Predicting the Future Price of S&P 500 index

David Meiners
Summer 2021
https://github.com/dmonies/dsc680Portfolio

Abstract

Investing in financial markets is essential for investors to maintain and increase their wealth. This is because inflation can average around 3% each year and failure to invest wisely can lead to loss of wealth. The S&P 500 is one of the best indices to invest in as it tends to double every ten years and represents the composite value of the 500 largest companies in the United States. The amount of time a person has until retirement is dependent on how they will invest in financial markets with those nearing retirement having less margin for error. To combat loss of wealth, a prediction model can be leveraged to understand if one should invest in the S&P index fund, sell out their position, or hold current positions. The following case study will ingest the historical closing values of the S&P 500 and make a prediction on what the future values of the index will be by leveraging deep learning techniques.

Introduction

The objective of this research is to understand if an investor should buy, hold, or sell investments in the S&P 500 index. There are several companies which make these recommendations although the methods leveraged to make these assessments vary. The case study will be leveraging historical closing prices of the S&P 500 index to make predictions on the future price valuation of the index fund. A deep learning model will be leveraged to make these predictions. The Prophet deep learning model is an open source package developed by Facebook that makes time series predictions based on two inputs.

Data Sources and Cleaning

The data source which was leveraged to conduct the analysis included the historical metrics associated with the S&P 500 index derived from Yahoo. The Yahoo dataset provided a complete historical view of the index for the last 90 years and did not contain any missing or inaccurate data points. The metrics which were available are the date, closing price, opening price, trading volume, low price, and high price. The data did not require any cleaning other than transferring the data points from the website into a CSV file and then appending the data types to match the field values.

Data Preparation

The data was prepared by ingesting csv file into Python and then creating a new data frame that contained the date and closing price field values. This is because the model leveraged only required two data inputs to be fed into the model for the time series forecasting.

Prediction Model

The prediction model that was leveraged to make the predictions was the Neural Prophet open-source package developed by Facebook. This model has many advantages as it is easy to implement by only requiring y and date values as inputs as well as many features to effectively handle outliers, missing values, and seasonality. The package is still in development however and has some limitations specific to

what date intervals can be input into the model. The package also allows for tuning parameters to enhance the fit of a model so that it is not under or overfit when making predictions.

Research Questions

The distinct research questions I wanted to ask after compiling the data is to understand if a investor should sell, hold, or buy into the S&P 500 index based on current trends.

These questions include:

- Will the S&P increase in value over the next two months?
- Based on current trends should an investor buy, hold, or sell their current investment?

