timespans. Three parameters, namely Average Movement, Maximum Movement and Minimum Movement were used [14]. These parameters are calculated in equations 1-3 for the mentioned timespan after news publication and a threshold is applied.

$$Avg = \frac{Average\ Price - Publication\ Price}{Publication\ Price} \qquad (1)$$

$$Max = \frac{Maximum\ Price - Publication\ Price}{Publication\ Price} \qquad (2)$$

$$Min = \frac{Minimum\ Price - Publication\ Price}{Publication\ Price} \qquad (3)$$

$$\begin{aligned} &IF\ Avg > ThresholdA\ AND \\ &Max > ThresholdB\ THEN\ News = Positive \end{aligned} \qquad (4)$$

$$\begin{aligned} &IF\ Avg < -ThresholdA\ AND \\ &Min < -ThresholdB\ THEN\ News = Negative \end{aligned} \qquad (5)$$

In Equations 4 and 5, *ThresholdA* is 0.01 and *ThresholdB* is 0.02.

2.1.2. Feature Extraction

Stop words were removed since they made no contribution to the meaning and classification of the article. Stop words are a list of the most common Farsi words. Afterwards, each word was stemmed. Only after these steps, the actual features were extracted from text.

There are many types of features which are routinely extracted from text such as single word, n-gram and word combination. In this part we use the sample sentence "They have not achieved good results" to further explain each feature type.

Single word feature type is considered as a single word, extracted from corpus e.g. a news article. By using this feature type we would achieve the following features: "They", "have", "not", "achieved", "good" and "results".

Each n-gram feature contains n consecutive words. By using this feature type as 2-grams we would achieve the following features among others: "They have", "have not", "not achieved", etc.

Word combination is a feature type that places a window frame of size *W* on the text and extracts each possible combination of size *S* words in this window [10]. Based on the window size, this feature type allows a limited amount of distance between words in a feature. Word combination is depicted in Fig. 2.

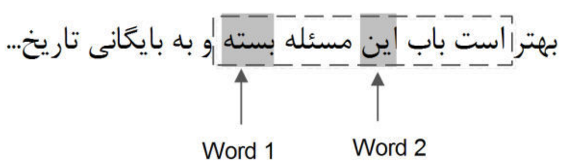

Fig. 2: Word combination applied to a sample sentence. Window size is 5 and feature size is 2.

As the features become more complex, they would also become more descriptive. For instance, features of "They have not achieved good results" can be extracted in at least two different ways. If it is attempted to extract the single word features from this sentence, the single word feature "good" would be a candidate feature. This word alone conveys a positive meaning, whereas the complete sentence conveys a negative meaning. Therefore, this feature type is not descriptive of the example sentence. A more suitable and more complex feature would be "not good". Achieving more complex and multiword features can be accomplished by using features such as word combination or n-grams. Both of these feature types would extract "not good" from the sample sentence. We have compared n-gram, word combination and single words in the results section.

2.1.3. Feature Selection

After feature extraction step, many features were obtained even though a high percentage of the data was not appropriate for further data consideration due to non-effectiveness of the features. Furthermore, because of the sheer number of the features, it was not practical to train the classifier with all the features, as this would lead to poor classification results. In order to reduce the features among others Chi Square analysis, Bi-Normal Separation (BNS) and Information Gain (IG) were used. A score was computed for each feature according to the distribution of that feature in negative and positive news articles. The main idea was to allocate low score to features with the same frequency in positive and negative classes of news articles. Afterwards, a threshold of N features was set to select the highest scores.

Chi square statistic is a metric, used in many studies [10], [14–17]. It measures the independence of the feature from the class values. In other words, it measures the independence of a feature from being negative or positive. Naturally, a value of zero would mean complete independence and would lead to lack of consideration of that particular feature. A higher value means the feature is dependent on negative or positive class (which is our desire). Chi square is obtained from Equation 6.

$$\chi^2 = \sum_{i=1}^{4} \frac{(O_i - E_i)^2}{E_i} \tag{6}$$

In Equation 6, i is one of four possible outcomes, namely the presence/absence of feature in negative/positive document, O is the observed frequency and E is the expected frequency. This equation was calculated for each feature and their respective scores were recorded.

