**Predicting Stock Prices in Python using Machine Learning**

**Abstract**

When investing in the United States stock market, people often claim they can detect trends in stock prices and use this information to predict future outcomes. However, there are an enormous amount of factors that contribute to the price of a stock, making it very challenging to forecast future stock prices. It is possible to succeed in the short term, but after repeated attempts we notice that our predictions become much less accurate.

To reduce the human bias and error in our predictions, we can use algorithms constructed with machine learning techniques. If we train a machine learning model in python with several years of historical stock market data, we can expect our future predictions to be much more accurate. Python is a powerful programming language with a wide selection of machine learning resources, making it a perfect candidate to carry out our algorithm.

Our proposed stock price predictor will not be immune to unforeseen stock market volatility, but removing the human element will greatly improve the efficiency and accuracy of the prediction process. The results of this algorithm are not intended to be financial advice, but rather a showcase of the capabilities of machine learning models.

**Introduction**

The idea behind predicting stock prices is that investors will be able to buy or sell their shares at the right time to make a profit. If we suspect a certain stock price will be increasing soon, this allows us time to purchase shares before they increase. In contrast, predicting a stock price will decrease gives us an opportunity to sell shares at their current price. This concept seems achievable, but the vast majority of short term investors end up losing money from their predictions. Why is this the case?

Four broad categories contribute to the fluctuation of stock prices: company performance, industry performance, investor sentiment, and economic factors. Each of these categories contain several components that play a role in the short and long term movement of stock prices. What makes stock price predictions incredibly difficult is that an investor could analyze multiple factors correctly, but if any single aspect of their forecast turns out to be incorrect, the stock can move in the complete opposite direction than expected. Often times the stock market goes through cycles, where one sector experiences tremendous gains and another sector rapidly declines. A company might introduce a new product, release its earnings, or announce other significant changes, each of which has a different impact on their stock price. With this many factors at play, short term stock price prediction becomes a dangerous task.

**Related Work**

Some of the existing approaches to stock price prediction have a certain level of credibility, but they come with other disadvantages. One common idea of stock market investing comes from the concept of momentum. This approach is used when the stock market is strongly trending in one direction, and investors are speculating that the current movement will continue. For example, in the past month of the United States stock market, stocks in the energy sector have noticed significant increases. If an investor wanted to use the momentum strategy, they could purchase energy stocks in hopes that the current increases will continue. At times this method can be effective, but stock market momentum is often short lived. Investors may find themselves buying in after the momentum has run its course, and soon the trend will balance itself out by moving the other direction.

Another way of predicting stock prices involves searching for stocks that have prices much lower than anticipated. There are times when stock prices overcorrect, and investors place a higher level of value on a certain stock than what the market reflects. This strategy is called value investing, which billionaire investor Warren Buffet is famous for. Although it can be difficult to correctly locate and time when a stock is undervalued, executing this strategy can bring great returns. It is important to properly research these value stocks, because an incorrect analysis of an undervalued stock can leave you with significant shares of a declining asset.

A third approach to stock price prediction is called mean reversion. Investors with years of experience might notice how after stocks experience several highs and lows, they tend to eventually revert back to the middle. Therefore if we notice outliers in stock prices, a wise decision could be to predict the stock will eventually settle back to a medium level of value. This approach is most effective in longer term scenarios, where the market has time to correct from both extremes. If an investor tried to use a mean reversion approach during short term momentum, they may not see any returns for an extended period of time.

**Proposed Approaches**

Our project takes a data-driven approach to stock price prediction. Many human investors are influenced by companies they like, people they know, and other emotional factors. With machine learning, all emotion is removed, and only the data is brought into consideration. The most important part of training our machine learning algorithm to predict stock prices will come from the data we use. A machine learning model is only as powerful as the data it is given. The more years of stock market data we provide to the model, the more opportunities it has to learn from and develop a stronger, more accurate prediction. Thankfully, with the power of modern computing, we have an abundance of stock market datasets at our fingertips.

One of machine learning’s greatest strengths is the ability to detect patterns and trends in the data. You could give a human a large dataset of stock prices from the last 100 years and ask them to point out what trends they notice. It might take them several minutes or even hours to get a grasp of the data and start to locate any noticeable trends. Whereas with machine learning, we can provide the same dataset, and it could locate much more trends, to a higher degree of accuracy, in a small fraction of the time a human takes to identify them.

In addition to pattern recognition, machine learning provides a level of automation that humans cannot replicate. Programming scripts can be scheduled to run at regular intervals as more data is collected, allowing predictions to become even more accurate. Humans, who have other responsibilities and obligations, are not always available to sit down and analyze financial data. Computers, however, are designed to perform exactly what we tell them. They do not need to eat or sleep, they don’t get sick, and they never get frustrated or bored. This level of consistency and reliability will significantly outperform the other approaches.