Model Tuning

To tune the model, the default parameters were first leveraged to see if the model was underfit when leveraging Neural Prophet. As can be observed below, these default parameters resulted in an underfit model with the fit being in a straight line and not properly considering the variance which can occur in the real world.
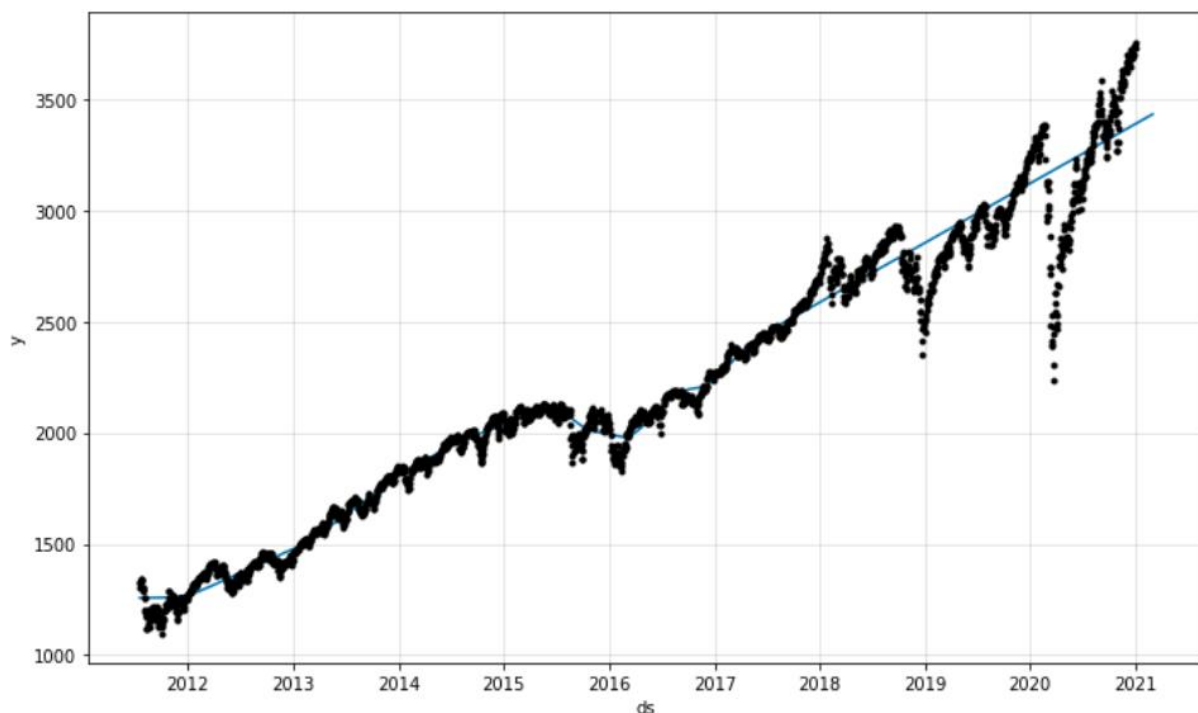

This resulted in a model which made predictions that were far below the current value of the index which would cause a false flag to sell the fund.
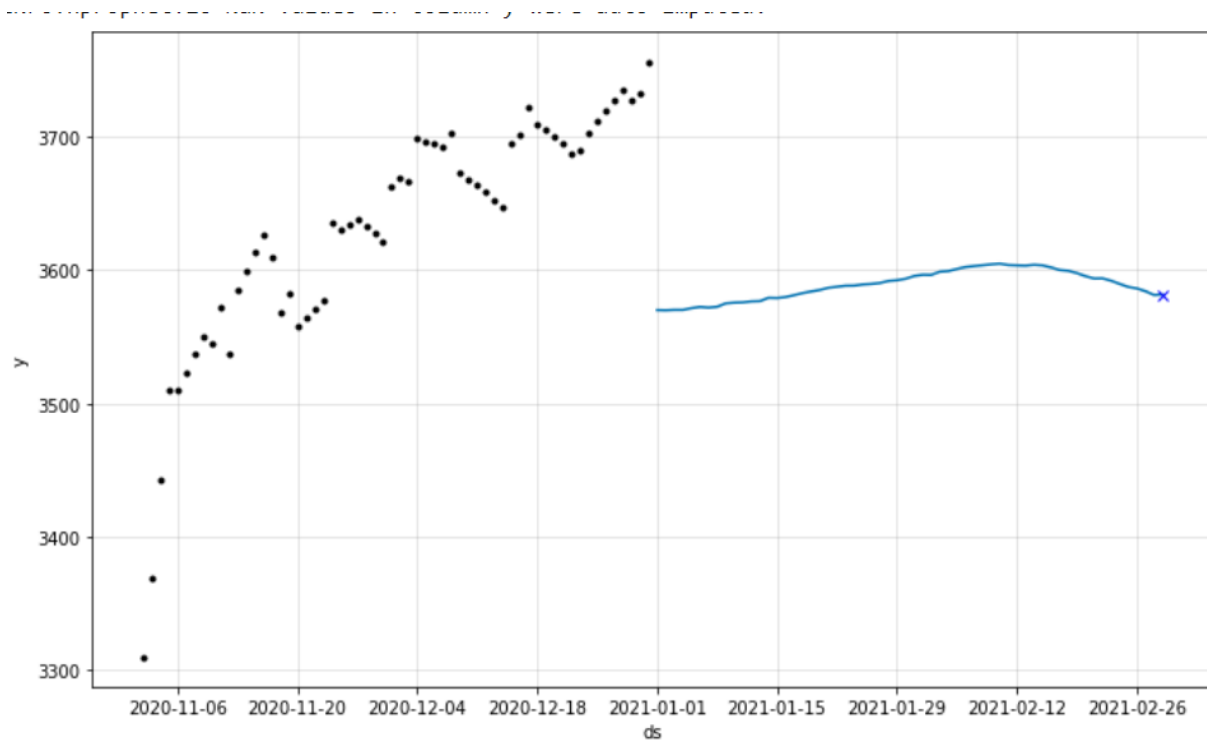
After observing that the model was underfit, the parameters of the model were tuned to have a better fit by leveraging an autoregressive neural network approach and implementing seasonality. This resulted in the model having a better forecast and reflect a real-world outcome to make predictions. These changes resulted in the fit indicated in the below snapshot.

