Sentiment Classification of Financial Texts for Stock Markets using LSTM Technique: A Survey

¹Kanchan Raipure, ²Prof. Mahendra Sahare, ³Prof. Anurag Shrivastava

M. Tech. Scholar¹, Associate Professor^{2,3} Department of Computer Science and Engineering NRI Institute of Information Science and Technology, Bhopal

Abstract: - Sentiment Analysis (SA) is the current field of research in text mining field. SA is detecting opinions, sentiments, and subjectivity of text. It is the application of natural language processing techniques and text analytics to identify and extract subjective information from the frequently used sources such as web and microblogs. The main objective of sentiment analysis is to analyze reviews of products and services, and determine the scores of such sentiments. The major problem is that the reviews are mostly unstructured and thus, need classification or clustering to provide meaningful information for future use. The study was conducted with different sectors like, IT sector, Automobile sector, Banking sector, Pharmaceutical sector and FMCG sector. All the sectors taken for the study is highly volatile compared to other sectors in NSE/BSE. Hence it is very essential to study on the nexus between the Indian Stock Market and selected companies behavior. The early stage of the share market was very familiar for average investor. Now the markets are wide enough to invest. There are different markets like bond market, forex market, derivative market and other specialty markets. Analysis of the stock price we take the price. By using the artificial neural network, we develop a model. within the neural network, we use a recurrent neural network that remembers each and each information through time.

Keywords: - National Stock Exchange (NSE), Bombay Stock Exchange (BSE), Neural Network

I. INTRODUCTION

Sentimental analysis is the study of people's ideas, opinions, feelings, attitudes, thoughts, and their associative features for determining the polarity of the reviews which can be used to improve products and services. Sentimental analysis is also called as opinion mining is a part of data mining activity that analyses text to topics. Sentimental analysis is classified into three types; document level, sentence level and aspect level. The sentence-level does not represent what the reviewers like or not and the Sentimental analysis is done at the document level to predict the opinion. A key problem in this area is sentiment classification, where a document is expressed as a positive or negative evaluation of a target object, the target object may be film, book, movie and business products. Sentiment Analysis supports both supervised and unsupervised learning approaches. Many machine learning approaches are used for effective sentiment classification. Naive Bayes algorithm is proposed for sentimental analysis. The classification result and the accuracy of Naive Bayes, and SVM are analyzed through a comparative experiment [1]. Techniques are also proposed to evaluate the scalability of Naive Bayes Classifier (NBC) in large scale datasets. Naïve Bayes classifier is implemented to achieve fine-grain control of the analysis procedure for different implementations. The valuable information is obtained by the typical method to extract the sentiment or opinion from a message which is useful for the business consumer industry was reviewed [2].

The financial markets in India especially stock markets have witnessed a number of changes in terms of trading environment, regulatory framework, increased competition, increase in range of instruments traded and presence of better informed market players. In a financial market investors are exposed to a variety of information on daily basis like updated corporate earnings, revised macroeconomic indices, policy maker's statements and political news. A public company initiates a number of actions, termed as corporate actions, which affect prices of securities issued by company. The corporate actions are - dividend announcements, rights issues, mergers and acquisitions, stock splits, bankruptcy, delisting, de-merger, initial public offer, liquidation, takeover, partial or final redemption, conversion of convertible bonds, warrants issue etc. These corporate actions may have direct or indirect impact on prices, volume, face value etc. of securities when information about the actions is disclosed through any channel of communication. A corporate announcement is information about the corporate action. It is important to understand which news is good or positive and which news is bad or negative for a company [3, 4]. Researchers are interested in understanding effect of different corporate announcements. Studies have been done in past to analyze impact of these corporate announcements on share prices. Stock split is a corporate announcement which has been subject of interest for academicians as well as practitioners and topic of interest in the current study. There is ample scope for academic contributions in this research area, both in terms of appropriately analyzing the impact of stock splits and in terms of extending literature by investigating new aspects of the theme [5].

The present chapter starts by introducing concept of stock splits along with brief discussion of different hypotheses relating to it formulated and tested in different empirical studies done in past covering various aspects of stock splits.

