**Introduction**

Stock market prediction is a challenging problem as there are many factors that affect stock market prices like industry performance, company news and performance, investor and social media sentiment and economic factors, this research paper explores the predictability of stock market trends using deep convolutional neural networks and candlestick charts. This outcome is used to design a decision support framework that is used by traders to provide suggested indications of directions of future stock prices.

The proposed method of work in this research paper is to use represented candlestick charts of Taiwan and Indonesian stock markets to predict the price movement. Three trading period times were used to analyse correlation between the time periods and stock market movements with and without daily volume stock data. These experiments were conducted in two kind of image sizes to analyse the correlation of hidden pattern in various image sizes. Thereafter our dataset will be feed as input to a CNN learning algorithm. The goal is to analyse the correlation of some parameters such as period time, image size, feature set with the movement of stock market to check whether it will be going up or going down in the next day.

**Data collection**

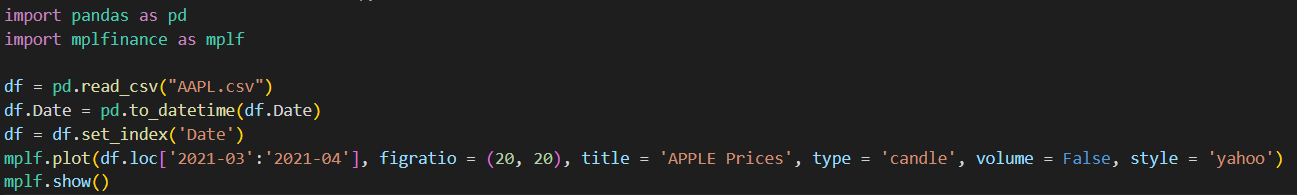
In the paper, the model was trained and evaluated on 2 different stock markets, Taiwan and Indonesia. For my implementation, I’ll be using the stock prices of Tata Motors, Apple, or more stocks as required over the years.

**Data pre-processing**

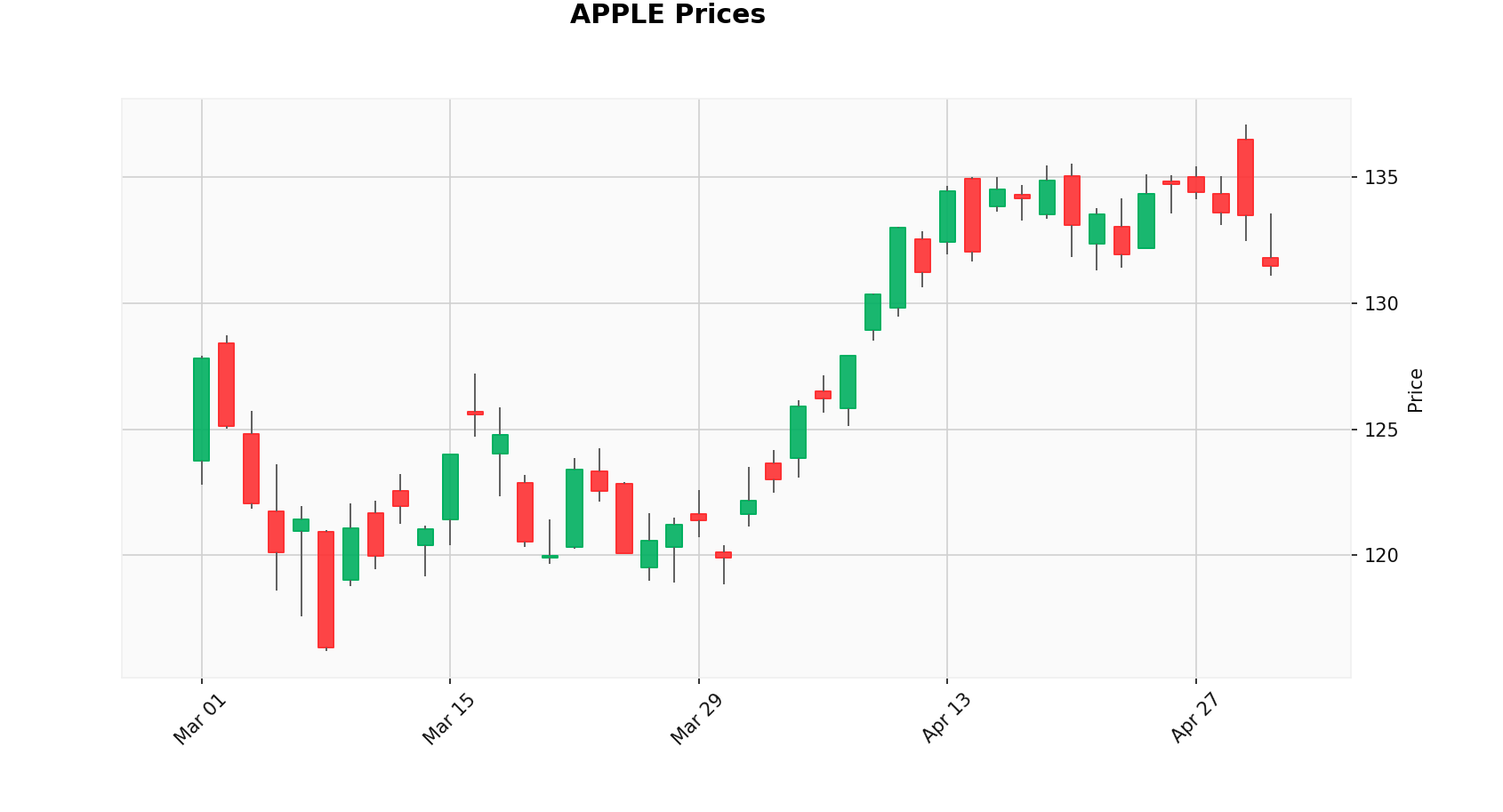
The data collected is to be converted into candlestick charts using Matplotlib, and the data is divided based on three period times, i.e., 5, 10, and 20 trading days data. Besides the time period, data is also divided with and without volume indicator.

