**Predicting Stock Market Movement from News Headlines using NLP Techniques**

**Introduction:**

Analyzing and predicting stock market movement for short term and long term is very important and difficult problem across investors. Large amount of research has been done in this space using wide range of technologies and wide range of data. But still the dynamically changing price of stocks easily getting affected by multi macro and micro economic factors makes it still a challenge to predict the movement accurately. The magnitude of change of each economic factor, and the nature of change (positive or negative change) create a strong dependency between the factors and the stock market movement.

In the recent past stock price studies are conducted through two methods - fundamental analysis by predicting stock price through analyzing underlying businesses and forecasting future business performances, and technical analysis that predicting future stock prices based on past and present price trends.

There has been a strong assumption too that Positive sentiment derived from political news will have a positive impact on the markets on the other hand negative sentiment will have negative impacts on the market. Also, publicly listed companies are required to disclose any information that might have an impact on its stock price as per the regulatory institutions, which are also known as company announcement whose topics will have impacts on its stock price.

NLP is one of the fields in Machine Learning that deals with technique that represents transformation of human-language and computer languages. Since the inception of NLP in the 1950s, previous research has been focusing on tasks such as machine translation, topic modeling, text summarization, information extraction, information retrieval, and more recently, opinion mining.

The stock price movements are not only affected by its past trends or correlation with the financial markets, but also news, investors comments, current events, or company announcements. For this research we will consider only top 25 news from reddit and study the correlation with stock market movement.

**Background:**

There has been lot of authors who have been working on more statistical methods and machine learning methods for predicting stock market prices. But in this paper we will focus on machine learning, ensemble and deep learning methods used for predicting stock market movements. Nabipour et al. used stock market data to predict future values of stock market groups using Machine Learning and Deep Learning Algorithms [1]. Infact the author predicted the future values for upto 30 days but failed to use external factors for stock market prediction which plays an important role. For stock price prediction direction, article by [Michel Ballings](https://www.sciencedirect.com/science/article/abs/pii/S0957417415003334?via%3Dihub" \l "!) et al.[2] is a notable paper and the paper benchmark ensemble methods against single classifier models and found Random Forest Outperforms all the other classifiers. But the article classifies the stock price growth within buckets of predetermined percent amount.

X.Wang et al proposes methods for sentiment analysis of short texts[4]. There has been lot of research using sentiment analysis of news articles and its correlation with stock prices prediction and almost all of them are using word vectors to compute the sentiments. All these research was based on the assumption that when there is positive sentiment the stock price would go up else it will come down. The assumption can’t be true always. Yang Liu et al.[5] has proposed prediction of stock price movement based on top news articles during pandemic and got accuracy of 62.9% for GRU and word vectors. In our research, we are going to use sentence embedding vectors instead of word vectors and study the performance of stock price movement of DJIA(Dow Jones Industrial Average) with respect to top 25 trending news from reddit.

Results

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **ML Algorithm** | **Embedder Name** | **Accuracy** | **Precision** | **Recall** | **F1-Score** |
| Logistic Regression | Tf-IDF | 0.47 | 0.45 | 0.47 | 0.44 |
| Sentence Encoder | 0.48 | 0.44 | 0.48 | 0.4 |
| Bag of Words Vector | 0.43 | 0.42 | 0.43 | 0.42 |
| Random Forest | Tf-IDF | 0.49 | 0.48 | 0.49 | 0.46 |
| Sentence Encoder | 0.5 | 0.49 | 0.5 | 0.46 |
| Bag of Words Vector | 0.49 | 0.48 | 0.49 | 0.45 |
| Gradient Boosting Machines | Tf-IDF | 0.49 | 0.48 | 0.49 | 0.45 |
| Sentence Encoder | 0.49 | 0.49 | 0.49 | 0.47 |
| Bag of Words Vector | 0.49 | 0.48 | 0.49 | 0.46 |
| Stochastic Gradient Descent Classifier | Tf-IDF | 0.44 | 0.44 | 0.44 | 0.44 |
| Sentence Encoder | 0.49 | 0.48 | 0.49 | 0.42 |
| Bag of Words Vector | 0.49 | 0.48 | 0.49 | 0.46 |
| MLP | Tf-IDF | 0.46 | 0.45 | 0.46 | 0.45 |
| Sentence Encoder | 0.49 | 0.24 | 0.49 | 0.32 |
| Bag of Words Vector | 0.49 | 0.41 | 0.49 | 0.36 |
| LSTM | Tf-IDF | 0.51 | 0.26 | 0.51 | 0.34 |
| Sentence Encoder | 0.51 | 0.26 | 0.51 | 0.34 |
| Bag of Words Vector | 0.51 | 0.26 | 0.51 | 0.34 |
| CNN | Tf-IDF | 0.51 | 0.26 | 0.51 | 0.34 |
| Sentence Encoder | 0.51 | 0.26 | 0.51 | 0.34 |
| Bag of Words Vector | 0.51 | 0.26 | 0.51 | 0.34 |

