CMM799 - MSc PROJECT 2023

FORECASTING SHARE PRICES LISTED IN COLOMBO STOCK EXCHANGE USING MACHINE LEARNING

Anojan Srikantha

RGU No: 2017590

IIT No: 20200710

Supervisor: Mr. Alroy Mascrenghe

A Dissertation Submitted in Fulfilment for the Requirement for MSc. Business Analytics

Informatics Institute of Technology, Sri Lanka in Collaboration with Robert Gordon University UK.

2023

Abstract

Investors in Sri Lanka who generally have low net worth and relatively have small portfolios, categorized as retail investors, do not have substantial insights regarding selecting the optimal assets or stocks in order to generate the best return for their portfolios. The investment decisions made by them are not optimal to generate the best return for their portfolios. Though there are many professional investment managers who manage portfolios of individuals and corporates, it is sensible for them to manage portfolios with large quantum of wealth since the fees are higher which allows them to manage their expenses. Professional investment managers have access to substantial amount of data and insights for them to make an informed decisions. Unfortunately, the retail investors mostly do not have access to those type of investment managers. Further, when it comes to forecasting stock prices for the future period, many crucial factors need to be considered. In addition there are several types of approaches to follow to forecast the share price. Most of the investors use the traditional way of forecasting share prices, either using fundamental analysis or technical analysis.

This study primarily focuses on forecasting share prices using machine learning models, as opposed to traditional forecasting models. Further the objective of the study is also to construct a portfolio with the forecasted share prices in order maximize returns of an investor. The machine learning models are used to forecast the share prices of 10 companies with the largest market capitalization in the Colombo Stock Exchange. With the support of past literature, the algorithms were chosen that suits share price forecasting scenarios. The algorithms that have been used are Random Forest Regressor, Long Short Term Memory, Lasso Regression and Auto Regressive Integrated Moving Average. While the closing price of the share of the selected stock are the dependent variable, the features for the model are daily opening price, daily high price, daily low price, share volume, 5 day moving average, 10 day moving average, 20 day exponential moving average and moving average convergence/divergence (MACD). Feature engineering process was run in order to develop the latter 4 variables, which are metrics used for standard technical analysis of a share price. The hypothesis testing, linear regression and trend analysis supports the fact that the features that are used for the model have a significant impact on the dependent variable.

3 algorithms were used to build a model to forecast share prices for the next 30 days. Random Forest, Lasso Regression and LSTM. Using Python language, individual models were built for the selected stocks under each algorithm. The conclusion is that the selected stocks can be forecasted at moderately accurate levels using the daily price and volume related factors. The models were tested by measuring Mean Absolute Error, Mean Squared Error and R squared. Out of the three algorithms LSTM performed relatively better in predicting the future share prices. LSTM had the lowest MAE (0.18) and MSE (0.27) which are indicators of a good model, while Random forest had showed the highest R Squared.