Investment Portfolio Prediction

Tanner Kunz
October 4th 2021 cohort

Issue

- Now is the best time to invest in the stock market.[1]
- Investing in the stock market can feel like gambling.
- Difficult to create a portfolio.
- Better to invest in an Exchange Traded Fund (ETF)?

Question

Can a small portfolio reliably perform as well as the market?

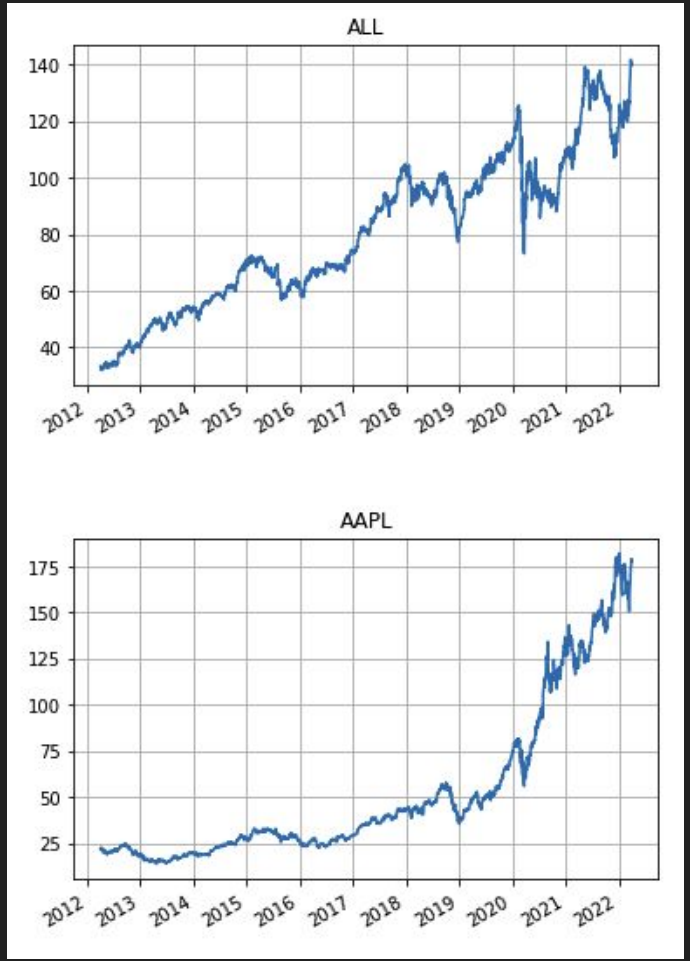
Can its future performance be accurately forecasted?

Plan

- Create 100 random portfolios.
- Compare the risk and returns against the S&P 500.
- Analyze the portfolios that historically have the most stable risk and higher returns.
- Create a model that accurately predicts the risk and returns 6 months into the future.

Data

- S&P 100 Stock data (Stocks of some of the largest companies)
- 10 years of data
- S&P 500 Index



Data Cleaning

- Returns are more comparable than closing price.
- Beta is a common measure for risk.

$$return = \frac{price - purchase\ price}{purchase\ price}$$

$$\beta = \frac{Cov(return, market\ return)}{Var(market\ return)}$$

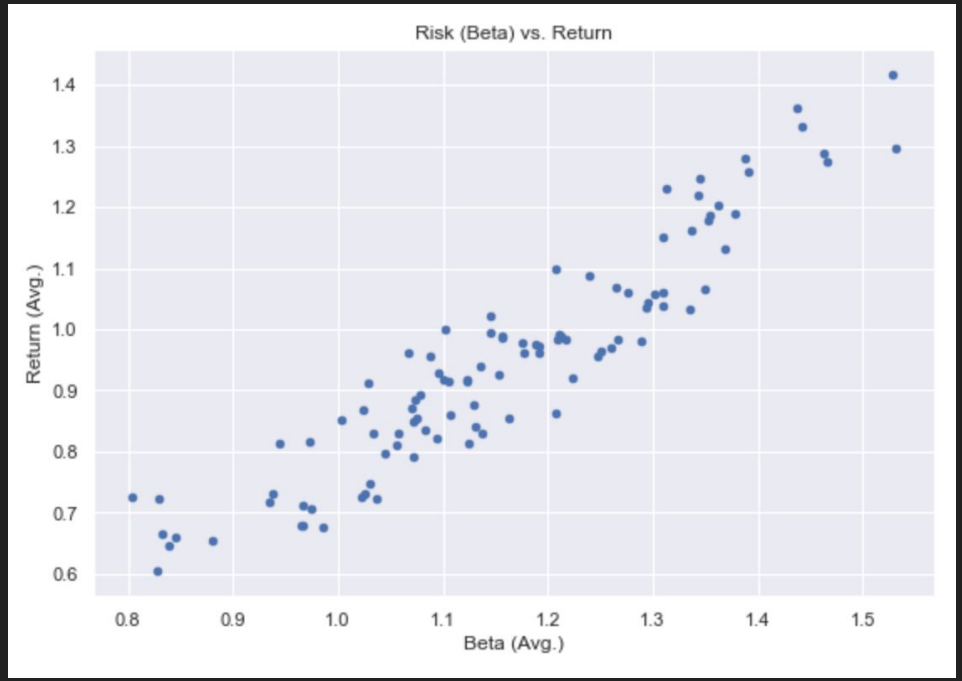
Portfolio Creation

- Randomly generate 100 different portfolios of 20 different stocks.
- Average return of the 20 stocks is the return of the portfolio.
- Average beta of the 20 stocks is the beta of the portfolio.

