Financial News Classification using SVM

Rama Bharath Kumar*, Bangari Shravan Kumar, Chandragiri Shiva Sai Prasad*****

* Department of Computer Science and Engineering, Jayamukhi Institute of Technological Sciences , India.
** Department of Computer Science and Engineering, University college of Engineering, Osmania University, India.

Abstract- Stock market prediction is an attractive research problem to be investigated. News contents are one of the most important factors that have influence on market. Considering the news impact in analyzing the stock market behavior, leads to more precise predictions and as a result more profitable trades. So far various prototypes have been developed which consider the impact of news in stock market prediction. In this paper, the main components of such forecasting systems have been introduced.

Index Terms- news mining, knowledge discovery, stock market prediction, data mining

I. Introduction

Data mining can be described as “making better use of data”. Every human being is increasingly faced with unmanageable amounts of data; hence, data mining or knowledge discovery apparently affects all of us. It is therefore recognized as one of the key research areas. Ideally, we would like to develop techniques for “making better use of any kind of data for any purpose”. However, we argue that this goal is too demanding yet. It may sometimes be more promising to develop techniques applicable to specific data and with a specific goal in mind.

Individuals, researchers, investors, financial professionals, are continually looking for a superior system which will yield them high returns. One of the best known concepts in finance is that markets are efficient. An efficient market adjusts prices without delay to reflect all available public information thus making it not possible to make excessive profits. The efficient market hypothesis was associated with the idea of a "random walk”. However, financial economists now believe that our securities markets are at least partially predictable.

As a part of this research to achieve this purpose, a software tool has been developed which has a simple graphical user interface. This can be used to perform the required analysis by an investor. Thirdly, accuracy of results of the model is compared against a traditional forecasting method, linear regression analysis and the probability of the model's forecast being correct is calculated.

Financial forecasting is still regarded as one of the most challenging applications of modern time series forecasting. Financial time series have very complex behavior, resulting from a huge number of factors which could be economic, political, or psychological. These are inherently noisy, non-stationary, and deterministically chaotic.

The number of proposed methods in financial time series prediction is tremendously large. These methods rely heavily in using structured and numerical databases. In the field of trading, most analysis tools of the stock market still focus on statistical analysis of past price developments. But one of the areas in stock market prediction comes from textual data, based on the assumption that the course of a stock price can be predicted much well by looking at appeared news articles. In stock market, the share prices can be influenced by many factors, ranging from news releases of companies and local politics to news of superpower economy.

Easy and quick availability to news information was not possible until the beginning of the last decade. In this age of information, news is now easily accessible, as content providers and content locators such as online news services have sprouted on the World Wide Web. Nowadays, there is a large amount of information available in the form of text in diverse environments, the analysis of which can provide many benefits in several areas. The continuous availability of more news articles in digital form, the latest developments in Natural Language Processing (NLP) and the availability of faster computers lead to the question how to extract more information out of news articles. It seems that there is a need for extending the focus to mining information from unstructured and semi-structured information sources. Hence, there is an urgent need for a new generation of computational theories and tools to assist humans in extracting useful information (knowledge) from the rapidly growing volumes of unstructured digital data. These theories and tools are the subject of the emerging field of knowledge discovery in text databases, known as text mining.

Knowledge Discovery in Databases (KDD), also known as data mining, focuses on the computerized exploration of large amounts of data and on the discovery of interesting patterns within them. Until recently computer scientists and information system specialists concentrated on the discovery of knowledge from structured, numerical databases and data warehouses. However, a lot of information nowadays is available in the form of text, including documents, news, manuals, email, and etc. The increasing number of textual data has led to knowledge discovery in unstructured (textual databases) data known as text mining or text data mining. Text mining is an emerging technology for analyzing large collections of unstructured documents for the purposes of extracting interesting and non-trivial patterns or knowledge. Text mining has a goal to look for patterns in natural language text and to extract corresponding information.

One of the applications of text mining is discovering and exploiting the relationship between the document text and an external source of information such as time stamped streams of data namely stock market quotes. Predicting the movements of stock prices based on the contents of news articles is one of many applications of text mining techniques. Information about company’s report or breaking news stories can dramatically affect the share price of a security. There have been many researchers conducted to investigate the influence of news articles on stock market and the reaction of stock market to press releases.