A candlestick chart is like a combination of a line and bar chart with each bar representing the open, close, high, low prices. If the opening price is higher than the closing price, then the real body will fill in red colour. Otherwise, the real body will be filler in green colour. The upper and a lower shadow represent the high and low-price ranges within a specified time period. The candlestick is called bullish candlestick when the close is greater than the open. Otherwise, it is called bearish candlestick.

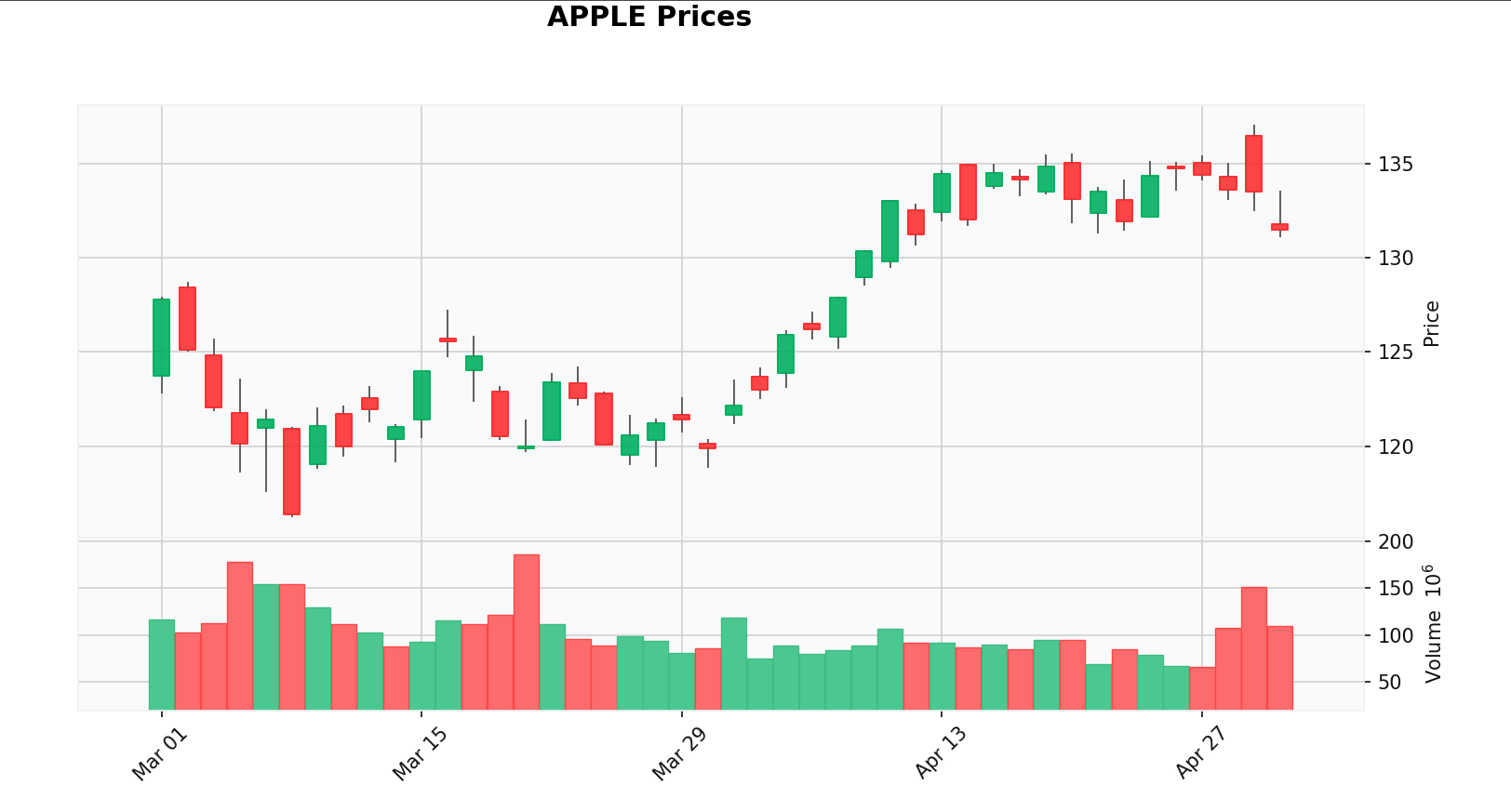
Below is a code to plot a candlestick chart for the stock prices of Apple for a given month. These charts can be saved locally using the savefig() method. It will take around a day to complete plotting the entire dataset according to different time periods.



