World Academy of Science, Engineering and Technology
International Journal of Computer and Information Engineering
Vol:16, No:11, 2022

# Stock Prediction and Portfolio Optimization Thesis

**Authors :** Deniz Peksen

**Abstract :** This thesis aims to predict trend movement of closing price of stock and to maximize portfolio by utilizing the predictions. In this context, the study aims to define a stock portfolio strategy from models created by using Logistic Regression, Gradient Boosting and Random Forest. Recently, predicting the trend of stock price has gained a significance role in making buy and sell decisions and generating returns with investment strategies formed by machine learning basis decisions. There are plenty of studies in the literature on the prediction of stock prices in capital markets using machine learning methods but most of them focus on closing prices instead of the direction of price trend. Our study differs from literature in terms of target definition. Ours is a classification problem which is focusing on the market trend in next 20 trading days. To predict trend direction, fourteen years of data were used for training. Following three years were used for validation. Finally, last three years were used for testing. Training data are between 2002-06-18 and 2016-12-30 Validation data are between 2017-01-02 and 2019-12-31 Testing data are between 2020-01-02 and 2022-03-17 We determine Hold Stock Portfolio, Best Stock Portfolio and USD-TRY Exchange rate as benchmarks which we should outperform. We compared our machine learning basis portfolio return on test data with return of Hold Stock Portfolio, Best Stock Portfolio and USD-TRY Exchange rate. We assessed our model performance with the help of roc-auc score and lift charts. We use logistic regression, Gradient Boosting and Random Forest with grid search approach to fine-tune hyper-parameters. As a result of the empirical study, the existence of uptrend and downtrend of five stocks could not be predicted by the models. When we use these predictions to define buy and sell decisions in order to generate model-based-portfolio, model-based-portfolio fails in test dataset. It was found that Model-based buy and sell decisions generated a stock portfolio strategy whose returns can not outperform non-model portfolio strategies on test dataset. We found that any effort for predicting the trend which is formulated on stock price is a challenge. We found same results as Random Walk Theory claims which says that stock price or price changes are unpredictable. Our model iterations failed on test dataset. Although, we built up several good models on validation dataset, we failed on test dataset. We implemented Random Forest, Gradient Boosting and Logistic Regression. We discovered that complex models did not provide advantage or additional performance while comparing them with Logistic Regression. More complexity did not lead us to reach better performance. Using a complex model is not an answer to figure out the stock-related prediction problem. Our approach was to predict the trend instead of the price. This approach converted our problem into classification. However, this label approach does not lead us to solve the stock prediction problem and deny or refute the accuracy of the Random Walk Theory for the stock price.

**Keywords :** stock prediction, portfolio optimization, data science, machine learning
**Conference Title :** ICBAIDS 2022 : International Conference on Business Analysis, Intelligence and Data Science
**Conference Location :** Istanbul, Türkiye
**Conference Dates :** November 29-30, 2022