**References**

1)M. Nabipour, P. Nayyeri, H. Jabani, and A. Mosavi, “Deep learning for stock market prediction,” arXiv preprint arXiv:2004.01497, 2020.

2) Ballings, D. Van den Poel, N. Hespeels, and R. Gryp, “Evaluating multiple classifiers for stock price direction prediction,” Expert Syst. Appl., vol. 42, no. 20, p. 7046–7056, Nov. 2015. [Online]. Available: <https://doi.org/10.1016/j.eswa.2015.05.013>

3) Kalyani Joshi , Prof. Bharathi H. N. , Prof. Jyothi Rao, “STOCK TREND PREDICTION USING NEWS SENTIMENT ANALYSIS”, Department of Computer Engineering, KJSCE, Mumbai <https://arxiv.org/ftp/arxiv/papers/1607/1607.01958.pdf>

[4] X. Wang, W. Jiang, and Z. Luo, “Combination of convolutional and recurrent neural network for sentiment analysis of short texts,” in Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers. Osaka, Japan: The COLING 2016 Organizing Committee, Dec. 2016, pp. 2428–2437. [Online]. Available: <https://www.aclweb.org/anthology/C16-1229>

[5] Machine Learning for Predicting Stock Market Movement using News Headlines, Yang Liu; Jelena Trajkovic; Hen-Geul Henry Yeh, California State University, Long Beach, CA; Wenlu Zhang, [2020 IEEE Green Energy and Smart Systems Conference (IGESSC)](https://ieeexplore.ieee.org/xpl/conhome/9284968/proceeding).

[6] Universal Sentence Encoder, Daniel Cera , Yinfei Yanga , Sheng-yi Konga , Nan Huaa , Nicole Limtiacob , Rhomni St. Johna , Noah Constanta , Mario Guajardo-Cespedes ´ a , Steve Yuanc , Chris Tara , Yun-Hsuan Sunga , Brian Stropea , Ray Kurzweila, Google Research.

[7] Nti, I., Adekoya, A., & Weyori, B. (2019). A systematic review of fundamental and technical analysis of stock market predictions. Artificial Intelligence Review, 53(4), 3007-3057. doi: 10.1007/s10462-019-09754- z

[8] Picasso, A., Merello, S., Ma, Y., Oneto, L., & Cambria, E. (2019). Technical analysis and sentiment embeddings for market trend prediction. Expert Systems With Applications, 135, 60-70. doi: 10.1016/j.eswa.2019.06.014

[9] Paramanik, R., & Singhal, V. (2020). Sentiment Analysis of Indian Stock Market Volatility. Procedia Computer Science, 176, 330-338. doi: 10.1016/j.procs.2020.08.035

[10] Möller, R., & Reichmann, D. (2021). ECB language and stock returns – A textual analysis of ECB press conferences. The Quarterly Review Of Economics And Finance, 80, 590-604. doi: 10.1016/j.qref.2021.04.003

[11] Suleman, M. T. (2010). Stock Market Reaction to Good and Bad Political News. SSRN Electronic Journal. <https://doi.org/10.2139/ssrn.1713804>

[12] Ratku, A., Feuerriegel, S., & Neumann, D. (2014). Analysis of How Underlying Topics in Financial News Affect Stock Prices Using Latent Dirichlet Allocation. SSRN Electronic Journal. doi: 10.2139/ssrn.2529457