

**Forecasting SPDR
S&P500 ETF (SPY)
Buy and Sell Trend
Using Logistic
Regression**

**By Karim
Elbahloul**

Abstract- Predicting buy/sell trends is quite a desirable research topic. Stock brokers currently rely on predictive models and analysis to help them make better financial decisions on the stock market exchange. This research will discuss the results when using Logistic Regression to forecast a buy/sell strategy where buy(1) and sell(-1). The following methods that will be utilized is only the Logistic Regression model that would be outputted.

INTRODUCTION

For the purpose of this research, Logistic Regression would be examined and studied to forecast the trend whether it is heading towards buy or sell. Historical dataset would be used to extract past stock closing prices derived from Yahoo Finance. The stock that would be conducted for analysis and forecast would be the SPDR S&P500 ETF (SPY). This is an Exchange-Traded-Fund (ETF) that holds 500 companies. Logistic Regression would assist this research to examine closely about the trend and behavior. This paper will extensively look at Logistic Regression as well as accuracy score that would show how accurate is the model.

BACKGROUND

Using Machine Learning methods, one way to forecast market direction and buy/sell trends is to utilize Logistic Regression.

Logistic Regression is one of the common regression models used for machine learning. As a forecast model, it is like linear regression except that both have different output models. Logistic Regression will assist in the deciding whether to buy or sell the stock based on the given historical dataset that would be tested.

Accuracy score would be utilized as well, this will show how accurate is the outputted model, as stated before this will depend on the historical dataset that is used in this research.

Deploying the above-mentioned machine learning methods would assist this research by extracting historical datasets as well as examining the predictive model as a result of the dataset that is used.

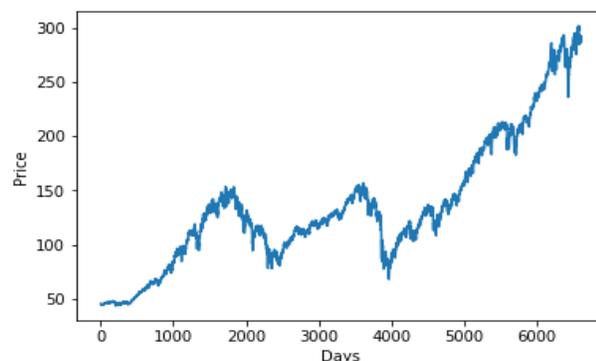


Figure 1: Historical SPDR S&P500 ETF (SPY) Stock Market Data. 06/07/1993 – 08/29/2019.

DATASET DESCRIPTION

For the purpose of this research, the dataset from SPDR S&P500 Exchange-Traded-Fund called SPY will be analyzed. The SPDR S&P500 (SPY) (Formally Standard & Poor's 500 Index) is a market-capitalization-weighted index of the 500 biggest U.S. publicly-traded companies on the New York Stock Exchange (NYSE). The index is widely regarded as the best single measure of large-cap U.S. equities. Furthermore, it is an excellent gauge for the overall performance of the stock market.[1]

The SPDR S&P500 ETF (SPY) dataset is publicly available on Yahoo Finance. It consists of historical daily opening and closing prices along with a few other additional market parameters. Figure 1 shows the daily closing of the SPY from June 7, 1993 to August 29, 2019.[2]

EXPERIMENTAL ENVIRONMENT

The following python libraries were used in conducting the logistic regression and accuracy analysis:

- Numpy: used to perform matrix operations, such as flip, reshape, and create random matrices.[3]
- Pandas: used for data manipulation and analysis.

- Pandas-DataReader: used for the purpose of extracting historical data from Yahoo Finance.
- Talib: used for technical analysis for stock trading.
- Sklearn: is a machine learning python library for regression, classification, and other machine learning algorithms
- Matplotlib: Used for plotting the graphs for the actual time-series as well as predicted trends.

METHODOLOGY

The machine learning tools used in this study are Logistic Regression and Accuracy Model.

First, Logistic Regression is used as a forecast analysis to describe the data as well as using binary values in which case buy(1) and sell(-1). This method is useful to forecast buy or sell trends.

Lastly, Accuracy score is a classification score by examining the buy and sell trend rate.

Accuracy scores are utilized to tell how precise the model can learn from the actual data as well as being able to correctly follow the actual stock market model.

RESULTS AND DISCUSSION

LOGISTIC REGRESSION-



Figure 2: Logistic Regression, SPDR S&P500 (SPY) Stock Trading Trend. Sell(-1) and Buy(1)

The results of the logistic regression showed it had fluctuated overtime. Furthermore, The forecasted trading strategy (blue line) showed it had followed the actual stock market data (green line). Thus, according to the outputted model it is trending toward a forecasted buy from the trading strategy. First, the blue line showed from the start it had learn by following the actual model until it went toward a downward trend, as result a sell prediction. Second, after the downward trend it has predicted a buy trend because the predictive line kept going in an upward trend which could resulted in positive news or stock brokers kept buying the stock overtime to which came to the point that an upward trend suddenly took that path. the stock.

ACCURACY SCORE-

	precision	recall	f1-score	support
-1	0.47	0.15	0.22	610
1	0.54	0.86	0.66	708
micro avg	0.53	0.53	0.53	1318
macro avg	0.50	0.50	0.44	1318
weighted avg	0.51	0.53	0.46	1318

Figure 3: This is the classification report which shows the accuracy score and accuracies for both buy and sell.

The results of the overall accuracy score showed the model has achieved a 52.8% or 53%. This score is not a bad score, but it is saying the model is fine and that it is learning. The classification report shows that the accuracy score findings show that the buy achieved a 66% which is why the prediction line kept going towards an upward trend. However, the sell has achieved a 22% which is low and not a good score to earn. Findings show that the sell prediction could not be used and would be rejected for such a low score presentation. Overall, scores that show low accuracy should not be considered as a prediction to rely and should be rejected.

CONCLUSION

This research has assisted in making predictions to forecast buy or sell approaches. Stock trading is intense especially in the financial sector. Overall, predicting the stock market's direction is unpredictable because it fluctuates overtime.

REFERENCES

[1]

“Novel Deep Learning Model with CNN and Bi-Directional LSTM for Improved Stock Market Index Prediction”
<https://vermaabhi23.github.io/publication/2019CWC1.pdf> [Accessed August 30, 2019]

[2]

“SPDR S&P500 (SPY) Stock Historical Prices & Data.” *Yahoo Finance*. [Online]
<https://finance.yahoo.com/quote/SPY?p=SPY>

[3]

Numpy [Online], (2018). Available:
<http://www.numpy.org/>. [Accessed August 30, 2019]