Using News Articles to Predict Stock Movements Based on Different Forecasting Techniques Statistical, Regression and Text Mining

[1]Fakhre Alam, [2]Dr. Ashish Kumar, [3]Arun Vela

[1]M. Tech (CSE Dept), Manipal University Jaipur, [2]Professor, Manipal University Jaipur, [3]Professor, C-DAC, Pune

Abstract: The presence of data today enables investors at any scale to make better investment decisions. The challenge is ingesting and interpreting the data to determine which data is useful, finding the signal in this sea of information. Question is can we use the content of news analytics to predict stock price performances. This paper shows the analysis of news data to predict stock prices. This unique opportunity will advance the state of research in understanding the predictive power of the news. This power if harnessed could help financial outcomes and generate significant economic impact all over the world. Here target is to predict the 10-day market residualized returns depending upon various price movements recorded at the close and open time of the market associated with that stock using news articles.

Keywords - **Stock, Investment, News, Market close time, Market open time, Market residuals return.**

I. INTRODUCTION

The stock market follows very random walk and is dynamic because of this instability the price of stock has very complex behavior, in financial market profit opportunities are exploited as soon as they arise hence there are very important need to explore this enormous amount of data of stock market. Investor decision are based how a news articles is influencing the market. These financial news articles have information about organization vision, mission, coming project, ongoing project, activity in which it is involved and expectation from the other competitor and also looks for financial market information like the trading volume inflation, demand for product or services offered by the organization opening price and closing price, calculated return etc. It has been shown in the previous research work. The fluctuation in the stock prices is strongly related to the publication of related news article and financial market information of the stock. So, for most of the research papers either it is determined the best time to buy or sell the stock or whether the stock price will increase or decrease or may be near about same. So, this research paper we are predicting the signed confidence value between (-1, +1) which will be multiplied by the market adjusted value of the given stock over a 10-day market residualized return value. The output of this research will be a model which will be used for short term prediction.

STOCK MARKET – AT INDIAN VIEWPOINT

The idea of stock markets came to India in 1875, when Bombay Stock Exchange was established as 'The Native Share and Stockbrokers Association 'a voluntary non - profit making association. The stock market as a place where stocks shares are buying and selling. The stock market manages the day's price for stock through a process of bid and offer. You have the right to bid and buy a stock share and offer to sell the stock shares at a high price. Buyers competed for the best bid and got their highest price quoted to purchase a particular Stock Market Shares. So, sellers compete for the minimum price requested to sell the stock. When a coordinate is made between the finest offered and the best offer a trade is executed. In electronic exchanges, high-performance computers do this entire process. Stocks of different companies are listed on stock exchanges. Currently, there are 28 stock markets In India. The Bombay Stock Exchange, the National Stock Exchange and the Calcutta Stock Exchange are the three large stock exchanges in India. There are many small regional exchanges located in state capitals and other major cities. Though it is called the stock market or equity market and is primarily known for trading stocks.

SHORT TERM TRADING

Short-term trading can be beneficial and profitable, but it can also be unsafe. Short-Term trading can last for as little as some minutes to as long as several days. To achieve this strategy, traders must understand the risks and rewards of each trade. They must not only know how to spot good short-term opportunities but also how to protect themselves. So,

examine the basics of spotting good short-term trades and how to profit from them and some basic step must be followed to find the right trades at the right times [30].

Step 1: Watch the Moving Averages

A moving average is the average stock price over a particular period. The first common time outlines are 15, 20, 30, 50, 100 and 200 days. So, the overall concept is to show whether a stock price trend is increasing or decreasing. Usually, a good candidate will have a moving average that is sloping to upward, and if you are viewing for a good stock to short, you usually find a moving average that is flattening out or declining.

Step 2: Understand Overall Cycles

Usually, the markets exchange in cycles, which makes it essential to view the calendar at particular times period. Since 1950, most of the stock exchange markets gains have occurred in the November to April time period, and during the May to October period, the averages have been relatively static. So, cycles can be used to traders' advantage to determine good times to enter into long or short positions.

