Time-series Stock Price Predictive Analytics

# **Introduction**

The stock market began in 1792 at Wall Street, New York with a few stockbrokers on the street (Posner, 2020). This platform has allowed investors to buy and sell shares of a public company or better well known as stocks. Other investments include bonds and commodities such as gold and oil. The main benefits from an investor’s perspective are to stay ahead of inflation with compound interest and personal financial growth. With pros come cons and the main disadvantages are competition and risk (Amadeo, 2020).

On average “90% of traders fail to make money when trading in the stock market (Gillham, 2021)." As a result, traders tend to lean towards safer well-known options such as the S&P 500. The S&P 500 is based on the top 500 companies in America and is considered a benchmark for annual stock market returns. Many investors know that the long-term investment into the S&P 500 has an average return on investment of 10% and as a benchmark investor’s goal is to outperform it. With technology growing, algorithm trading has found a place in the stock market and like everyone their goal is to outperform the S&P500 using data driven technologies. Using data from the market and time-series predictive analytics can a model be built to outperform the S&P500 for individual investors.

# **Literature Review**

Recently the stock market has seen a major crash in 2020 due to the COVID-19 pandemic. This has caused the volume of stock trading to go up and lead the market to be flooded with new cash resulting in a quick recovery. With emerging technology, the stock market is adapting towards algorithm trading. “Algorithmic trading is the process of buying or selling a security basing on some pre-described set of rules tested on historical data” (Turner, 2019). This technology was used to help Tesla buy 1.5 billion dollars’ worth of bitcoin at a good average price. The popularity of algorithm trading grew massively over the years to 85% of market volume in 2012 (Turner, 2019). This method of trading has proven to be beneficial and eliminates poor trading ideas (Turner, 2019).

# **Dataset**

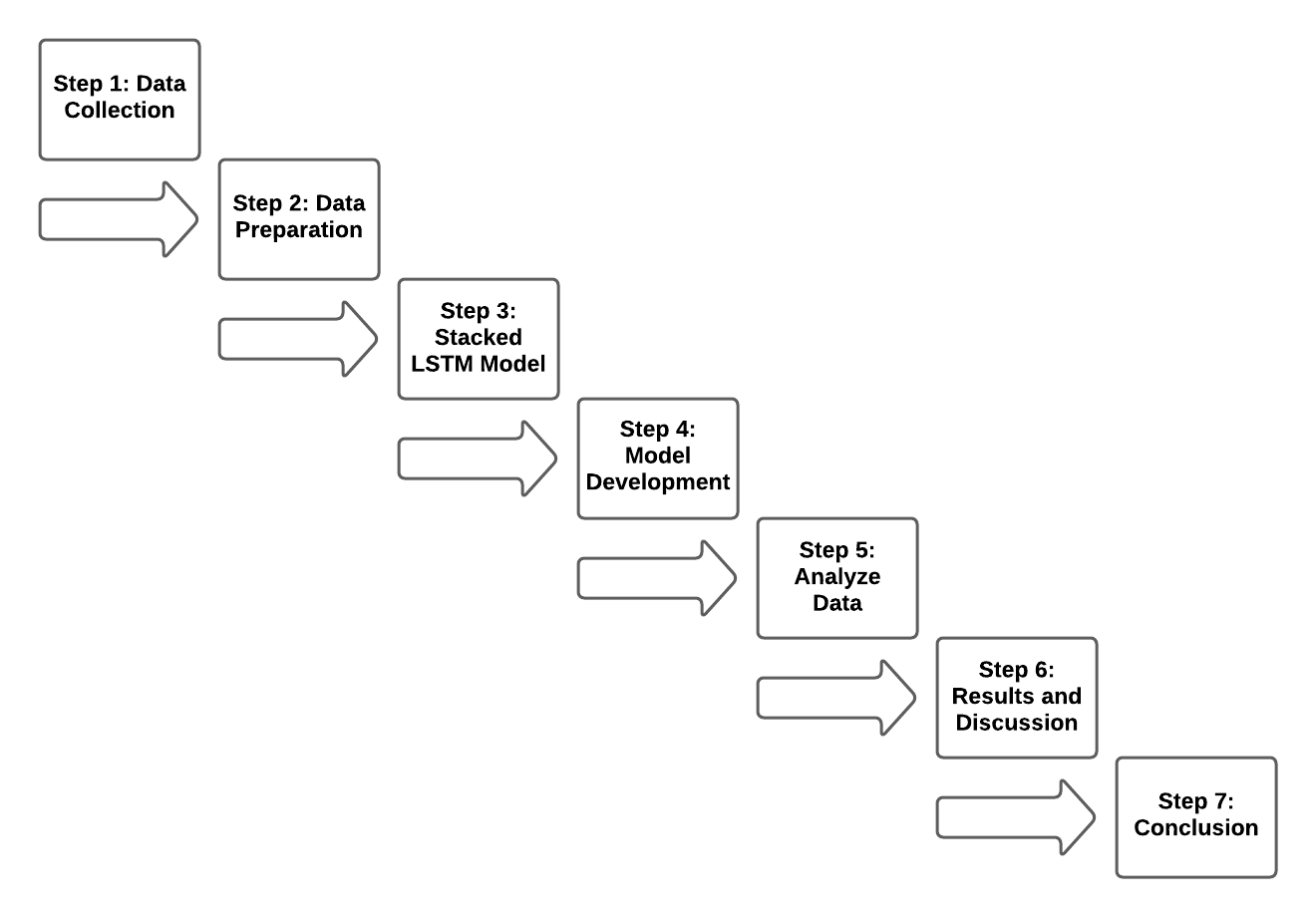
The dataset used in this research is sourced from Yahoo Finance. It is from the SPDR S&P 500 ETF Trust which replicates the S&P 500 on a lower expense ratio and is traded under the ticker SPY. It includes over 7075 instances and 6 attributes. The data attributes outline the date, volume, and various prices.