Bi-Normal Separation is a metric which has recently gained attention for text classification purposes [10]. BNS score of a feature, as shown in Equation 7, is the difference in prevalence of a feature in negative and positive classes under the assumption of normal distribution [13]:

$$BNS = \left| F^{-1}\left(\frac{O_{pos}}{pos}\right) - F^{-1}\left(\frac{O_{Neg}}{neg}\right) \right| \tag{7}$$

In Equation 7, F^{-1} is a standard normal distribution's inverse cumulative probability function also known as z-score, pos/neg is the number of positive/negative articles and O_{pos}/O_{neg} is the number of observed feature in positive/negative articles.

Information Gain (IG) measures decrease of entropy when the feature is present compared to when the feature is absent. It can be calculated from Equations 8 and 9.

$$\begin{aligned} IG = e(pos, neg) - \\ [\frac{O_{pos} + \overline{O_{pos}}}{N} e(O_{pos}, \overline{O_{pos}}) \\ + \frac{O_{neg} + \overline{O_{neg}}}{N} e(O_{neg}, \overline{O_{neg}})] \end{aligned} \tag{8}$$

$$e(x, y) = -\frac{x}{x+y} \log_2^{\frac{x}{x+y}} - \frac{y}{x+y} \log_2^{\frac{y}{x+y}} \tag{9}$$

In Equation 8, $\overline{O_{neg}}/\overline{O_{pos}}$ is the feature's frequency of not being present in negative/positive documents and N is the document count. Document count is the total count of the news articles.

2.1.4. Feature representation

After thresholding the features in the feature selection phase, N features were selected. Therefor each article was converted to a TF·IDF vector of size N.

$$TF.IDF = TF * IDF \tag{10}$$

$$IDF = \log \frac{all}{df} \tag{11}$$

In Equation 10, Term Frequency (TF) is the frequency of a particular feature in a document. In Equation 11, Document Frequency (df) is the sum of documents containing any number of a particular feature.

2.1.5. Classification

Two classifiers are considered for classification: Support Vector Machines (SVM) and Artificial Neural networks (ANN).

ANN is a nonlinear, multiclass classifier which has been inspired from biological neural networks. The input of ANN is TF·IDF vector and the output is a single neuron with bipolar sigmoid function which produces number between -1 to 1. This output would be thresholded to become a binary value, which will classify "Negative" and "Positive" articles.

The SVM is a binary classifier which aims to maximize functional margin and was first introduced by Cortes et. al. [18]. By maximizing functional margin, better generalization is achieved. Unlike ANNs, training an SVM involves analytical solving of an equation. Thus, SVM solutions are global optima and they do not change in each run. The main parameters of SVM to adjust, are the cost parameter and the kernel parameter (in case of nonlinear kernel). Cost parameter controls the generalization of the model. If it is set too high, it will basically classify all the training data correctly but the model is not generalizable for test data. For unbalanced data, it is possible to tune different cost parameters for different classes. Therefor this approach results in a biased specificity and sensitivity [19].

