STOCK PREDICTION AND ANALYSIS USING R LANGUAGE AND ARIMA

Dheeraj.R[1], Hariprasath.R[2], Akshay Kannan.V[3], Nishanth Kumar.S[4]
Ms.V.Sahaya Sakila[5]

[1] *Under graduate, computer science & engineering, SRM institute of science and technology, Tamil Nadu, INDIA*
[2] *Under graduate, computer science & engineering, SRM institute of science and technology, Tamil Nadu, INDIA*
[3] *Under graduate, computer science & engineering, SRM institute of science and technology, Tamil Nadu, INDIA*
[4] *Under graduate, computer science & engineering SRM institute of science and technology, Tamil Nadu, INDIA*
[5]*Assistant Professor, computer science & engineering SRM institute of science and technology, Tamil Nadu, INDIA*

ABSTRACT

 Abstract: The project is aimed at developing a software model that predicts the proceeding of a stock using ARIMA (Auto Regressive Integrated Moving Average), combining it with the twitter sentiment analysis of the stockholder thus allowing the inclusion of human sentiments in the prediction process. Twitter sentiment analysis is achieved through the use of R language for its efficient ability to visualise raw data in terms of stock analysis. Thus stock analysis using R-language and ARIMA is channelled by taking raw data and numbers and converting it into tangible and readable format for the increase in the probability of economic welfare of the user.

Keywords: - *ARIMA, Stock analysis, Sentiment analysis, and R language.*

1. INTRODUCTION

The Stock Exchange is where you can buy, sell, and trade stocks any business day. The stock price is based on the corporation's earnings. This project is designed to develop a stock prediction system that allows the users to target best opportunities of investment in various securities; It is aimed at developing a software model that predicts the proceeding of a stock using ARIMA (Auto Regressive Integrated Moving Average), combining it with the twitter sentiment analysis of the stockholder thus allowing the inclusion of human sentiments in the prediction process. As the progression of stocks is highly dynamic in nature it is impossible for humans to predict the proceeding of a stock, this is where the stock analysis and prediction system comes into use.

1.1 Existing system
The existing systems that are employed to predict stock market are fundamental analysis, technical analysis and data mining technologies such as decision trees and vector models. These conventional methods of analysing the stock market are time consuming and lack accuracy.

1.2 Issues in Existing System
Following are the issues:

- Human intervention causes opportunities for error.
- ARIMA is not used in stock analysis.
- R language is not used for its efficient ability to visualize raw data in terms of stock analysis.
- Human sentiment has not been included in stock analysis so far.

1.3 Proposed System
The proposed system is to build a statistics model using Big Data, Data Processing and data science techniques to predict the Stock Exchange. Auto-Regressive Integrated Moving Average (ARIMA) is combined with R language to create a prediction system based on previous stock. It reduces the error caused by human intervention in stock analysis and increases the probability of economic welfare of the user.

Following are the advantages:

- ARIMA is used to calculate the moving average of stock data as it is dynamic in nature which is complemented by ARIMA model, and as a result increases the efficiency and the accuracy of the prediction cycle.

- Human sentiment is included in this system which tells other users about what the general reaction of the population is towards a particular stockholder which enables the user to invest in a better stock to obtain maximum profits.

- R language has been used to its full potential in data mining algorithms in this prediction system which has not been done before.

- Human intervention is minimized to avoid errors in the system upon failure which can potentially decrease the accuracy and efficiency of the prediction system.