The chapter advances by throwing light on motivation of current study and framing of objectives. The chapter presents research hypotheses developed for the analysis. Structure of the thesis brings this chapter to an end.

II. RELATED WORK

Shanshan Dong et al. [1], for predicting stock markets and financial crises, sentiment classification in financial texts is crucial. The use of automatic text classification and text-based sentiment classification has grown in popularity as more and more applications in the field of natural language processing (NLP) adopt deep learning. However, due to a lack of labeled samples, financial text-based sentiment classification applications are limited. This paper proposes a domain-adaptation-based financial text sentiment classification method that can use a large amount of unlabeled target domain (TD) financial text data and sentiment-labeled source domain (SD) text data as training samples for the proposed neural network. The proposed method is based on cross-domain transfer learning. The domain classification sub network and the domain classification loss function are both added to the original neural network. As a result, the network is able to complete the classification task while simultaneously adjusting to the target domain. An open-source dataset is used in the experiment of the proposed sentiment classification transfer learning method. The reviews of Amazon books, DVDs, electronics, and kitchen appliances are used as the source domain for the proposed cross-domain learning method in this paper. The classification accuracy rates can reach 65.0%, 61.2 percent, 61.6 percent, and 66.3%, respectively.

Shravan Raviraj et al. [2], securities exchanges structure the biggest roads of interest in India basically through two stock trades: Bombay Stock Exchange) BSE and National Stock Exchange (NSE). Examiners and financial backers investigate different factors and attempt to anticipate the patterns in stock costs in these trades. Being incredibly unpredictable in nature, share value forecast is a genuinely complicated task. Regardless of the overflow of information, innovation has not had the option to complete right expectations up to an ideal exactness more often than not. The new improvements in profound learning innovation have demonstrated to be a helpful asset in working on the precision of forecasts. The proposed profound learning based expectation calculations utilize Recurrent Neural Network, Long Short Term Memory and Gated Recurrent Unit over the long haul series information acquired on the web. The created calculations predict the patterns five days ahead of time. The consequences of the expectation on stocks from different enterprises are investigated to determine important experiences.

J. J. Duarte et al. [3], market members utilize a wide arrangement of data before they choose to put resources into risk resources, like stocks. Financial backers regularly follow the news to gather the data that will assist them with choosing which procedure to follow. In this review, we dissect how public news and recorded costs can be utilized together to expect and forestall monetary misfortunes on the Brazilian securities exchange. We remember a broad arrangement of 64 protections for our examination, which address different areas of the Brazilian economy. Our examination looks at the customary Buy and Hold and the moving normal systems to a few trials planned with 11 AI calculations. We investigate day to day, week by week and month to month time skylines for both distribution and bring windows back. With this approach we had the option to survey the most significant arrangement of information for financial backer's choice, and to decide for how long the data stays pertinent to the market. We observed a solid connection between news distributions and stock cost changes in Brazil, recommending even transient exchange amazing open doors. The review shows that it is feasible to anticipate stock cost falls involving a bunch of information in Portuguese, and that text mining-based approaches can defeat conventional methodologies while determining misfortunes.

Jingqi Liu et al. [4], forecast of monetary time series is difficult for factual models. By and large, the financial exchange time's series present high unpredictability because of its aversion to monetary and political variables. Besides, as of late, the Coronavirus pandemic has caused an extreme change in the stock trade times series. In this difficult setting, a few computational procedures have been proposed to work on the presentation of foreseeing such occasions series. The fundamental objective of this article is to think about the forecast execution of five brain network structures in anticipating the six most exchanged supplies of the authority Brazilian stock trade B3 from March 2019 to April 2020. We prepared the models to anticipate the end cost of the following day utilizing as sources of info its own past qualities. We thought about the prescient execution of various direct relapse, Elman, Jordan, spiral premise work, and multi-facet perceptron designs in view of the base of the mean square blunder. We prepared all models utilizing the preparation set while hyper-boundaries, for example, the quantity of information factors and secret layers were chosen utilizing the testing set. In addition, we utilized the managed normal of 100 bootstrap tests as our expectation. Hence, our methodology permits us to quantify the vulnerability partner with the anticipated qualities. The outcomes showed that for all times series, thought about all structures, with the exception of the spiral premise work, the organizations tunning give appropriate fit, sensible forecasts, and certainty spans.