Researchers have shown that there is a strong relationship between the time when the news stories are released and the time when the stock prices fluctuate. This made researchers enter to a new area of research, predicting the stock trend movement based on the content of news stories. While there are many promising forecasting methods to predict stock market movements based on numeric time series data, the number of predicting methods concerning the application of text mining techniques using news articles is few. This is because text mining seems to be more complex than data mining as it involves dealing with text data that are inherently unstructured and fuzzy.

The main objective of this design is to answer the question of how to predict the reaction of stock market to news article, which are rich in valuable information and are more superior to numeric data. The influence of news articles on stock price movement, different data and text mining techniques are implemented to make the prediction model. With the application of these techniques the relationship between the news features and stock prices are found and a prediction system would be learned using text classifier. Feeding the system with upcoming news, it forecasts the stock price trend. Moreover in this design aims to show that how much valuable information exists in textual databases which with the help of text mining techniques can be extracted and used for various purposes. The following questions rise.

- How to classify the textual financial news?
- How data and text mining techniques help to generate this predictive model?

In order to investigate the impact of news on a stock trend movement, this has to make a prediction model.

II. PREDICTION SYSTEMS

Keyword tuples contains over four hundred individual sequences of words such as “bond strong”, “dollar falter”, “property weak”, “dowrebound”, “technology rebound strongly”, etc. These are sequences of words (either pairs, triples, quadruples or qu~ntup~epsr~ov ided once by a domain expert and judged to be influential factors potentially moving stock markets.

Prediction is done as follows:

The number of occurrences of the keyword tuples in the news of each day is counted.

1. The occurrences of the keywords are then transformed into weights (a real number between zero and one). This way, for each day, each keyword gets a weight.
2. From the weights and the closing values of the training data, probabilistic rules are generated [14,151.
3. The generated rules are applied to today's news. This predicts whether a particular index such as the Dow will go up (appreciates at least 0.5%), moves down (declines at least 0.5%) or remains steady (changes less than 0.5% from its previous closing value).
4. From the prediction whether the Dow goes up, down or remains steady and from the latest closing value also the expected actual closing value such as 8393 is predicted.

In categorizing the stock market prediction systems different dimensions can be considered:

Input data:

Some prediction methods are based on historical market prices and use technical analysis to predict the market. Some other methods are based on analyzing the news content; however combination of historical market data and news can also be used.

Prediction goal:

The possible market prediction goal can be the future stock price or the volatility of the prices or market trend. Market trend is the general direction of the stock’s prices which is upward or downward. Market volatility is defined in [7] as: “the amount of uncertainty or risk about the size of changes in a security's value”. A higher volatility means the higher fluctuation of the corresponding stock prices.

Prediction horizon:

Prediction horizon is the time span in which the prediction would be valid. It can be short-term or long-term prediction. Short-term prediction starts from 5 minutes to 1 hour after the news release and long term starts from 24 hours and can last longer.

2.1 A News Based Prediction System Processes

News based stock market prediction can be considered as a text classification task. Generally the goal is to forecast some aspects of the stock market such as price or volatility based on the news content. Based on prediction goal described in previous section, a set of final classes are defined, such as “Up” (which means this news cause the prices to go up), “Down” (which means this piece of news is probable to causes decrease in prices) and etc. the prediction system is supposed to classify the incoming news into one of these classes.

News based market prediction can be divided into two main phases. “Training phase” and “Operational phase”. In operational phase, one of the predefined classes will be assigned to incoming news; however, to make the system ready for the operational phase a classifier should be trained in the training phase. Machine learning techniques are widely used to automate such processes. As a part of the training phase, a set of training data shall be prepared which in our case the train data are the pre-classified news and market information such as market prices. These labeled (pre-classified) news and possibly market numerical data will be processed to be fed into the classifier for training. The trained classifier would be ready to get a piece of news and assign a class to it in operational phase of the system.

Generally, such predictive systems consist of following components which are depicted in Figure 1:

- News labeling
- Classifier input generation
- Classification

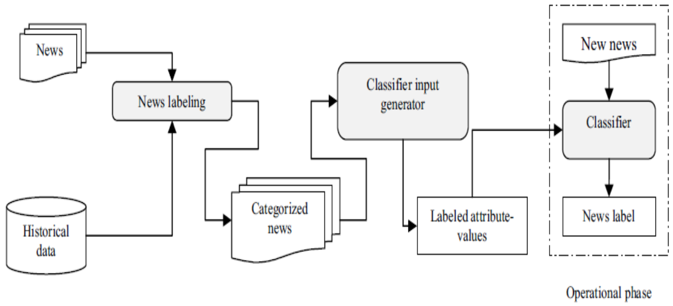


Figure 1: Overall view of a news based Stock Prediction System

2.1.1 News Labeling

Based on the selected prediction goal, a set of classes are defined for the news and the attempt is to identify the class that each news article belongs to and label them accordingly.

