ST494 FINAL PROJECT

Stock Market Predictions Using Statistical Learning

Maheep Jain and George Mandl
203386460, 200655330

Contents

Abstract

This project focuses on using technical analysis and technical indicators to make predictions and future price movements in financial markets. The data used for this project is obtained from yfinance and cpi, including historical price and volume data for various equities. Technical indicators, such as moving averages and trend lines, are created using this data and used to make the future price movements predictions. Support Vector Machine (SVM) model and Recurrent Neural Network (RNN) model were used in attempt to predict future price movements. The RNN models proved to be more accurate and flexible to different equities.

The following link will send you to the GitHub page of the project. Run the interactive.py file and give our models a try! LINK

Introduction

Technical Analysis is a method of evaluating various financial products based on statistical trends and patterns in their historical price and volume data. It involves analyzing charts and technical indicators to identify buy and sell signals. Technical analysis is primarily used by short-term traders such as day traders, swing traders, and other active traders who aim to profit from short-term price fluctuations in the market. However, technical analysis can also be used by longer-term investors who use technical indicators to identify entry and exits points for their positions.

Technical indicators are mathematical calculations based on price and/or volume data of a security. These indicators are typically displayed on charts and can include moving averages, oscillators, trend lines, and other graphical representations of the market data. The motivation of this project is to use technical indicators to make predictions about future price movements. In essence, these predictions can be used by anyone who manages an investment portfolio. It is important to note that there is significantly more risk in short term trading which is why retail investors should be cautious before incorporating the results of this report in their investment strategies.

Data

The data is retrieved using two python APIs. The first, yfinance, is an API that can connect to the yahoo finance website and scrapes information regarding a manually inputted equity into dictionaries and pandas DataFrames. The yfinance API is used to collect the historical price and volume data for the equities on which this project was built on.

The second API used is cpi. The purpose of the cpi API is to transform the historical prices to their present-day values. The inflate function in cpi library compounds the historical prices to their theoretical present-day value which then can be compared to all the other prices. By compounding the prices to today's dollars, have reduced the variability in the price by removing the noise caused by inflation.

Now that the APIs have collected the historical data and inflated it into equal value, the technical indicators can be created. Each technical indicator has its own interpretation, formula, scale, and trading signals. The description of each of the technical indicators used can be found in the Appendix. There are both discrete and continuous variables representing most technical indicators. The continuous variables are the outputs of the formulaic calculation of the indicators and the discrete variables are trading signals produced based on the continuous variables of the indicator. Standardization was applied to the continuous technical indicator variables.

The rate of change in prices and indicators can be crucial predictors for short term price fluctuations. A rate of change transformation has been applied to the standardized price and indicators.

Exploratory data analysis was performed on the prices, technical indicators, and the rate of change approximations. The results of EDA showed that the many of the variables followed a normal distribution or log normal distribution (See EDA.ipynb for EDA results). This allows for the conversion of many of the observed variables into continuous random variables with normal distributions. The continuous random variables make room for simpler approximation of conditional probability mass functions and conditional expected values which can be used in future scopes.

Technical indicators can be grouped into the different attributes of a trend that they observe. The most common types of indicators are momentum indicators, strength indicators, and direction indicators. The description of each type can be found in the Appendix. An underlying assumption of technical analysis is that the indicators directly reflect the trends of the equity. If this holds true, then I can be assumed that there is correlation between indicators that represent the same attributes of a trend. In the EDA process, clustering was performed but unfortunately there was no decisive K value found from the elbow plot and

silhouette plot (see Appendix for plots). Principal components analysis was also performed, and the principal components were extracted into a data frame to be used in the model training.

The target variables of this dataset can come in different forms. There are several characteristics of prices that can be predicted which allow for a lot of flexibility in the choice of model. Future price, future price changes, and future price movement classifier columns were mutated through the shifting of pricing columns in the dataset. Future price and future price change are continuous variables and future price movement is a discrete column taking on 0 for a fall in price and 1 for a rise in price.

Methods

Support Vector Machines (SVM)

SVMs are a good choice for stock market prediction because they make use of a risk function that is made up of empirical error and regularized term that is derived from structural risk minimization principle. We implemented SVM using the sklearn library. Here, the indicators were the Open – Close and High – Low price of the stocks. These variables are based on which the model will predict tomorrow's trend. Our target parameter is storing the correct trading signal which our model will try to predict. If tomorrow's price is greater than today's price then we will buy the particular stock else not. We trained our model using Grid Search to find the best hyper tuning parameters for our model. Three parameters we tuned are Kernels, Gamma, and C (Regularization). The main function of kernels is to take low dimensional input space and transform it into a higher-dimensional space. Gamma defines how far influences the calculation of plausible line of separation. C represents misclassification or error term. It tells the SVM optimization how much error is bearable.

