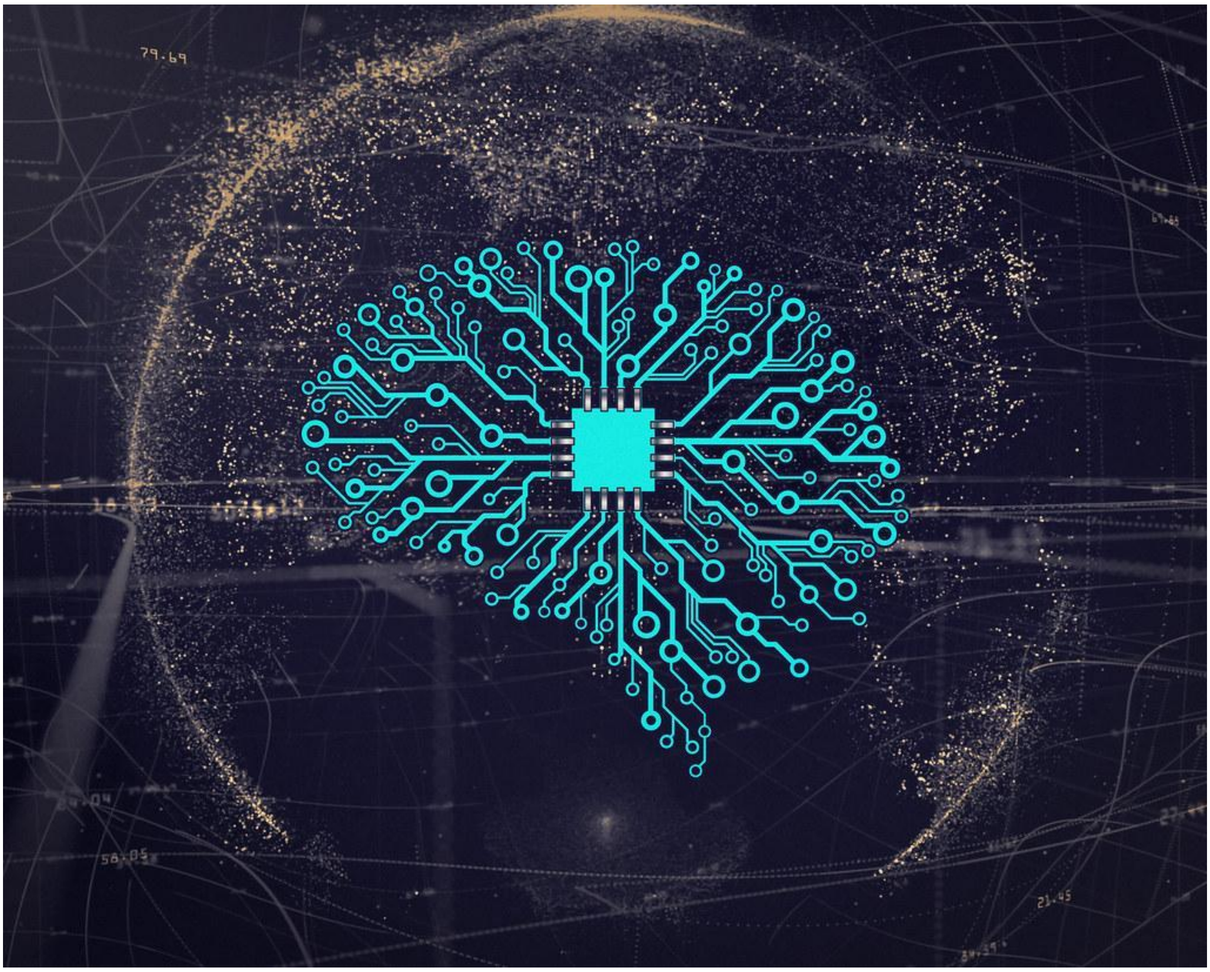
BF3210 Issues in Global Financial Markets

FutureSight App White Paper

Invest Smart with Machine Learning



Executive Summary

Prediction of stock prices is a complicated process but with more information and advances in financial technologies, we are beginning to understand the stock market well enough to make educated deductions of future prices. In this paper, we make use of a practical machine learning approach to predict stock's forward quarter performance. In this paper, we modified one of the most well-known machine learning stock prediction algorithms to create a financial application out of it, FutureSight. The paper will explain the economic theory, followed by a machine learning explanation and the backtest results, and how our financial application brings about this futuristic technology to retail investors and their portfolio performance. We would like to credit Professor Hongyang (Bruce) Yang, Xiao-Yang Liu, Qingwei Wu for pioneering this machine learning approach, which our team relies heavily on.

Introduction

The stock market can be broadly defined as the aggregation of buyers and sellers of stocks which represent ownership claims on those stocks. Stock prices reflect both the intrinsic value of a company in terms of its current performance as well as expectations of future performance. When a company performs well, its stock price adjusts favourably, and investors stand to obtain profits in the form of both capital appreciation and increased dividend pay-outs. Hence, investors are incentivised to select the right companies to invest in, but this is no easy feat. Prediction of stock prices is a complicated process but with more information and advances in financial technologies, we are beginning to understand the stock market well enough to make educated deductions of future prices.

Every stock investor is familiar with the *efficient market hypothesis.* This hypothesis states that stock prices are completely random and unpredictable. While it is true that the stock market is a bustling environment of people buying and selling thousands of stocks with little coordination, due to human nature of investors, certain behaviours occur in reaction to market events and these correlations have been studied by financial institutions worldwide. Analysts manage to deduce certain hidden patterns in market movements and trade based on these patterns to generate profit for their firms. Furthermore, predictive models are rapidly being developed to consistently and accurately pinpoint prices such that big firms consistently outperform their competitors.

