

# Take Stock of Your Life: Designing a Stock Portfolio

Team 33 - Alpha Seekers

Abhirup | Beenapreet | Daniel | Darnesh | Kimi | Somrita

## 1. MOTIVATION

To invest successfully in the world of stocks, it is important to understand and keep track of multiple factors. The complexities of these factors along with the required time commitment often act as deterrents for many potential investors.

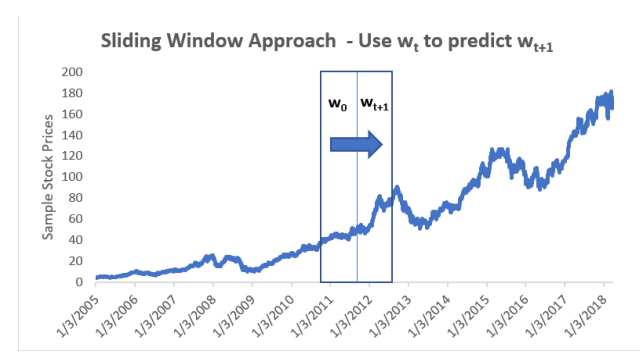
We have designed an interactive web-based application to address this problem. Our website will suggest a portfolio based on user inputs such as investment budget, expected returns and time horizon. The objective is to minimize risk given the user provided tolerance on the return



## 4. SHORT TERM STRATEGY

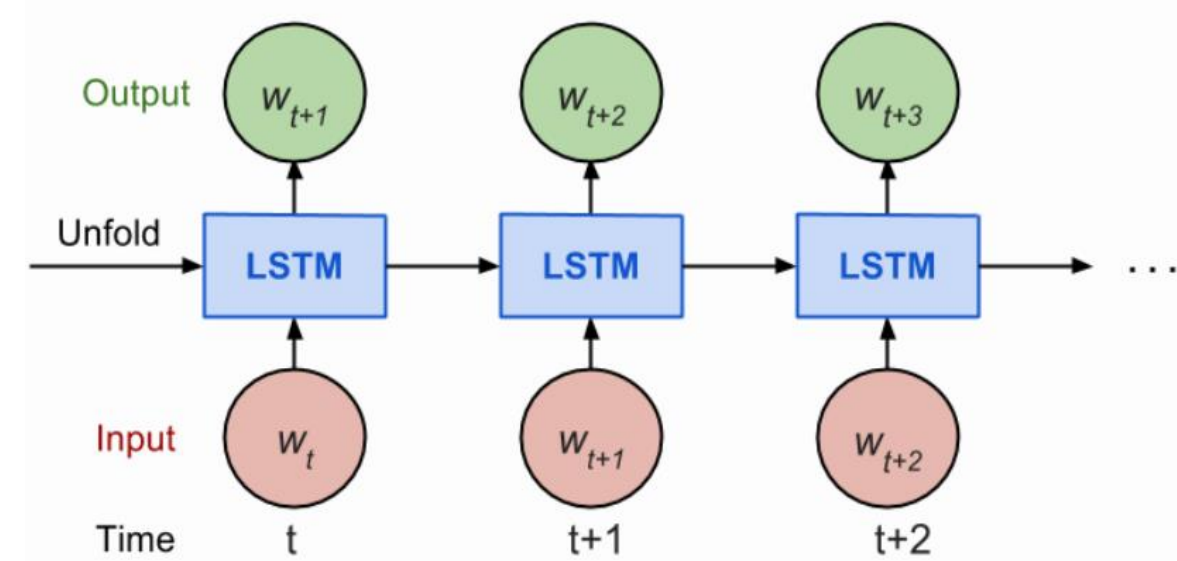
We have used an **RNN-LSTM** model to predict stock prices for the next **5 days** using Adjusted average OHLC prices for the previous 5 days ( $t, t-1, t-2, t-3$  &  $t-4$ ) to predict average OHLC price for the next day ( $t+1$ )

We used values from the very beginning in the first sliding window ( $w_0$ ) to the window in time  $t$  ( $w_t$ )



$$W_0 = (P_0, P_1, \dots, P_{W-1})$$
$$W_1 = (P_W, P_{W+1}, \dots, P_{2W-1})$$
$$\dots$$
$$W_t = (P_{tW}, P_{tW+1}, \dots, P_{(t+1)W-1})$$

- RNNs consist of states, which are updated every time step
- The state, at time step  $t$ , is a summary of the information in the input sequence till  $t$
- Hence to capture the price changes of the stock from the start an RNN is used
- To prevent unwanted information to be retained we use **LSTM to enable selective forgetting or remembering** information from sequence