Recurrent Neural Network (RNN)

RNNs are a good choice for stock market prediction because they capture the sequential nature of trends in the markets. The feedback loop of RNNs allows them to maintain a memory of past inputs and learn from them. The RNN models are implemented using the Keras library that is built on TensorFlow. They are sequential models that contain long & short term memory (LSTM) layers. Between the LSTM layers there are Dropout regularization layers that promote the reduction of variance and overfitting. At the end of the LTSM and dropout cycle, the outputs are fed to a single Dense layer to create a singular linear output. Hyperparameter tuning was performed on the number of LSTM layers in the model, the number of neurons in each LSTM layer, and the learning rate for the model. Once the hyperparameters are tuned, a basic model will be built, and more features will be added to see if they improve the performance.

Results

Support Vector Machine Results

In our SVM model, we used hyper parameter tuning to find the best parameters for our model using Grid Search. After finding the best parameters, we were able to see a test accuracy of approximately 53%. It's a good accuracy but not what we expected out from our model.

Recurrent Neural Network Results

The first initial model produced a test loss of 0.0042 (see Appendix 3.1 to view the results). Once the initial model was built tuning was preformed on the number of layers, number of neurons, and learning rate on the RNN using 5-fold cross validation. There seems to be no concrete correlation between number of layers and MSE, and number of neurons and MSE but MSE seems to increase as the learning rate increases (see Appendix 3.2 for the cross validation for hyperparameter tuning results).

Using the optimal hyperparameters, the training of the final models began. The initial final model uses only the previous closing price as a predictor variable and produced a test loss of 0.0082 (see Appendix 3.3 to view the results). The assumption of technical analysis is that technical indicators can predict the future price movements, so another model was trained which included the first two principal components of the technical indicators input space (selected using the scree plot in Appendix 2). The principal components reduced the test loss by 0.0005, producing a test loss of 0.0077 (see Appendix 3.4 for results). The final model using the closing price and the first two components was validated using a completely different stock. The validation loss is 0.0006 and the results can be seen in Appendix 3.5.

Conclusion

In conclusion, Support Vector Machines didn't show us the results we expected for this problem of technical analysis, that is why we moved to deep learning model of RNN for better results. Recurrent Neural Networks are very powerful at predicting stock prices and are very flexible across different stocks. Although it does require a large sample size and can be quite computationally intense, its results are fantastic. The financial markets are being ever more dominated by artificial intelligence and the performance of RNNs in this project shows just why.

Appendix

Appendix 1.1 Domain Definitions

Equity An equity is the stake which shareholders of a company hold. Equity's which are listed to the public can be traded one exchanges such as the Toronto Stock Exchange (TSX) and the NASDAQ stock market. For this report the term equity is synonymous with the terms stock, share, security, and asset.

Overbought A condition that indicates that the equity is trading at a higher price than it is worth. If an equity is overbought, then it is a signal to sell or short it and can indicate a potential price pullback.

Oversold A condition that indicates that the equity is trading at a lower price than it is worth. If an equity is oversold, then it is a buy signal and can indicate a potential price bounce.

Overvalued A condition that indicates that the equity is trading at a higher price than it is worth. If an equity is overvalued, then it is a signal to sell or short it and can indicate a potential price pullback.

Undervalued A condition that indicates that the equity is trading at a lower price than it is worth. If an equity is undervalued, then it is a buy signal because you are purchasing it for a bargain. An undervalued condition could also indicate a potential price bounce.

Bullish A condition that indicates that the equity is trading at a lower price than it is worth. A bullish signal would lead an investor to believe that the equity will go higher and indicate that they should purchase the equity.

Bearish A condition that indicates that the equity is trading at a higher price than it is worth. A bearish signal would lead an investor to believe that the equity will go lower and indicate that they should sell or short the equity.

Note: For the purpose of this project the terms *Bullish, Undervalued,* and *Oversold* are synonymous and the terms *Bearish, Overvalued,* and *Overbought* are synonymous.

Appendix 1.2 Technical Indicators Definitions

Moving Average (MA) A moving average is a calculation done over a time series and is used in investing to smooth out the price of an equity. It updates the average price while it moves along the time series, hence the name moving average. A rising moving average indicates that an equity is in an uptrend and a declining moving average indicates than an equity is a downtrend.

Simple Moving Average (SMA) The simplest way of calculating a moving average is to take the arithmetic mean of the stock prices over the lookback period.

$$SMA = \frac{Price_1 + \ Price_2 + \cdots + Price_n}{n}$$

Price is the price from i days ago from i equals 0 to n-1

n is the length of the lookback period

Exponential Moving Average (EMA) The EMA gives more weight to recent prices. It multiplies the previous days EMA by its weight and adds it to the price of the equity that day.

$$EMA = Price * \frac{Smoothing\ Factor}{1+Lookback\ Length} + Previous\ Day\ EMA * [1 - \frac{Smoothing\ Factor}{1+Lookback\ Length}]$$

Oscillator An oscillator is a tool in technical analysis of equities used to measure overbought and oversold conditions. It is set up to be bounded between two extreme values and fluctuates between the bounds. As the oscillator trends towards the upper extreme, it indicates that the equity is overbought and as it trends towards the lower extreme, it is indicating that the equity is oversold.