Economic Theory

Many fundamental financial ratios are used today by investors and portfolio managers as an indicator of its intrinsic value. In value investing, stock's long run prices are expected to regress towards its intrinsic values, thus making these financial ratios to have powerful predictive power on the future performance. Additionally, financial ratios provide normalization to allow for comparison across companies given the scale of data. Therefore, by using key financial ratios as inputs for machine learning modelling, we can make educated and reliable stock's future price return.

Machine Learning

Machine learning is the study of computer algorithms which improve automatically through experience. The automation component is especially important with the vast number of stock movements and relative signals one has to take into account when attempting to trade systematically.

The particular concept we based FutureSight upon is "feature selection", considered to be a core aspect of feature engineering and in turn, Machine Learning. Features to be selected for our machine learning are based on top twenty most popular financial ratios in the following table:

20 Financial Indicators	
Revenue Growth	Price to cash flow ratio
Earnings per share (EPS)	Cash ratio
Return on asset (ROA)	Enterprise multiple
Return on equity (ROE)	Enterprise value/cash flow from operations
Price to earnings (P/E) ratio	Long term debt to total assets
Price to sales (P/S) ratio	Working capital ratio
Net profit margin	Debt to equity ratio
Gross profit margin	Quick Ratio
Operating margin	Days sales of inventory
Price to book (P/B) ratio	Days payable of outstanding

The data for this project is taken from the Compustat database accessed through Wharton Research Data Services (WRDS). The dataset consists of the data over the period of 27 years (from 1st June 1990 to 1st June 2017)

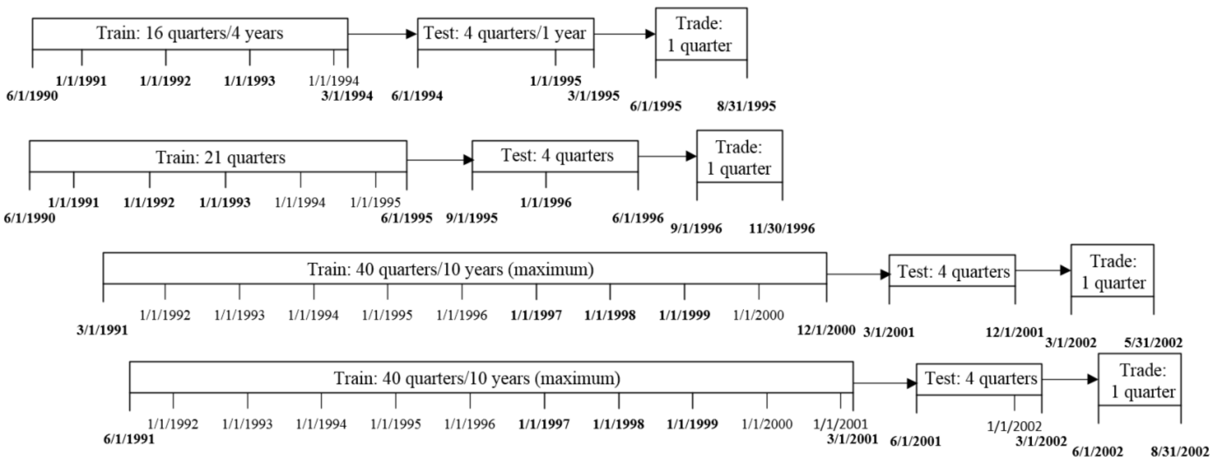


Fig. 1. Rolling Window Based Data Seperation: Training rolling window is followed by a testing rolling window. There is a two-month delay after the end of the training rolling window.

The goal is to predict the stock's forward quarter log-return given predictors constructed from historical data of the twenty financial factors over a particular quarter. The machine learning algorithm will make use of the 4 most popular prediction models mainly: linear regression, regularized linear OLS estimator ridge regression, tree based nonlinear model random forest and generalized boosted regression model (GBM) using gaussian distribution which implements AdaBoost algorithm and Friedman's gradient boosting machine. The reason for using these 4 models is that we need feature selection models to remove undesirable features, thus reducing the overfitting issues, improving model accuracy and expediting the training procedure.

Creating an application for the purposes of stock price prediction is fundamentally a time-series problem where we aim to predict the next point in the time series. The similarity of our prediction efforts to linear regression allows us to solve the problem by means of linear regression, hence our solution would be an algorithm targeting regression models.

For our application, we will be selecting the ensemble algorithm, RandomForestRegressor to develop our baseline model. A 'Random Forest' is a large collection of individual decision trees which each produce a class prediction. From this cluster, the class prediction that obtains the most votes becomes our model's prediction. The science behind the success of the RandomForestRegressor is that each decision tree in the Random Forest is highly uncorrelated.

After selecting the model to use, we will proceed to train it by inputting the relevant data which we plan to make predictions of. Mean Squared Error (MSE) is used as the metric for the evaluation.