2.1.6. Classification Evaluation

In order to make classification results comparable, they must be evaluated. Classification results are spread in 4 different categories: True Positive (*TP*), True Negative (*TN*), False Positive (*FP*) and False Negative (*FN*). These correspond to correctly classified positive news, correctly classified negative news, negative labeled news which are classified as positive and positive labeled news which are classified as negative. Using these 4 categories, the main evaluation measures are defined in Equations 12 through 15.

$$\text{Accuracy} = \frac{TP + FP}{TP + TN + FP + FN} \tag{12}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{13}$$

$$\text{Sensitivity} = \text{Recall} = \frac{TP}{TP + FN} \tag{14}$$

$$\text{Specificity} = \frac{TN}{FP + TN} \tag{15}$$

2.2. Proposed Model

The news dataset consisted of 32343 news articles, extracted from boursenews.ir [20]. The stock symbol, publication date and time were all extracted from each news article to build the dataset. Since the majority of the news articles did not have recognizable stock market symbol in their content they were subsequently removed, resulting in 3866 news articles. These filtered news articles ranged from August 2008 to October 2013, spanning 26 months. These were divided into 2406 training articles, 960 test articles and 500 validation articles. The validation data was only used in the Artificial Neural Network.

Overall, in 3866 inspected news articles we found 289 stock symbols, accounting for the most liquid stock symbols in the market. Each time series data point consisted of a closing rate. In order to make intraday predictions possible, data points were gathered in 15 minute time intervals.

In the feature extraction step, the words were stemmed. Farsi is inherently different from European languages and is more similar to Arabic and Urdu. Therefore, the stemmer was adjusted by considering these differences. Due to the lack of an implemented Farsi stemmer, a simple stemmer was implemented which did not incorporate part of speech tagging. Despite the fact that the mentioned stemmer was very simple, it was capable of removing certain suffix and prefixes.

In the feature selection step, three feature selection methods were used. These were IG, Chi Square and BNS. In section 3, these are compared in detail. After this step each feature had a score.

In feature representation, scores of the previous step were thresholded such that only top *N* features would remain. Afterwards, each news article was represented by a TF·IDF vector of size *N*. The effect of *N* on the accuracy and precision is reported in more detail in section 3.

Each vector was finally fed to the classifier in order to determine the trend which each news article would cause. For classification ANN and SVM were used. These are further discussed in section 3.

The applied ANN algorithm contained *N* inputs, a hidden layer with 20 hidden neurons and one neuron in its output. All the layers had a bipolar sigmoid function working as ANN's activation function. All alpha parameters of sigmoid activation functions were set to 2. The used training algorithm was the resilient backpropagation [21]. ANN is inherently prone to random behavior, meaning that its output vary according to the initial state (initial weights) of the ANN. This is caused whenever the algorithm is stuck on a local optima. In order to address this phenomenon, ANN results were averaged on 5 runs.

The SVM which we apply, uses a linear kernel, therefore it does not have any parameters. Historically, Gaussian kernel have achieved good results, but it has been shown by Joachims [22] that in problems with many features linear kernel performs best and usually these problems are linearly separable. The only tuned parameter was the *C* parameter of the SVM. Due to unbalanced data, we have penalized negative and positive classes differently, resulting in two C_is. This method of training for unbalanced datasets have been further discussed in Wang and Japkowicz [19].

Finally, in order to evaluate the effectiveness of our proposed method, a trading engine was implemented. This trading engine would trade on the basis of the SVM classifier's predictions. In order to trade, the engine was given a defined amount of fund. After each trade the engine would reinvest the fund along with all previous profits or losses. It would immediately buy/sell shares after the classifier's signal and would sell/buyback the shares the next morning at the first available time series data point. In order to simplify the trading, many assumptions were made. For example, the availability of the presumed amount of shares at the start and end point is assumed. Actually, in the real world scenario, in order to stop the engine from excessive losses, a complex strategy is required that is beyond the scope of this study [9].

TABLE I show a list of applied parameters and their respective values. This list is provided for to help reproduce the results.

TABLE I: A list of various parameters and their respective values.

News Alignment	ThresholdA (Average)	0.01
	ThresholdB (Max and Min)	0.02
Feature Extraction	Word Combination	Window size:5 Feature size:2
	N-Gram size	2
Classification (ANN)	Layers	2 layers (1 hidden)
	Neuron count	20-1
	Activation function	Sigmoid
Classification (SVM)	Kernel	Linear
	Cost	Different cost for each class, tuned to the ratio of training dataset's negative to positive count

TABLE III: Classification results. The classification results were obtained using 50 top features.