Trend Indicators A technical indicator that is used to measure the direction and strength of a trend in a stock by comparing prices to an established benchmark.

Average Directional Index (ADX) - Trend Indicator The ADX is a technical indicator that measures the strength of a trend using moving averages of the directional movements usually over 14 days. It is a popular indicator because investors reduce risk and increase profit potential by trading the direction of a strong trend. The values of ADX range from 0 to 100.

0 - 25	Absent/Weak Trend
25 - 50	Strong Trend
50 - 75	Very Strong Trend
75 - 100	Extremely Strong Trend

Momentum Indicators These indicators are used to identify the speed of price movement by comparing the current closing price to previous closes.

Relative Strength Index (RSI) - Momentum Indicator RSI is a momentum indicator that measures the magnitude of an equity's recent price change to evaluate oversold or overbought conditions. RSI is an oscillator that fluctuates between 0 and 100. An RSI value over 70 indicates an overbought condition and can indicate that the equity is primed for a pullback. An RSI under 30 indicates an oversold condition.

RSI is calculated in two steps. The first step uses the average percentage gain and loss over the period in which you are looking back. The standard lookback period is 14 days. Days where the stock gained value are counted as 0 in the average loss calculation and days where the stock lost value are counted as 0 in the average gain calculation. The formula for the first step is:

$$RSI1 = 100 - \frac{100}{1 + \frac{Average\ Gain}{Average\ Loss}}$$

The second calculation smooths the results over the lookback period similar to a moving average.

$$RSI2 = 100 - \frac{100}{1 + \frac{Previous\ Average\ Gain * 13 + Current\ Gain}{Previous\ Average\ Loss * 13 + Current\ Loss}}$$

Moving Average Convergence Divergence (MACD) - Momentum Indicator The MACD is a trend-following momentum indicator that shows the relationship between two EMAs. Generally, the MACD is equal to the 26-day EMA subtracted from the 12-day EMA. For trade signals, the MACD is plotted against its 9-day EMA which we call the signal line. The buy signal is

when the MACD crosses above the signal line and the sell signal is when the MACD crosses below the signal line. This strategy is called MACD crossover.

Stochastic Oscillator - Momentum Indicator The stochastic oscillator is a momentum indicator which compares the closing price to a range of prices over a period of time. It is bounded between 0 and 100 and is used to generate overbought or oversold signals.

$$Stochastic\ Oscillator\ = \frac{Close\ -\ Lowest\ Low}{Highest\ High\ -\ Lowest\ Low}$$

A value greater than 80 is an indication of overbought and a value less than 20 is an indication of oversold.

Relative Vigor Index (RVI) - Momentum Indicator The RVI is a momentum indicator that measures the strength of a trend by comparing the equity's close to its trading range while using a SMA to smooth results. The RVI is calculated in two steps. First the numerator and denominator are calculated separately and then their SMA is taken.

$$Numerator\ = \frac{a\ +\ 2b\ +\ 2c\ +\ d}{6}$$

a = close - open

b = close - open one bar prior to a

c = close - open one bar prior to b

d = close - open one bar prior to c

$$Denominator\ = \frac{e\ +\ 2f\ +\ 2g\ +\ h}{6}$$

e = high - low of bar a

f = high - low of bar b

g = high - low of bar c

h = high - low of bar d

$$RVI = \frac{SMA\ of\ numerator}{SMA\ of\ denominator}$$

The RVI is also plotted against a signal line and then uses crossovers as a trading signal. If the RVI crosses above the signal line it is a buy signal and if the RVI crosses below its signal line then it is a sell signal.

$$Signal\ Line = \frac{RVI + 2i + 2j + k}{6}$$

i = RVI one bar prior

j = RVI one bar prior to i

k = RVI one bar prior to j

Another trading signal is RVI-price divergence. Divergence between the RVI and price suggests there will be a near-term change in the trend in the direction of the RVI's trend. For example, if the equity is rising and RVI is falling, then the equity will correct itself in the near term.

Know Sure Thing Indicator (KST) - Momentum Indicator KST is a momentum indicator that takes the SMA of four rate-of-change (ROC) periods and adds them together. KST is then compared against its signal line to indicate overbought and oversold trends.

KST = 10-day SMA of 10-day ROC + 2 (10-day SMA of 15-day ROC) +
3 (10-day SMA of 20-day ROC) + 4 (15-day SMA of 30-day ROC)

KST fluctuates above and below the 0 line. A positive KST indicates a bullish trend and a negative KST indicates a bearish trend. KST is also plotted against its signal line which is the 9-day SMA of the KST. When KST crosses over the signal line it is a buy signal and when KST crosses below the signal line it is a sell signal.