There are two ways to assign labels, manually and automated. In the first, financial market experts will read the news and assign a class based on their opinion. In automated assigning, time stamped numerical market data is analyzed to determine the right class for a piece of news [11][12][13]. Usually a time interval around the news release will be selected and the prices are analyzed in that interval to determine the news impact. For example Fung et al. [14] divided the time series data into independent segments and labeled the segments according to its average slope. Mittermayer and Knolmayer [15] labeled the news based on the percentage change of the price 15 minutes after the news release. If a news article caused at least 3 % increase in the price, it is labeled as “BUY” and labeled “Short” if decreases 3% respectively.

Prediction goal, number of pre-defined classes and prediction horizon are the aspects that affect the method applied in news labeling.

Prediction goal: If the interest is to detect the impact of news on market volatility, the price changes should be analyzed and the corresponding labels such as “High impact”, “Low impact” and etc are assigned to news. On the other hand, if the classification is based on market trends (whether the market will go up or down), prices are analyzed accordingly to discover the news impact on the market direction.

Number of predefined classes: Number of final classes can affect determining the criteria for assigning a class to a piece of news. E.g. if the number of final classes are 2, then usually the prices in a time range after or before the news release will be compared with the price at the time of news release to decide about the news label [15][16]. On the other hand, if the number of final classes are more than two the percentage of price change in the defined range is used to determine the correct label [14].

Prediction horizon: Prediction horizon affects selecting time interval to process the market prices e.g. if the goal is to predict the news impact on tomorrow close price, then the horizon will be 24 hours[17], [12] which means during 24 hours of the news release, the prices or other market technical indicators are watched to analyze the impact of the news; On the other hand for short term prediction the time interval can be 5 to 20 minutes after the news release[15],[13].

2.1.2 Classifier Input Generation

The success of any classifier in generating accurate results is highly dependent to the way that the input data are presented. Specific features of the input data are selected to represent the whole document. In classifying the news two important factors should be considered; the news content and the numerical market data such as stock prices. Both of these factors shall be considered for classifier input generation.. Regarding the comprising elements in the input vector in vector based classifiers, two main approaches are followed.

- News Content
- Combination of news and numerical market data

In the first approach the news content is used as input data source, while in the second approach market data such as stock price at the time of news release [13], closing price and change indicator values [18] are included in the classifier input.

Numerical data are the representatives of the real stock market situation around publication of news. By comprising more expressive data about the stock in the input vector, the classification accuracy can be improved. In fact the classifier is expected to find the answer of this question: the current situation of the market is x, and the current news has the content y, so what would be the possible impact of this piece of news on the selected stock?

Representing the news section consists of two main tasks: feature selection and feature weighting. First a set of features will be selected to represent a piece of news and next step is to assign weights to theses selected features. These weighted vectors would be the inputs to classifier.

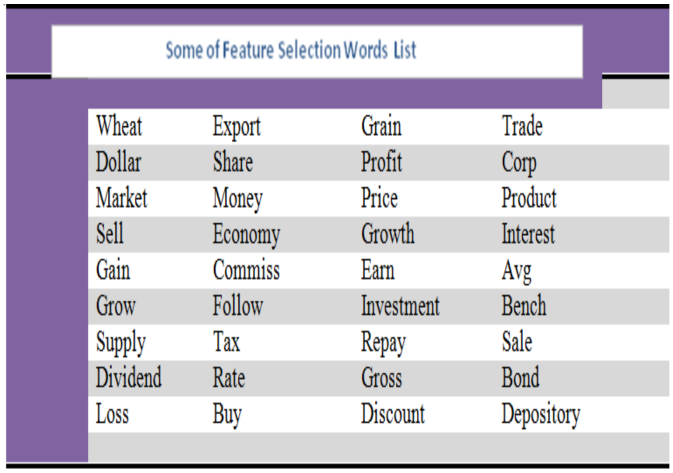
2.1.2.1 Feature selection

Features are the representatives of a document in a classification problem. Based on the classification goal a setoff features from document should be selected which best convey the document content in feature selections two main approaches has been followed.