**Plan**

This project involves using the Least Square Support Vector Machine (LSVM) and other associated machine learning techniques on historical datasets imported from Google Finance and Quandl/Quantpedia for predicting the future stock prices. In addition to LSVM we will also implement the Linear Regression technique to the dataset and the SVM models (linear/Poly/RBF kernels). We will incorporate the aforementioned ML techniques and compare the various methods involved in assessing the data and predicting the trends from it. A key element of our model will be the dynamic response it has to the introduction of new data in our system. The model would be trained to reflect any changes it perceives from stock market price fluctuations. We also aim to ensure that our model would have efficient checks and balances to accommodate outliers and predict the future trends accurately. As mentioned previously, our approach would be a data driven one with emphasis on reliably sourced data and past historical trends for predicting the future trends. We will also strive to acknowledge, source, and give any due credit to references, materials, studies, and sources we plan to use in the development of our machine learning model.

**Tools Used**

We mainly implemented our ML models using python coding language and the use of Yahoo finance historical datasets. For implementation of the models we made extensive used of python modules like scikit learn, keras and pandas.

**Observations and Results**

Before implementing our project we had expected the LSVM and linear ML models(especially Linear Regression and Linear SVM) to perform better than the rest of the bunch. As stock prices and time generally follow a linear relationship the linear models are expected to perform comparatively better. However we made sure that no biases were involved in the implementation and all models had been trained with the same dataset and resources.

A brief overview of our plots and results include:

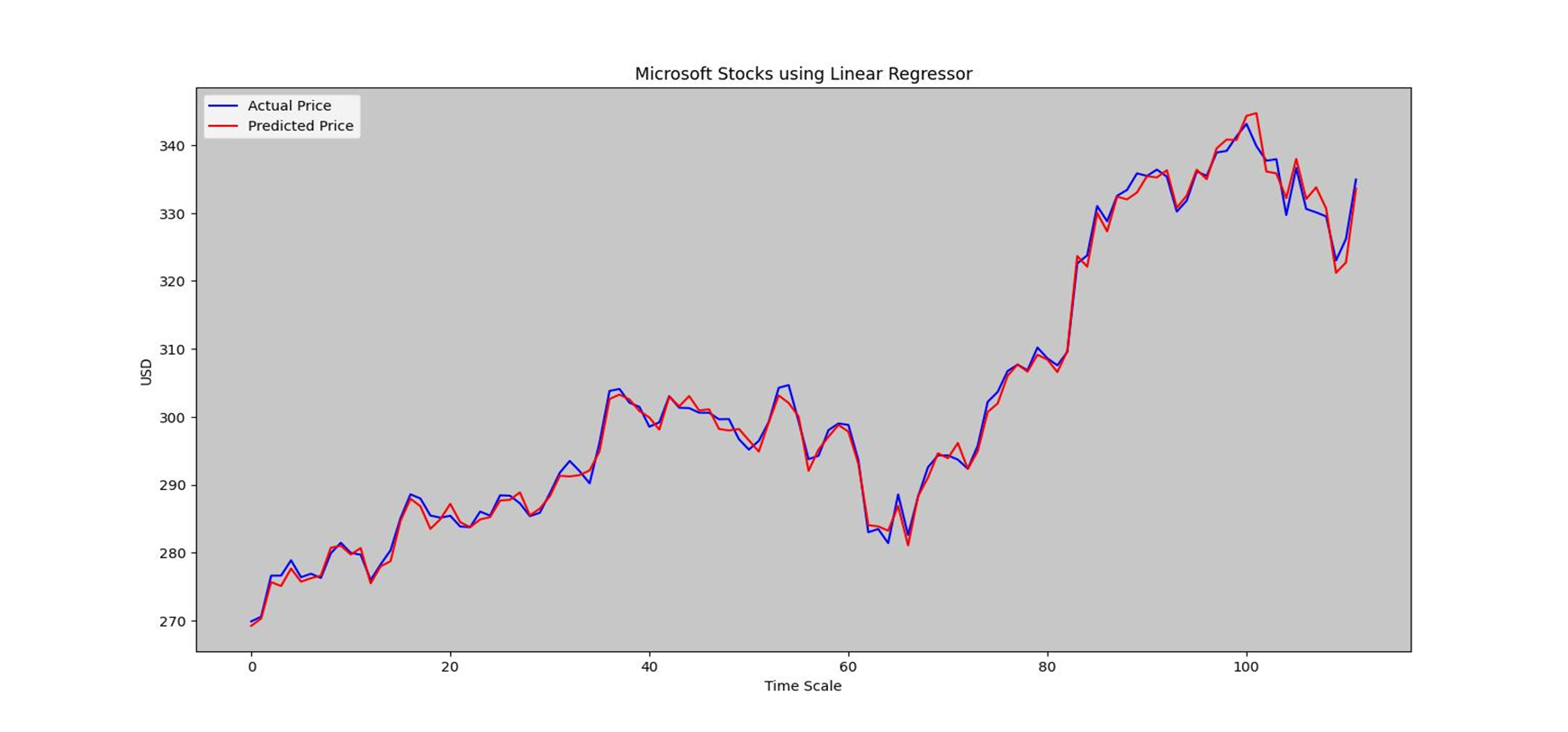
LSTM(Model 1)