Disparity Index - Momentum Indicator The disparity index is a momentum indicator that measures the relative position of the closing price to a selected moving average.

$$Disparity = \frac{Close - SMA}{SMA * 100}$$

When disparity is greater than zero the equity has upward momentum, when the disparity is less than zero the equity has downward momentum.

Commodity Channel Indicator (CCI) - Momentum Indicator CCI is a momentum based oscillator used to signal overbought and oversold trends by assessing the equity's direction and strength.

$$CCI = \frac{Typical\ Price - SMA\ of\ typical\ price}{0.015 * |\ Typical\ Price - SMA\ |}$$

Overbought and oversold levels are not fixed since the indicator is unbounded. Once the signal levels are established a crossover technique is used for trading. When the CCI crosses above the overbought level it is a sell/short signal and when the CCI crosses below the oversold it is a buy signal.

Volume Indicators Volume indicators measure the strength of a trend based on the volume of shares traded.

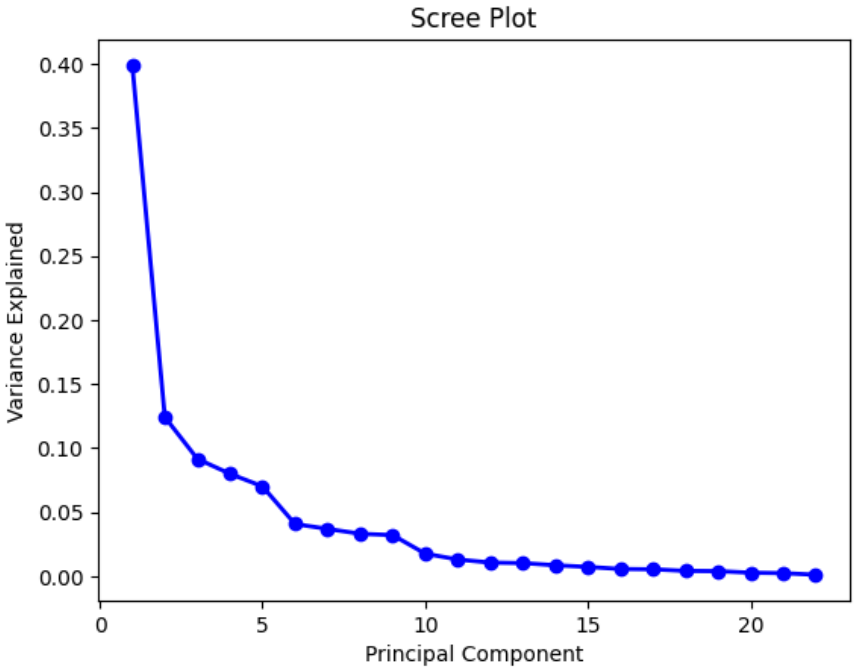
Aroon Indicator - Volume Indicator The Aroon indicator is a technical indicator that is used to identify changes in the equity's price and the strength of that trend. It measures the time between when highs are formed and when lows are formed. A strong uptrend will have regular new highs and a strong downtrend sees regular lows. The Aroon indicator has two components - Aroon up which measures the strength of the uptrend and Aroon down which measures the strength of the downtrend. The general time frame of the Aroon indicator is 25 days.

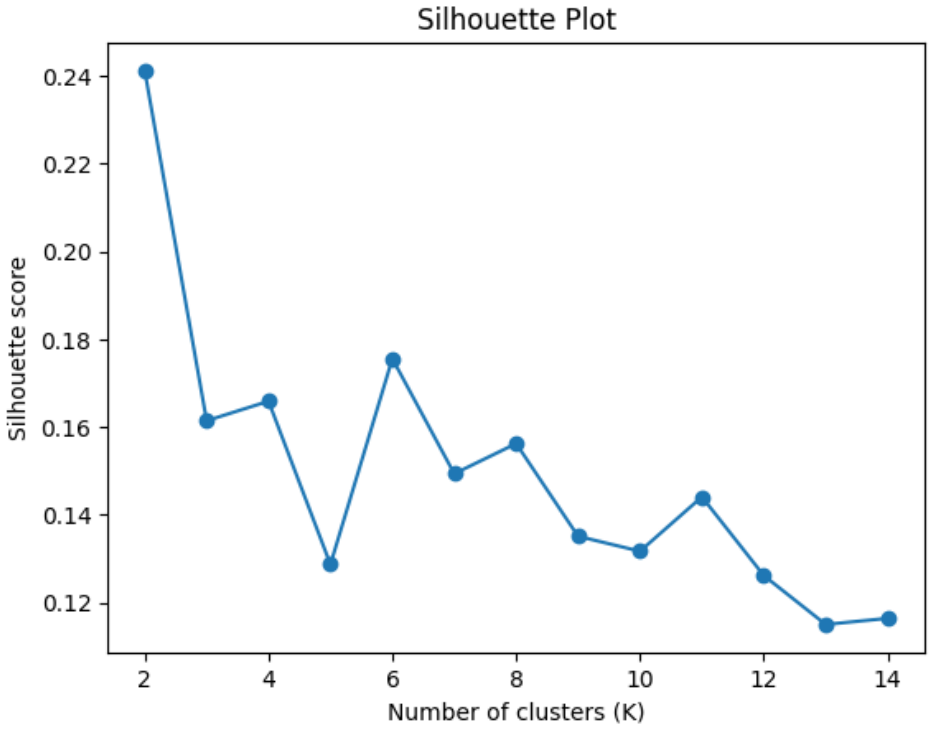
$$Aroon\ Up = \frac{25 - periods\ since\ 25\ period\ High}{25}$$