Without volume indicators,

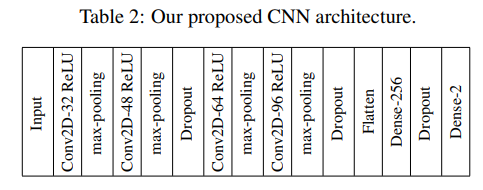
****

With volume indicators,

****

**Learning algorithm**

In my implementation, I’ll be focusing on the proposed CNN architecture in the paper. CNN is a feed forward artificial neural network which includes an input layer, hidden layers, and output layers. The hidden layers usually consist of pooling layers, convolution layers, and full connected layers. It is similar to ordinary neural networks, the difference being in convolutional layers which use a convolution operation to the input then transfer the result to the next layer. The proposed CNN architecture consists of 4 layers of max pooling 2d, 4 layers of convolutional 2d, and 3 dropouts.

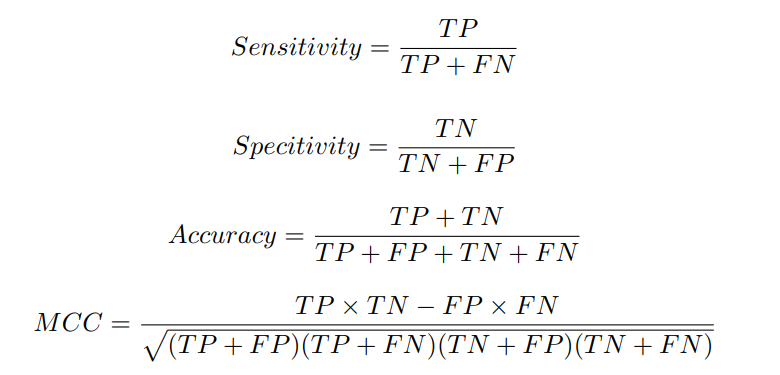


Typically, training a CNN model feeding our dataset to the network and adjusting the parameters accordingly to obtain the minimum loss function. This usually consists of two phases, a forward phase, where the input is completely passed through the network, and a backward phase where the gradients are back propagated and the weights are updated. During the forward phase, each layer will cache any data that will be needed for the backward phase, and during the backward phase each layer receives the gradient of loss with respect to output and outputs the gradient of loss with respect to input. Being the most intensive part, it could take around 3 – 4 days.

**Performance evaluation**

There are some statistical measures of the performance evaluation like sensitivity, specitivity, accuracy and Matthew’s correlation coefficient (MCC).

In general, TP is true positive or correctly identified, FP is false positive or incorrectly identified, TN is true negative or correctly rejected and FN is false negative or incorrectly rejected.



**Experimental results**

Then we have to perform classification based on the CNN learning algorithm and then evaluate its performance.

In the paper for both the Taiwan and Indonesia datasets, the method using CNN model with longer trading days period without volume indicator achieved the best result based on the values obtained for sensitivity, specitivity, accuracy, and MCC.

Testing the model independently for certain stocks also showed that the CNN model with longer trading period had the best results.

This model was found to perform better than other state of the art models already in place. Analysing these results could take me around a day or two.