K. M. El Hindi et al. [5], the naive Bayes (NB) learning algorithm is frequently used in text classification, among other areas. However, when used in situations where its naive assumption is broken or the training set is too small to accurately estimate probabilities, its performance suffers. To solve both issues, we propose a lazy fine-tuning naive Bayes (LFTNB) approach in this research. In order to fine-tune the probability terms utilized by NB, we propose a local

fine-tuning algorithm that makes use of the query instance's closest neighbors. The independence assumption is only strengthened when the nearest neighbors are used, but the fine-tuning algorithm is used to obtain more precise estimations of the probability terms. 47 UCI datasets were used to evaluate the LFTNB approach's effectiveness. The findings demonstrate that the LFTNB approach outperforms the traditional NB, eager FTNB, and k-nearest neighbor algorithms in terms of performance. Multinomial NB, complement NB, and one-versus-all NB, all powerful NB-based text classification algorithms, are also offered in eager and lazy versions for fine-tuning. The proposed approaches perform better than untuned versions of these algorithms, as demonstrated by empirical results utilizing 18 UCI text classification datasets.

G. Ding et al. [6], securities exchange has gotten boundless consideration from financial backers. It has generally been a problem area for financial backers and venture organizations to get a handle on the change routineness of the securities exchange and anticipate its pattern. At present, there are numerous strategies for stock cost forecast. The forecast strategies can be generally isolated into two classifications: factual techniques and man-made consciousness strategies. Factual strategies incorporate calculated relapse model, ARCH model, and so on Man-made brainpower techniques incorporate multi-facet perceptron, convolutional brain organization, gullible Bayes organization, back proliferation organization, single-layer LSTM, support vector machine, repetitive brain organization, and so on Yet, these investigations foresee just a single worth. To anticipate various qualities in a single model, it need to plan a model which can deal with numerous sources of info and produces different related yield values simultaneously. For this reason, it is proposed a related profound intermittent brain network model with numerous data sources and various results in view of long momentary memory organization. The related organization model can foresee the initial value, the most reduced cost and the greatest cost of a stock at the same time. The related organization model was contrasted and LSTM network model and profound intermittent brain network model. The tests show that the precision of the related model is better than the other two models in foreseeing different qualities simultaneously, and its forecast exactness is more than 95%.

Y.-T. Tsai et al. [8], compare a white-box attack algorithm called "Global Search" with "Greedy Search," a more sophisticated and common white-box attack method, and a simple misspelling noise. The CNN sentiment classifier, trained on the IMDB movie review dataset, is used to evaluate the attack methods. The perplexity of the sentences is used to measure the degree of distortion of the generated adversarial examples, and the attack success rate is used to evaluate the effectiveness of the attack methods. According to the findings of the experiment, the proposed "Global Search" approach produces more potent adversarial examples with less distortion or modification of the source text.

Z. Zhang et al. [9], financial exchange is one of the main pieces of the speculation market. Contrasted and different businesses, the securities exchange not just has a higher pace of profit from speculation yet in addition has a higher gamble, and stock cost forecast has forever been a nearby worry of financial backers. Consequently, the exploration on stock cost expectation techniques and how to lessen the mistake of stock cost forecast has turned into an interesting issue for some researchers at home and abroad. As of late, the improvement of PC innovation, for example, AI and econometric strategy makes the stock cost forecast more solid. Because of the secret Markov nature of stock value, this paper proposes a stock cost expectation strategy in light of stowed away Markov model (HMM). To be explicit, since the information of stock cost have coherence in time series, it is important to stretch out the discrete HMM to the constant HMM, and afterward set forward the all over pattern forecast model in view of the nonstop HMM. The first-request nonstop HMM is stretched out to the second-request ceaseless HMM, and the stock cost is anticipated by joining the expectation technique for change range. Accordingly, the proposed second-request nonstop HMM-based stock cost forecast model is recreated on Hang Seng Index (HSI), one of the earliest financial exchange files in Hong Kong. The assessment results on a half year HSI show that the anticipated worth of the proposed model is exceptionally near the genuine worth and beats three benchmarks concerning RMSE, MAE, and R2.

