QMBU 450 Final Project: Stock Price Analysis and Forecasting using ML

Introduction & Hypothesis

Starting in 2000's financial institutions have relied increasingly on different modeling techniques to increase trading revenues to make more accurate forecasts. This is usually a quite difficult task as prices theoretically move randomly and according to the Efficient Market Hypothesis, which states that markets already price in all relevant information and only react to information that has not been released to the public. However, several institutions have relied heavily on algorithmic trading and modeling techniques using machine learning and AI with phenomenal returns over 2 decades. Among the machine learning methods used are linear regression and decision tree regression models. This project aims to make a basic analysis of returns and look at which method predicts prices more accurately using historical data of ASELSAN, a Turkey's largest defense corporation.

From the two methods used I believe the Decision Tree Regression will outperform due to the model evaluating multiple possibilities until it is comfortable assigning a specific value. Linear regressions attempt to find a linear relationship between its dependent and independent variables; a Decision Tree Regressor seems to be more suited to make more accurate decisions.

Data

All the data used was obtained using Yahoo Finance's API; the historical data used contained price data such as Open, High, Low, Close and Volume. The time frame used for all the data used was 1/6/15 – 29/5/20, an exact 5year period in trading days. The data used was monthly data not daily in order to obtain more meaningful return data. Only the Close price data was used so most of the other data was deleted. While analyzing returns, data of competitors was also used in order to observe correlations and how ASELSANS returns compare in performance to its competitors.

Models

Decision Tree and Linear Regression models were used in order to make predictions. First, a variable was created to predict a specific number of months in the future (24 months) to shift the data up and created a feature and target data set, after obtaining the last rows of the feature data set, the found the predictions. The data was split into 70% training and 30% testing.

Conclusion

The graphs clearly display the advantage the Decision Tree Regressor had on the Linear regression Model. It also shows that by using machine learning it is possible to obtain accurate predictions by using historical data and challenges the Efficient Market Hypothesis. This also proves that more complex and sophisticated models can predict prices with high accuracy.

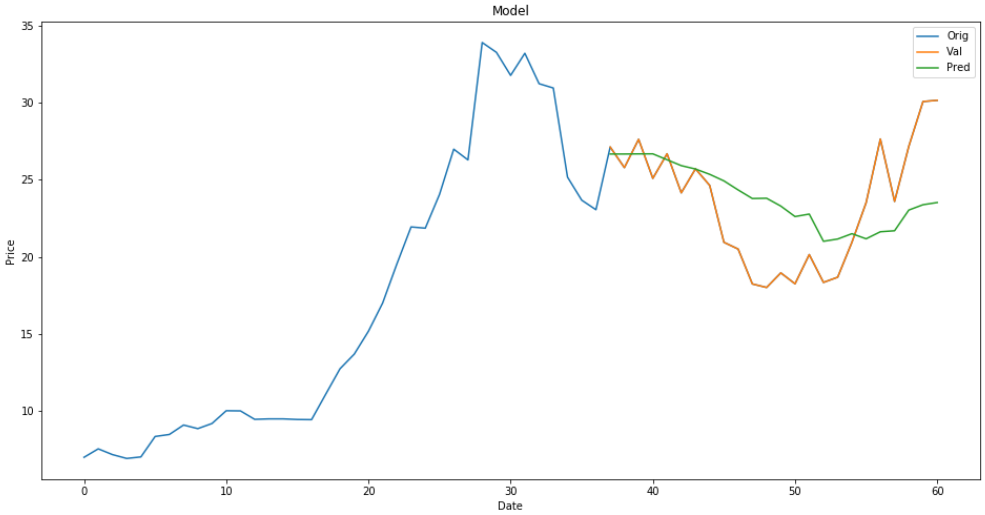

Figure 1:Linear Regression Model

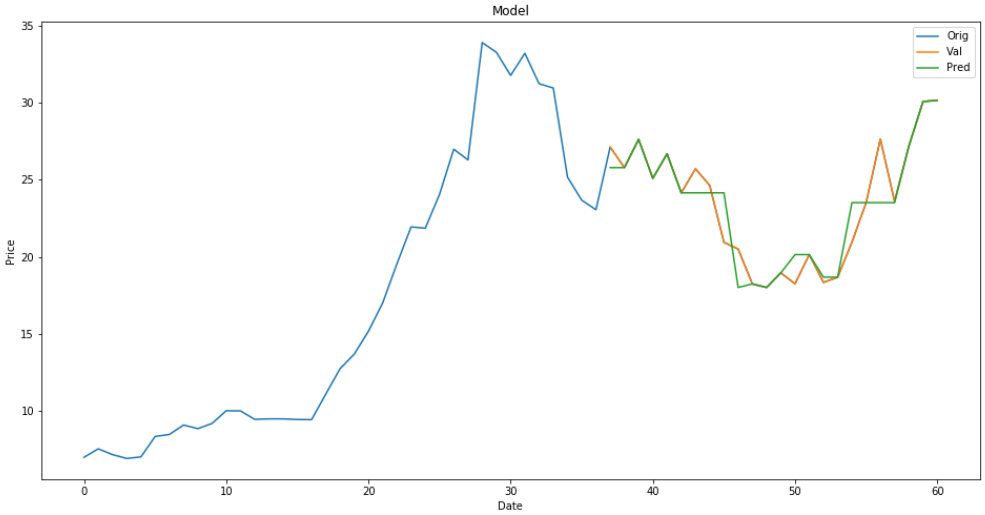

Figure 2: Decision Tree Regression Model