Generating a term dictionary a set of terms are gathered and used as the fixed vector elements. Usually a group of financial experts select the representative terms, for each category there exists a set of special vocabulary that if exist in a document the probability of belonging the document to that category would be higher.

Bag of words:

In this method all the training news words are extracted. The stop words are removed. Sometimes stemming is done in which the stem of each word is replaced with the original word. Usually by applying *tf-idf* the terms with highest meaning contribution will be selected as representatives.



Some of Feature Selection Words List

Wheat	Export	Grain	Trade
Dollar	Share	Profit	Corp
Market	Money	Price	Product
Sell	Economy	Growth	Interest
Gain	Commiss	Earn	Avg
Grow	Follow	Investment	Bench
Supply	Tax	Repay	Sale
Dividend	Rate	Gross	Bond
Loss	Buy	Discount	Depository

2.1.2.2 Feature Weighting

Feature weighting is the process of assigning values to the selected terms. The Boolean values are used for weighting. The words with higher degree of membership as input features and binary representation for weighting is applied.

However by assigning non Boolean values, classification can be more accurate. Usually *tf-idf* is used to calculate the weights.

Term Frequency:

$$\mathrm{tf}_{\mathrm{i,j}} = \frac{n_{i,j}}{\sum_k n_{k,j}}$$

where *ni,j* is the number of occurrences of the considered term (*ti*) in document *dj*, and the denominator is the sum of number of occurrences of all terms in document *dj*, that is, the size of the document | *dj* |

Inverse Document Frequency:

$$\mathrm{idf_i} = \log \frac{|D|}{|\{j : t_i \in d_j\}|}$$

IDF = log (total-number-of-documents / number-of-documents-containing-t)

| *D* |: Cardinality of D, or the total number of documents in the corpus |{j :t(i) belongs to d(j)}| number of documents where the term *ti* appears (that is n(i.j) not equals to 0). If the term is not in the corpus, this will lead to a division-by-zero. It is therefore common to use 1+ |{j :t(i) belongs to d(j)}| .

Then (tf-idf$)_{(i,j)}$ = tf$_{(i,j)}$ * idf $_{(i)}$

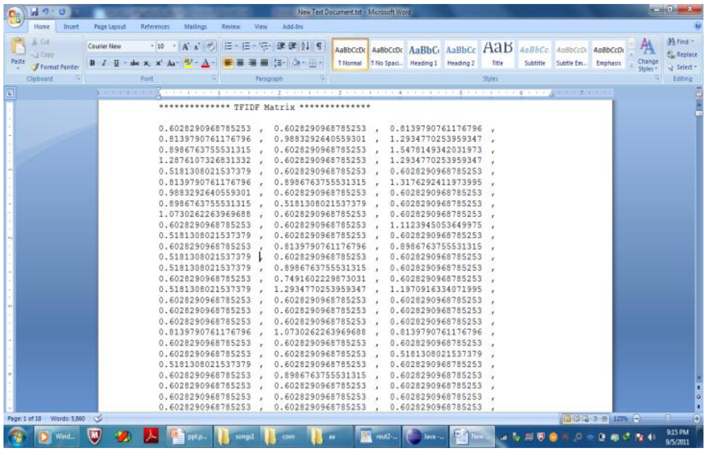
Example:

Consider a document containing 100 words wherein the word *profit* appears 3 times. Following the previously defined formulas, the term frequency **(TF)** for *profit* is then (3 / 100) = **0.03**.

Now, assume we have 10 million documents and *profit* appears in one thousand of these.

Then, the Inverse Document Frequency is calculated as log (10 000 000 / 1 000) **= 4**.