Step 3: Get a Sense of Market Trends

If the stock market trend is negative, we might consider shorting and do very little buying, and if the stock market trend is positive, we may consider buying very little shorting and when the overall stock trend market is against, the odds of having a successful trade drop.

II. LITERATURE REVIEW

In 2013 Ashwini et al. studies with the increasing accuracy and precision of the analytical measuring methods show that all result that is of interest cannot be described by simple univariate and even not by the linear multivariate correlations precise. However, in recent researches, it is found a set of methods said to be the artificial neural networks gives the optimized result [1]. Ayman et al. (2017), building a useful model to predict stock market future trends with small error ratio and improve the prediction accuracy. Using data mining techniques divide the task into two important part descriptive and predictive. So, the author takes predictive part, and classification task is used to predict stock movement behavior Naive Bayes, and KNN algorithm is used to build the predictive model [2]. Yauheniya et al. (2015) proposed a model how to improve stock price movements result using different categories of the news article and using the multiple kernel learning techniques for predicting the price movements and how the simultaneous use of financial news articles with different levels of relevance to the target stock can give an advantage in financial news-based forecasting. To achieve this goal, the Global Industry Classification Standard is employed for dividing stocks by industries and sectors and for assigning their corresponding news articles to five categories: stock-specific, sub-industry-specific, industry-specific, group-industry-specific and sector-specific news items [3].Hiral R. et al. presented a model for stock market price prediction in 2016.Principle goals of this model for better investment decision for stock market price. The decision policy would be driven by analyzing stock price change based on sector choice and news released and model works with three different types of forecasting techniques: Predictive, Statistical and Regression. These methods are selected by performing comparative analysis that published in the earlier research paper. As a Regression Analysis, the Multiple Linear Regression is used and as a Predictive method neural network is applied. As a Statistical Technique the Exponential smoothing, Double exponential smoothing and moving average are used [4].

Kazutaka et al. (2016) proposed the method to predict the one-day stock price performance from web news. The target is the Nikkei Stock Average, which is one of the most important stock market indexes and focuses on the evaluation method of the prediction [5]. Sahaj et al. (2017) used an idea is to predict stock market trends using machine learning models like Random forest and Support vector machine give the better performance and optimized the result. These two models are used to forecast whether the price of a stock in the future will be higher than its price on a given day based on the historical data. Using sentiment analysis of text from the content derived from news sources for understanding the natural language and inferring whether the message it provides is positive, negative or neutral towards its influence on the stock price [6].

Ramalingam et al. (2018) used statistical model for predicting the future stock market values. Using Quindl, it provides alternative data platform and the method comprises of completing the Natural Language Processing of the data and then making it easier for the system to understand, finds and identifies the correlation in between this data and the stock market fluctuations [7]. Lee et al. (2014) have introduced a system that forecasts companies' stock price changes in response to financial events. They reported that predicting next day's price movement was improved by 10% if the

text is considered [8]. Ichinose et al. (2015) have proposed a combination method with filtering and a machine learning technique for a stock market prediction task. They reported that the filtering method improved the classification accuracy. These studies handled financial documents and web news about financial topics. On the other hand, some researchers focused on Social Media, such as Message boards and Twitter [9].

Nguyen et al. [10] have proposed a stock exchange market prediction model with a topic- sentiment feature and a new topic model. By using the feature and topic model, the accuracy of the proposed method increased by 6% as compared with some baselines. Bollen et al. [11] have proposed a method with six mood dimensions on Twitter for the stock market prediction. They obtained a high accuracy rate; 86.7%. These studies only focused on the accuracy rates of the classification models. However, there was no mention whether the improvement of the accuracies contributed to making a profit. One of the most critical points for personal investors is whether they eventually gain a profit by using the prediction model.