TABLE II
MODEL ERROR AND SELECTED MODEL FOR SECTOR 10, ENERGY

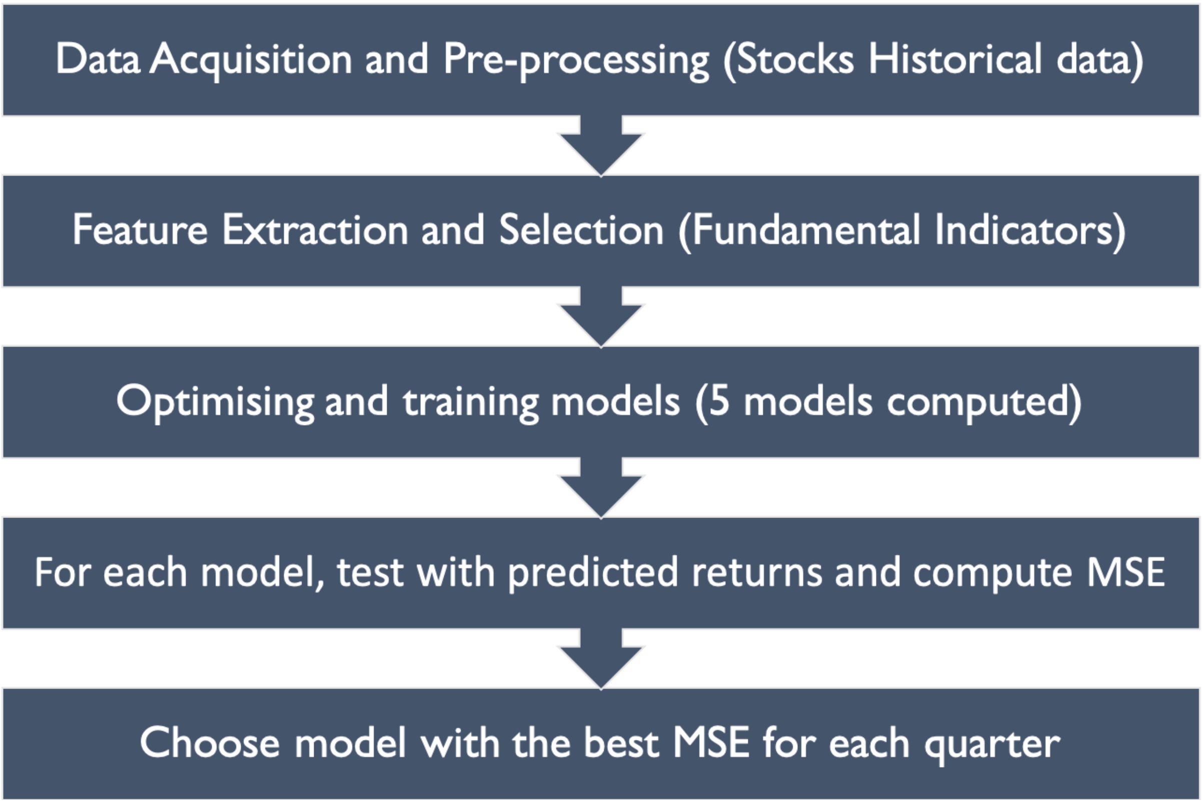
trading date	MSE linear	MSE RF	MSE ridge	MSE step	MSE gbm
19950601	0.02238	0.02180	0.02161	0.02205	0.02443
19950901	0.01908	0.01828	0.01870	0.01841	0.02098
19951201	0.01852	0.01641	0.01820	0.01855	0.01996
19960301	0.02040	0.01822	0.01981	0.01879	0.02192
19960603	0.02442	0.01885	0.02394	0.02340	0.02210

TABLE III
PREDICTED RETURN ON TRADE DATE: 1995/06/01 SECTOR 10, ENERGY

	Linear return	RF return	Ridge return	Step return	GBM return
WMB	10.42%	5.12%	9.24%	9.01%	3.51%
OKE	7.55%	4.12%	7.42%	8.53%	2.56%
RRC	4.16%	7.01%	3.74%	4.84%	1.83%
PXD	4.63%	0.96%	3.66%	3.91%	0.23%
VLO	3.48%	2.99%	3.47%	4.04%	2.56%
EQT	2.47%	4.78%	2.34%	2.36%	1.83%
HES	1.80%	4.30%	1.61%	1.81%	0.38%
BHI	1.33%	−0.70%	1.15%	1.99%	−0.27%
MUR	1.11%	1.07%	1.01%	0.49%	0.38%
NE	1.16%	−6.33%	0.94%	0.85%	−2.21%

Implementation

Our implementation can be summarized in the following steps:



Prediction of stock price performance will be the main core of FutureSight.

TABLE IV
PREDICTED RETURN ON TRADE DATE: 1995/09/01 SECTOR 10, ENERGY

	Linear return	RF return	Ridge return	Step return	GBM return
CHK	0.97%	12.45%	2.17%	4.86%	0.03%
SFS.1	4.69%	7.35%	4.27%	4.49%	0.78%
WMB	5.87%	6.37%	5.15%	5.46%	0.77%
RRC	1.78%	6.13%	1.52%	1.76%	0.09%
VLO	2.50%	4.83%	2.65%	3.01%	0.77%
BJS.1	5.36%	4.23%	5.28%	6.38%	−1.71%
MDR	1.54%	3.96%	1.26%	1.14%	0.77%
PZE.1	0.96%	3.78%	0.55%	0.78%	−1.71%
HP	−1.63%	3.50%	−1.44%	−1.62%	0.77%
CVX	−0.73%	3.19%	−0.71%	−1.12%	0.09%

The other feature that will add value to retail investors is the portfolio optimization tool to allow retail investors to allocate the stocks that have picked. There will be two portfolio construction methods: max-Sharpe ratio and equally weighted optimization approach.