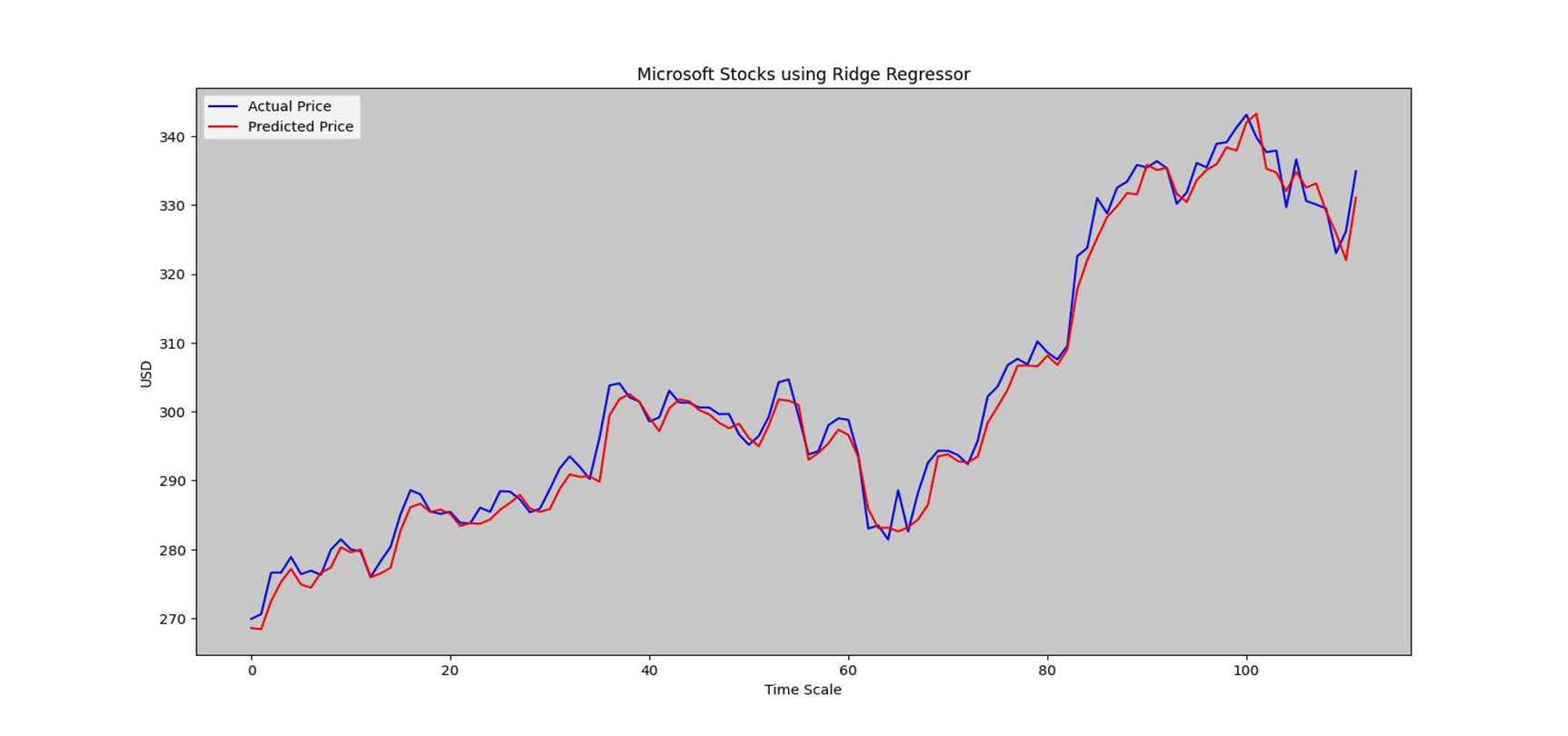
LSTM(Model 2)