	ANN				SVM			
	Accuracy %	Precision %	Specificity %	Sensitivity %	Accuracy %	Precision %	Specificity %	Sensitivity %
BNS	79.5	55.2	80.3	77.1	86	67.9	88.2	79.2
Chi^2	64	33	69.1	47.9	69	36	78.9	37.5
IG	59.5	26.8	65.8	39.6	68	37.5	73.7	50

3. Evaluation and Simulation Results

TABLE II compares ANN and SVM statistics. Meanwhile, it depicts the effects of various feature selection metrics on classification. The results were obtained using 50 top features.

TABLE III compares various feature types and their performances, using SVM classifier with 100 top features. The evaluated feature types were word combination, N-gram and bag of words. The performance of word combination is superior both in classification results and simulation.

TABLE IV shows the effect of feature count on classifier's accuracy and precision. The results were obtained by the use of SVM classifier for 100 top features.

TABLE III: Feature type comparison. SVM was used for classification with 100 top features. W and S depict window size and feature size of word combination respectively.

	Accuracy %	Precision %	Simulation profit %
Word Combination W:5 S:2	87	76.1	76.1
2-Gram	82	65	-10.4
Bag of Words	61	24.1	-72.4

TABLE IV: Effect of feature count on classification. SVM was used for classification.

Top feature count	Accuracy %	Precision %
50	86%	67.9%
100	87%	73.9%
200	71%	38.1%
400	55%	16.1%
800	60%	23.3%

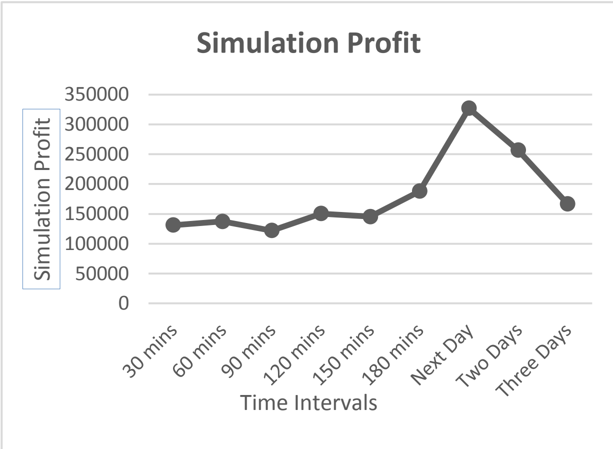

Fig. 3: Simulation profit of various new alignment timespans. Next Day indicates the first opening rate on the next working day. The results were obtained using SVM classification.

Fig. 3 shows the achieved simulation profit of various timespans in which the articles were aligned. News alignment was performed in defined time interval. These time intervals immediately started after a news publication and ended at defined time points. Profit of 326915 for Next Day significantly outperforms all other time intervals.

Although traditional evaluation metrics such as accuracy and precision are useful, the ultimate benchmark would be a simulation. Our simulations demonstrated a 76.1% profit in a two month period between 1/10/2013 to 5/8/2013, whereas the random trader reaches 3.08% profit.

4. Discussion

Promoting the efficiency of stock market trading necessitate the use of different classification methodologies. These classification methodologies are routinely based on stock market indicators, micro and macro-economic and news articles [1]. The latter has recently gained lots of attention due to the fact that news articles are routinely written by experts, and, therefore they implicitly report the most liquid share symbols. Hence, news-based stock trading is oriented towards potentially short-term profitable opportunities [7].

