



e-ISSN: 2319-8753 | p-ISSN: 2347-6710

IJIRSET

International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 12, Issue 3, March 2023

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 8.423

9940 572 462

6381 907 438

ijirset@gmail.com

www.ijirset.com

Stock Market Prediction Using Sentiment Analysis of Twitter Data

Ketan Patil, Ritesh Zurange, Swarali Kshirsagar, Nihar Ghate, Prof.A.G. Ghule

Department of Computer Engineering, KJ College of Engineering and Management Research, Pune, India

ABSTRACT: Stock exchange is a subject that is highly affected by economic, social, and political factors. There are several factors e.g. external factors or internal factors which can affect and move the stock market. Stock prices rise and fall every second due to variations in supply and Demand. Various Data mining techniques are frequently involved to solve this problem. But technique using machine learning will give more accurate, precise and simple way to solve such issues related to stock and market prices.

Stock Price Prediction Using Twitter Sentiment Analysis is a method for predicting stock prices using the tweets. The changes in stock prices of a company, the rises and falls, are correlated with the public opinions being expressed in tweets about that company. Understanding author's opinion from a piece of text is the objective of sentiment analysis. Positive news and tweets in social media about a company would definitely encourage people to invest in the stocks of that company and as a result the stock price of that company would increase.

In this project a method for predicting stock prices is developed using Twitter tweets about various companies. Sentiment analysis of the collected tweets is used for prediction model for finding and analyzing correlation between contents of tweets and stock prices and then predictions for future prices will be developed by using machine learning.

I. INTRODUCTION

1.1 Need of the system

We have chosen to work with twitter since we feel it is a better approximation of public sentiment as opposed to conventional internet articles and web blogs. The reason is that the amount of relevant data is much larger for twitter, as compared to traditional blogging sites. Moreover the response on twitter is more prompt and also more general (since the number of users who tweet is substantially more than those who write web blogs on a daily basis). Sentiment analysis of public is highly critical in macro-scale socioeconomic phenomena like predicting the stock market rate of a particular firm. This could be done by analyzing overall public sentiment towards that firm with respect to time and using economics tools for finding the correlation between public sentiment and the firm's stock market value. Firms can also estimate how well their product is responding in the market, which areas of the market is it having a favorable response and in which a negative response (since twitter allows us to download stream of geo-tagged tweets for particular locations. If firms can get this information they can analyze the reasons behind geographically differentiated response, and so they can market their product in a more optimized manner by looking for appropriate solutions like creating suitable market segments. Predicting the results of popular political elections and polls is also an emerging application to sentiment analysis

This project proposes text mining of financial news and public sentiments and opinions from social media such as twitter. The combination of market data and news features together helps improve the accuracy of predictions. Regardless, already existing systems have failed to effectively integrate news features together with market data. With this, the results obtained are converted into numeric forms that feed the prediction process.

1.2 Sentiment analysis of twitter data

Sentiment Analysis is a technique used in text mining. It may, therefore, be described as a text mining technique for analyzing the underlying sentiment of a text message, i.e., a tweet. Twitter sentiment or opinion expressed through it may be positive, negative or neutral. However, no algorithm can give you 100% accuracy or prediction on sentiment analysis. As a part of Natural Language Processing, algorithms like SVM, Naive Bayes is used in predicting the polarity of the sentence. sentiment analysis of Twitter data may also depend upon sentence level and document level.

II. LITERATURE SURVEY

2.1 Stock Market Forecasting Using Machine Learning Algorithms

Author S. Shen, H. Jiang, and T. Zhang have tried predicting the movement of stock market using machine learning algorithms such as support vector machine (SVM) and reinforcement learning.

In this paper, they have proposed a new prediction algorithm that exploits the temporal correlation among global stock markets and various financial products to predict the next-day stock trend with the aid of SVM. The stocks were predicted using SVM as well as MART(a decision tree based boosting algorithm) we note that SVM algorithm is very sensitive to the size of training data. When the size of training set is not large enough, hyper-plane found by SVM algorithm might not be able to split the data properly. Thus, feature selection is essential when using SVM. In contrast, Multiple Additive Regression Trees (MART) algorithm requires less training data and prefers high dimensional feature set.

Numerical results indicate a prediction accuracy of 74.4% in NASDAQ, 76% in S&P500 and 77.6% in DJIA. The same algorithm is also applied with different regression algorithms to trace the actual increment in the markets. Finally, a simple trading model is established to study the performance of the proposed prediction algorithm against other benchmarks.

2.4 Sentiment Analysis of Twitter Data for Predicting Stock Market Movements

Author VenkataSasankPagolu, Kamal Nayan Reddy Challa, Ganapati Panda and BabitaMajhi have co-written this thesis whose work is to observe how well the changes in stock prices of a company, the rises and falls, are correlated with the public opinions being expressed in tweets about that company.

