



# INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

## STOCK PRICES PREDICTION USING NEWS CATEGORIZATION, SENTIMENTAL ANALYSIS AND DEEP LEARNING

Vedant D Shirodkar <sup>[1]</sup>

Department of Information Technology in Engineering

Padmabhushan Vasantdada Patil  
Pratishthan's College of Engineering, Mumbai, India

Nasir J Kadri <sup>[2]</sup>

Department of Information Technology in Engineering

Padmabhushan Vasantdada Patil  
Pratishthan's College of Engineering, Mumbai, India

Vikas V Yadav <sup>[3]</sup>

Department of Information Technology in Engineering

Padmabhushan Vasantdada Patil  
Pratishthan's College of Engineering, Mumbai, India

**Abstract:** In this paper, we have illustrated an idea for predicting stock prices using Sentimental Analysis and Deep Learning. Showcased method mainly uses Natural Language Processing (NLP) & Information Retrieval (IR) as its core components coupled with Artificial Neural Network (ANN). The predictions are based upon two things which are historical stock data and financial news that we come across daily. Financial news includes news articles and datasets which are available on Yahoo Finance and could be obtained with the help of API(s). The Paper illustrates prediction of stock prices based on individual sentimental ratings of stocks extracted by applying sentimental analysis on news from various data sources. The sentimental analysis is specifically on the news that belongs to that companies' stocks rather than determining a general stock market sentiment. Moreover, the categorization of stocks based on industry helps in determining the organizations getting affected by news articles if a particular news is based on overall specific areas of industry and not on companies, in general; this is achieved by feature extraction and categorizing them based on the industry they belong to. Further, the Industry reveals the companies that comes under that industry thus, predicting the future prices of stocks.

**Keywords-** *Stock Market Prediction, Sentimental Analysis, Natural Language Processing, Data Mining, Feature Extraction.*

### I - INTRODUCTION

Starting from the digitalization of data there is been continuous improvement in data science and machine learning fields as massive amount of data was readily available to draw conclusion, further as the machines became computationally strong all the mathematical theories was practically implemented to draw information from the data. Several areas were embracing the new era of Machine Learning and Stock Market is one of the areas where Machine Learning and AI has achieved unremarkable growth. Stock Market Prediction aims to predicting the future value of stocks of companies or another financial instrument. Precise prediction of future prices of stock could produce significant profits; while inaccuracy in prediction may lead to huge monetary losses. Over 50% of US population invests in stocks while this percentage is less than 2% in India and nearly 10% of the overall world's population. Most people think stock market as a gamble due to its highly volatile nature but continuous research is been carried out to predict the stock prices more accurately.

Here, in this paper we will be showcasing a way to predict stock prices using Sentimental Analysis and ANN which are subsets of artificial Intelligence. Daily, we come across several financial articles which directly impact on future stock prices. By coming across such articles, an investor or a person involved in stock market can research a bit and can easily build conclusions and predict the stocks movement whether prices will be climbing northwards or southwards; but for a non-technical person it may be difficult to predict. Our paper focuses on predicting stocks using such articles.

## II – RELATED WORK

Several research intellectuals have ventured to envision the prices of stock market. They have built and carried out several techniques in the field of economics and data science which contribute towards the upcoming technologies to predict stock prices more accurately and make extraordinary fortune out of it. Hundreds of papers are been designed which provides different ways to predict stocks but majority of them focuses on stock prediction on Historical data only. Although they provide decent accuracy but still improvements could be done with the use of data that exists around us in the form of news articles.

A survey on stock market prediction was carried out by Mohit and Ritika [1] wherein their paper suggested that Sentimental Analysis helps in better forecast of stocks prices and pattern of the market. The same paper also suggested that Sentimental Analysis when combined with data mining techniques helps in finding an unknown pattern from the data collected over news articles and which have the ability to predict the stock probably. It utilizes updated news to measure its effect on the stock in real time. In the paper The Effects of Twitter Sentiment on Stock Price Returns [2] by Darko, Gabriele, Guido, Miha and Igor the authors proved that there exists a correlation between the stock prices and the stocks news data on Twitter. Ishita Parmar's [3] model included Regression and LSTM and was quite successful but they didn't account the news articles or any twitter data which also acts as an important factor. Z Lei and W Lin [4] suggested prediction of prices using outlier method, which draws better insights on long term prices of stocks. The prediction of Chinese stock market with the help of public moods is extracted from micro blog feeds [5] by D. Yan, G. Zhou, X. Zhao, Y. Tian and F. Yang showed positive correlation between sentiment of articles published by people. D Shah, H Isah and F Zulkermine[6] predicted the effect of news sentiments from news with the help of n-gram model. K. V. Sujatha and S. M. Sundaram [7] suggested insightful techniques on handling abnormal situations which commonly arises during the working of the system and cause disruptions or lead to inaccurate predictions. Shyi-Ming Chen and Yuan-Kai Ko [8], represented prices prediction system in the form of fuzzy interpolative reasoning method via cutting and transformation techniques for sparse fuzzy rule-based systems. This system model required through writing of each logic.