We set the following constraints for max-Sharpe ratio:

1. Expected return: predicted return of next quarter

2. Covariance matrix: use 1-year historical daily return
3. Long only portfolio
4. Fully invested capital: sum of weights = 100%
5. No leverage
6. Risk-free rate assumed to be at 0.005
7. Time window/frequency: 252 (Annual)

By establishing the above constraints, FutureSight's portfolio optimization tool has incorporated risk management in terms of decision rules. Fundamentally, due to the nature of the long-only strategy, the risk is managed internally through the means of no leverage and maximum limits of position sizes.

Since Sharpe ratio in this case is the portfolio's return less the risk-free rate, per unit risk (volatility). Therefore, FutureSight will calculate the optimal weights based on the user's selected stocks to arrive at a portfolio with the maximum Sharpe Ratio, also called the tangency portfolio.

From here on, users can compare the difference in portfolio performance between a max-Sharpe ratio optimized one and an equally weighted one.

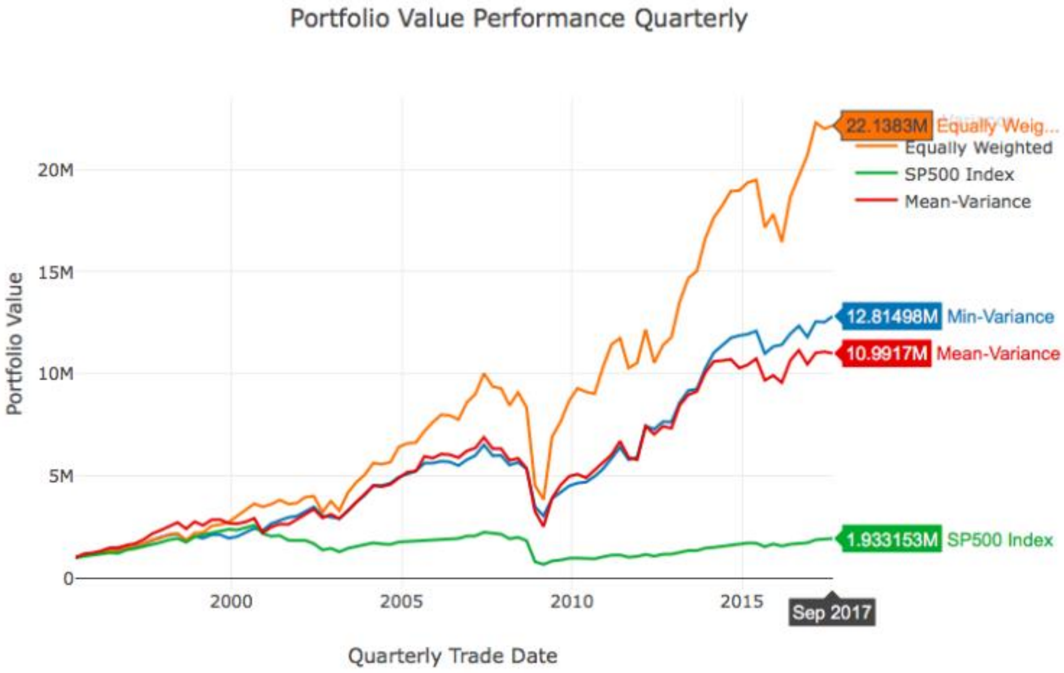


Fig. 5. This figure shows Portfolio Value starts with 1 million.

t

Solution

Retail investors lack knowledge on stock performance and portfolio optimization. Most importantly, they have a lack of access to information about machine learning and how to utilize it in their own personal trading portfolios. FutureSight simplifies the process and brings machine learning to the doorsteps of retail investors within a user-friendly user interface.

Appendices A-H provides some screenshots of our app's user interface to show how easy and user-friendly our app is.

Future Plans

We look to add the capability for syncing our application with the top brokerages to provide a comprehensive real time view at a glance for all our users' portfolios. Our application could also generate periodical summaries of account performance to keep our users abreast of any ongoing developments with users being able to set how frequent they would like to receive these summaries.

Additionally, news-scanning capabilities could be integrated in the future to alert our users to potential impacts on their account holdings from relevant news releases. The news function could also be used to conduct sentiment analysis which would also be factored into our performance prediction process and provide another leading indicator.

Finally, we would look to implement the watch-list function for our users to track the movement of stocks they would like to add to their portfolio. The function would provide users with alerts at users' predetermined price points such that they would be able to enter a position at a favourable price.

In conclusion, our application aims to be a one-stop mobile portfolio management service. While we are currently focused on stock performance prediction, we believe that enabling our users to conveniently transition from educating themselves about their portfolios to swiftly enacting the recommended changes within our application itself would provide for a positive user experience and hence, represent an attractive proposition for existing and incoming retail investors.

References

BruceYanghy. (n.d.). *BruceYanghy/Dynamic-Stock-Recommendation-Machine_Learning-Published-Paper-IEEE*. GitHub. Retrieved May 4, 2020, from https://github.com/BruceYanghy/Dynamic-Stock-Recommendation-Machine_Learning-Published-Paper-IEEE

constverum. (n.d.). *constverum/Quantdom*. GitHub. Retrieved May 4, 2020, from https://github.com/constverum/Quantdom