Although this model appears to be slightly overfit, it is a much more reliable model when observing the predictions that were output in the below forecast.
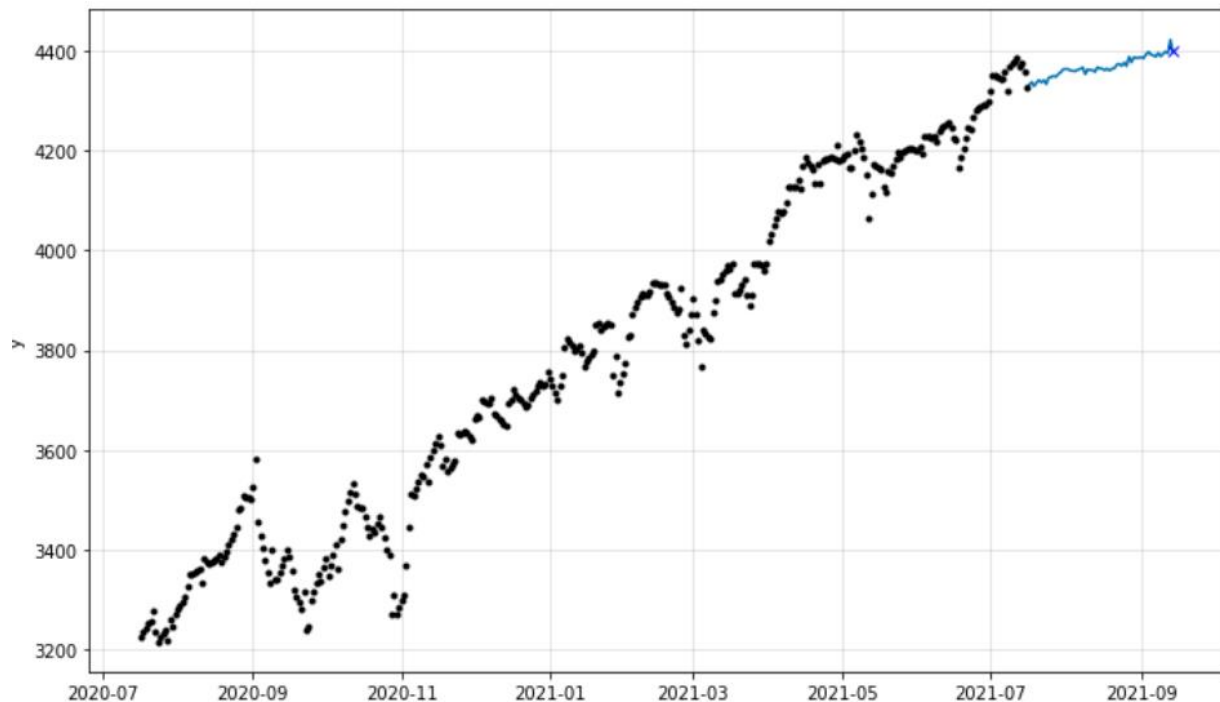
Results

Will the S&P increase in value over the next two months?