Izumi et al. [12] have proposed a text-mining approach for long-term market analysis using market reports published by financial professionals and institutions on the web. For Japanese government bond two-year, five-year, and ten-year markets, the proposal method forecast in higher accuracy about both the level and direction of long-term market trends. The task was forecasting long-term market trends of Japanese government bond. On the other hand, the target and resources of our task are the Nikkei Stock Average and news articles on the Web.

Schumaker and Chen [13] have proposed an SVR-based method that made a discrete prediction of what +20-minute stock should be. The method, AZFinText, used financial news articles and stock price quotes. They compared AZFinText against the top 10 quants for one year. As a result, AZFinText outperformed six of the top 10 quant funds. They considered the effectiveness of the method regarding real profit-and-loss. However, they evaluated the investment performance at the end of the period. We also discuss a criterion based on property changes during the testing. Various studies have been carried out to understand the intricate relationship between sentiment and price on the financial market. Wang et al. [14] investigated the correlation between stock performance and user sentiment extracted from Stock Twits and Seeking Alpha3. Ding et al. [15] proposed a deep learning method for event-driven stock market prediction and achieved nearly 6% Improvements on S&P 500 index prediction. Arias et al. [16] investigated whether information extracted from social media Twitter, it can improve time series prediction, and could help predict the trend of volatility indices and historical volatilities of stocks.

The textual data pre-processing is an essential part of the text Mining. Concerning financial forecasting based on news, the target of the pre-processing is to extract important information from a given set of news articles that signals a change in a price, and to represent it in a machine learning friendly format. Mittermayer (2004) suggested to divide the preprocessing procedure into three preparatory steps: feature extraction, feature selection and feature representation. The feature extraction step involves the generation of the list of features that sufficiently describe the documents [17]. This terminology was applied in later works [19].

Schumaker et al. [18] compared several textual analyses techniques applied to financial articles, including the Bag-of-Words, Noun Phrasing, Named Entities and Proper Nouns Approaches. The authors claimed that Proper Nouns showed Better results than the others. Hagenau et al. [19] explored the performance of the Bag-of-Words, Noun Phrases, N-Grams (the sequence of N words) [20] and word combinations techniques moreover, found that the word combinations approach significantly outperformed the others. In this study, the Bag-of-Words the approach is used for feature extraction, where symbols such as pronouns, articles and prepositions as well as numbers, punctuation marks and stops words are removed from the data. Then word stemming techniques are usually applied to every word. Semantically clear terms are eliminated, and the remaining words are utilized to represent the article. This method is often preferred in research studies for its intuitive meaning and simplicity.

When powerful features are selected from the list of all extracted features, those containing the least information are neglected [18]. In some research works, a dictionary containing a list of terms selected by domain experts is utilized [22]. Others employ statistical information about the articles, e.g. term frequency-inverse document frequency (TF*IDF) [23], [17], [24], [25]. Recently, the use of external market feedback was proposed in some research works. Wang, Liu and Dou [10] used the Chi-square test to select features for predicting volatility. Hagenau et al. [1] examined the efficiency of the Bi-normal separation and Chi-square test methods for evaluating the analytical ability of a word. The external market feedback was employed in both methods and showed promising results. In this study, the Chi-square test is selected for feature selection.

In feature representation, the whole set of documents is represented in a machine learning friendly format [19]. For example, a feature vector of n elements is formed from each document, where n is the number of selected features [17]. The fact that a feature appears in a document is usually regarded as an essential factor. In [26], the membership value was computed for each term, and binary representation was used for assigning weights in the developed trading system. In other research works, real values are assigned to the weights. Luss and d'Aspremont [24] employed TF*IDF calculations for feature weights when predicting abnormal returns. In [27], the direction of changes in volatility was forecast with TF*IDF values used as weights. Because of their popularity in the reviewed literature, the TF*IDF values are used in this study for computing feature weights for each data point.