Problem Formulation

Following are the problems which is to be consider as identify from the Base Paper:

- 1. The number of stocks share price is quite high in some specific cases, which can be further improved.
- 2. The time complexity of the algorithm is high if dimension of data increases.
- 3. Algorithms for Stocks Share Price Analysis and Prediction have much to be researched.

III. NEURAL NETWORK

MLP is a feed-forward neural system with at least one layers among information and yield layer. MLP will make locations sets of information into suitable yields. Feed forward implies that information streams in a single course form contribution to yield layer (forward). An MLP comprises of various coatings of center points in a blended outline, with each layer completely connected with the accompanying one. Beside the information centers, each center is a neuron (or taking care of part) with a nonlinear commencement work. This kind of framework is ready with the back designing learning estimation. MLPs are for the most part used for test request, affirmation, forecast, and estimation. Multi-Layer Perceptron can take care of issues which are not straightly detachable. MLPs make distinct categories through the

utilization of Hyper planes. MLPs utilize disseminated training. MLPs contain at least one shrouded layers [12]. Effective Market Hypothesis (EMH) is a hypothesis that states it is hard to "beat the market" since securities exchange capability makes the current stock expenses continually intertwine and mirror all pertinent information present on the lookout. As per the Efficient Market Hypothesis theory, stocks are constantly traded at their reasonable esteem hence making it in intelligible for examiners to either purchase underestimated stocks or sell stocks at extended expenses. Subsequently, it makes us to be hard to beat the market. The Efficient Market Hypothesis was created by Eugene Fama 1970 [13, 14].

It incorporates three structures: weak, semi-strong and strong efficiency. Under the weak EMH, future costs can't be anticipated by essentially examining historical costs. Speculators can't yield returns over the long run by utilizing speculation methods focused around historical costs or other verifiable data. As it were, under weak efficiency technical analysis techniques won't have the capacity to reliably create abundance returns. Then again, a few types of fundamental analysis provide excess returns [15]. The semi-strong form of EMH goes above and beyond by consolidating all historical and current information into the cost. The strong form of EMH incorporates historical, public, and private data, for example, insider data, in the share cost [16].

Stock Market (SM)

A SM commonly known as Equivalent market is a public market which operates virtually to ensure smooth buying and selling of stocks at a reasonable price and are often governed by an Apex body that regulates all the transaction and sets up procedures for effective trade. Stocks are the units of ownership in a company and are also known as equities. In order to impact development these SM act as a facilitators for companies to access funds from the common public investment. The stock market can serve two very important objectives:

- Providing Capital to companies which they can utilize to expand their business flow. Lets assume a company issues 10 million stocks priced at Rs 10 each, then this has the potential to make 100 million for the company in form of public investment and this can add huge amount of working capital to the business, it can easily be used to fund operations and affect growth.
- Giving Investors an opportunity to trade public stocks and share company profitability, making this trade a win win for both Investors and companies.

Although stock trading is an old age concept, the modern day stock trading began to spread its roots with the formation of East India company from London. A stock bazaar or a stock market or even equivalent market can be defined as the assemblage of those getting goods or service and those rendering this or making this possible. (Or a loose connection of transactions which are only economic and there is no physical institution or even digital entrance of goods and services); some of these could be safely written on goods exchange and still to those not sold in public. Supplies could be classified in different ways [17]. One major method is by the residential country approach in which the case study is Nestle and another one called Novartis from Switzerland, and thus they could be regarded as a portion of Swiss supply market, though this supply could also be exchanged with neighboring or even countries from other regions.