As you can see in the forecast snapshot, the model predicts a continued upward trend in index fund valuation over the next two months. The final estimated value that is forecast is 4400 points for the S&P 500 index compared to the current value of 4327. In other words, an estimated gain of 1.5% can be expected for those who have invested in the S&P 500.

Based on current trends should an investor buy, hold, or sell their current investment?

A 1.5% increase in valuation is a modest gain but is on par with the index increasing on average of 10% annually although falling slightly short at 9%. Based on this information, there may be better opportunities to invest in during this period that would have larger gains. Those with a low-risk investment profile should consider buying into the index during this time, however. The forecast also indicates those who have already invested in the index should hold and not sell as the estimated value of the index is expected to increase.

Considerations

Although this analysis resulted in a recommendation based the most recent data that is available, things can change very rapidly. This is because there are many factors that can change the predicted outcome in a very short time. Investors who are leveraging forecast models should revisit the forecast daily and recompile it to have the best information based on the current trends being observed. If trends continue as they are, the predicted outcomes are highly likely, but a catastrophic event can sink financial markets rapidly resulting in significant losses especially when factoring in inflation at 3%.
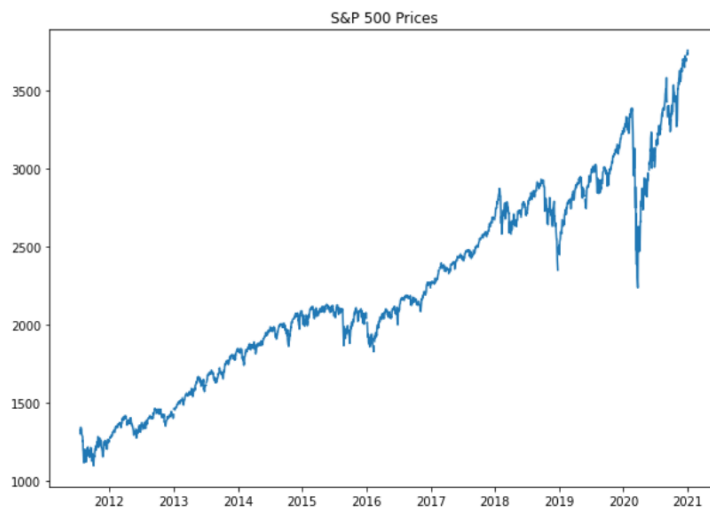
Conclusion

For investors to maintain their wealth and guarantee a stable financial future, they must invest wisely to grow their wealth. With the S&P doubling in value every ten years, it is a great fund to invest in while minimizing risk due to diversifying in the 500 largest companies in the United States. Based on the analysis and answers to the research questions, an investor should feel confident in investing in the S&P 500 based on current trends and forecasts for the next 60 days. The Neural Prophet model is a very reliable way to effectiuvely tune a model to output a reliable forecast. The forecast anticipates a 1.5% increase in value which is on par with the value increasing 10% annually. Investors should hold onto any investments in this fund and not sell over the next 60 days based on current estimations. With any financial asset, there are many unforeseen events that can impact the value of an asset, so it is important to keep a careful eye on the forecast based on current trends. I would recommend recompiling the model daily to prevent catastrophic loss due to an event which the model was not able to observe in its current analysis.
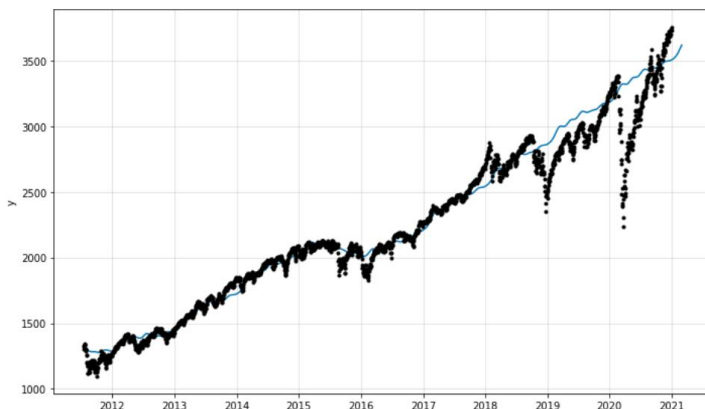
Appendix