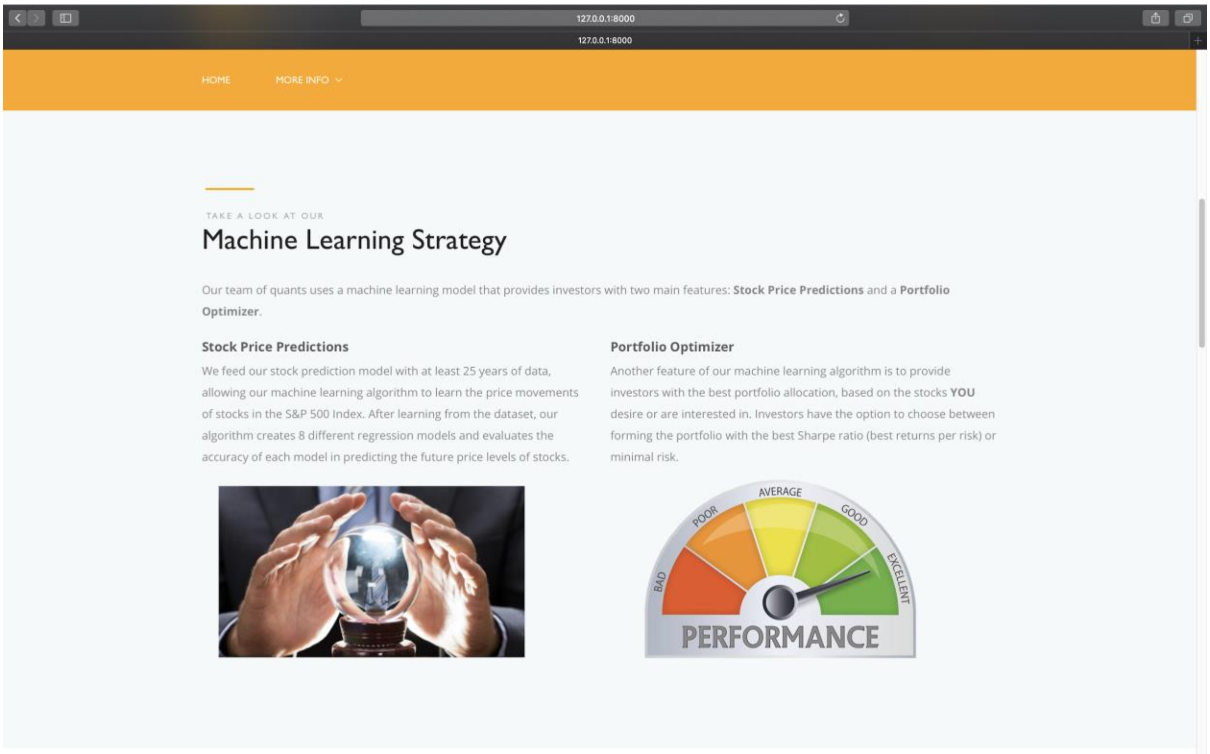
Machine Learning for Stock Selection (In Practice). (n.d.). CFA Institute. Retrieved May 4, 2020, from https://www.cfainstitute.org/research/financial-analysts-journal/2019/ip-machine-learning-for-stock-selection

stefan-jansen. (n.d.). *stefan-jansen/machine-learning-for-trading*. GitHub. Retrieved May 4, 2020, from https://github.com/stefan-jansen/machine-learning-for-trading

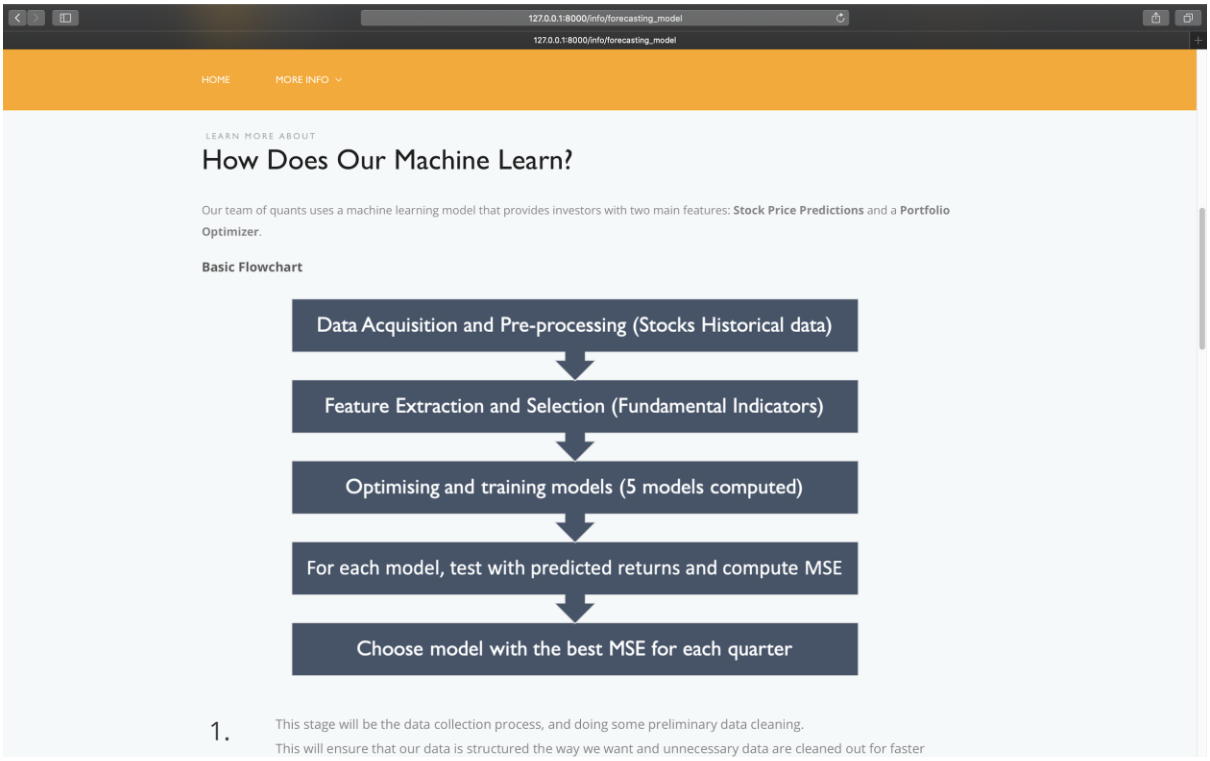
Xu, J. (2019, August 30). *How To Use Machine Learning To Possibly Become A Millionaire: Predicting The Stock Market?* Medium; Towards Data Science. https://towardsdatascience.com/how-to-use-machine-learning-to-possibly-become-a-millionaire-predicting-the-stock-market-33861916e9c5