After all steps of the pre-processing are completed, the processed articles need to be aligned with the price data and subsequently labelled. The news articles are generally classified into two (positive or negative) or three (positive, negative or neutral) classes. Each class corresponds to the effect of a published news article on an asset price. Rachlin et al. [26] specified five classes to highlight the degree of influence of a news item. In this study, news articles were classified into having a positive or negative effect.

The Naive Bayes and SVM techniques were employed by Atelier and Frank [27] for classifying the messages extracted from Yahoo Finance also, Raging Bull websites into bearish, bullish or neutral. SVM slightly outperformed Naive Bayes regarding out-of-sample accuracy. Hagenau et al. [19] utilized SVM for classifying the effect of a message on the market price into positive or negative. The authors state that a pilot comparison study showed that SVM performs better than Naive Bayes and ANN. Considering previous findings, the SVM method is recognized as the most promising machine learning technique for text classification [19]. SVM employs the principle of structural risk minimization that minimizes the upper limit of the usual risks and construct a robust model to avoid the over-fitting problem. Many machine learning approaches use the empirical risk minimization principle to minimize the error of training that may lead to over-fitting.

Additionally, the ensemble methods are actively used for forecasting financial markets. The results obtained from base learners can be comparatively enhanced using these methods. An ensemble learning algorithm is a computational intelligence approach that integrates a set of base learners into a single model [29].

Table 1: Findings

Algorithm / Technique	Advantage	Finding
Feed-forward neural network, Back Propagation Algorithm	Using feed forward Neural Network to train the dataset and back propagation algorithm is used for prediction. It will give better and accurate result.	With increasing accuracy and precision of the analytical measuring methods show that all result that is of interest cannot be described by simple univariate and even not by the linear multivariate correlations precise. However, in recent researches, it is found a set of methods said to be the artificial neural networks gives the optimized result [1].
Naive Bayes, KNN	Naive Bayes classifier is used to classify the stock news to be either positive or negative sentiment based on TF-IDF values and KNN classifier is used to predict the stock trend up and down.	Building a useful model to predict stock market future trends with small error ratio and improve the prediction accuracy. Using data mining techniques divide the task into two important part descriptive and predictive. So, the author takes predictive part, and classification task is used to predict stock movement behavior Naive Bayes, and KNN algorithm is used to build the predictive model [2].
Multiple kernel learning techniques	It allows us to combine different kernels when it is better to use different kernels for different input features.	Improve stock price movements result using different categories of the news article. Using the multiple kernel learning techniques for predicting the price movements.

	MKL mitigates the risk of erroneous kernel selection to some degree by taking a set of kernels and deriving a weight for each kernel such that predictions are made based on the weighted sum of the kernels	The MKL approach was utilised for learning from different news categories; independent kernels were employed to learn from each subset. A number of different types of kernels and kernel combinations were used. It gives highest prediction accuracy and accurate result [3].
Multiple Linear Regression, Neural Network, Statistical Technique	Advantage of using MLR and NN algorithm for developing the model gives better result.	Principle goals of this model for better investment decision for stock market price. The decision policy would be driven by analyzing stock price change based on sector choice and news released [4].
Random Forest, Support Vector Machine	Advantage of using Random Forest and Support Vector Machine are providing a reliable prediction of stock market trends based on historical data.	The primary approach of this paper is towards prediction of stock market trends using machine learning models like Random forest and Support vector machine give the better performance model and optimised the result. These two models are used to forecast whether the price of a stock in the future will be higher than its price on a given day based on the historical data [6].

III. DISCUSSION

A stock market or equity market is the aggregation of buyers and sellers of stocks these may add securities listed on a stock exchange as well as those only traded confidentially. Stocks can be categorised in different ways. One general way is by the country where the company is domiciled. For example, Novartis and Nestle are domiciled in Switzerland, so they may be acknowledged as part of the Swiss stock market, although their stock may also be traded at exchanges in other countries. At the close of 2012, the volume of the world stock market was around US$55 trillion. By country, the biggest market was the United States (about 34%), followed by Japan (about 6%) and the United Kingdom (about 6%). This went up more in 2013. A stock market is a place or organisation by which stock traders (people and companies) can sell or buy stocks and companies may want to get their stock listed on a stock exchange. Different stocks may be traded "over the counter", that is, by a dealer.

