Time Series Data Analysis for Stock Market Prediction using Data Mining Techniques with R

Mahantesh C. Angadi
M.Tech Student, Dept. of ISE,
Acharya Institute of Technology,
Bengaluru, Karnataka, India.

Amogh P. Kulkarni
Assistant Professor, Dept. of ISE,
Sai Vidya Institute of Technology,
Bengaluru, Karnataka, India.

Abstract: Nowadays, the stock market is attracting more and more people's notice with its high challenging risks and high return over. A stock exchange market depicts savings and investments that are advantageous to increase the effectiveness of the national economy. The future stock returns have some predictive relationships with the publicly available information of present and historical stock market indices. ARIMA is a statistical model which is known to be efficient for time series forecasting especially for short-term prediction. In this paper, we propose a model for forecasting the stock market trends based on the technical analysis using historical stock market data and ARIMA model. This model will automate the process of direction of future stock price indices and provides assistance for financial specialists to choose the better timing for purchasing and/or selling of stocks. The results are shown in terms of visualizations using R programming language. The obtained results reveal that the ARIMA model has a strong potential for short-term prediction of stock market trends.

Keywords: Time Series Data, Stock Market, Prediction, Analysis, Data Mining, ARIMA, R.

I. INTRODUCTION

Today we live and breathe data. Forecasting the stock exchange data is an important financial subject which involves an assumption that the fundamental information publicly available in the past has some predictive relationships to the future stock returns. Stock market forecasting contains uncovering the market trends, planning investment tactics, identifying the best time to purchase the stocks and which stocks to purchase. A stock exchange or equity business sector is a non-direct, non-parametric framework that is difficult to model with any sensible exactness. It is the mix of speculators who need to purchase or offer or hold a share at a specific time. Prediction will continue to be an exciting locale of research, making scientists in the analytics field always desiring to enhance the existing forecasting models. The motivation is that companies and individuals are empowered to make investment decisions to develop viable system about their future endeavors.

Stock trend forecasting is considered as one of the most difficult tasks to achieve in money related gauging because of the difficulty in the multifaceted world of stock market. Many of the investors in the stock market are finding a technique that could guarantee easy profiting by forecasting the stock trends and minimize the risk of investing. This motivates the researchers in the domain field to delve and develop new forecasting models. Time series data analysis techniques use verifiable information as the premise for evaluating future results. Time series data can be defined as numerical data collected in a particular sequence over a period of time at regular intervals. The time series data can include the values collected at the end of every week, month, quarter, or year etc. The intention is to find if there is any link between the data collected so far and in what way does the data changes. In order to reduce the risk of investment, exchange of securities between the seller and buyer are facilitated by the stock exchanges. A stock exchange is an organization or a place where the stock traders or investors can deal with stocks. Some of the examples for stock market organizations include NASDAQ, NYSE, BSE, NSE etc.

R is a programming language and environment for statistical processing and graphics. The R dialect is generally utilized among analysts and data excavators for statistical programming and data analysis. R language was created by Ross Ihaka and Robert Gentleman at the University of Auckland, New Zealand and is presently maintained by the R development core team. Amid the most recent decade, the energy originating from both the scholarly world and industry has lifted the R programming dialect to turn into the most essential tool for computational insights, perception and data science. Around the world, a large number of analysts and information researchers use R language to take care of their most difficult issues in the fields going from computational science to quantitative promoting.

R-Studio is a free and powerful integrated development environment for R language. R-Studio allows the user to run R scripts in a more user-friendly environment. R-Studio is available in two editions: R-Studio Desktop, where the program is run locally as a regular desktop application; and R-Studio Server, which allows accessing R-Studio using a web browser while it is running on a remote Linux server. R-Studio is written in the C++ programming language and uses the Qt framework for its graphical user interface. R-Studio has a general console where we can type commands and see output, it has an editor that supports direct code execution and has features of highlighting the syntax, debugging and managing the workspace. R-Studio also has a workspace tab which includes all the active objects, a history tab which shows a list of commands used so far, and the files tab that shows all the files and directories in the default workspace.

The remainder of the paper is organized as follows. Chapter 2 discusses about the literature review of previously published research papers related to forecasting the stock trends. Chapter 3 discusses about the system analysis. Chapter 4 discusses the implementation methodology which is used to predict the stock trends. Chapter 5 includes some of the screen shots of predicted results. We conclude the paper in chapter 6 by providing the ideas for future enhancements in chapter 7. Finally references have been included.

II. LITERATURE REVIEW

To attain the objective of forecasting the stock returns, fundamental analysis and technical analysis techniques are used