In our work, a key point in feature extraction was to choose the right set of stop words. The included stop words not only covered the usual language specific stop words, but also covered some domain specific words which only appear in stock market context such as "year", "dollar" and "share". The fact is that domain specific stop words can appear in top N features as part of a multi-word feature such as "year increase". These stop words would cause anomalies in the results in the form of fluctuated accuracies and precisions. Indeed, there was no particular concern when the feature type was bag of words or simple single-word features since the feature selection step would rule these out. On the other hand, the use of multi-word features caused the entrance of domain specific stop words into the top *N* features.

In our research, news alignment was applied within defined time intervals. Each time interval consisted of two data points. The first data point was immediately acquired after news publication. The second data point was set to 30, 60, 90, 120, 150, 180 minutes, next day, two days and three days post first data point. Among all second data points, inspected in TSE, Next Day resulted in the most simulation profit with 76.1% profit. Next Day indicates the first opening rate on the next working day. This interval might seem to have relatively long time span even though time interval is inherently variable for different stock markets. Indeed, the length of the time interval is dependent on the efficiency of the stock market. Indeed, the studied TSE stock market had lower efficiency than European and American counterparts.

In the next step by the use of ANN and SVM classifiers, we classified the news articles into positive and negative trends. The results indicated that SVM performs better than ANN in most cases. Indeed, this finding was in alignment with other study results in which they showed that not only SVMs are more resistant to overfitting but also are a good match for financial predictions [15,23]. Hence, we continued all other experiments by the use of SVM algorithm.

Finally, In order to demonstrate effectiveness of the proposed algorithm in the real world, the classification results were simulated. To this end, a simple trading engine was implemented. To the best of our knowledge, the simulation outcomes are one of the best results in similar context [10,7]. The trading engine reached 76.1% profit in two months. Indeed, in comparison with other similar studies, this profit rate is considerably high. However, it is worth noting that the average market growth in the same period was 24.54%. In other words, the buy-and-hold strategy would result in the average profit of 24.54%.

Collectively, in this study we showed that by classifying news articles in TSE, it was possible to achieve profits which were higher than the average market growth. In spite of the fact that our simulation results were achieved in TSE by the use of Farsi news articles, the proposed algorithm can be easily transferred to markets which are located in countries with similar languages. Indeed, Arabic and Urdu are very similar to Farsi, linguistically dominating a considerable proportion of stock markets in the world, and, therefore our proposed method paves the way for news based trading in these markets.

5. Conclusion

This paper demonstrated the prediction of stock markets on TSE by news articles. Features were extracted and reduced by feature selection. Afterwards, TF·IDF of the words in the news were used to represent the features and SVM and ANN were used as classifier. Experiment results showed that the best feature selection method for the application is BNS and for classification SVM outperforms ANN. Moreover, it was shown that trading on the basis of these predictions was profitable and the profit was higher than a random trader and even better than average market growth of TSE.

Furthermore, it is concluded that the best time interval for news alignment is immediately after the news publication to the next day morning's first data point. This is a market dependent variable.

As future works, the provided approach, can be used for other markets such as forex and gold. Of course the response times for news alignment might greatly vary. Although each language is inherently different, Middle Eastern language differences is minimal. Finally, there are several feature types (e.g. noun phrases and verb phrases) which can be tested and might show promise.

Acknowledgement

This research was supported in part by Excellence for Robust Intelligent Systems of Yazd University. We appreciate this support.

References

[1] G. S. Atsalakis and K. P. Valavanis, "Surveying stock market forecasting techniques – Part II: Soft computing methods," *Expert Systems with Applications*, vol. 36, no. 3, pp. 5932–5941, 2009.

[2] W. Leigh, R. Purvis, and J. M. Ragusa, "Forecasting the NYSE composite index with technical analysis, pattern recognizer, neural network, and genetic algorithm: a case study in romantic decision support," *Decision Support Systems*, vol. 32, no. 4, pp. 361–377, Mar. 2002.

[3] H. Kim and K. Shin, "A hybrid approach based on neural networks and genetic algorithms for detecting temporal patterns in stock markets," *Applied Soft Computing*, vol. 7, no. 2, pp. 569–576, Mar. 2007.