The **TF-IDF** score is the product of these quantities: 0.03 × 4 = **0.12.**



```
************** TFIDF Matrix **************

0.6028290968785253 , 0.6028290968785253 , 0.8139790761176796 ,
0.8139790761176796 , 0.9883292640559301 , 1.2934770253959347 ,
0.8986763755531315 , 0.6028290968785253 , 1.5478149342031973 ,
1.2876107326831332 , 0.6028290968785253 , 1.2934770253959347 ,
0.5181308021537379 , 0.6028290968785253 , 0.6028290968785253 ,
0.8139790761176796 , 0.8986763755531315 , 1.3176292411973995 ,
0.9883292640559301 , 0.6028290968785253 , 0.6028290968785253 ,
0.8986763755531315 , 0.5181308021537379 , 0.6028290968785253 ,
1.0730262263969688 , 0.6028290968785253 , 0.6028290968785253 ,
0.6028290968785253 , 0.6028290968785253 , 1.1123945053649975 ,
0.5181308021537379 , 0.6028290968785253 , 0.6028290968785253 ,
0.6028290968785253 , 0.8139790761176796 , 0.8986763755531315 ,
0.5181308021537379 , 0.6028290968785253 , 0.6028290968785253 ,
0.5181308021537379 , 0.8986763755531315 , 0.6028290968785253 ,
0.6028290968785253 , 0.7491602229873031 , 0.6028290968785253 ,
0.5181308021537379 , 1.2934770253959347 , 1.1970916334071995 ,
0.6028290968785253 , 0.6028290968785253 , 0.6028290968785253 ,
0.6028290968785253 , 0.6028290968785253 , 0.6028290968785253 ,
0.6028290968785253 , 0.6028290968785253 , 0.6028290968785253 ,
0.8139790761176796 , 1.0730262263969688 , 0.8139790761176796 ,
0.6028290968785253 , 0.6028290968785253 , 0.6028290968785253 ,
0.6028290968785253 , 0.6028290968785253 , 0.5181308021537379 ,
0.5181308021537379 , 0.6028290968785253 , 0.6028290968785253 ,
0.6028290968785253 , 0.8986763755531315 , 0.6028290968785253 ,
0.6028290968785253 , 0.6028290968785253 , 0.6028290968785253 ,
0.6028290968785253 , 0.6028290968785253 , 0.6028290968785253 ,
0.6028290968785253 , 0.6028290968785253 , 0.6028290968785253 ,
```

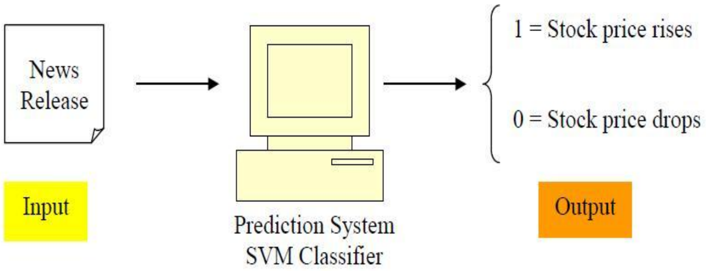
2.1.3 Classification

Classification is analyzed from two aspects:

- Number of classes
- Classification algorithm

In processing the news generally the goal is to classify the news into two classes either good news or bad news regarding the selected stock. Sometimes this classification is extended and another category indicating neutral news is added. If the degree of news influence is important to be identified more final categories will be defined

In most of the methodologies the Support Vector Machine (SVM) selected as their classification algorithm. SVM is a binary classifier which tries to classify the input data by defining a hyper plane or a set of hyper planes in high-dimensional space. SVM tries to maximize the distance of the hyper plane with the nearest data points of each class.



Prediction Model

Evaluations of Performance:

		Predicted	
		negative	positive
Actual	Negative	a	b
	Positive	c	d

Accuracy (AC) is the proportion of the total number of predictions that were correct.

AC = (a + d) / (a + b + c + d)

Recall is the proportion of positive cases that were correctly identified.

R = d / (c + d)

Precision is the proportion of the predicted positive cases that were correct.

III. Experimental Results

Result of Prediction Model:
The decision made by the classifier can be represented in a structure known as a confusion matrix or contingency table. This is used for the evaluation of classifier performance.

	Predicted by Model			
Actual		Rise(1)	Drop(0)	
	Rise(1)	TP = 2	FN = 1	3
	Drop(0)	FP = 1	TN = 2	3
		3	3	Total = 6

Result of Prediction Model

Accuracy, Precision and Recall Evaluation
According to the confusion matrix, among the 6 pieces of news 3 of them are actually labeled as rise and 3 of them are actually labeled drop. From the 3 rise labeled news, the model predicts 2 of them correctly as rise and the remaining 1 are incorrectly labeled as drop. On the other hand, from the 3 drop labeled news, 2 of them are correctly labeled as drop and 1 of them are incorrectly labeled as rise. From these calculate the model total accuracy, true positive rate (recall for rise category), true negative rate (recall for drop category), precision for rise, and precision for drop category.