Exploratory Data Analysis

Data Analysis

- The returns and the betas have a positive linear trend.
- Higher risk = high reward
- Beta of 1 is the risk of the market (S&P 500)



Risk test

- Look for portfolios with an average beta that isn’t statistically different from 1.
- One portfolio falls into this category.

Returns test

- Look for portfolios that have an average return higher than the market.
- 75 portfolios meet this criteria.
- Portfolio 65 is the only portfolio that meets both criteria.

Feature Engineering

Prediction method

- Predict 6 months into the future.
- Train set: 95%
- Test set: 5% (6 months of data)

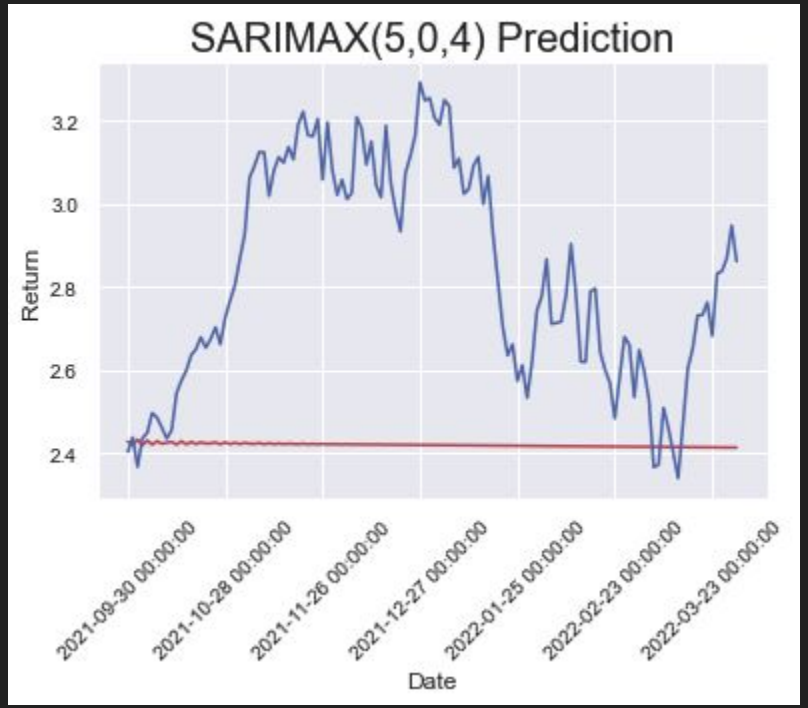
Machine Learning preparation

- Used Featuretools to use 6 months prior data for prediction
- Created rolling:
  - Min
  - Max
  - Mean
  - Standard deviation

Modeling

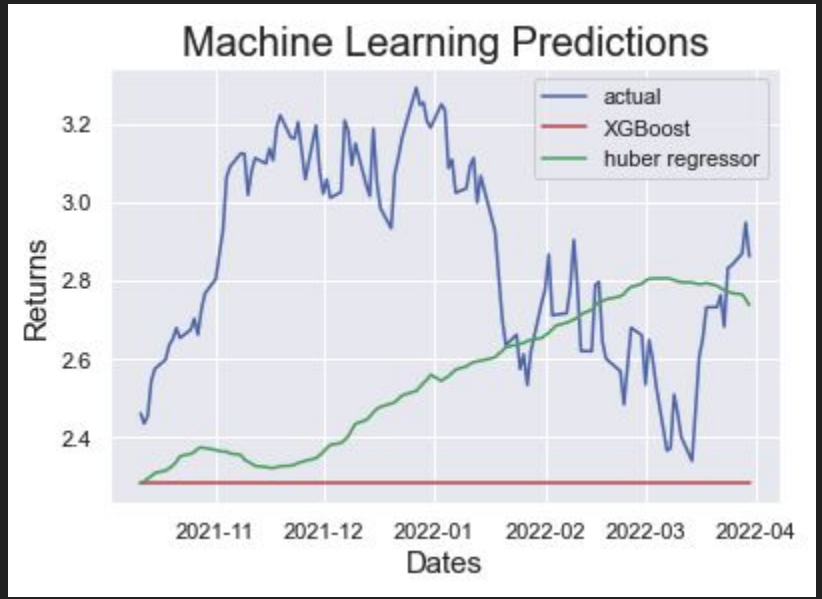
Traditional Modeling

- The data needs to be stationary to perform well.
- Transform the data by taking the square root and first difference.
- To detransform the data need the previous period data.
- Performs well, but unrealistic given the scope of the project.



