

Prediction of Stock Market Trends Using News Headlines

Mr. Saransh Surana¹, Ms. Sahithi Ankareddi²

^{1,2}Dept of Electronics and Communication Engineering

^{1,2}Gayatri Vidya Parishad College for Degree and PG Courses(A), Visakhapatnam, India

Abstract- *In today's world a single source of income is not sufficient, so people opt to various methods of financing themselves, and one such source is Stock market. However, the stock market today possesses a lot of menace. Therefore, already a lot of works have to predict the stock market by building various models using machine learning. Today we will predict the stock market sentimentation via news headlines using machine learning algorithms like Random Forest Classifier, Naïve Bayes. By analyzing stock market trends, and therefore predicting a market sector that is currently growing will continue to grow in the future, or whether a market trend will fall steep and trend in another, and so on. The use of technical analysis in stock market prediction is growing by the day, as more people strive to gain technical analysis in order to make big money.*

Keywords- Machine Learning, Technical Analysis, Random Forest Classifier, Naïve Bayes.

I. INTRODUCTION

The stock market nowadays has become an obsession among people. This is a place where anyone can join and earn big bucks. There are many questions which arises in one's mind such as when should we buy and sell stock, what are the factors affecting the stock market, when does the price of a particular stock reduce, how will the market change after a certain event. The major factor that changes the price of the stock can be news. After a certain event, the prices can fall or rise. The best scenario is to buy a stock at a low price and sell it a high price which can result in profit. This might look easy, but one has to be sure of the current position of the stock market, which can be derailed by technical analysis.

These days' technical analysis is given a lot of importance to understand the stock market. As the days are turning out to be more eventful, the rise and fall have also become similar. These analyses can be given more importance and when combined with knowledge, it can turn out to be a right decision.

With the booming market and a lot of history of data, one can predict the result of enhancement in stock market

using machine learning algorithms. There are many machine learning algorithms out of which we consider Naïve Bayes, Random Forest Classifier using TF-IDF Classifier.

The rest of the paper consists of complete work as section II summarizes the related work. Section III describes the approach and implementation of various machine learning algorithms. Section IV shows the results and section V gives the conclusion and prospect for the future work.

II. RELATED WORK

There a lot of surveys and works done on stock market prediction following news articles. There have been the various algorithms used like Logistic Regression, K-Means Clustering, KNN-Classifier etc. Some recent works that have the best accuracy and have achieved high accuracy are mentioned

S.Coleman [1] presents us with an approach to use a news article for anticipating the price trend rather than predicting a value. He has used his approach of news labelling. The algorithm they have used is Support Vector Machines (SVM). When the results were compared they scored an accuracy of 83% whereas the random accuracy was 50%

Gaurav Jariwala [2] used the method of naïve bayes, K-Means Clustering and SVM. For the 3 different methods used they have been able to produce an accuracy of highest order of 83%.

III. METHODS

We are using sentiment analysis methods on news headlines for the stock whose next day's movement is to be predicted. The news headlines are extracted from NDTV.in and TIMES NOW. Extracted using selenium. The data is clean and pipelined. After that, each headline is given an indication as of in positive, negative and neutral. Scikit-learn library is used for data analysis and Natural Language Toolkit(NLTK) library is utilized for language processing.

A. Random Forest Classifier

A random forest is a meta estimator that fits a number of decision tree classifiers on various sub-samples of the data set and uses averaging to improve the predictive accuracy and control over-fitting.

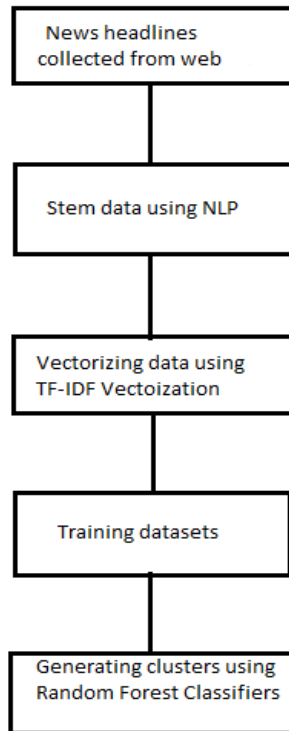


Fig 1:Flowchart for classification using Random Forest Classifier

Word Tokenization helps in splitting the sample sentences into words. Here we use TF-IDF Vector to similarize words and to avoid duplicates. It also helps in count of words, all at once. Finally, headlines are assigned to various clusters based on calculations and data passed.

B. Naïve Bayes

Naïve Bayes(NB) classifier is based on Bayes Theorem. They are among the simplest Bayesian network models, but coupled with kernel destiny estimation, they can achieve high accuracy levels. Naïve Bayes learners and classifiers can be extremely fast compared to other methods.