Appendices

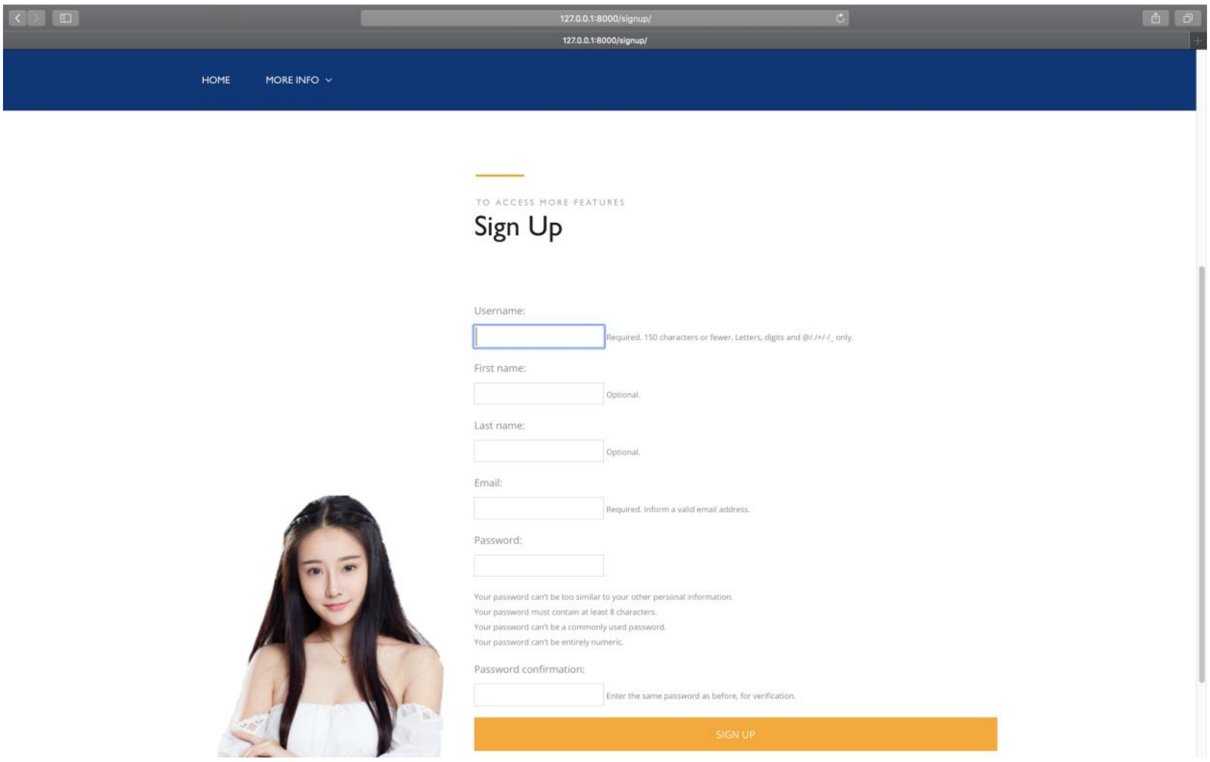
Appendix A: App Home Screen (Pre-login)



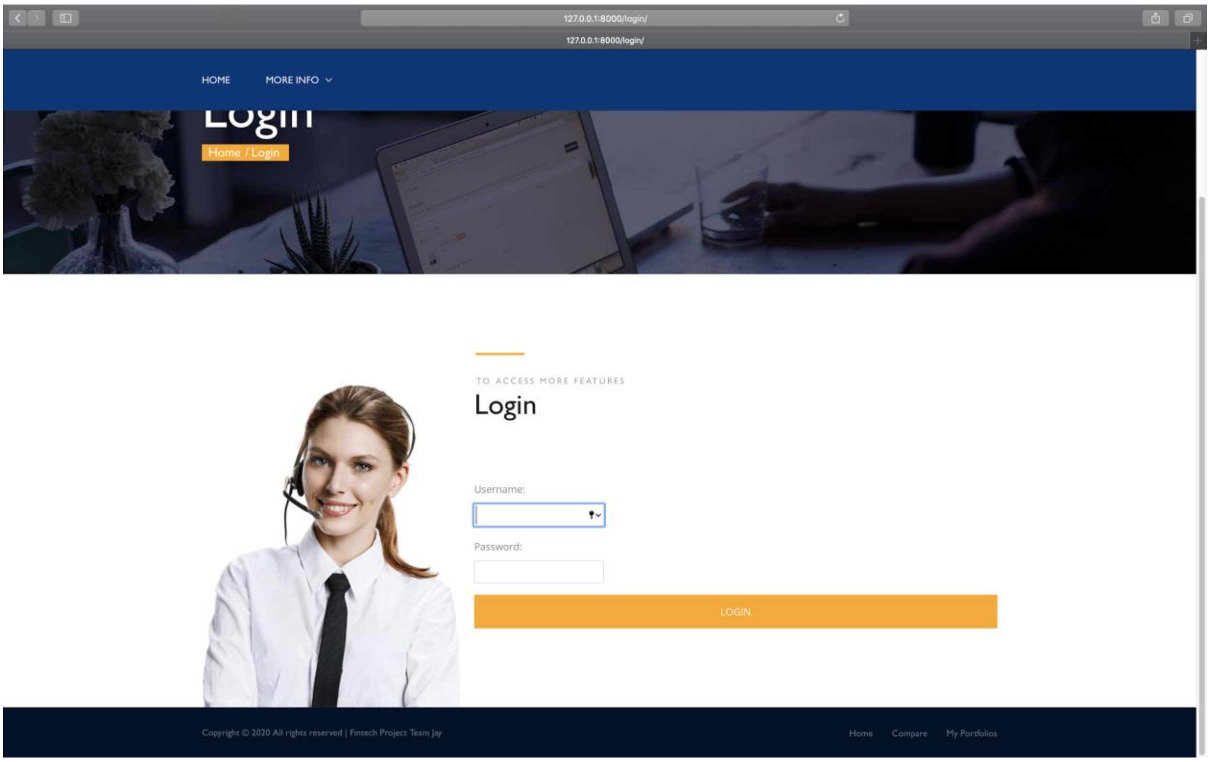
Appendix B: App Home Screen (Pre-login), Short Introduction of App



