Stock price prediction using sentiment analysis of business domains and company-related news data

Aditya Kadam
adityahk2000@gmail.com
Pune Institute of Computer Technology, Pune, Maharashtra

Akshay Hake
akshayhake4@gmail.com
Pune Institute of Computer Technology, Pune, Maharashtra

Pravin Salunkhe
salunkhepravin41@gmail.com
Pune Institute of Computer Technology, Pune, Maharashtra

Piyush Agrawal
agrawalpiyush2505@gmail.com
Pune Institute of Computer Technology, Pune, Maharashtra

ABSTRACT

Stock price of a company is volatile in nature but when considering long term prediction of stock price, it is dependent on the company's business model and feasibility of the environment for the company to work. This data related to company and domains in which company works is available on the internet through news media but it is an arduous task for one to keep track of all this data and perform predictions on stock prices. Proposed model in this paper takes sentiment from this available data and provides prediction of stock price in the future with deliberation of the market environment.

*Keywords***:** *Sentiment Analysis, Deep Learning, Stock Market, Data Mining, Web Scraping*

1. INTRODUCTION

Stock market is an area where people can own the part of the business of a corporation and even sell those parts. Companies make part of their assets public on stock exchanges to gain liquidity from the market which provides company or owners funding to invest in their business. Along with companies cash flow, trust of investors also performs a major role in stock prices of companies. It is believed that stock market prices are erratic in nature because individuals who are investing in the stock market are risk takers and they tend to avoid any losses by any means and sometimes decisions taken by investors are irrational, taken by their instincts and decisions taken encompassing other investors. Market risks affect all the stocks present in the market and can bring some abrupt changes in the market. Most of the stock market is occupied by Multinational banks, National banks, hefty investing firms. Hence any losses in these banks or firms shakes the market. Any of the rumors about the company or stock market have a potential to down or raise the price of stocks . But avoiding this short period shakes in the stock market if trying to predict how stock will perform in the long term? The answer can only be found from how business is run by a company. Under which domain a corporation is working and does those domains have potential to rise or ability to survive,

these factors affect companies cash flow and hence it affects stock price of company. Sometimes Idiocratic risks taken by corporations which are specific to that corporation only affect the share price of the company. Some big investors' opinions about a company provokes small investors to invest or sell their equity.

Thinking about all these major factors, the stock price of a corporation is determined by investors sentiment, companies business and flow of market. This is also stated in [2] with correspondence to the Hong Kong stock exchange. Major source to provide all the information about companies business and investors sentiment is news media. To channelize this news data into useful information in prediction of stock price can be done by sentiment analysis of news data. Previous models [4], [1] only took into consideration the company related data for predicting the stocks. Current situation of the domain in which that company works was not taken into consideration by previous models because of which models could not predict some rational changes in the market. Any company's long term success is in the core of its business and business strategies. Even if a company is affected by market fall or any random information about the company after market consolidation how the company will rise up and recover its assets depends upon sentiment of investors toward the company and stability of the domain in which the company works.

Domain specific data and company specific data can be extracted from business focused newspapers such as economics times. Instead of manual reasoning of this data our model provides an automated system which recognizes sentiments. Model represented in this paper is a tailor-made model which predicts the share price of companies which belong to a particular domain. These domains are predefined in the proposed model.

Rest of the paper gives information about work done in stock price prediction using artificial intelligence and how it is done in online sources present on the internet at II. Architecture of our proposed model is explained at III.

2. RELATED WORK

2.1 Stock market prediction using news data

Yujie wang, hui liu et al [4] proposed that news and stock price reactions have an intimate connection. The news data is preprocessed and its word vector is generated which gives features of news data. This is then fused with Quantitative data model results that give the volatility of a particular stock price. This model achieved 69.51% of accuracy in predicting stock market prices in Hong Kong. This model achieved superior performance in terms of balancing the prediction effect and running time. Segmentation length for news and index sequences are fixed 3. The effect of different events on stock fluctuation may be different.

Qing Li member, yan chen, jun wang et al [2] explained how different social media, discussion boards and news data affect the stock prices, which kind models are giving better results for each of the data sets. Paper well explains how the data should be filtered for using it for stock analysis. The method explained in paper contains following steps:

- Extract the media content from CNET/google trends (news data), twitter/wikipedia (social media), Discussion board (yahoo finance).
- Perform sentimental syntax analysis on data.
- Use statistical, regression or ML based models to predict final result.