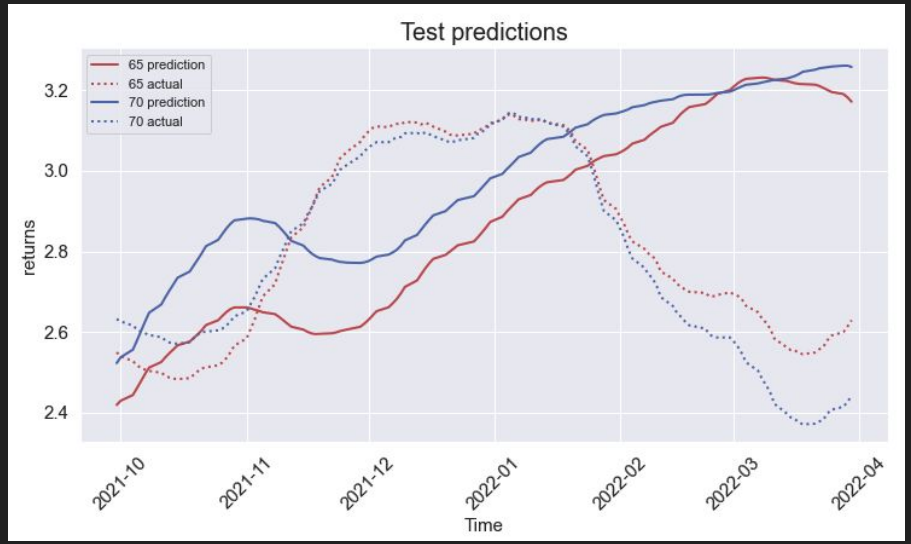
Machine Learning Modeling

- Models: Huber Regressor, XGBoost
- Data seems to follow an increasing trend, but high variance.
- Huber regressor more robust to outliers and performs the best.



LSTM Recurrent Neural Network

- The Long Short Term Memory neural network selectively uses the previous data to make predictions.
- It can also incorporate many portfolios data during training.
- It creates predictions that are less linear in trend and turn out to be more natural.
- Sudden drops seem to throw off the predictions.



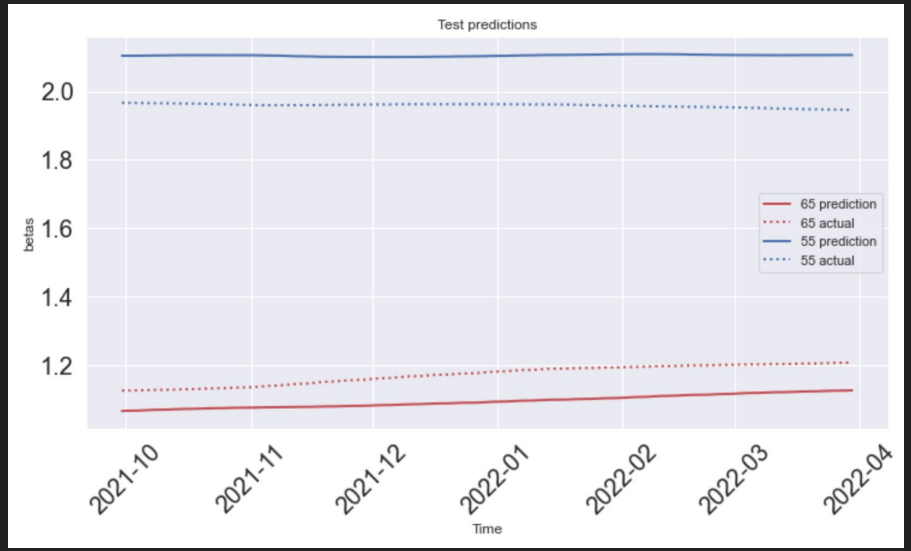
Comparison

- Traditional modeling: performs well when using the most recent data, but performs poorly when predicting into the future.
- Machine Learning: Huber Regressor shows the general trend, but is unable to make very accurate predictions.
- LSTM RNN: Creates accurate predictions and is able to follow the trend well. This model is difficult to gain insights from and is more like a black-box model.

	Traditional Modeling	XGBoost	Huber Regressor	LSTM
MSE	8.473240	0.385633	0.231171	0.166148
MAE	2.875522	0.566812	0.397977	0.348054

Beta Modeling

- The beta models results were similar to that of the returns.
- LSTM performed the best.



Next Steps

- Create a pipeline that regularly updates with the most recent stock data.
- Make a leaderboards of the portfolios that are performing the best. (highest returns, lowest risk, and a hybrid risk/return metric)
- Make a dashboard that tracks these portfolios and projects their performance with the LSTM model created
- Look into the correlation between stocks and portfolio optimization.
- Create a portfolio optimization model.
- Create a tool that allows a user to input a portfolio of stocks, date purchased, weights each stock, and how long to forecast, then outputs a dashboard with performance, future forecast, and provides suggestions on how to optimize the portfolio.
- Look into adding other exogenous variables such as NLP analysis of news and social media to add to the model.
- Etc. (I can think of so many more things that I would like to keep adding to this project.)