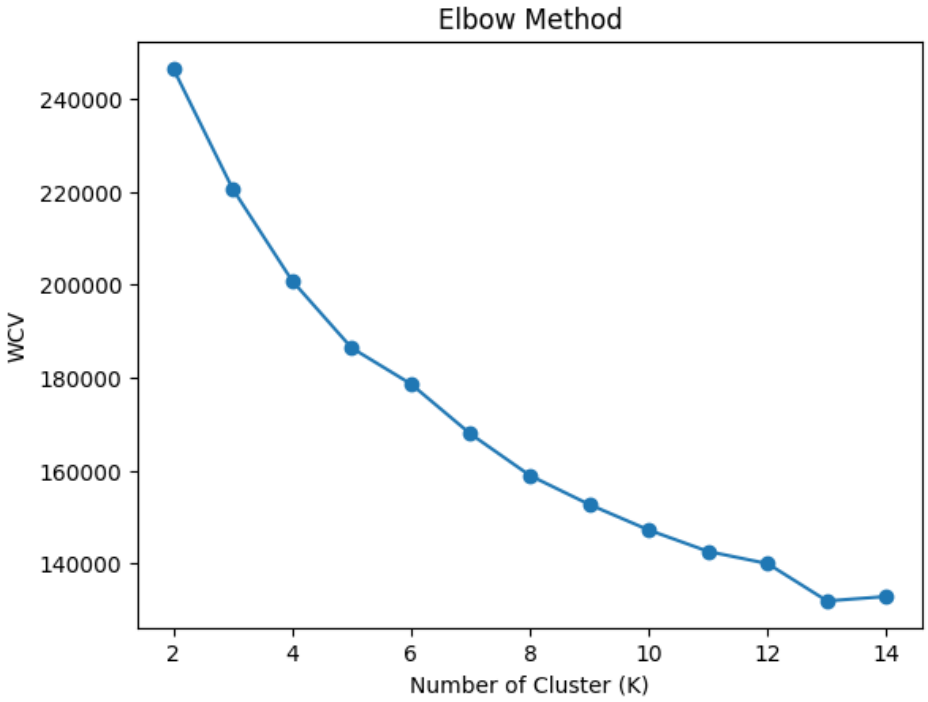
$$Aroon\ Down = \frac{25 - periods\ since\ 25\ period\ Low}{25}$$

The trade signal for the Aroon indicator is the crossover between the Aroon up and the Aroon down. When the Aroon Up line crosses above the Aroon Down line it is a buy signal and when the Aroon Down line crosses above the Aroon Up line it is a sell/short signal.

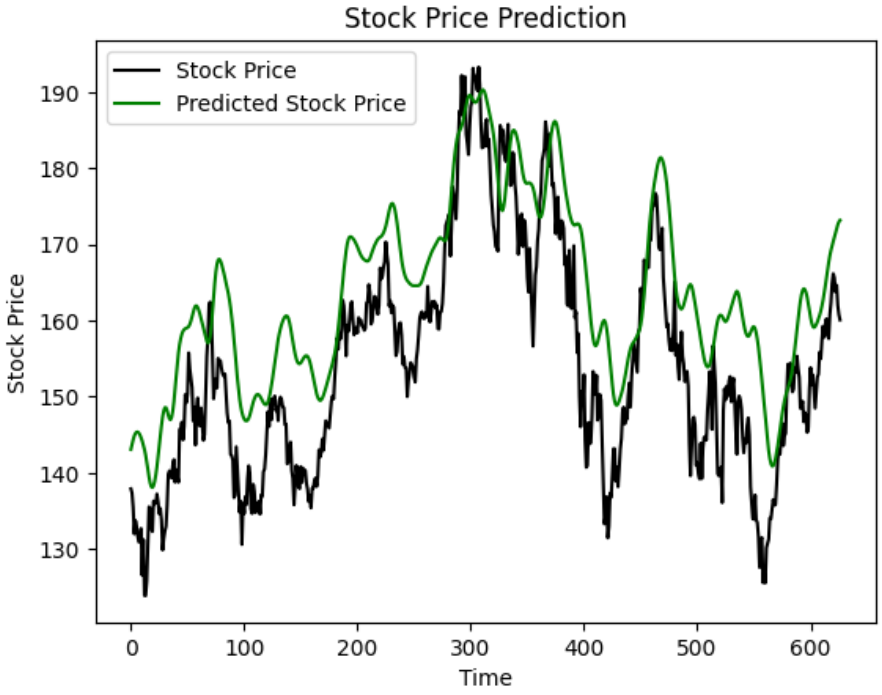
Appendix 2.1 Exploratory Data Visualization