The appendix contains additional visualizations regarding the Neural Prophet Model and Historical S&P 500 index values.

Historical S&P 500 Index Composite Valuation

Intermediate Model Fit by adjusting seasonality parameters

Neural Prophet Model Details (https://facebook.github.io/prophet/)

Accurate and fast.

Prophet is used in many applications across Facebook for producing reliable forecasts for planning and goal setting. We've found it to perform better than any other approach in the majority of cases. We fit models in Stan so that you get forecasts in just a few seconds.

Fully automatic.

Get a reasonable forecast on messy data with no manual effort. Prophet is robust to outliers, missing data, and dramatic changes in your time series.

Tunable forecasts.

The Prophet procedure includes many possibilities for users to tweak and adjust forecasts. You can use human-interpretable parameters to improve your forecast by adding your domain knowledge.

Available in R or Python.

We've implemented the Prophet procedure in R and Python, but they share the same underlying Stan code for fitting. Use whatever language you're comfortable with to get forecasts.

References

Budiharto, W. (2021, March 11). *Data science approach to stock prices forecasting in Indonesia during Covid-19 using Long Short-Term Memory (LSTM)*. Journal of Big Data. https://link.springer.com/article/10.1186/s40537-021-00430-0.

Vijh, M., Chandola, D., Tikkiwal, V. A., & Kumar, A. (2020, April 16). *Stock Closing Price Prediction using Machine Learning Techniques*. Procedia Computer Science. https://www.sciencedirect.com/science/article/pii/S1877050920307924.

Predicting Stock Prices Using Machine Learning Techniques. IEEE Xplore. (n.d.). https://ieeexplore.ieee.org/abstract/document/9358537.

Enhancing Profit by Predicting Stock Prices using Deep Neural Networks. IEEE Xplore. (n.d.). https://ieeexplore.ieee.org/abstract/document/8995189.

Zahedi, J., & Rounaghi, M. M. (2015, July 9). *Application of artificial neural network models and principal component analysis method in predicting stock prices on Tehran Stock Exchange*. Physica A: Statistical Mechanics and its Applications. https://www.sciencedirect.com/science/article/abs/pii/S0378437115005932.

Henrique, B. M., Sobreiro, V. A., & Kimura, H. (2018, April 27). *Stock price prediction using support vector regression on daily and up to the minute prices*. The Journal of Finance and Data Science. https://www.sciencedirect.com/science/article/pii/S2405918818300060.

Kalyoncu, S., Jamil, A., Karataş, E., Rasheed, J., & Djeddi, C. (n.d.). *Stock Market Value Prediction using Deep Learning*. Data Science and Applications. http://www.jdatasci.com/index.php/jdatasci/article/view/42.

Applying long short term momory neural networks for predicting stock closing price. IEEE Xplore. (n.d.). https://ieeexplore.ieee.org/abstract/document/8342981.

Kim, T., & Kim, H. Y. (n.d.). *Forecasting stock prices with a feature fusion LSTM-CNN model using different representations of the same data*. PLOS ONE. https://journals.plos.org/plosone/article?id=10.1371%2Fjournal.pone.0212320.

Zhang, J., Li, L., & Chen, W. (2020, July 20). *Predicting Stock Price Using Two-Stage Machine Learning Techniques*. Computational Economics. https://link.springer.com/article/10.1007/s10614-020-10013-5.