2. SYSTEM ARCHITECTURE

In this chapter we are going to discuss about the system architecture of the stock analysis and prediction system. The stock data and the twitter sentiment data is uploaded as a training set into R language and the ARIMA model is applied upon them to get the predicted outcome. This predicted outcome is taken as input for Tableau and image magick software as a combination, then the predicted stock outcomes are presented to the user in the form of graphical representation. The architecture of this system is kept as simple as possible to make it accessible to a wide range of consumers and to maintain a simple user interface.
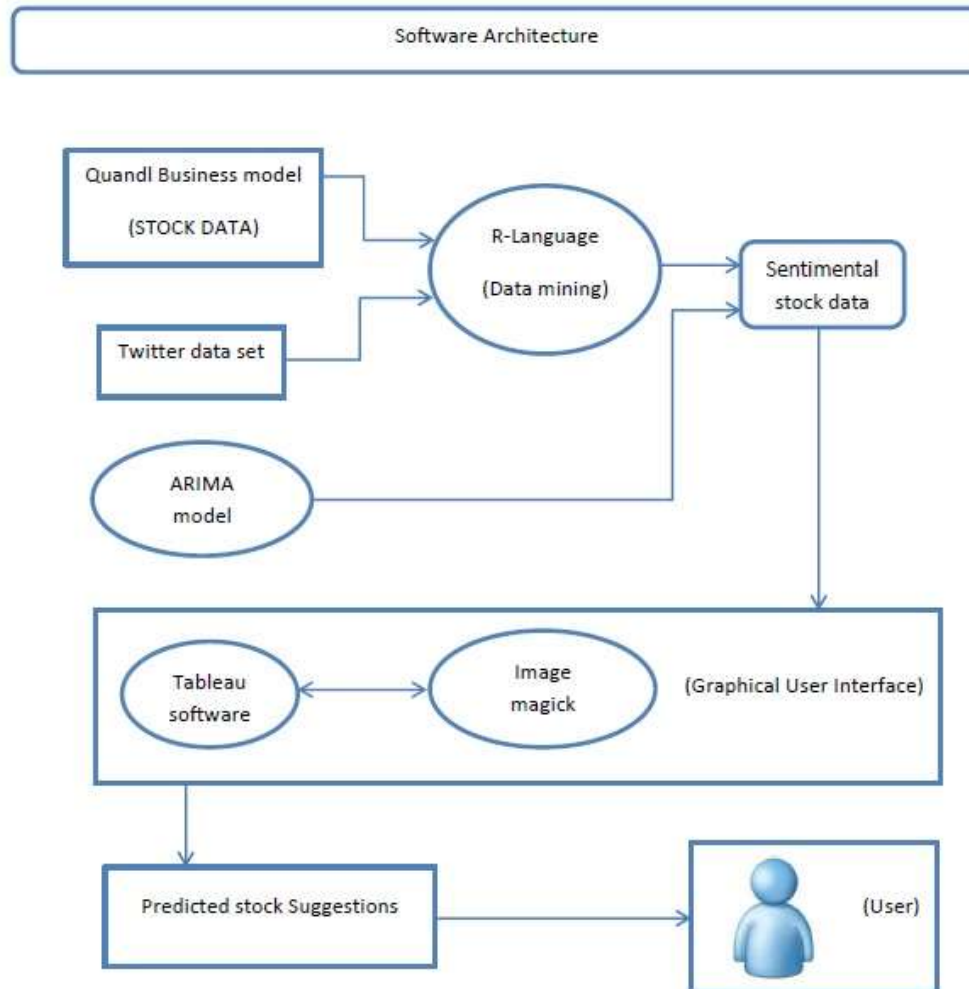
Fig -1 System Architecture Diagram

3. IMPLEMENTATION
The project named Stock analysis and prediction using ARIMA and R language has been built using R language in R studio. System has satisfied all the proposed work. This project is implemented successfully. This platform will give an excellent result in terms of user interaction accuracy of prediction of stocks in the market. These are the following modules that have been implemented in this project:

1. Data collection
2. Data aggregation
3. Analysis of data
4. Prediction of stock
5. Graph Plotting

3.1 Data collection
The Quandl package in R language is imported. Quandl unifies over 20 million financial, economic and alternative datasets from over 500 publishers on a single platform. All data is available directly in R, using the Quandl R package, from the Quandl module; a certain stock market data set is taken to be trained, this set of data is given as the input to the prediction algorithm.
3.2 Data aggregation

In this module all the variables are sorted into a super variable. The data from input stock module is gathered and expressed in a summary form, for statistical analysis.

3.3 Analysis of data

In this module the aggregated data is analyzed using different mathematical algorithms used in financial calculations. The entire algorithm's that are necessary are implemented here.

3.4 Prediction of stock

In this module the data that has been processed so far is subjected to ARIMA (Auto Regressive Integrated Moving Average) model and the predicted results are displayed in numerical format.

3.5 Graph plotting

In this module all the analyzed data is converted into graphical representations with the help of Tableau software and Image magicks platform. The plotting is based upon Bollinger's band. Bollinger Bands are volatility bands placed above and below   a moving average. Volatility is based on the standard deviation, which changes as volatility increases and decrease. The bands automatically widen when volatility increases and   narrow when volatility decreases.

4. CONCLUSIONS

Thus, hereby we conclude that the proposed system removes all the drawbacks of the existing system, decreases human intervention greatly and increases the efficiency and accuracy of the stock analysis and prediction system. R Language and ARIMA model has been implemented upon raw data and prediction is carried out. The system provides an user friendly interface to provide ease of access to a wide range of users.

5. REFERENCES

[1] S. Navale, Pawan Gabda, Nishant Dudhwala, Kunal Jadhav, "Prediction of Stock Market using Data Mining ", International Journal of Computer Applications (975–887) Volume 134 – No.12, January 2016.

[2] Varunesh Nichante, Prof. Sulabha Patil , "Analysis of stock market by using big data analytic technology", International Journal on recent and Innovation Trends in computing and communication Volume: 4 Issue:1.

[3] Aditya Bhardwaj, Yogendra Narayan, Vanraj, Pawan, Maitreyee Dutta ,"Sentiment Analysis for Indian Stock Market Prediction Using Sensex and Nifty", 4th International Conference on Eco-friendly Computing and Communication Systems.

[4] Prapanna Mondal, Labani Shit and Saptarsi Goswami, "STUDY OF EFFECTIVENESS OF TIME SERIES MODELING (ARIMA) IN FORECASTING STOCK PRICES", International Journal of Computer Science, Engineering and Applications (IJCSEA) Vol.4, No.2, April 2014.

[5] Banhi Guha, Gautam Bandyopadhyay, "Gold Price Forecasting Using ARIMA Model", Journal of Advanced Management Science Vol. 4, No. 2, March 2016.

[6] K.Arun, A.Srinagesh, M.Ramesh, "Twitter Sentiment Analysis on Demonetization tweets in India Using R language "International Journal of Computer Engineering In Research Trends Volume 4, Issue 6, June-2017.