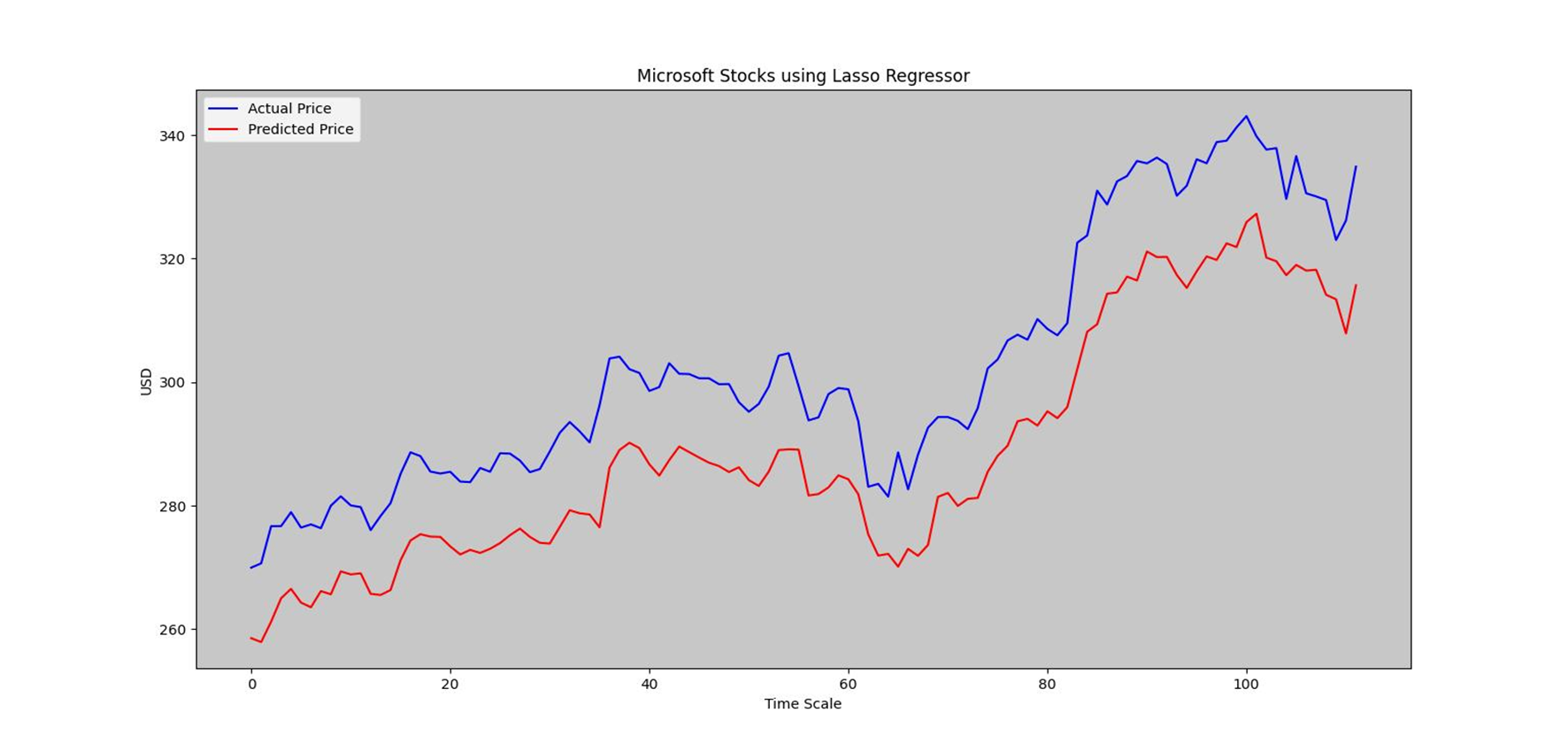
Linear Regressor



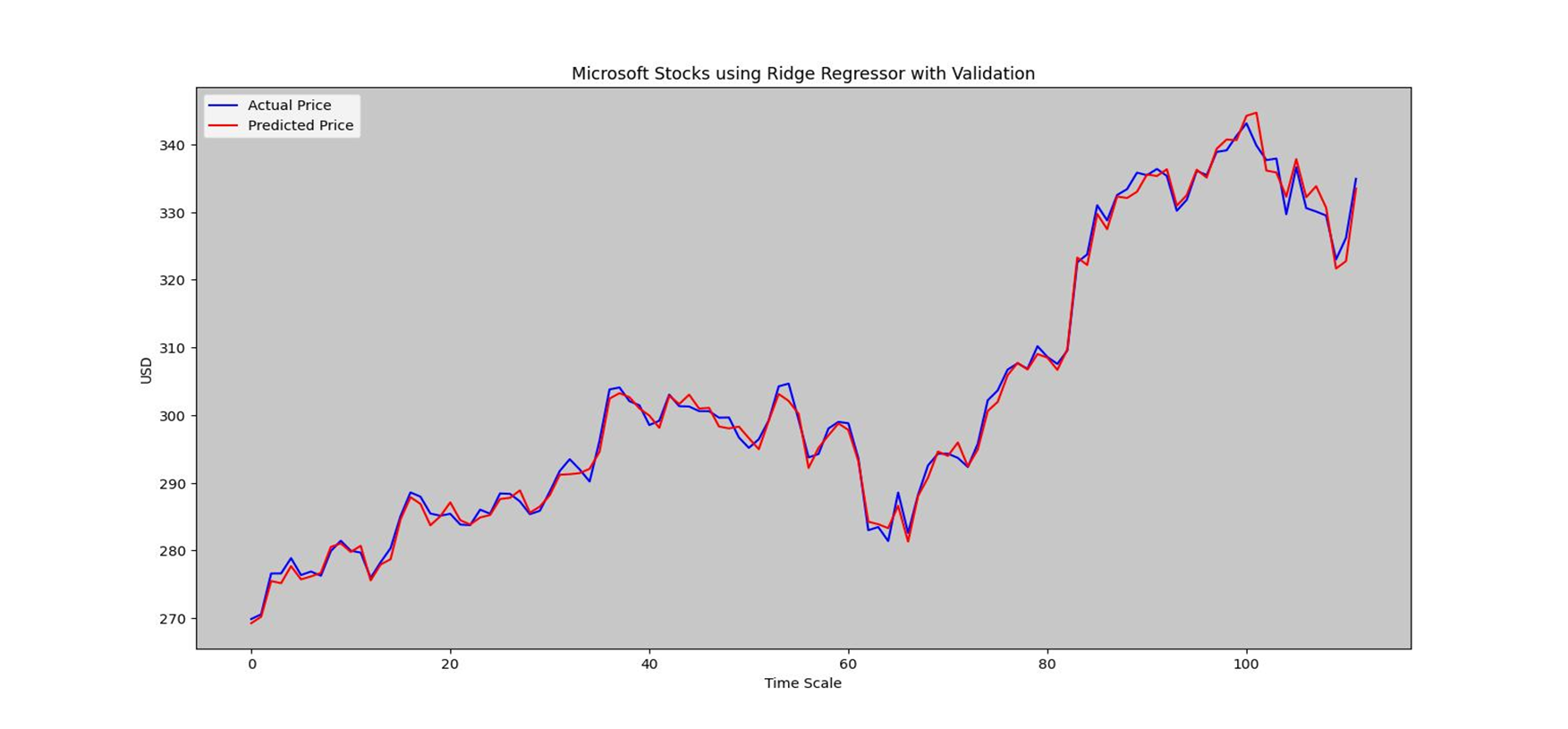
Ridge Regressor



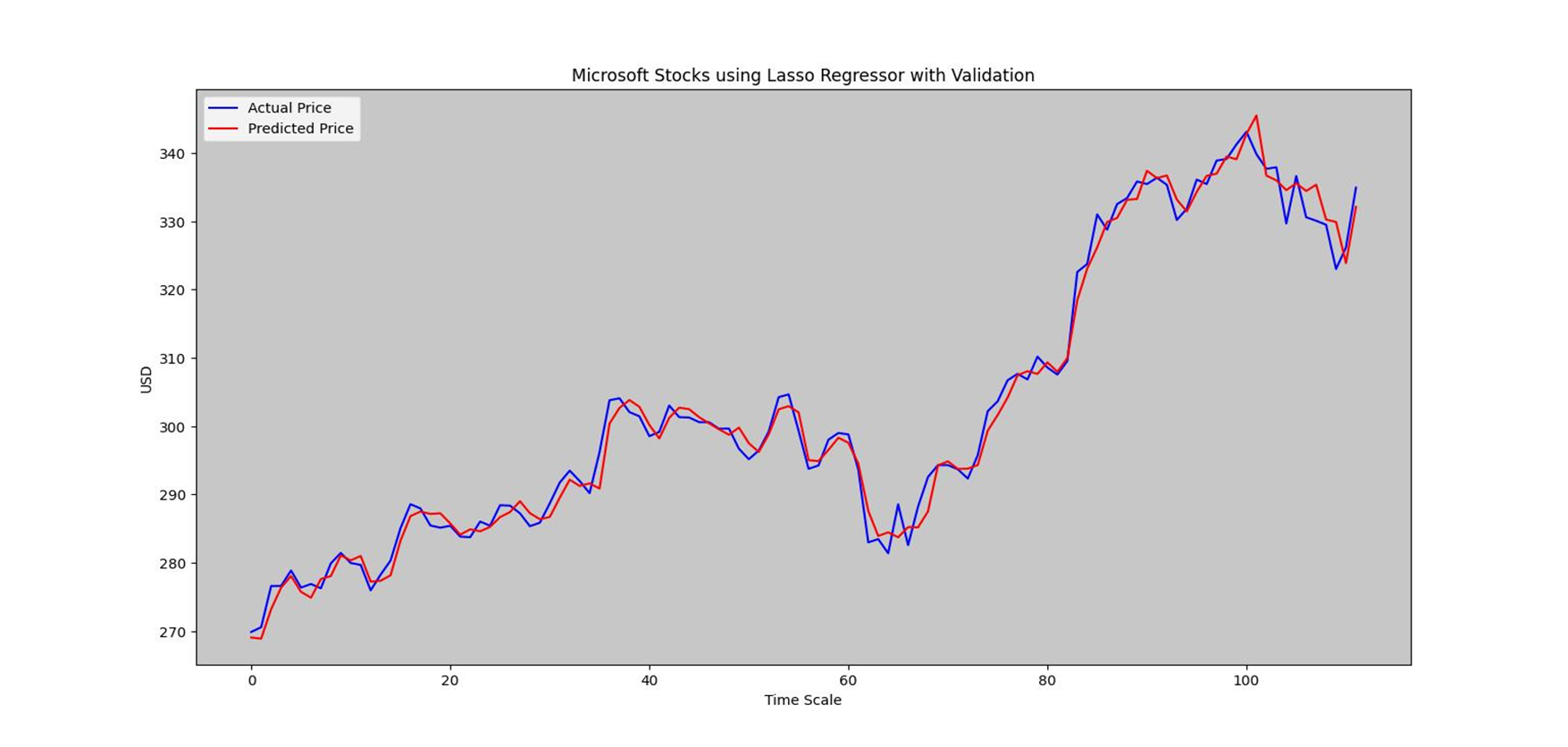
Lasso Regressor



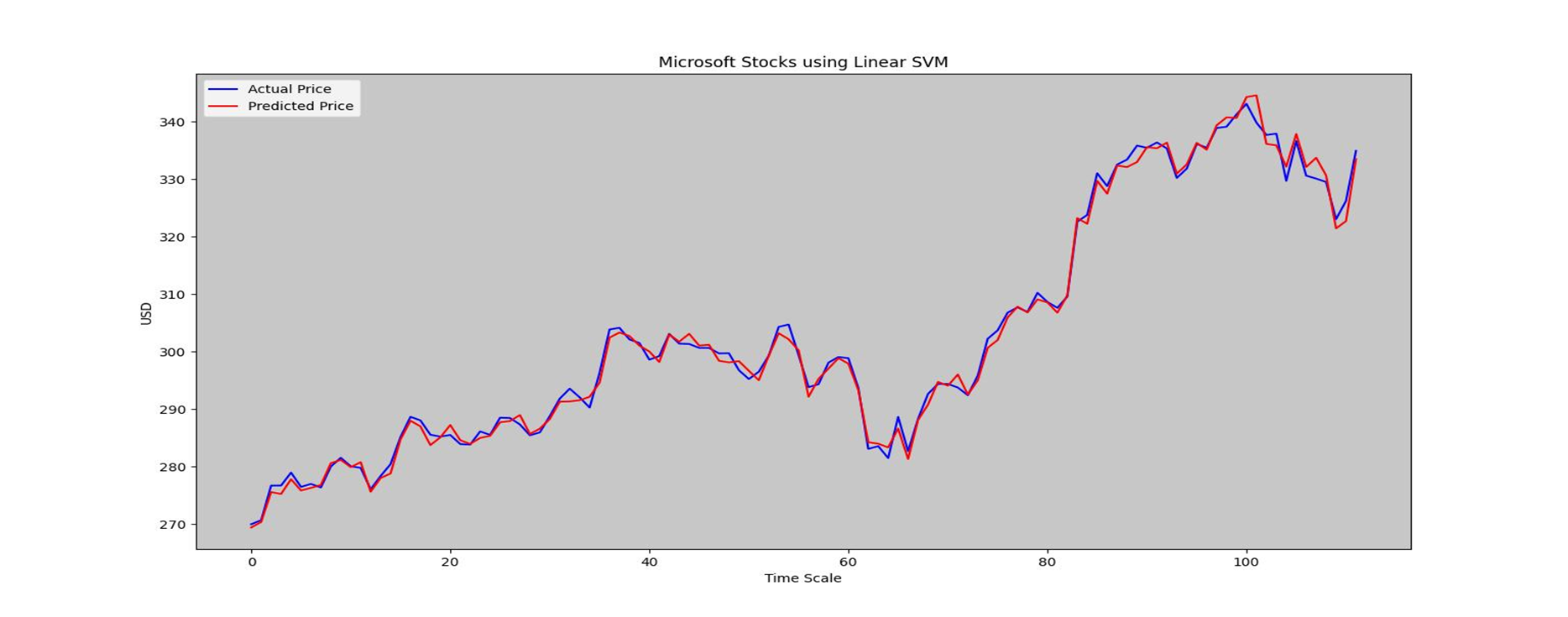
Ridge Regressor with Validator



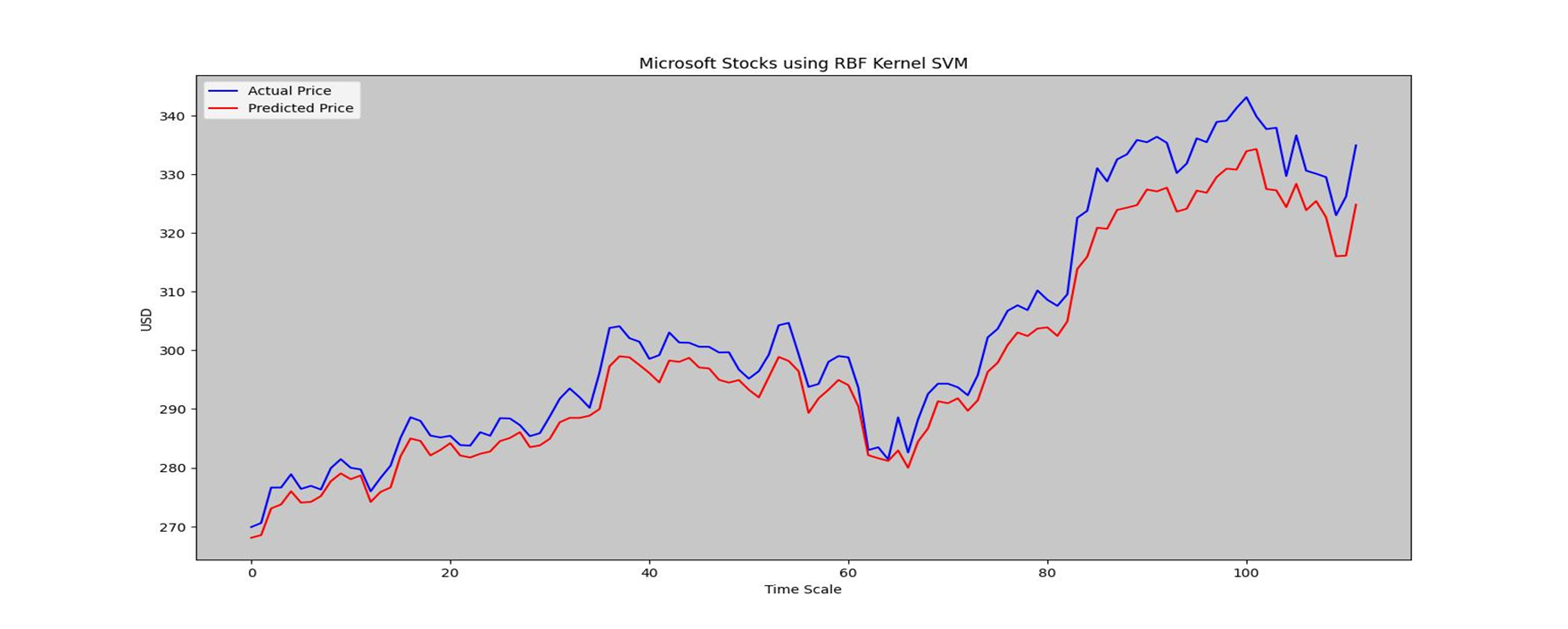
Lasso Regressor with Validator



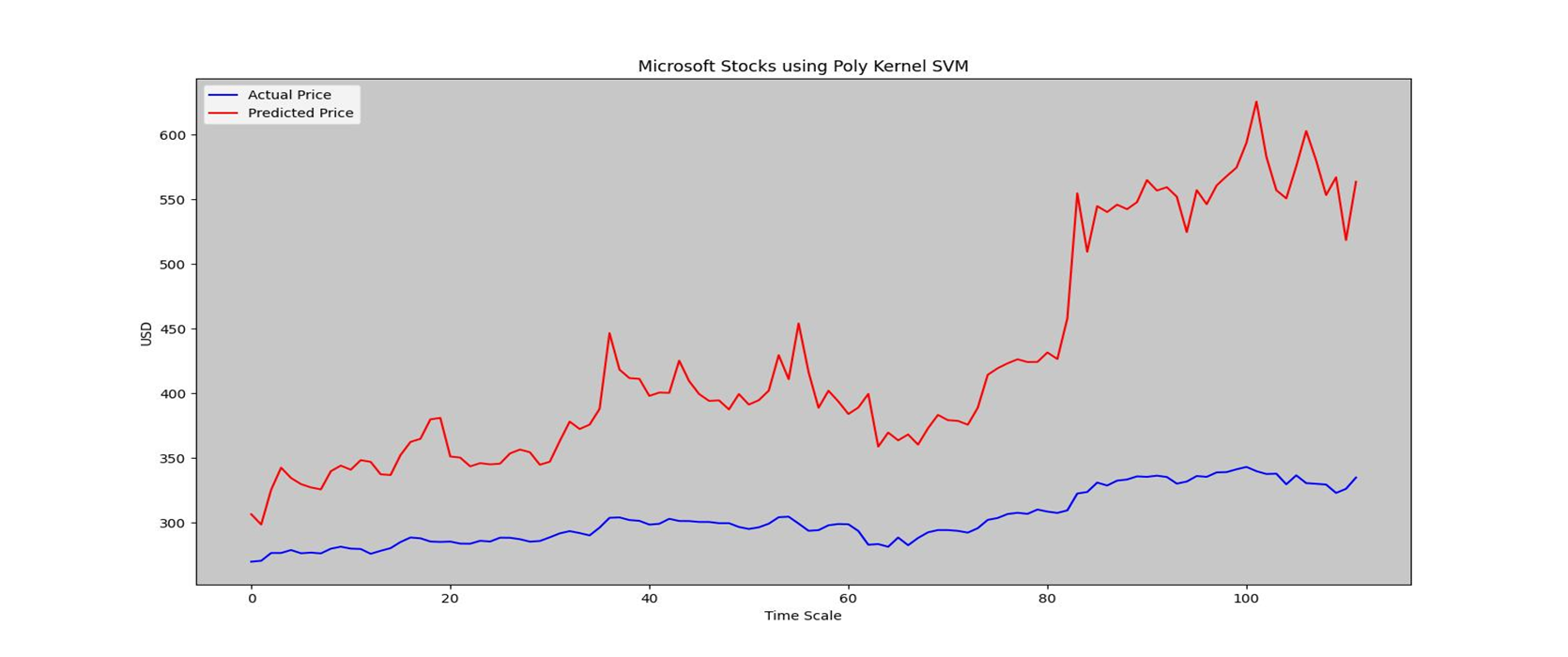