A large company will generally have its stock listed on many exchanges across the world. Stock Exchanges may also cover different types of security such as fixed interest securities or interest derivatives. Trade in stock markets means the transfer for money of a stock or security from a seller to a buyer. This needs these two parties to agree on a price. Equities (stocks or shares) confer an ownership interest in a particular company.

Participants in the stock exchange market range from small individual stock investors to larger traders' investors and the participants can be based anywhere in the world, and may include banks, insurance companies or pension funds, and hedge funds. Their buy or sell orders may be performed on their behalf by a stock exchange trader. So, the example of such an exchange is the New York Stock Exchange. The other type of stock exchange is a virtual kind, composed of a network of computers where trades are made electronically by traders. So, the example of such an exchange is the NASDAQ [21].

This paper shows the analysis of news data to predict stock prices. This unique opportunity will advance the state of research in understanding the predictive power of the news. This power if harnessed could help financial outcomes and generate significant economic impact all over the world. We have two dataset stock market and news article dataset. Stock market data contain financial market information such as opening price, closing price, trading volume, calculated return and news data set contains information about news articles/alerts published about assets such

as article details, sentiment and other commentary. So, what we will do first, preprocessing the both dataset and after that we will join features of news and market data and then make structured data mart. Using this structured data, we will make the model by using different machine learning algorithm. So, the objectives of this modeling practice are to predict the signed confidence value (-1,1) which will be multiplied by the market adjusted value of the given stock other 10 days market residualised return value.

IV. CONCLUSION

The stock market follows very random walk and is dynamic because of this instability the price of stock has very complex behavior, in financial market profit opportunities are exploited as soon as they arise hence there are very important need to explore this enormous amount of data of stock market. Trader decision are based how a news articles is influencing the market. These financial news articles have information about organization vision, mission, coming project, ongoing project, activity in which it is involved and expectation from the other competitor and also looks for financial market information like the trading volume inflation, demand for product or services offered by the organization opening price and closing price, calculated return etc. So, analyzing of news data to predict stock prices. This unique opportunity will advance the state of research in understanding the predictive power of the news. This power if harnessed could help financial outcomes and generate significant economic impact all over the world.

V. REFERENCES

[1]   Ms. Ashwini Pravin Navghane, Dr. Pradeep Mitharam Patil, "*Forecasting Stock Market Movement: A Neural Network Approach*", IJECET, Volume 4, Issue 5, September - October 2013, pp.117-125.

[2]   Ayman E. Khedr, S.E. Salama, Nagwa Yaseen, 2017. *Predicting Stock Market behavior using Data Mining Technique and News Sentiment Analysis*: I.J. Intelligent Systems and Applications,2017,7,22-30 Published Online July 2017 in MECS (http://www.mecs-press.org/)
DOI: 10.5815/ijisa.2017.07.03.

[3]   Yauheniya Shynkeyich, T.M. Micginnity, Sonya Coleman, Ammar Belatreche, "*Predicting Stock Price Movements Based on Different Categories of News Article*" IEEE 2015.

[4]   Hiral R. Pate, Satyen M. Parikh, Dhara N. Darji, "*Prediction Model for Stock Market using News based different Classification, Regression and Statistical Techniques*", IEEE 2016.

[5]   Kazutaka Shimada, Ko Ichinose," *Stock market prediction from news on the Web and a new evaluation approach in trading*," IEEE 2016

[6]   Sahaj Singh Maini, Govinda.K," *Stock Market Prediction using Data Mining Techniques*" IEEE 2017.

