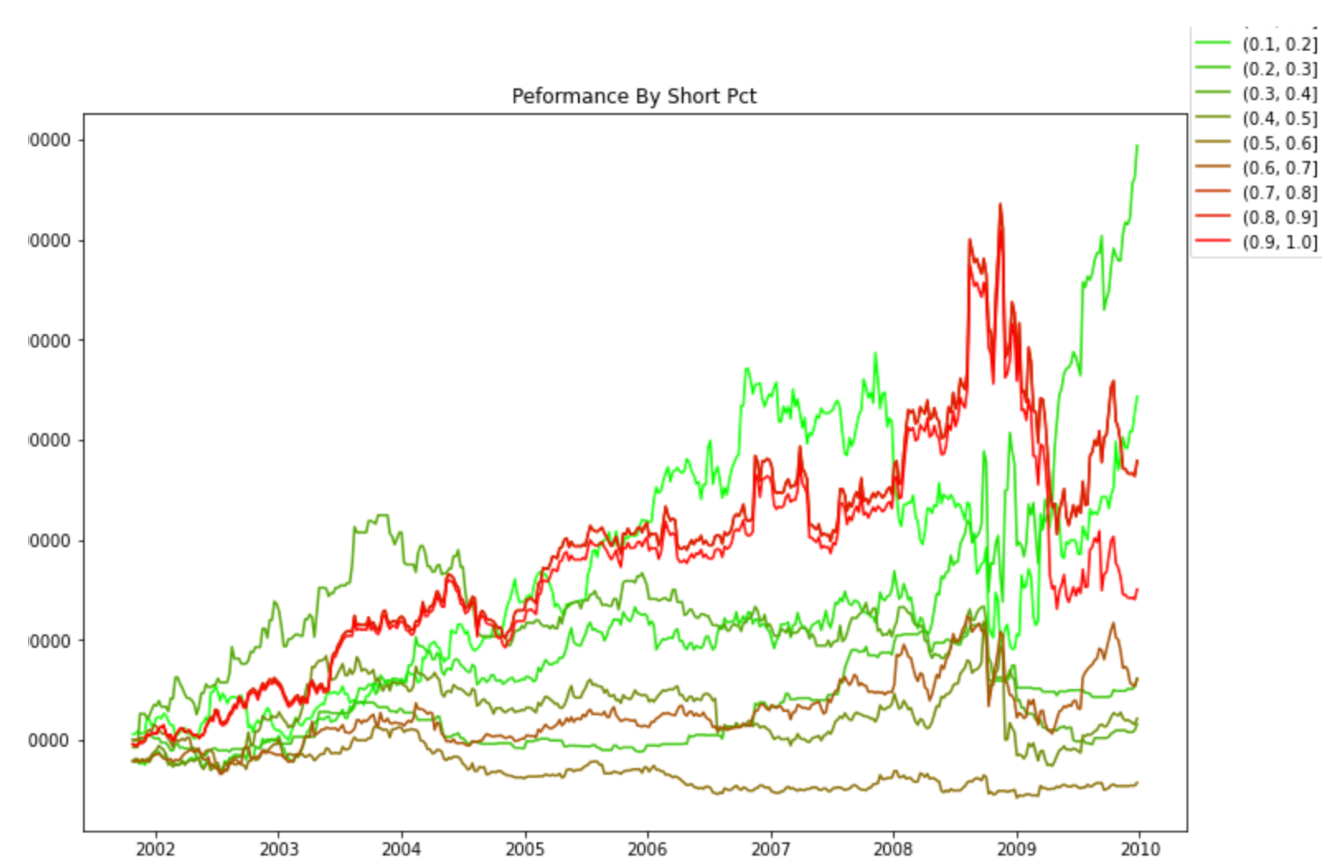
Stock Market Analysis With Machine Learning

CAL POLY

Kyle Thompson, Lubomir Stanchev: The purpose of this project is to investigate the profitability of various algorithmic trading strategies. As part of this investigation, we predict stock prices on both weekly and monthly intervals using multiple machine learning methods. These methods perform inference over technical and fundamental market indicators.

Given predictions for the change in value of S&P companies, we investigate strategies for when to buy a security, and when to sell a security. The top ten strategies as measured by median weekly change in portfolio value are shown above.



We can also analyze the performance of a strategy with respect to the ratio of long positions versus short positions. In the image above, green indicates a strategy that takes long positions more often, and red indicates a strategy that takes short positions more often.

Project Philosophy

Many researchers have worked on the problem of applying machine learning algorithms to market data. Often, they evaluate their methods with the accuracy of their models' predictions. However, if the purpose of the model is to facilitate an intelligent trading strategy, then the performance of the strategy should be used in conjunction with accuracy to evaluate the model. Therefore, we use both accuracy and back-testing to evaluate our models.