This analysis is intended to provide a framework for dividing large challenges into three main issues i.e media content, media representation, and analysis model. But only theoretical implementation is done.

Xi zhan , Siya Qu et al [1] proposed model which can predict the stock market movement and identify the importance of the information simultaneously. Paper also proposes a novel event representation learning process that can effectively capture the event information. The aim is on stock market direction prediction. All content used for sentiment is extracted from the web. Its features are extracted using a separate model and MSMI model takes input of all news data and quantitative data.

The model effectively integrates heterogeneous information, that is, the events, sentiments and historical quantitative features into a comprehensive framework, n considers the consistencies among different data source to make a better prediction. Window size is fixed. In this model Social media data did not put extra accuracy in predicting market value.

2.2 Stock market prediction using Historical data

Kyoung-jae Kim et al [7] studied applying SVM predicting the stock price index. In addition, this study examines the feasibility of applying SVM in forecasting by comparing it with back propagation neural networks and case-based reasoning.

Samuel Olusegun Ojo et al [8] studied that by making use of a stacked long short term memory network model, future stock market behavior can be predicted with better accuracy than other machine learning models working on historic events.

Other than the use of machine learning algorithms stock market price is predicted with the help of moving averages , moving averages crossover and technical indicators such as RSI , MACD , Stochastic , ROC , CCI , Williamson%R , MFI , ATR , ADX.

3.PROPOSED METHODOLOGY

3.1. Sentiment Analysis of domain wise data

- **Data collection:** Data related to a particular domain is available on online news sites such as economics times which will be extracted using web scraping tools such as scrapy and beautiful soup, selenium drive. Window size is 30 days.
- **Sentiment analysis:** Sentiment analysis of domain data will be done independent of stock prices variation. There are models trained on bigger datasets which can be used for transfer learning and train domain specific models. Results of these models and stock price variation are further used for training of ANN to combine all these results. Word vector representation of each word will be done by averaging the word embeddings created by the glove model. Diabiasing will be performed on this word embeddings.
- **Combining results of domain wise sentiment models:** A feature vector as input for a company will be created by training a simple neural network with stock price variation of a particular company.

3.2 Sentiment Analysis of Companies data

- **Data collection:** Data related to particular companies is available on online news sites such as economics times which will be extracted using web scraping tools. Window size is 30 days.
- **Sentiment analysis:** Sentiment analysis of company data will be done with respect to stock prices variation. There are some pretrained models trained on bigger datasets which can be used for transfer learning and train company specific models. Word vector representation of each word will be done by averaging the word embeddings created by glove model. Diabiasing will be performed on this word embeddings.
- **Pretrained models:** XLNET (attentive),ERNIE,T5,BPT,NABOE(attentive),Google's BERT, OpenAL's GPT-Z .
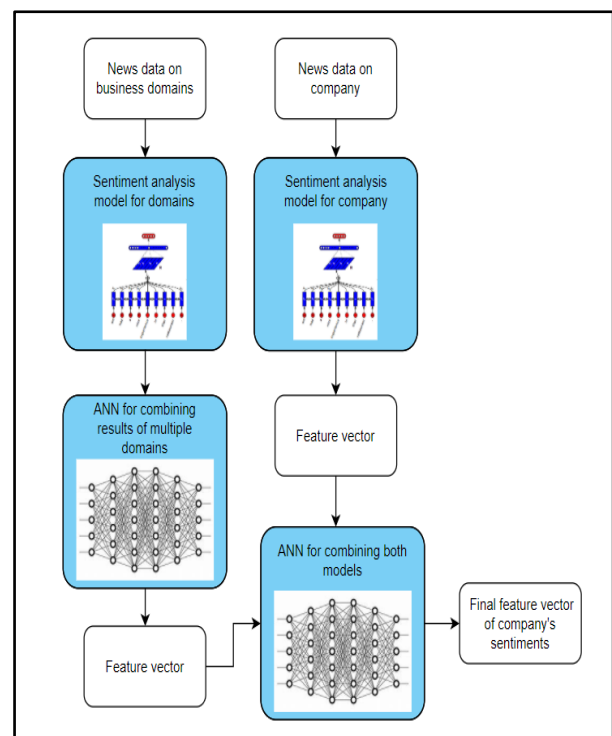- **Word to vector conversion:** Glove model, Word2vec model.