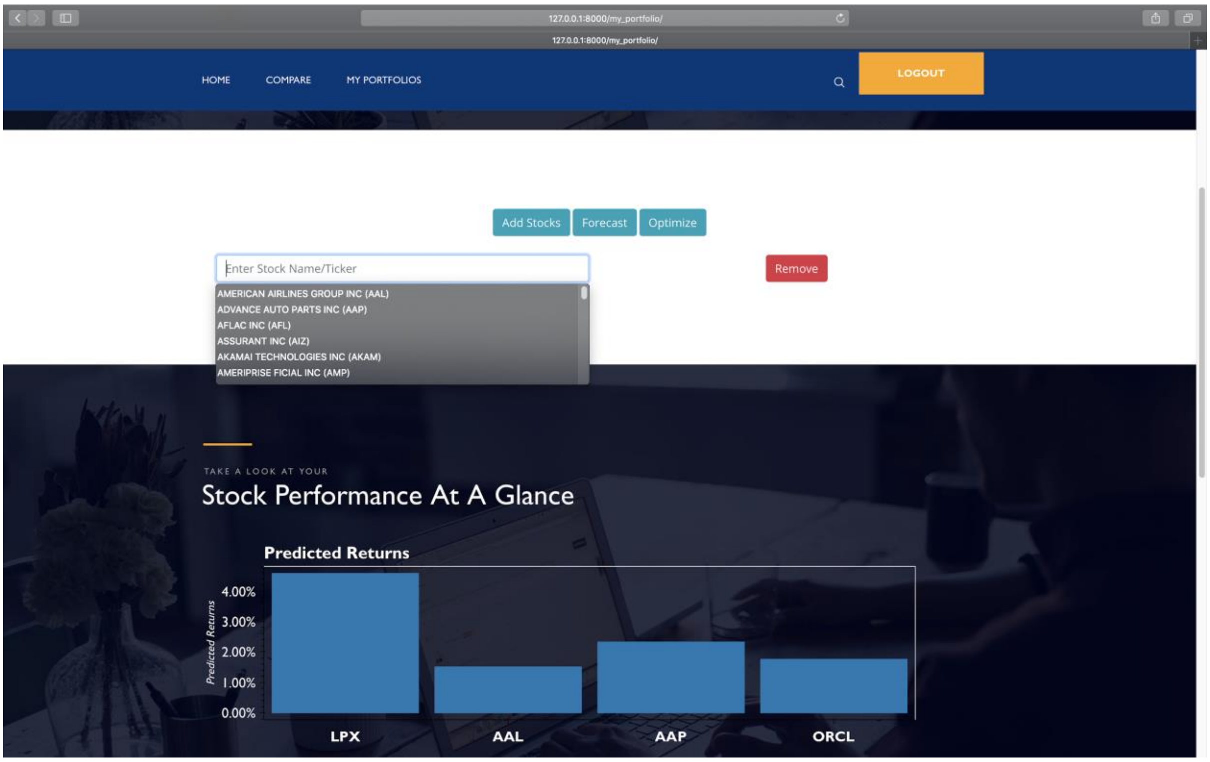
<u>Appendix C: App More Info (Pre-login), Detailed Explanation behind our Machine Learning Model</u>



