MACHINE LEARNING-BASED SHARE PRICE PREDICTION MODEL FOR THE HEALTHCARE SECTOR IN SRI LANKA

C.H.H.K.M.P. Ruparatne

A dissertation submitted in partial fulfilment of the requirement for

Master of Science degree in Business Analytics

Department of Computing

Informatics Institute of Technology, Sri Lanka

in collaboration with

Robert Gordon University, UK

2022

Abstract

Background

Coronavirus outbreak has impacted many domains including the financial equity markets around the world. Sri Lanka being one of the leading south Asian countries has been facing a challenging time in the Colombo Stock Exchange. Attracting the right investment in the most crucial sector is important. Hence this study will focus on developing a machine learning based share price prediction models specified for healthcare sector companies listed in Colombo Stock Exchange.

Methods

This study employs machine learning based algorithms in predicting the share prices related to healthcare companies in Colombo Stock exchange. Daily share prices of the selected companies were collected covering the coronavirus impacted period to build these models. Then data is preprocessed, and feature engineered before implementing the models. Share price prediction models were developed based on support vector regression, linear regression, lasso regression, decision tree regression, random forest regression, artificial neural network, k nearest neighbor regression and ARIMA model. Then the models were tested for the performance and accuracy and selected the most efficient machine learning model.

Discussion and Conclusion

All the machine learning models were trained using the trained data set and tested with the testing data set to evaluate the prediction each model made. All the models showed accurate results for predictions and some of the model's accuracy was improved after the feature selection done with the principal component analysis. Artificial Neural Network model was the most accurate model in predicting the share prices for all the stocks during high volatile period.

Key words – Artificial Neural Network, Support Vector Regression, Lasso Regression, Decision Tree regression. Random Forest Regression, ARIMA, Stock market.