Fig. 1: Sentiment analysis model for one company which is scaled for all the companies.

Combining results of Company specific sentiment model with its domains feature vector: Final result of sentiment will be created by training a simple neural network with stock price variation of a particular company.

3.3 Historical data analysis

Historical data such as open, close, low, high price and volume will be used to predict stock price of data after a certain period of time. As this is sequential data we will be using the LSTM algorithm which we discovered from literature gives better performance.

- **Fusion of both models:** Both models will predict the result related to stock price of company after 10,20,30 days based on 30 days of previous data from the current date. For combining these models, we are using SNN which will be trained on variation of stock price.
- **Fully connected layer:** This layer will give final output as stock prices after 10,20,30 days, which will be visible to the user.
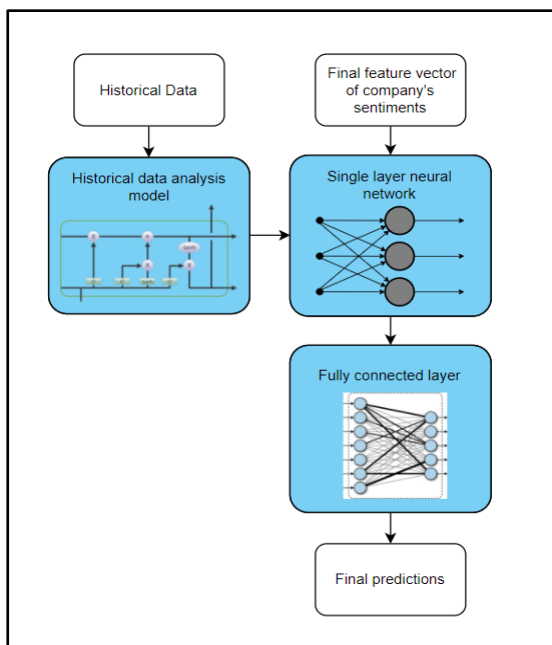

Fig. 2: Combining both sentiment and historical data model for final prediction

4. CONCLUSION

Thus, we propose a model that studies and performs sentiment analysis on the business models of the companies and then effectively integrates heterogeneous information, that is, the sentiments and historical quantitative features into a comprehensive framework, and considers the consistencies among different data sources to make a better prediction.

5. ACKNOWLEDGEMENT

6. REFERENCES

[1] Xi Zhang, Stock Market Prediction via Multi-Source Multiple Instance Learning, IEEE Access (Volume: 6), 13 September 2018

[2] Li, Q., Chen, Y., Wang, J., Chen, Y., & Chen, H. (2018). Web Media and Stock Markets: A Survey and Future Directions from a Big Data Perspective. IEEE Transactions on Knowledge and Data Engineering, 30(2)

[3] Shi, L., Teng, Z., Wang, L., Zhang, Y., & Binder, A. (2018). DeepClue: Visual Interpretation of Text-based Deep Stock Prediction. IEEE Transactions on Knowledge and Data Engineering, 1–1.

[4] YUJIE WANG, HUI LIU, Stock Volatility Prediction by Hybrid Neural Network, October 23, 2019

[5] R. Ren, D. D. Wu and T. Liu, "Forecasting Stock Market Movement Direction Using Sentiment Analysis and Support Vector Machine," in IEEE Systems Journal, vol. 13, no. 1, pp. 760-770, March 2019.

[6] D. D. Wu, L. Zheng and D. L. Olson, "A Decision Support Approach for Online Stock Forum Sentiment Analysis," in IEEE Transactions on Systems, Man, and Cybernetics: Systems, vol. 44, no. 8, pp. 1077-1087, Aug. 2014.

[7] K. jae Kim, "Financial time series forecasting using support vector machines, " Neurocomputing, vol. 55, 2003.

[8] S. O. Ojo, P. A. Owolawi, M. Mphahlele and J. A. Adisa, "Stock Market Behaviour Prediction using Stacked LSTM Networks*," 2019 International Multidisciplinary Information Technology and Engineering Conference (IMITEC), Vanderbijlpark, South Africa, 2019, pp.1-5,doi:10.1109/IMITEC45504.2019.9015840.