[4] F. a. de Oliveira, C. N. Nobre, and L. E. Zárate, "Applying Artificial Neural Networks to prediction of stock price and improvement of the directional prediction index – Case study of PETR4, Petrobras, Brazil," *Expert Systems with Applications*, vol. 40, no. 18, pp. 7596–7606, Dec. 2013.

[5] A. Esfahanipour and W. Aghamiri, “Adapted neuro-fuzzy inference system on indirect approach TSK fuzzy rule base for stock market analysis,” *Expert Systems with Applications*, vol. 37, no. 7, pp. 4742–4748, 2010.

[6] P.-C. Chang, D. Wang, and C. Zhou, “A novel model by evolving partially connected neural network for stock price trend forecasting,” *Expert Systems with Applications*, vol. 39, no. 1, pp. 611–620, Jan. 2012.

[7] T. Geva and J. Zahavi, “Empirical evaluation of an automated intraday stock recommendation system incorporating both market data and textual news,” *Decision Support Systems*, vol. 57, pp. 212–223, Jan. 2014.

[8] M. T. Leung, H. Daouk, and A. S. Chen, “Forecasting stock indices: a comparison of classification and level estimation models,” *International Journal of Forecasting*, vol. 16, no. 2, pp. 173–190, 2000.

[9] M.-A. Mittermayer, “Forecasting intraday stock price trends with text mining techniques,” in *Proceedings of the 37th Annual Hawaii International Conference on System Sciences*, 2004.

[10] 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.

[11] P. S. M. Nizer and J. C. Nievola, “Predicting published news effect in the Brazilian stock market,” *Expert Systems with Applications*, vol. 39, no. 12, pp. 10674–10678, 2012.

[12] K. Daniel, D. Hirshleifer, and A. Subrahmanyam, “Investor Psychology and Security Market Under- and Overreactions,” *Journal of Finance*, vol. 53, no. 6, 1998.

[13] G. Forman, “An extensive empirical study of feature selection metrics for text classification,” *The Journal of Machine Learning Research*, vol. 3, pp. 1289–1305, 2003.

[14] K. G. Aase and P. Öztürk, “Text mining of news articles for stock price predictions,” M.Sc. Thesis, Dep. of Computer and Information Science, Norwegian University of Science and Technology, Norway, 2011.

[15] 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.

[16] S. S. Groth, M. Siering, and P. Gomber, “How to enable automated trading engines to cope with news-related liquidity shocks? Extracting signals from unstructured data,” *Decision Support Systems*, vol. 62, pp. 32–42, Jun. 2014.

[17] P. Falinouss, D. M. Sepehri, and D. M. Limayem, “Stock trend prediction using news articles,” M.Sc. Thesis, Dep. of business administration and social sciences, Lulea University of Technology, Sweden, 2007.

[18] C. Cortes and V. Vapnik, “Support-vector networks,” *Machine Learning*, vol. 20, no. 3, pp. 273–297, Sep. 1995.

[19] B. X. Wang and N. Japkowicz, “Boosting support vector machines for imbalanced data sets,” *Knowledge and Information Systems*, vol. 25, no. 1, pp. 1–20, 2009.

[20] “BourseNews.” [Online]. Available: http://www.boursenews.ir/. [Accessed: 06-Oct-2014].

[21] M. Riedmiller and H. Braun, “A direct adaptive method for faster backpropagation learning: the RPROP algorithm,” in *IEEE International Conference on Neural Networks*, 1993, pp. 586–591.

[22] T. Joachims, “Text categorization with support vector machines: Learning with many relevant features,” in *Proceedings of the European Conference on Machine Learning*, 1998.

[23] R. Choudhry and K. Garg, “A hybrid machine learning system for stock market forecasting,” *World Academy of Science, Engineering and Technology*, vol. 39, pp. 315–318, 2008.