## III - METHODOLOGY

The existing systems are highly focussed on few parameters of historical data (previously available

company's sale data) but there exist other significant factors of stocks which also plays crucial role in deciding the future prices of stocks. Predicting stocks prices for short term may be easy but for long term stock prediction it requires processing of not only the historical data but it also needs to pay attention to the important indicators such as current scenario of the market (trends in the market) and the products served to satisfy this current need in the market will definitely be on the profitable site of the scale. This kind of data is available in abundance and should be used more wisely to predict the stock prices more thoroughly. There is abundant of data available on Twitter and other social medias but we have avoided that as they constitute of fake data as well as irrelevant data.

We gathered news data from 3-4 financial news providing sources and applied. This data is then pre-processed by applying the NLP techniques onto it. i.e., removing punctuations, stop words and tokenization and then Naïve based Algorithm is applied to the corpus to find the term frequency and bag of words representing positive, negative and neutral sentiment score along with it. Here we used Multi-Hot Coding to separate to categorize the news in three different classes. Based on the company the news belongs to, the data is segregated into company it belongs to. The news which doesn't includes any company name(s) but includes a specific industry like the Smartphone Industry, Automobile Industry, Steel Industry, etc. are then formed a separate table and as these news doesn't belong to any one particular company but a group of companies; so, the sentiment score associated with that news represents a group of companies. We then combine add the positive and negative score of the news and calculate the overall sentiment associated with the news. This sentiment score is then normalized to range it between -1 to 1, with -1 being the maximum possible negative sentimental score and +1 being the maximum positive sentimental score. This score is stored for further usage.

On the other side using the historical data of stocks we feed to deep learning model. Here we have used Long Short Term Memory (LSTM) model, which is a type of Recurrent Neural Network (RNN) as this model overcomes the Vanishing Gradient Problem and suits best for time series data as it has a loop over itself which acts as memory to store previous data. Using Keras library of TensorFlow we calculate the estimated stock prices values. Now, this estimated price of stock is multiplied with the with the square root of overall calculated sentimental score of the financial news. This is the predicted stock price value for the particular stock. If instead a news belongs to an industry, then the sentimental score is multiplied with all those company's estimated stock price that we got from LSTM model.