IV. CONCLUSION

Another issue of improving accuracy is also addressed through the ensemble method. Sentiment Analysis is an evolving research field in the area of data mining text mining and web mining. With the advent of social networking, the proliferation of online surveys can be leveraged to infer better opinions of people about products and services. This paper aimed to contribute to the research on stock/Forex market prediction through an analysis of the above different deep learning prediction models. Through the review, it can be identified that there is a lack of studies on the combination of multiple deep learning methods, especially with respect to other deep learning methods. The hybrid networks are showing promising signs for future research. In the future, we will design a specific hybrid model based on the above analysis, incorporating the latest technologies such as advanced genetic algorithms and self-attention neural networks to predict the stock/Forex market.

REFERENCES

- [1] Shanshan Dong and Chang Liu, "Sentiment Classification for Financial Texts Based on Deep Learning", Hindawi, Computational Intelligence and Neuroscience, Volume 2021.
- [2] Shravan Raviraj, Manohara Pai M M. and Krithika M Pai, "Share price prediction of Indian Stock Markets using time series data A Deep Learning Approach", IEEE Mysore Sub Section International Conference (MysuruCon), IEEE 2021.
- [3] J. J. Duarte S. M. Gonzalez and J. C. Cruz "Predicting stock price falls using news data: Evidence from the brazilian market", Computational Economics vol. 57 no. 1 pp. 311-340 2021.
- [4] Jingqi Liu;Xinzhen Pei;Junyan Zou, "Analysis and Research on the Stock Volatility Factors of Chinese Listed Companies Based on the FA-ANN-MLP Model", International Conference on Computer, Blockchain and Financial Development (CBFD), IEEE 2021.

- [5] K. M. El Hindi, R. R. Aljulaidan, H. AlSalman, and H. AlSalman, "Lazy fine-tuning algorithms for na"ive Bayesian text classification," Applied Soft Computing, vol. 96, p. 106652, 2020.
- [6] G. Ding and L. Qin "Study on the prediction of stock price based on the associated network model of lstm" International Journal of Machine Learning and Cybernetics vol. 11 no. 6 pp. 1307-1317 2020.
- [7] S. T. Z. De Pauli M. Kleina and W. H. Bonat "Comparing artificial neural network architectures for brazilian stock market prediction" Annals of Data Science vol. 7 no. 4 pp. 613-628 2020.
- [8] Y.-T. Tsai, M.-C. Yang, and H.-Y. Chen, "Adversarial attack on sentiment classification," in Proceedings of the 2019 ACL Workshop BlackboxNLP: Analyzing and Interpreting Neural Networks for NLP, Florence, Italy, 2019.
- [9] Z. Zhang, Z. Wang, C. Gan, and P. Zhang, "A double auction scheme of resource allocation with social ties and sentiment classification for Device-to-Device communications," Computer Networks, vol. 155, pp. 62–71, 2019.
- [10] Zhihao PENG, "Stocks Analysis and Prediction Using Big Data Analytics", International Conference on Intelligent Transportation, Big Data & Smart City (ICITBS), IEEE 2019.
- [11] A. Site D. Birant and Z. Isik "Stock market forecasting using machine learning models" 2019 Innovations in Intelligent Systems and Applications Conference (ASYU) pp. 1-6 2019.
- [12] A. J. Balaji D. H. Ram and B. B. Nair "Applicability of deep learning models for stock price forecasting an empirical study on bankex data" Procedia computer science vol. 143 pp. 947-953 2018.
- [13] A. Dingli and K. S. Fournier "Financial time series forecasting-a machine learning approach" Machine Learning and Applications: An International Journal vol. 4 no. 1/2 pp. 3 2017.
- [14] S. Ot'alora, O. Perdomo, F. Gonz'alez, and H.M''uller, "Training deep convolutional neural networks with active learning for exudate classification in eye fundus images," Lecture Notes in Computer Science, vol. 10552, pp. 146–154, 2017.
- [15] Z. Li, "End-to-End adversarial memory network for crossdomain sentiment classification," in Proceedings of the 26th International Joint Conference on Artificial Intelligence, Melbourne, Australia, August 2017.
- [16] H. Sagha, N. Cummins, and B. Schuller, "Stacked denoising autoencoders for sentiment analysis: a review," Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, vol. 7, no. 5, p. e1212, 2017.