Informatics Institute of Technology
In Collaboration With

University of Westminster, UK

Helpinwest - Stock Prediction Based on Financial Statement Data and Past Trading Statistics

A dissertation by
Mr. Rachitha Caldera

Supervised by
Mr.Banuka Athuraliya

Submitted in partial fulfillment of the requirements for the
BEng (Hons) Software Engineering Degree Department of Computing

May 2021

Abstract

Colombo stock exchange is a marketplace situated in the capital of Sri Lanka where traders or investors buy and sell shares of stock of companies through Brokers on electronic trading platform. Investors wish to buy shares when the prices of shares decrease and sell shares when their prices increase. One of the major barriers of the investors is to determine whether the share prices further decrease or increase.

Currently, investors heavily depend on Brokers in making their buying/selling decision which is error prone as the brokers only uses their experience for the recommendation rather more complex insight in stock market business. Investors desperately need the best return on their investment and therefore, look for the robust system for their decisions. The Share market prediction is a process that attempt to predict the share prices of stocks using various techniques which varies from statistical analysis, technical analysis to Machine Learning to Data Science. Most of the researchers have conducted range of Machine learning techniques that falls on Supervised Learning (Classification) such as Linear Regression, Non-Linear regression, Neural Network for the prediction using trading statistics publicly available.

The use of Support Vector Machine (SVM) and the Fundamental analysis have not been well experimented. Unlike Empirical Risk Minimization (ERM) principle, the Structural Risk Minimization (SRM) principle deals with the complexity of the models against overfitting problem on the training data set. In Empirical Risk Minimization (ERM) principle, the models deal with the theoretical aspects of the performance and does not provide how well they work in practice. SVM is based on SRM. SVM inherits low variance error and works well even with small data set. In this thesis, the author investigates the SVM-SVR machine learning technique using both the statistical data from CSE and financial data from annual reports of companies. Three models are trained by using SVR Linear Regression, SVR Polynomial and SVR Gaussian Radial Basis Function and evaluated for their performances. It shows that SVR Gaussian Radial Basis Function outperforms the other two. It is also evaluated against a Neural Network Model and performs better than it.

Keywords: Helpinwest, Machine learning, Hyperparameters, Fundamental data, Technical data, Hybrid model, Stock Market