[7]   V. V. Ramalingam, A. Pandian, Shivam Dwivedi and Jigar P. Bhatt "*Analyzing News for Stock Market Prediction*", National Conference
on Mathematical Techniques and its Applications, 2018.

[8]   H. Lee, M. Surdeanu, B. MacCartney, and D.   Jurafsky, "*On the importance of text analysis for stock price prediction*," in Proceedings of the 9th edition of the Language Resources and Evaluation Conference (LREC), 2014, pp. 1170–1175.

[9]   K. Ichinose and K. Shimada, "*Stock market prediction from web news using filtering and machine learning,*" in IEICE technical report SIGNLC (in Japanese), vol. 115, no. 70, 2015, pp. 91–96.

[10]   T. H. Nguyen and K. Shirai, "Topic modeling based sentiment analysis on social media for stock market prediction," in *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing of the Asian Federation of Natural Language Processing*, 2015, pp.1354–1364.

[11]   J. Bollen, H. Mao, and X.-J. Zeng, "Twitter

mood predicts the stock market," *Journal of Computational Science*, vol. 2, no. 1, pp. 1–8, 2011.

[12]   K. Izumi, T. Goto, and T. Matsui, "Trading tests of long-term market forecast by text mining," in *Proceedings of 2010 IEEE International Conference on Data Mining Workshops (ICDMW)*, 2010, pp. 935–942.

[13]   R. P. Schumaker and H. Chen, *"A discrete stock price prediction engine based on financial news," Computer, vol. 43, no. 1, pp. 51–56, 2010.*

[14]   Gang Wang, Tianyi Wang, Bolun Wang, Divya   Sambasivan, Zengbin Zhang, Haitao Zheng, and Ben Y Zhao. 2015. *Crowds on wall street: Extracting value from collaborative investing platforms. In Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*. ACM, 17–30.

[15]   Xiao Ding, Yue Zhang, Ting Liu, and Junwen Duan. 2015. Deep Learning for Event-driven Stock Prediction. *In Proceedings of the 24th International Conference on Artificial Intelligence (IJCAI'15)*. AAAIPress,2327–2333. http://dl.acm.org/citation.cfm?id=2832415.2832572.

[16]   Marta Arias, Argimiro Arratia, and Ramon Xuriguera. 2014. *Forecasting with Twitter Data. ACM Trans. Intell. Syst. Technol*. 5, 1, Article 8 (Jan.2014), 24pages.https://doi.org/10.1145/2542182.2542190.

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

[18]   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.

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

[20]   M. Butler and V. Keselj, "Financial Forecasting using Character NGram Analysis and Readability Scores of Annual Reports," in *Proc. Of the 22nd Canadian Conference on Artificial Intelligence: Advances in Artificial Intelligence*, 2009, pp. 39–51.

[21]   Vivek Rajput, Sarika Bobde, "*STOCK MARKET FORECASTING TECHNIQUES*", IJCSMC, Vol. 5, Issue. 6, June 2016, pg.500 – 506.

[22]   B. Wiithrich, D. Permunetilleke, and S. Leung, "Daily prediction of major stock indices from textual www data," in *Proc. of the 4th Intl Conference on Knowledge Discovery and Data Mining*, 1998.

[23]   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.

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

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

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

[27]   F. Wang, L. Liu, and C. Dou, "Stock Market Volatility Prediction: A Service-Oriented Multi-kernel Learning Approach," in *Proc. of IEEE 9th Intl Conference on Services Computing*, 2012, vol. d, pp. 49–56.

[28]   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.

[29]   C. Cheng, W. Xu, and J. Wang, "A Comparison of Ensemble Methods in Financial Market Prediction," in *Proc. of the 5th Intl Joint Conference on Computational Sciences and Optimization*, 2012, pp. 755–759.

[30]   Lakonishok, J., Shleifer, A., & Vishny, R. W. (1992). The impact of institutional trading on stock prices. Journal of Financial Economics, 32(1), 23–43. doi:10.1016/0304-405x (92)90023-q.