The tweets were collected using Twitter API and filtered using keywords like \$ MSFT, # Microsoft, #Windows etc. Stock prices data collected is not complete understandably because of weekends and public holidays when the stock market does not function. The missing data is approximated using a simple technique by Goel. Data is pre-processed before inputting it into the machine learning algorithm. For pre-processing of tweets we employed three stages of filtering: Tokenization, Stopwords removal and regex matching for removing special characters.

N-gram representation and **Word2vec representation** were used for sentiment analysis. Both methods were fed to the classifier and trained using **Random Forest Algorithm**. Out of the two, model trained with word2vec representation is picked because of its sustainability of meaning and promising performance over large datasets.

In conclusion, the paper shows that a strong correlation exists between rise/fall in stock prices of a company to the public opinions or emotions about that company expressed on twitter through tweets. The classifier with features as Word2vec representations of human annotated tweets trained on Random Forest algorithm with a split percentage of 90 for training the model and remaining for testing the model showed an accuracy of 70.2%. With N-gram representations, the classifier model with same algorithm and with same dataset showed an accuracy of 70.5%. Though the results are very close, model trained with word2vec representations is picked to classify the nonhuman annotated tweetsbecause of its promising accuracy for large datasets and the sustainability in word meaning.

III. PROPOSED SYSTEM

3.1 Objective

The aim of this project is to

1. **Predict the movement of stock market on the basis of public sentiments expressed on Twitter-**
The project observes how well the stock prices of a company changes, and how the rise and fall are correlated with the public opinions being expressed in tweets about that company.
2. **Test and find a most efficient and accurate classifier for carrying out the above task-**
We will be testing 6 different ML classifier to find out the best alternative for the job. All 6 will be tested with the data and their outputs will be compared with the actual real world events to determine the most optimal solution.

3.2 Methodology and Functionalities

To achieve this the system will follow below steps.

- Data Collection
- Preprocessing
- Sentiment Classification

- Sentiment Analysis
- Classifier Selection

3.2.1 Data Collection –

- Stock Prices obtained using Yahoo! Finance API. This dataset consists of the Open, Close, High and Low values for each day.
- We obtained also a collection of tweets using Twitter’s Search API. For each tweet these records provide a tweet id, the timestamp and tweet text, which is by design limited to 140 characters and needs to be filtered from noise.

3.2.2 Preprocessing –

Data is not in the ideal state when extracted from Twitter. The structure of the data has to be converted and changed to the acceptable form which can be fed into the classifier for further processing. The data goes through the following steps:

- Cleaning
- Removal of unnecessary features
- Aligning dates throughout the timeseries
- Replacing missing data