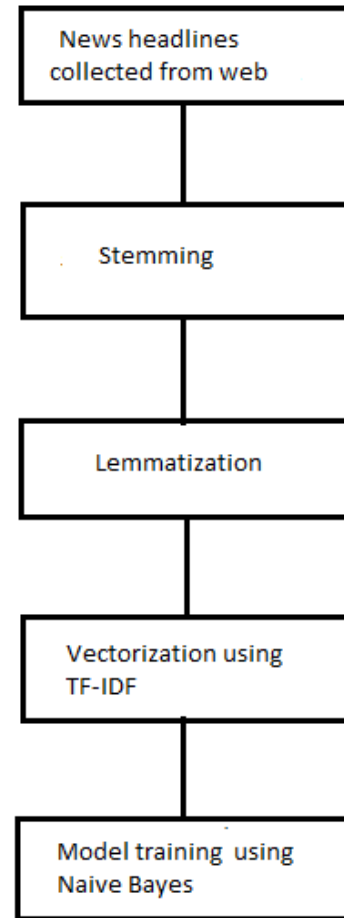


Fig 2: Flowchart for classification using Naïve Bayes.

Here, we first pipeline the data, then use stemming and lemmatization to convert words into base form and analyze similar set of words in a group. Then we use TF-IDF Vectorization for the similar process used in Random Forest. Then it is trained for a dataset to predict the outcome.

III. RESULTS

In this project, we have trained Random Forest Classifier, and Naïve Bayes. We have used various headlines from web on various topics. The dataset consists of various recessive headlines and plots indicated with labels.

Accuracy = 0.8518518518518519				
	precision	recall	f1-score	support
0	1.00	0.70	0.82	186
1	0.77	1.00	0.87	192
accuracy			0.85	378
macro avg	0.89	0.85	0.85	378
weighted avg	0.89	0.85	0.85	378

Table 1: Classification of accuracy for Random Forest Classifier

0.8465608465608465				
	precision	recall	f1-score	support
0	0.93	0.74	0.83	186
1	0.79	0.95	0.86	192
accuracy			0.85	378
macro avg	0.86	0.84	0.84	378
weighted avg	0.86	0.85	0.84	378

Table 2: Classification of accuracy for Naïve Bayes

Here we have checked accuracy for both Naïve Bayes and Random Forest Classifier. The accuracy for Random Forest Classifier for 4021 headlines is about 85 percent. Whereas for Naïve Bayes the accuracy for 4321 headlines is about 84 percent.

V. CONCLUSION

Unlike the conventional stock market prediction system, the approach of using news headlines sentiment through market participation and the average rise or fall of stock market. The algorithms of Random Forest Classifier and Naïve Bayes are easy enough to analysis news headlines for the stock market. Both these algorithms are equal to each other with Random Forest Classifier little better. The final accuracy of Random Forest Classifier is 85.15 percent, whereas the final accuracy of Naïve Bayes is 84.6 percent. The better accuracy rates will define the better results.

VI. FUTUREWORK

For future work we can try to further improve the accuracy, we can also try to implement the unsupervised learning. The accuracy can be further derailed for a longer timeline. The other ,way to improve accuracy can be taken by the complete news instead of headlines. Also, by reading and confirming from two or three sources, instead of confirming from just one.

REFERENCES

- [1] Y. Shynkevich, T. M. McGinnity, S. Coleman and A. Belatreche, "Predicting Stock Price Movements Based on Different Categories of News Articles," 2015 IEEE Symposium Series on Computational Intelligence, Cape Town, 2015, pp. 703-710
- [2] Gaurav Jariwal, Harshit Agarwal, and VrajJadhav, "Sentimental Analysis of News Headlines for Stock Market" 2020 IEEE International Conference for Innovation n Technology, Bengaluru India.
- [3] Kameshwari S, Kaniskaa. S, Kaushika. S, and R. Anuradhaa "Stock Trend Prediction Using news headlines" 2021 IEEE India Council International Subsections Conference (INDISCON), Nagpur, India.
- [4] Laree K, Shaiqu S. "Sentimental Analysis of News Headlines for Stock Market" 2019 IEEE Xplore, Kualalampur, Malaysia.
- [5] H. Zhang " the Optimality of Naïve Bayes" , New Brunswick, Canada: University of New Brumsick, 2004
- [6] YuqiaoGuo " Stock Price Prediction Based on LSTM Neural Network: the Effectiveness of News Sentiment Analysis" 2020 2nd International Conference on Economic Management and Model Engineering (ICEMME), Chongqing, China.
- [7] SashankSridar, SowmyaSanagavarapu, "Analysis of the Effect of News Sentiment on Stock Market Prices through Event Embedding"2021 16th Conference on Computer Science and Intelligence Systems (FedCSIS), Sofia, Blugaria.
- [8] PisutOncharoen, peerapunVateekul," Deep Learning for Stock Market Prediction Using Event Embedding and Technical Indicators" 2018 5th International Conference on Advanced Informatics: Concept Theory and Applications (ICAICTA), Krabi, Thailand,18282554
- [9] AndryAlamsyah, SiskaPrasetyaAyu, Brady Rikamahu, "Exploring Relationship between Headline News Sentiment and Stock Return" 2019 7th International Conference on Information and Communication Technology (ICoICT), KualaLampur, Malaysia,18995403.
- [10] S.Bird, E.Klein,"Natural Language Processing with Python", O'Reilly Media Inc, 2009