**Data Collection**

* Date: Year-Month-Day
* Open – Opening Price
* High – Day High Price
* Low – Day Low Price
* Close – Closing Price
* Volume – Volume of Stock

Each of the attributes will be used to create a time series stacked LSTM model to predict the future trends of the stock market.

# **Approach**



**(Fig. 1) Approach**

## **Step 1: Data Collection**

* Import data from Yahoo Finance.
* Understand the attributes.
* Visualize the correlation to find connections and attribute importance.

## **Step 2: Data Preparation**

* Check for missing values and outliers.
* Divide dataset into training and test sets.

## **Step 3: Stacked LSTM Model**

* Using time-series predictive analytics to create a stacked LSTM Model.

## **Step 4: Model Development**

* Predict the test data and the future stock performance.

## **Step 5: Analyze Data**

* Compare the model’s performance with actual data.

## **Step 6: Results and Discussion**

* Display results.
* Analyze the research findings.
* Answer the project question.

## **Step 7: Conclusion**

* Summarize the research and project question.

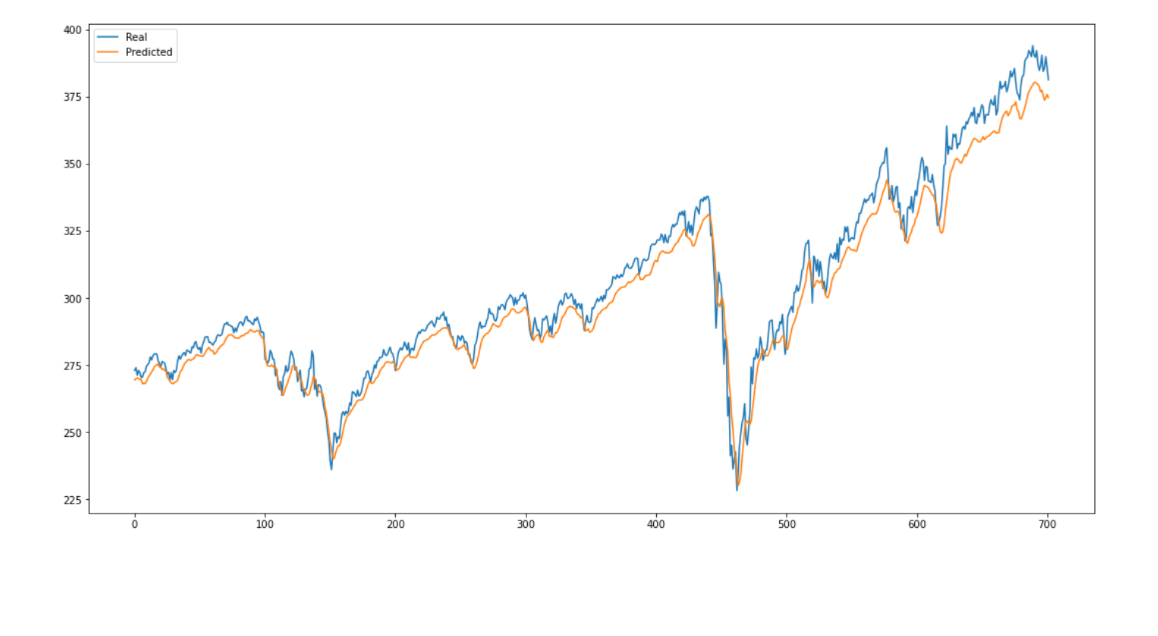
# Results

In figure 2, we divide the dataset in to training and test sets for the model. Here we can see the shape of these data frames. Indicating a ratio of 9 to 1 resulting in 6322 instances for training and 703 instances for testing. The dataset has a second dimension which includes the previous 50 data points to help predict the next traded price. The third dimension holds the open, high, low, close, and volume data points.

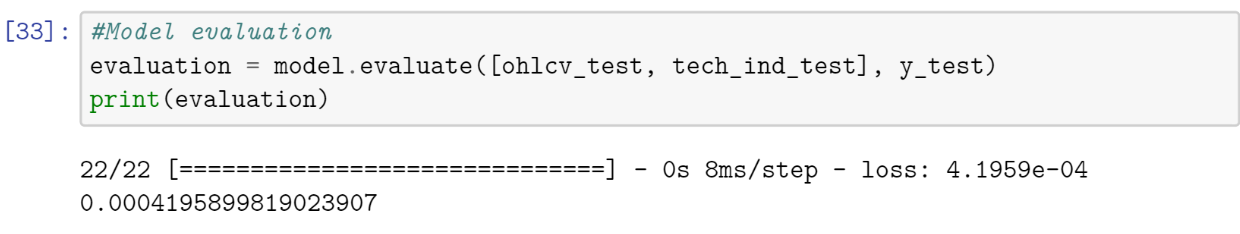


**(Fig. 2) Training and Testing Data Shapes**

As illustrated in figure 3 we can see the last 700 days of stock closing prices. The blue indicating the actual prices and the yellow showing the LSTM models predicted prices. We can see that the model can accurately predict the trend of the stock over a longer period.



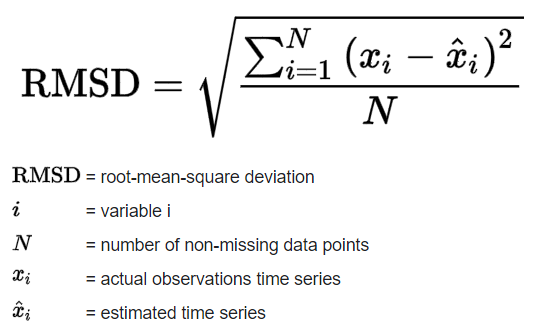
**(Fig. 3) Real vs Predicted Stock Prices**