Technical Analysis

We first constructed a model that predicted the stock price of a security after one week based solely on technical data. We made each prediction based on 40 weeks of the daily closing price and volume for a company. Then we fed our predictions to a back-testing framework that used the predictions to make informed investments. We tried various models for making our predictions including Random Forest models, Long-Short Term Memory models, and Transformers. We found the most success by using Transformers, achieving a mean squared error of 0.0019.

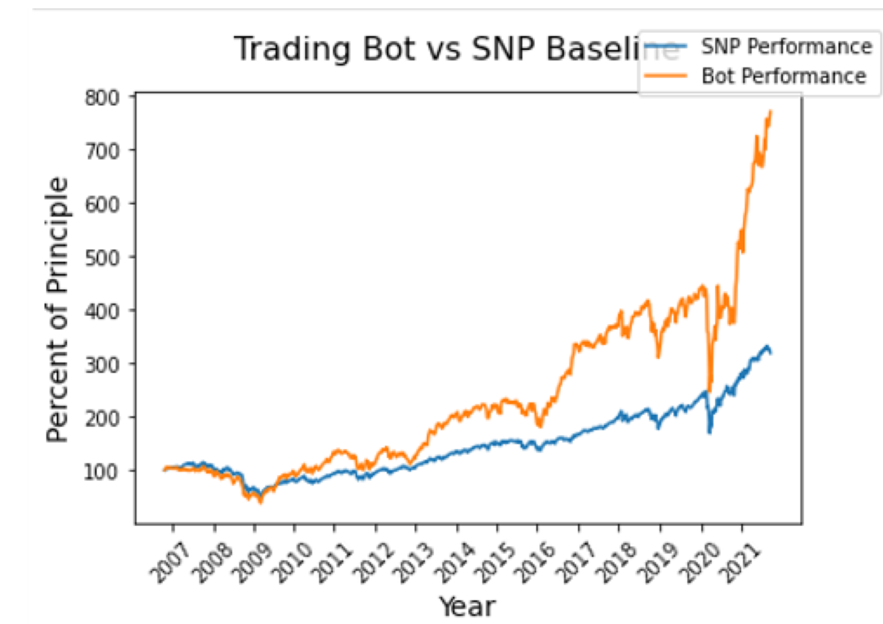
Fundamental Analysis

Methods

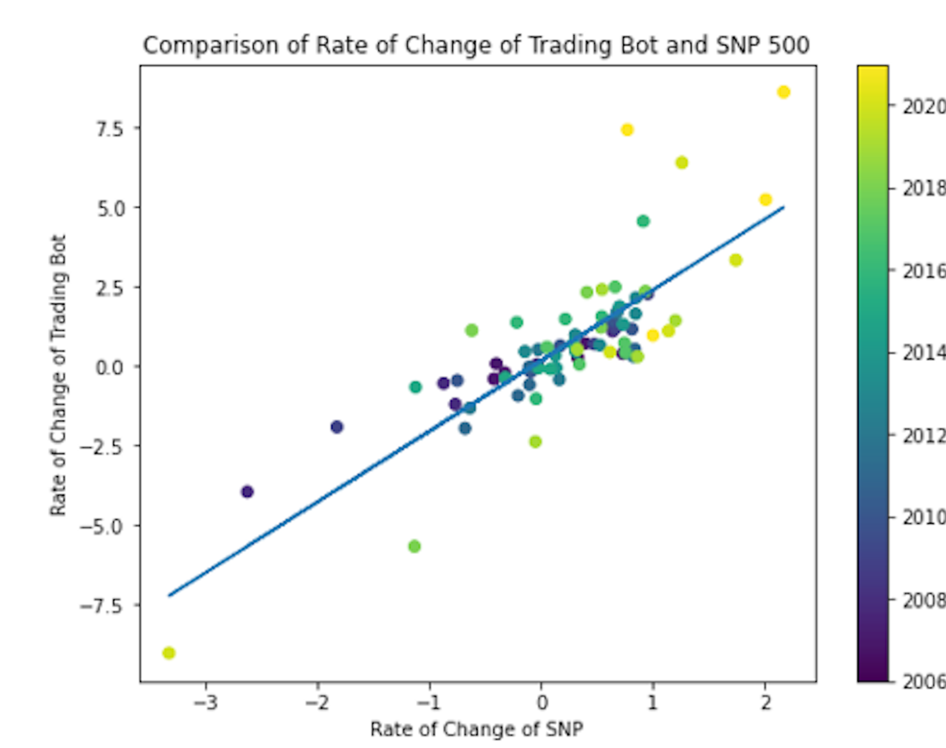
We then investigated models that predicted the price of a security one quarter in advance based solely on fundamental data. Our fundamental data included features such as a company's quarterly revenue, its liabilities, its earnings per share, its liquid assets and so on. We used two years of quarterly data to make each prediction. Because we only had quarterly data since 2010, we used a less data-intensive model to make our predictions. Specifically, we made our predictions with a Random Forest. We then used a back testing framework very similar to the one used in our Technical Analysis.