Linear SVM



RBF SVM



Poly Kernel SVM



**Preliminary Observations**

•Linear Regression provides best fit for future stock prices

•Ridge and Lasso without validation perform worse than Linear regression

•With validation, ridge performs nearly as well as Linear regression

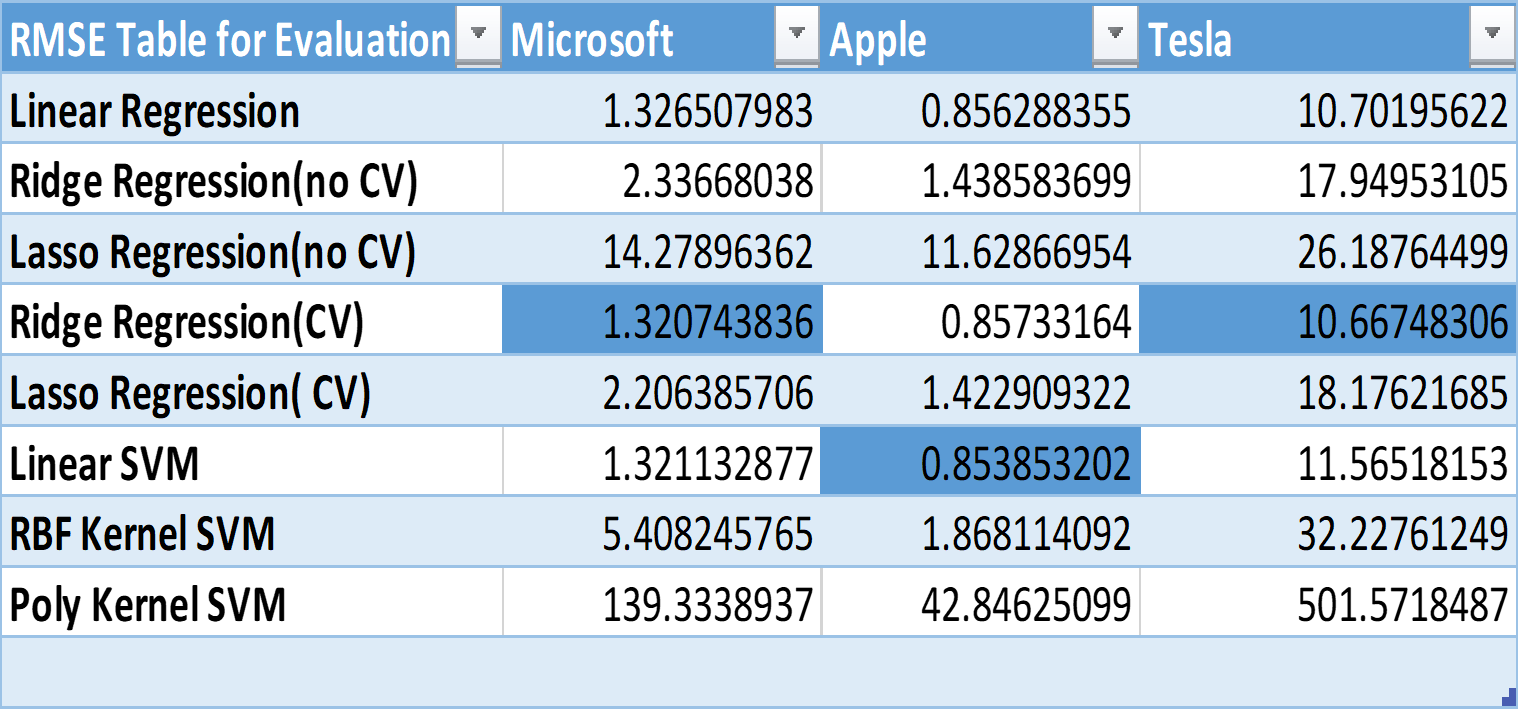
•Lasso regression with validation also shows improved performance

•Linear SVM seems to provide the best modelling for the stock prices

•The second best modelling is provided by RBF kernel SVM

•Linear SVM as it is similar to Linear Regression is a good choice for predicting stock prices

**Performance Evaluation**



**Conclusion of the Report**

•Our conclusion at the end of this project is that for stock price datasets we can implement Ridge Regression with cross validation, Linear Regression, Linear SVM and LSTM 2 modelling to forecast future stock prices

•Among these modelling Linear Regression provides a good compromise between cost, simplicity and implementation

•The other models also perform quite well in forecasting the future stock prices and the trader can pick any one of them depending on his personal convenience

•A word of caution is that stock prices are highly volatile and seem to break trends and predictions so no forecasting model should be taken as absolute.

**References**

• https://scikit-learn.org/stable/search.html?q=regression

•https://www.cs.princeton.edu/sites/default/files/uploads/saahil\_madge.pdf