1 Accuracy = (2+2)/6 = **66.66%**
2 True Positive Rate (Recall - Rise) = 2/3 = 66.66%
3 True Negative Rate (Recall - Drop) = 2/3 = 66.66%
4 Precision (Rise) = 2/3 = 66.66%
5 Precision (Drop) = 2/ 3 = 66.66%

	Predicted by Model			
Actual		Rise(1)	Drop(0)	
	Rise(1)	TP = 40	FN = 19	59
	Drop(0)	FP = 6	TN = 84	90
		46	103	Total=149

Values of Prediction Model

Accuracy, Precision and Recall Evaluation
According to the confusion matrix, among the 149 pieces of news 59 of them are actually labeled as rise and 90 of them are actually labeled drop. From the 59 rise labeled news, the model predicts 40 of them correctly as rise and the remaining 19 are incorrectly labeled as drop. On the other hand, from the 90 drop labeled news, 84 of them are correctly labeled as drop and 6 of them are incorrectly labeled as rise. From these calculate the model total accuracy, true positive rate (recall for rise category), true negative rate (recall for drop category), precision for rise, and precision for drop category.

1 Accuracy = (40+84)/149 = **83%**
2 True Positive Rate (Recall - Rise) = 40/59 = **67%**
3 True Negative Rate (Recall - Drop) = 84/90 = **93%**
4 Precision (Rise) = 40/46 = **87%**
5 Precision (Drop) = 84/ 103 = **81%**

The total accuracy of prediction model is equal to 0.83, which means 83% of time the model predicts correctly both the rise and the drop trends. But if you notice the true positive (67%) and true negative (93%) rates, you will realize that the model predicts the drop trend 1.38 times (93/67) better than the rise trend and the recall of drop category outperforms the recall of rise category.

Inversely, precision of rise category do better than the precision of drop category. It means that among the total number of rise labeled by the model, more of them are actually labeled as rise (40 out of 46)

But among the total number of drop labeled by the model, fewer of them (in compare with rise category) are actually labeled as drop.

For evaluating our prediction model, compare it with news random labeling (labeling without the application of prediction model). In order to do so, It labeled the 149 news randomly as either 1(rise) or 0 (drop) by generating Bernoulli trial. The random labeling is then compared with the actual labels and the confusion matrix for news random labeling generated. The result of news random labeling is provided in Table.

	News Random Labeling			
Actual		Rise(1)	Drop(0)	
	Rise(1)	TP = 23	FN = 36	59
	Drop(0)	FP = 36	TN = 54	90
		46	103	Total=149

Confusion Matrix for News Random Labeling

Like the previous case, among the 149 news selected for prediction, 59 of them are actually labeled as rise and 90 of them are actually labeled as drop. Among the 59 actually rise labeled news, 23 of them are correctly predicted as rise, and 36 of them are incorrectly predicted as drop which indicates that most of the actually rise labeled news are predicted incorrectly as drop resulting in the true positive rate of less than 50% (TPR = 38%) which is no good at all. Among the 90 actually drop labeled news, 54 of them are correctly predicted as drop, and 36 of them are incorrectly predicted as rise.

1 Accuracy = (23+54)/149 = **51%**
2 True Positive Rate (Recall - Rise) = 23/59 = **38%**
3 True Negative Rate (Recall - Drop) = 54/90 = **60%**

4 Precision (Rise) = 23/59 = **38%**
5 Precision (Drop) = 54/90 = **60%**

Comparing the results obtained from labeling with classification model and labeling randomly, one can realize that using the classification model improves the prediction to a great extent. The total accuracy for random news labeling is equal to 51% and compared to accuracy of prediction model, It can conclude that prediction model predicts 1.62 times (83/51) better than the random prediction and the model improves the prediction 30% from 51% to 83%. The other measure (recall and precision) of random prediction is also much lower than the prediction model.

IV. CONCLUSION

Stock markets have been studied over and over again to extract useful patterns and predict their movements. Financial News Classification based solely on the technical and fundamental data analysis. Textual data such as news articles have richer information, hence exploiting textual information especially in addition to numeric time series data increases the quality of the input and improved predictions are expected from this kind of input rather than only numerical data.

The main objective of this thesis is to predict the Classify the Financial News based on the contents of relevant news articles which can be accomplished by building a prediction model which is able to classify the news as either rise or drop. Making this prediction model is a binary classification problem which uses two types of data: past intraday price and past news articles.