3.4 Basic architecture of system components

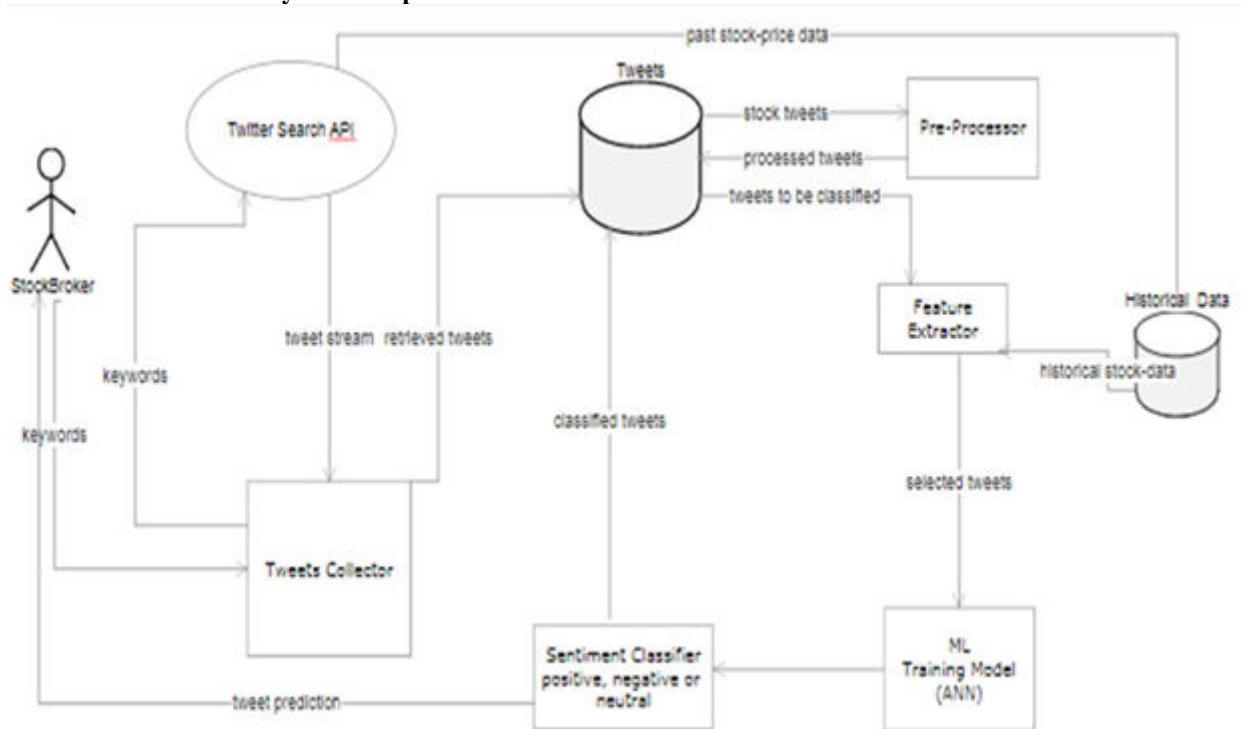


Fig. 3.1 Basic Architecture of Proposed System

System architecture outlines the structure of a system and its behavioral components. The proposed stock prediction system comprises of the classifier, the machine learning predictor, pre-processor component and historical stock prices data. The only users of the system will be the stockbrokers for the different companies. The stock price prediction begins with the user or stockbroker entering keywords to retrieve tweets related to stock prices of a particular company. Once the tweets have been obtained, the twitter search API matches the keywords and sends them to the pre-processor for cleanup. The processed tweets are transformed into a document-term matrix that is suitable for

machine learning algorithms and the tweet classifier. Based on the Machine Learning algorithm employed in the study; the tweets are classified as either positive (1), negative (-1) or neutral (0). Based on the classification, the tweets are matched against historical stock prices in the database hence the prediction of the stock price for the period under review. The result is then presented to the stockbroker for analysis or decision making.

3.5 Work flow of proposed system

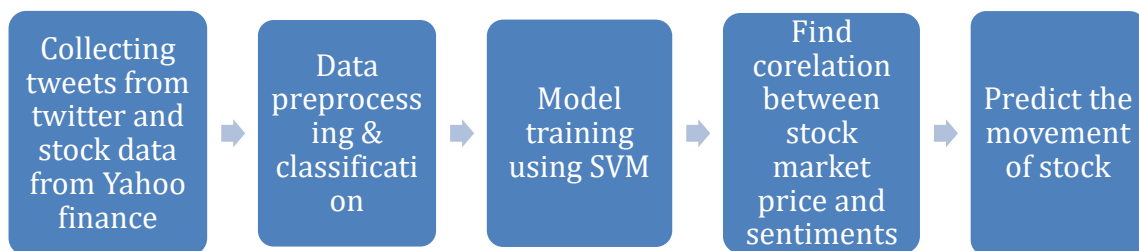


Fig. 3.2 Work flow of Proposed System

IV. CONCLUSION

Thus the “Human Computer Interaction using Artificial Intelligence Language Modelling” has been designed and tested successfully. It has been developed by Python Speech Recognition Using Google Speech API Model. Presence of every module has been reasoned out and placed carefully, thus contributing to the best working of the unit.

The above project can be implemented in Linux system. This project implementation can reduce the complexity of executing commands in Linux system. It is also helpful for computer illiterate people or to those who are beginner in Linux system. This system will also be helpful to execute daily tasks by voice module (just by giving instructions verbally).

V. FUTURE SCOPE

We can make the admin side application more interactive by adding moreover tasks for analysis such as providing C or Java compilers, providing syntax in help window to execute other tasks. We can also track the live location of the bus by using the same hardware. We can also show the data graphically to administrator to understand better and clear.

Instead of taking only basic tasks as voice input we can modify it to other major tasks for more convenience by simple modifications in speech recognition and making necessary changes in codes.

In future we can implement the project in other operating systems.

REFERENCES

- [1] HyeJoeng Park and SedaMcKillingan, “Human-Computer Interaction and Design Thinking Process Integration” (2018)
- [2] Juan David Hincapie-Ramos, “Literature Reviews for HCI” (2018)
- [3] JesperKjeldskov and JeniPaay, “Mobile HCI research methods” (2012)
- [4] The Encyclopedia of Human Computer Interaction, Interaction Design Foundation [6] Interactions, Niamh McNamara, JurekKirakowski
- [5] R. Mahapatra, N. Sharma, A. Trivedi, and C. Arnan, "Adding interactive interlace to e-government systems using aiml based chatterbots," in Software Engineering (CONSEG), 2012 CSI Sixth International Conference on. IEEE, 2012, pp. 1-6.



INNO SPACE
SJIF Scientific Journal Impact Factor
Impact Factor: 8.118



ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 **9940 572 462**  **6381 907 438**  **ijirset@gmail.com**



www.ijirset.com

Scan to save the contact details