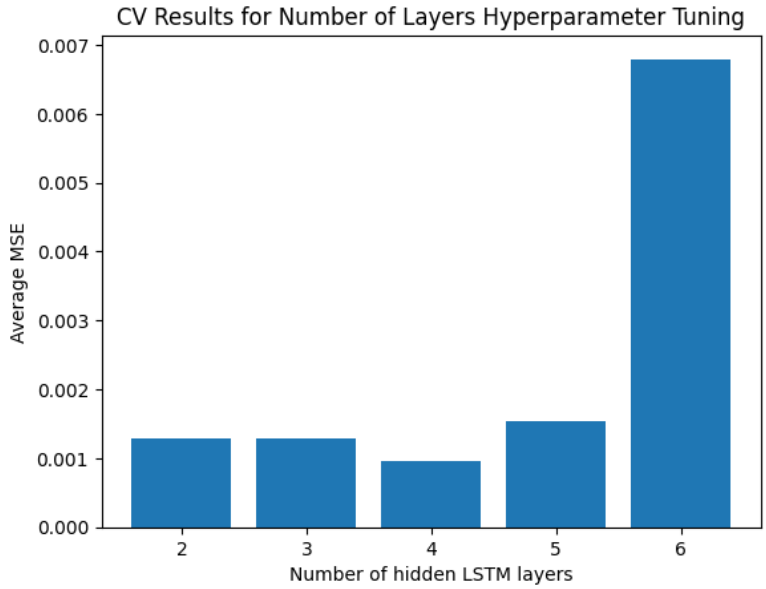


Appendix 3.1 Initial Model Results

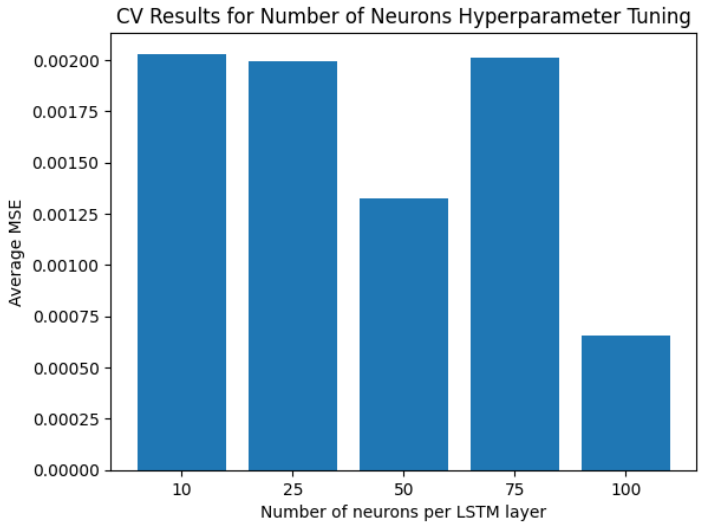


Appendix 3.2 Results from Hyperparameter Tuning

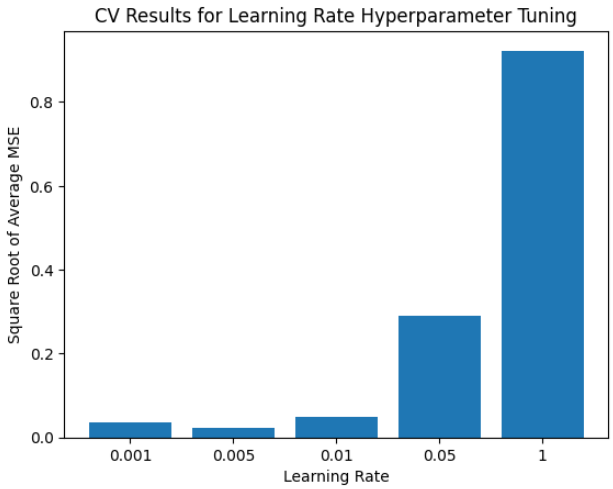
Appendix 3.2.1 Number of Layers Hyperparameter Tuning