Results

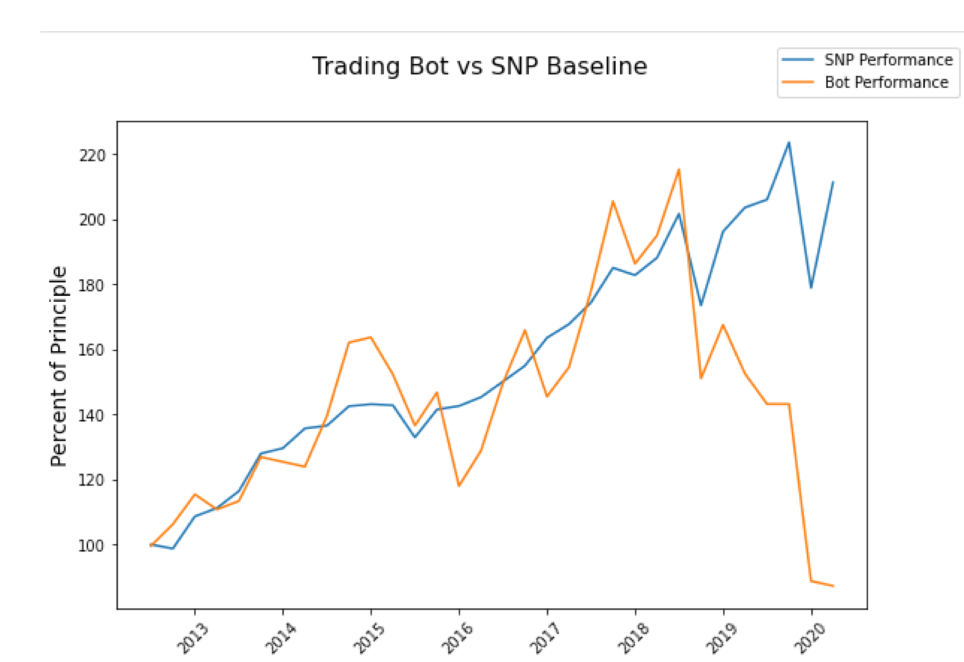
Our fundamental-based model reported a mean-squared error of 0.0237. This approach proved less effective than our technical approach. Although there were instances where our back testing reported strong profitability, the models' weak performance metrics indicate that any trading success can be attributed to luck rather than intelligence.



The image above shows the result of back-testing using our technical model. Back testing can introduce a large amount of randomness, so it is important to be weary of strong results. This test was not allowed to short securities which we have found to be a much more successful strategy with our models.



This is a plot of the rate of change of a portfolio trading with the technical strategy against the rate of change of the S&P 500. The plot indicates that the performance of the trading strategy is completely dependent on market conditions which indicates little value in using the model for trading.



This plot shows the performance of a fundamental-based trading strategy against the value of the S&P 500. The reason this chart is less granular than the technical chart is because the strategy trades every quarter rather than every week. On this back test, the strategy lost money over a span of 7 years.

Future Work

We are currently working on a model that takes as input time series for technical, fundamental, and interest rate features. We will extend the most successful model discussed in the Technical Analysis section to this new dataset to decide if a strategy that leverages these three types of data can outperform a strategy that makes decisions based on fundamental data alone.

We also recently gained access to a more complete fundamentals dataset which allows us to use the same methods that we did on technical analysis. Once we apply those methods to the larger dataset, we will be able to comment more accurately on the differences between technical and fundamental modeling.

Conclusions

Creating accurate models for stock market prediction has been more challenging than we had anticipated. It has been common for us to train large regression models to predict change in price, only for the model to predict the mean change in price for each input. Most of our time has been spent trying to combat this challenge with new models and feature engineering. However, this behavior seems to suggest some of the ideas reported in the Efficient Market Hypothesis which claims that there is no trading strategy that systematically outperforms the market without incurring extra risk.

Data Science Fellowship

This project was made possible by the Central Coast Data Science Fellowship. I am extremely grateful to have been part of this program. It has been wonderful getting to know the other students and faculty who are a part of the fellowship. I have gained valuable data science knowledge every quarter. In the fall, the readings and discussions gave me a much better understanding of the history of data science. In the winter, I learned data science tools, practices, and ideas that I have since used in my coursework and projects. This quarter I have gained a much deeper understanding of what my career in data science could look like with the speaker series.