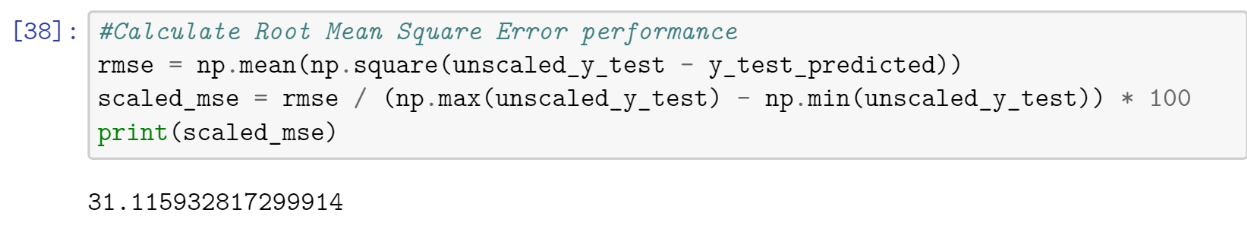
**(Fig. 4) Model Evaluation**

The given model evaluation indicates a loss of 0.04%. This is extremely good for a machine learning algorithm and demonstrates a high accuracy in the predictive model.

A root mean square error performance is a good indicator of the performance of the predicted model to the actual values. Calculated with the following function it uses the difference of the actual and predicted divided by the number of data points.



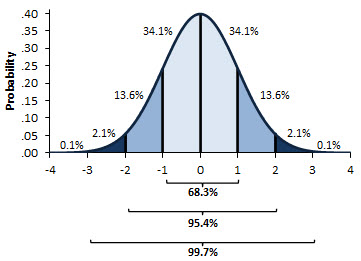
**(Fig. 5) Root Mean Square Error Formula**



**(Fig. 6) Model’s Root Mean Square Error**

The models root mean square error performance has a 31-standard deviation of residual. Based on the normal curve in figure 7 this tells us that:

* 68% of the data points will fall within ±0.31 standard deviation of our model.
* 95% of the data points will fall within ±2(0.31) standard deviation of our model.
* 99% of the data points will fall within ±3(0.31) standard deviation of our model.



**(Fig. 7) Normal Curve**

Considering the LSTM model was taught based on one dataset the root mean square performance is exceptional. If this was expanded to other datasets, we can expect the model to learn and predict prices more accurately. In theory we would see the root mean square performance result in a number less than 31 indicating a better and consist prediction of stock price.

# Conclusions

In conclusion, the research constructed using data from the market and time-series predictive analytics; a model can be built and trained to accurately predict the long-term prices. Remember each predicted value is based on the previous 50 days and it would be nearly impossible to predict prices without this information. Using this model, individual long term stock performance can be determined to find the annual return that outperforms the S&P500’s average of 10% in annual return.

Individual investors on average have never outperform the S&P500, by trying to time the market. A future and in progress implementation of an algorithm to accurately determine buying and selling points based on the model predictions. This will allow investors to time the market and outperform the long-term annual return. In theory this adaptation of the program will outperform long-term investors but lacks a random variable for outside noise. This can be adopted using the Monte Carlo simulation used in real world examples to introduce bull and bear cases for financial predictions. This implementation would work very similar to algorithm trading.

# **Citation**

Posner, J. (2020). Explained The Stock Market [Video]. Netflix.

Amadeo, K. (2020). Pros and Cons of Investing in Stocks. Retrieved 15 February 2021, from <https://www.thebalance.com/stock-investing-for-the-individual-investor-3306182>

Turner, C. (2019). The Future Of Algorithmic Trading. Retrieved 15 February 2021, from <https://www.experfy.com/blog/fintech/the-future-of-algorithmic-trading/>

Gillham, D. (2021). Trading the Stock Market – Why Most Traders Fail. Retrieved 30 March 2021, from <https://www.wealthwithin.com.au/learning-centre/share-trading-tips/trading-the-stock-market#:~:text=Anyone%20who%20starts%20down%20the,per%20cent%20make%20money%20consistently>

Gillham, D. (2021). Trading the Stock Market – Why Most Traders Fail. Retrieved 30 March 2021, from <https://www.wealthwithin.com.au/learning-centre/share-trading-tips/trading-the-stock-market#:~:text=Anyone%20who%20starts%20down%20the,per%20cent%20make%20money%20consistently>.