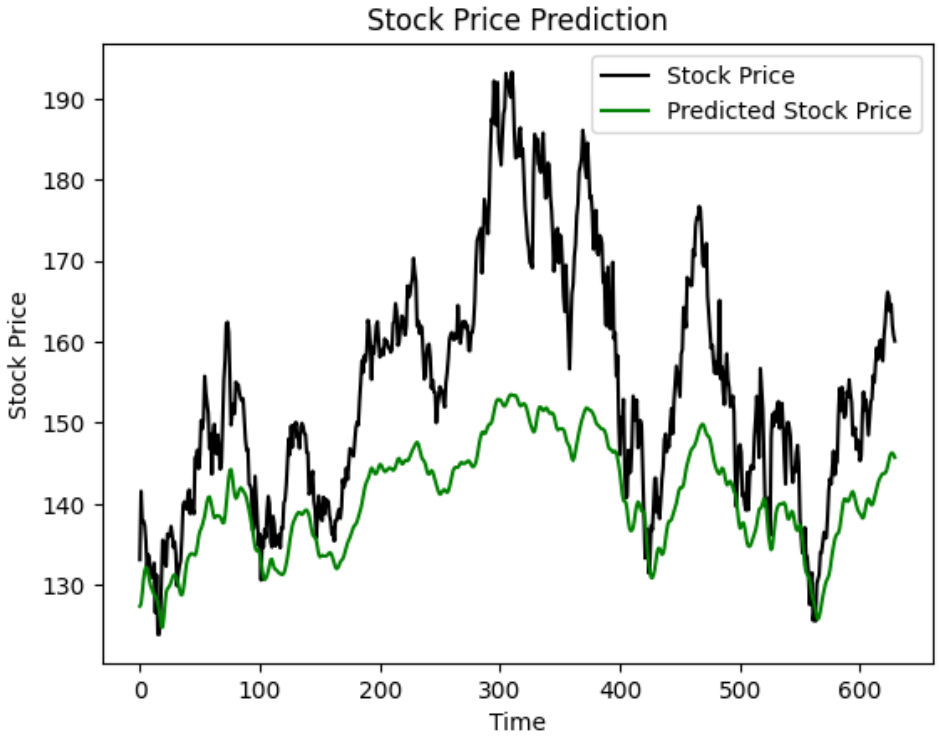
Appendix 3.2.2 Number of Neurons Hyperparameter Tuning



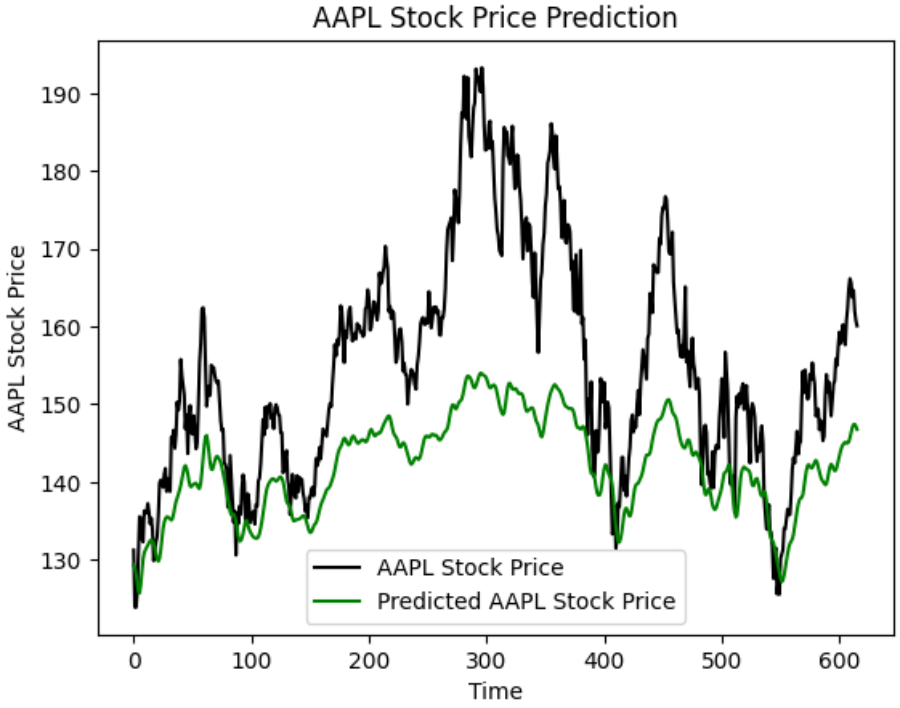
Appendix 3.2.3 Learning Rate Hyperparameter Tuning