The prediction model applying all the types of news related to auto industry in general and the ones related to competitors and compare the results with the current prediction model.

REFERENCES

[1] T. Hellström and K. Holmström, "Predicting the Stock Market," Technical Report Series IMATOM- 1997-07, 1998.

[2] K. Kyong-jae and I. Han, "Genetic algorithms approach to featurediscretization in aftificial neural networks for the prediction of stockprice index," Expert Systems with Applications, vol. 19, 2000, pp. 125-132(8).

[3] T.S. Quah and B. Srinivasan, "Improving Returns on Stock Investment through Neural Network Selection," Expert Syst. Appl.,vol. 17, 1999, pp. 295-301.

[4] W. Huang, Y. Nakamori, and S. Wang, "Forecasting stock market movement direction with support vector machine," Computers & Operations Research, vol. 32, 2005, pp. 2513-2522.

[5] S. Chun and Y. Park, "Dynamic adaptive ensemble case-based reasoning: application to stock market prediction," Expert Systems with Applications, vol. 28, 2005, pp. 435-443.

[6] E.F. Fama, "Market Efficiency, Long-Term Returns, and Behavioral Finance," Journal of Financial Economics, vol. 49, 1998, pp. 283-306.

[7] E.F. Fama, "Efficient Capital Markets: A Review of Theory and Empirical Work," The Journal of Finance, vol. 25, 1970, pp. 383-417.

[8] R.A. Haugen, The new finance : the case against efficient markets, New Jersey: Prentice Hall, 1999.

[9] M. Kaboudan, "Genetic programming prediction of stock prices," Computational Economics, vol. 16, 2000, p. 207–236.

[10] E. Faerber, "Fundamental Analysis," All about stocks: the easy way to get started, McGraw-Hill, 2000, pp. 129-168.

[11] G. Rachlin, M. Last, D. Alberg, and A. Kandel, "ADMIRAL: A Data Mining Based Financial Trading System," 2007 IEEE Symposium on Computational Intelligence and Data Mining, 2007, pp. 720-725.

[12] Y. Zhai, A. Hsu, and S. Halgamuge, "Combining News and Technical Indicators in Daily Stock Price Trends Prediction," Lecture Notes in Computer Science, 2007, pp. 1087-1096.

[13] 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, 2009, pp. 1-19.

[14] G. Fung, J. Yu, and W. Lam, "News sensitive stock trend prediction," Lecture Notes in Computer Science, vol. Volume 233, 2002, p. 481– 493.

[15] M. Mittermayer and G.F. Knolmayer, "NewsCATS: A News Categorization And Trading System," Proceedings of the Sixth International Conference on Data Mining (ICDM'06), 2006, pp. 0-5.

[16] M. Mittermayer, "Forecasting intraday stock price trends with text mining techniques," 37th Annual Hawaii International Conference on System Sciences, 2004. Proceedings of the, vol. 00, 2004, pp. 64-73.

[17] A. Soni, N.V. Eck, and U. Kaymak, "Prediction of stock price movements based on concept map information," IEEE Symposium on Computational Intelligence in Multicriteria Decision Making, 2007, pp. 205-211.

[18] A. Mahajan, L. Dey, and S.M. Haque, "Mining Financial News for Major Events and Their Impacts on the Market," 2008 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology, Ieee, 2008, pp. 423-426.

[19] N. Fuhr, "Probabilistic Models in Information Retrieval," The Computer Journal, vol. 35, 1992, pp. 243-255.

[20] M. Mittermayer and G. Knolmayer, Text mining systems for market response to news: A survey, 2006.

[21] T. Mitchelle, Machine Learning, McGraw Hill, 1997.

[22] H. Lu, N.W. Huang, Z. Zhang, T. Chen, and W. District, "Identifying Firm-Specific Risk Statements in News Articles," pp. 42-53.

[23] S.B. Achelis, Technical Analysis from A to Z, New York: McGraw Hill, 2001.

AUTHORS

First Author – R.Bharath Kumar, MSc, Email id - bharathrama1010@gmail.com
Second Author– B.Shravan Kumar, M.Tech, Email id - bangaaris@gmail.com
Third Author – Ch. Shiva sai Prasad, (M.Tech), Email id - chandragiri.shiva@gmail.com

Correspondence Author – R.Bharath Kumar, Email id - bharathrama1010@gmail.com