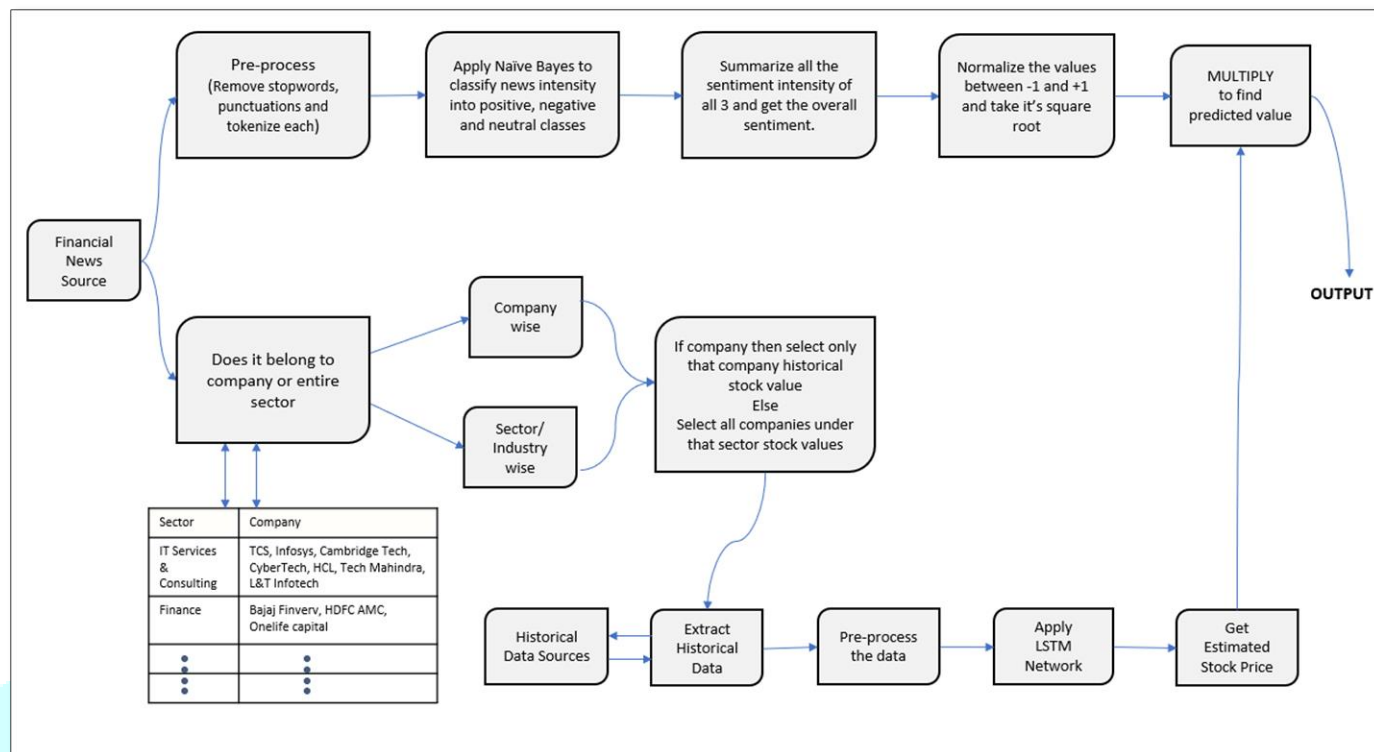


Fig: Conceptual View of Methodology

#### IV - CONCLUSION AND FUTURE SCOPE

This paper was an attempt to predict stock prices of a company with higher accuracy and greater reliability. Predicting stock prices is really challenging as it has lots of parameters involved within itself. While, historical data alone is good for prediction but not sufficient enough to get highly accurate results but historical data along with some other stock market indicators may prove to help the model predict more

accurately. This paper took historical data and sentiments of financial data into account and predicted the prices and sentiments based on particular company or domain.

Still there are scopes for improvement in predictions. We may use region specific data of news and determine the overall cause and effect of their sentiments combined. To the existing model one may even try adding more data regarding the shares a



company acquires in some other companies. Depicting this into a directed graphical structure would help us determining the rise or fall of a company's share based on others.

#### V – REFERENCES

- [1] M Iyer, R Mehra, "A Survey on Stock Market Prediction," 5th IEEE International Conference on Parallel, Distributed and Grid Computing (PDGC-2018), 20-22 Dec, 2018, Solan, India.
- [2] G. Ranco, D. Aleksovski, G. Caldarelli, M. Grčar and I. Mozetič, "The Effects of Twitter Sentiment on Stock Price Returns", *10.1371/journal.pone.0138441*.
- [3] I. Parmar, N. Agarwal, S. Saxena, R. Arora, S. Gupta, H. Dhiman, L. Chouhan, "Stock Market Prediction Using Machine Learning", 2018 First International Conference on Secure Cyber Computing and Communication (ICSCCC).
- [4] Z. Lei & W. Lin, "Price Trend Prediction of Stock Market Using Outlier Data Mining Algorithm", 2015 IEEE Fifth International Conference on Big Data and Cloud Computing.
- [5] D. Yan, G. Zhou, X. Zhao, Y. Tian and F. Yang, "Predicting Stock using Micro blog Moods", Journal on China Communications, Vol. 13, No. 10, pp.244-257, 2016.A.
- [6] D Shah, H Isah, F Zulkernine, "Predicting the Effects of News Sentiments on the Stock Market", 2018 IEEE International Conference on Big Data (Big Data).
- [7] KV Sujatha, S Meenakshi Sundaram, "Regression, Theil's and MLP Forecasting Models of Stock Index", International Journal of Computer Engineering and Technology.
- [8] Shyi-Ming Chen and Yu-Chuan Chang, "Multi-Variable Fuzzy Forecasting Based On Fuzzy Clustering and Fuzzy Rule Interpolation Techniques", Information Sciences, Vol.180, No.24, pp. 4772-4783.