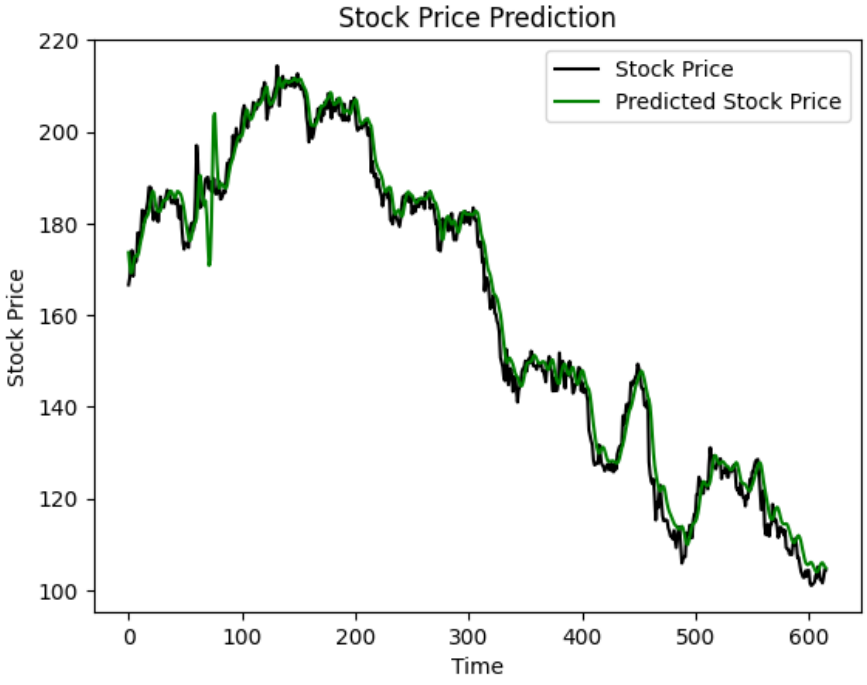
Appendix 3.3 "Vanilla" Model Using Only Closing Price



Appendix 3.4 Model Using Closing Price and First Two Principal Components



Appendix 3.5 Model Validation on MMM



References

Chen, J., 2022. Technical Indicator Definition. [online] Investopedia. Available at: <https://www.investopedia.com/terms/t/technicalindicator.asp> [Accessed 12 April 2023].

Desjardins, J., 2022. 12 Types of Technical Indicators Used by Stock Traders. [online] Visual Capitalist. Available at: <https://www.visualcapitalist.com/12-types-technical-indicators-stocks/> [Accessed 12 April 2023].

Hayes, A., 2022. Know Sure Thing (KST). [online] Investopedia. Available at: <https://www.investopedia.com/terms/k/know-sure-thing-kst.asp#:~:text=The%20Know%20Sure%20Thing%20(KST)%20is%20a%20momentum%20oscillator%20intended,for%20overbought%20or%20oversold%20conditions.> [Accessed 12 April 2023].

Hayes, A., 2022. Relative Vigor Index (RVI) Definition. [online] Investopedia. Available at: <https://www.investopedia.com/terms/r/relative_vigor_index.asp#:~:text=The%20Relative%20Vigor%20Index%20(RVI,simple%20moving%20average%20(SMA).> [Accessed 12 April 2023].

Hayes, A., 2022. Stochastic Oscillator. [online] Investopedia. Available at: <https://www.investopedia.com/terms/s/stochasticoscillator.asp#:~:text=A%20stochastic%20oscillator%20is%20a,moving%20average%20of%20the%20result.> [Accessed 12 April 2023].

Mitchell, C., 2022. Commodity Channel Index (CCI) Definition and Uses. [online] Investopedia. Available at: <https://www.investopedia.com/terms/c/commoditychannelindex.asp> [Accessed 12 April 2023].

Mitchell, C., 2022. What Is the Aroon Indicator?. [online] Investopedia. Available at: <https://www.investopedia.com/terms/a/aroon.asp#:~:text=The%20Aroon%20indicator%20is%20a,lows%20over%20a%20time%20period.> [Accessed 12 April 2023].

Mitchell, C., 2022. Coppock Curve Definition and Uses. [online] Investopedia. Available at: <https://www.investopedia.com/terms/c/coppockcurve.asp#:~:text=The%20Coppock%20Curve%20is%20a,of%20change%20for%20the%20index.> [Accessed 11 July 2022].

Schaap, C., 2022. ADX: The Trend Strength Indicator. [online] Investopedia. Available at: <https://www.investopedia.com/articles/trading/07/adx-trend-indicator.asp> [Accessed 12 April 2023].

Segal, T., 2022. Disparity Index Definition. [online] Investopedia. Available at: <https://www.investopedia.com/terms/d/disparityindex.asp#:~:text=The%20disparity%20index%20is%20a,the%20value%20as%20a%20percentage.> [Accessed 12 April 2023].

Delineation of Work

George Mandl: Introduction, Appendix Literature Review, Data Preprocessing, and Methods, Results, and Conclusions relating to Recurrent Neural Networks

Maheep Jain: Abstract, Methods, Results, and Conclusions relating to Support Vector Machines