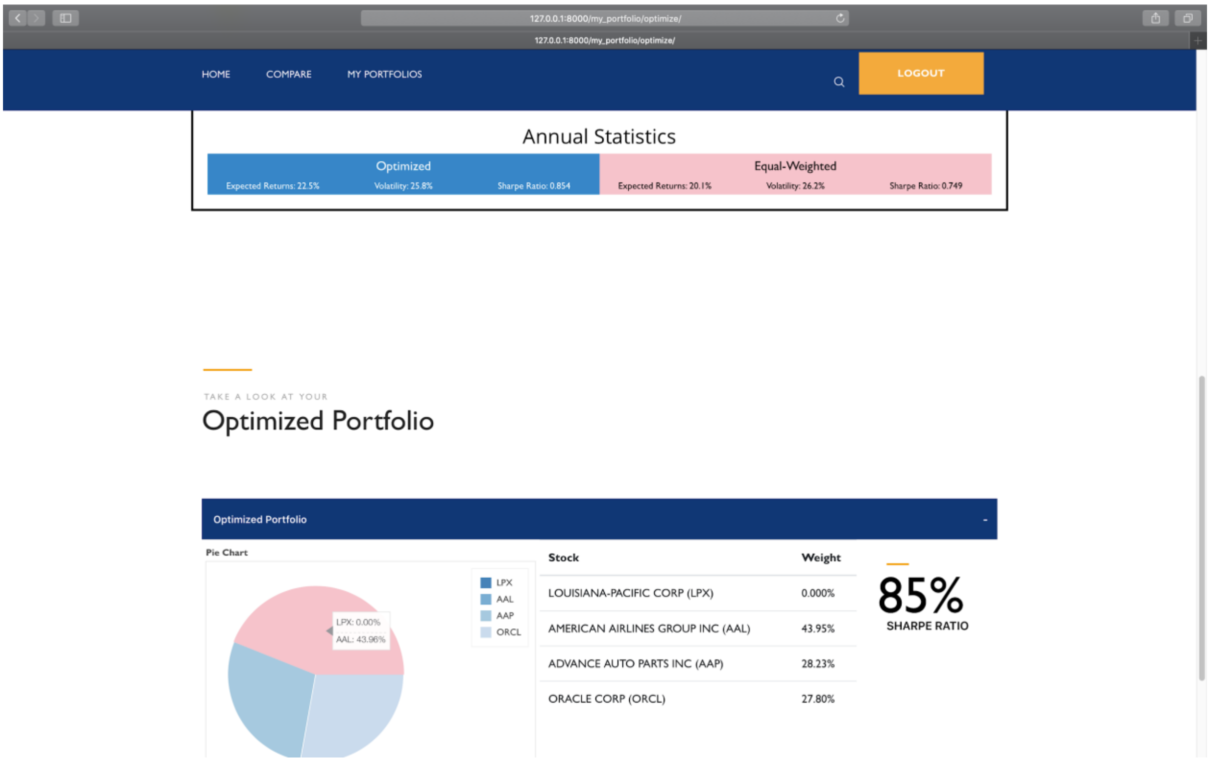
<u>Appendix D: App Sign Up Page</u>



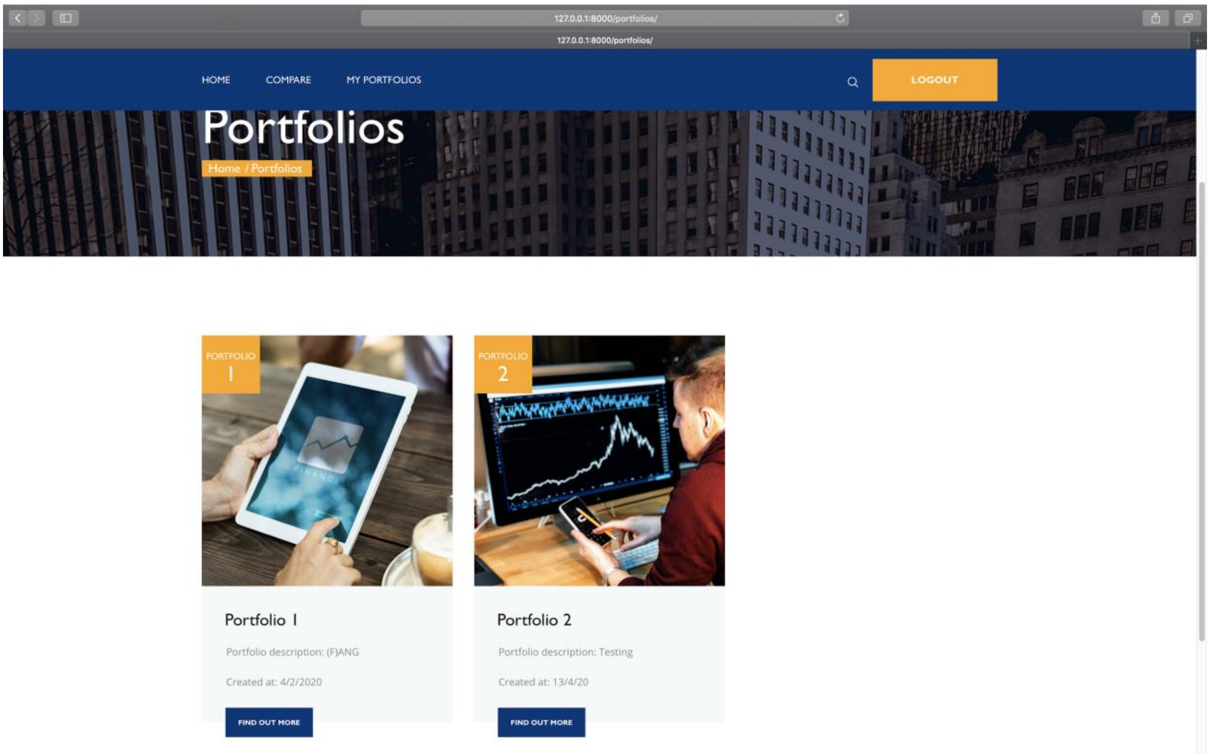
Appendix E: App Login Page



Appendix F: App Home Page (Post-login) / Stock Selection page



Appendix G: App Portfolio Optimizer page



Appendix H: App User Portfolio Management page