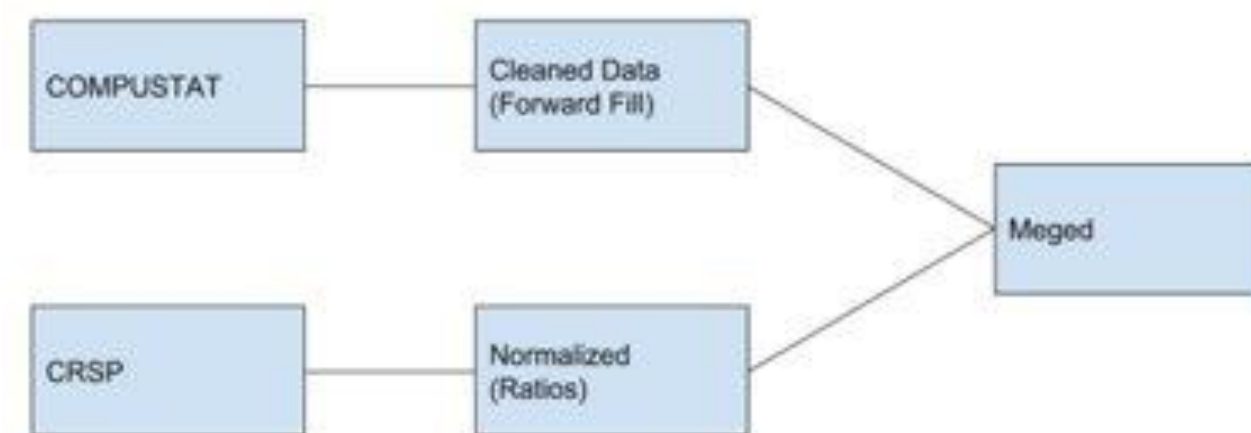
## 7. BACKTESTING

- Backtesting is defined as the process of testing a trading strategy on historical data to analyze whether the said strategy works on an out-of-sample dataset
- We backtested the model on our 1 week, 6 month, and 1 year investment strategies



## 2. DATA SOURCE

We sourced the stock closing price data from yahoo finance and "CRSP" database. The fundamental data was sourced from the "COMPUSTAT" database. The data were then filtered to limit the search to Russel's 1000 components. The historical stock quote data were then standardized to ensure that stock quotes were available for all the weekdays and were duly forward-filled in case of missing. The fundamental data were carefully normalized as ratios.



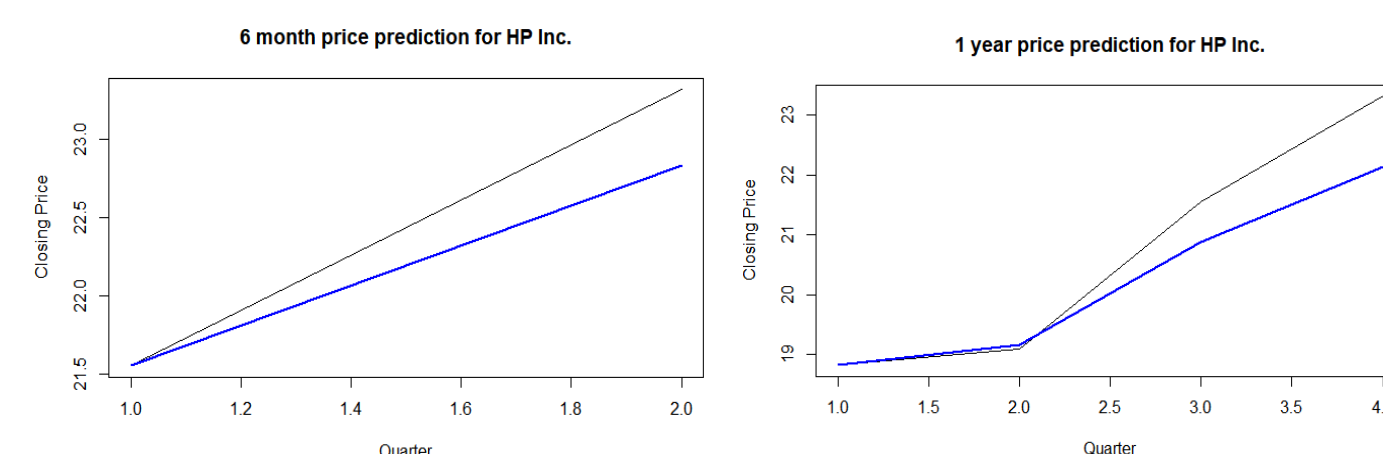
## 5. LONG TERM STRATEGY

**Two models to decide long-term strategy.**

(i) We developed linear regression models to predict returns for stocks (after they are ranked based on quality scores). The models predict 6/12 months' expected returns for each stock

- 6/12 month returns taken as dependent variable
- Fundamental variables (Earnings per share, Market Value, Return on assets, Return on equity, gross margin, payout etc.) considered as independent variables with present returns
- We used Lasso regression for variable selection

(ii) We used Autoregressive integrated moving average (**ARIMAX**) model to suggest best performing stocks for 6 months and 1 year horizon. Stock return prediction was based on exogenous variables including past closing price of the stock, long term debt, market value and net income of the firm



## 3. SCORING

We developed a quality score, that all else equal, an investor should be willing to pay a higher price for: stocks that are profitable, growing, safe and well managed. High quality stocks generally have a higher risk adjusted return. Fundamental data obtained from COMPUSTAT database is used to compute the score. It gives equal weightage to each of the factors namely "profitability", "growth", "safety" and "payout".

**Profitability:** It is profits per unit of book value. It can be measured in several way such as in terms of gross profits, margins, earnings and accruals.

**Growth:** It is measured as prior five year growth in each of our profitability measures.

**Safety:** We consider both return based measure of safety ( market beta and volatility) and fundamental based measures such as low leverage, low volatility of profitability and low credit risk.

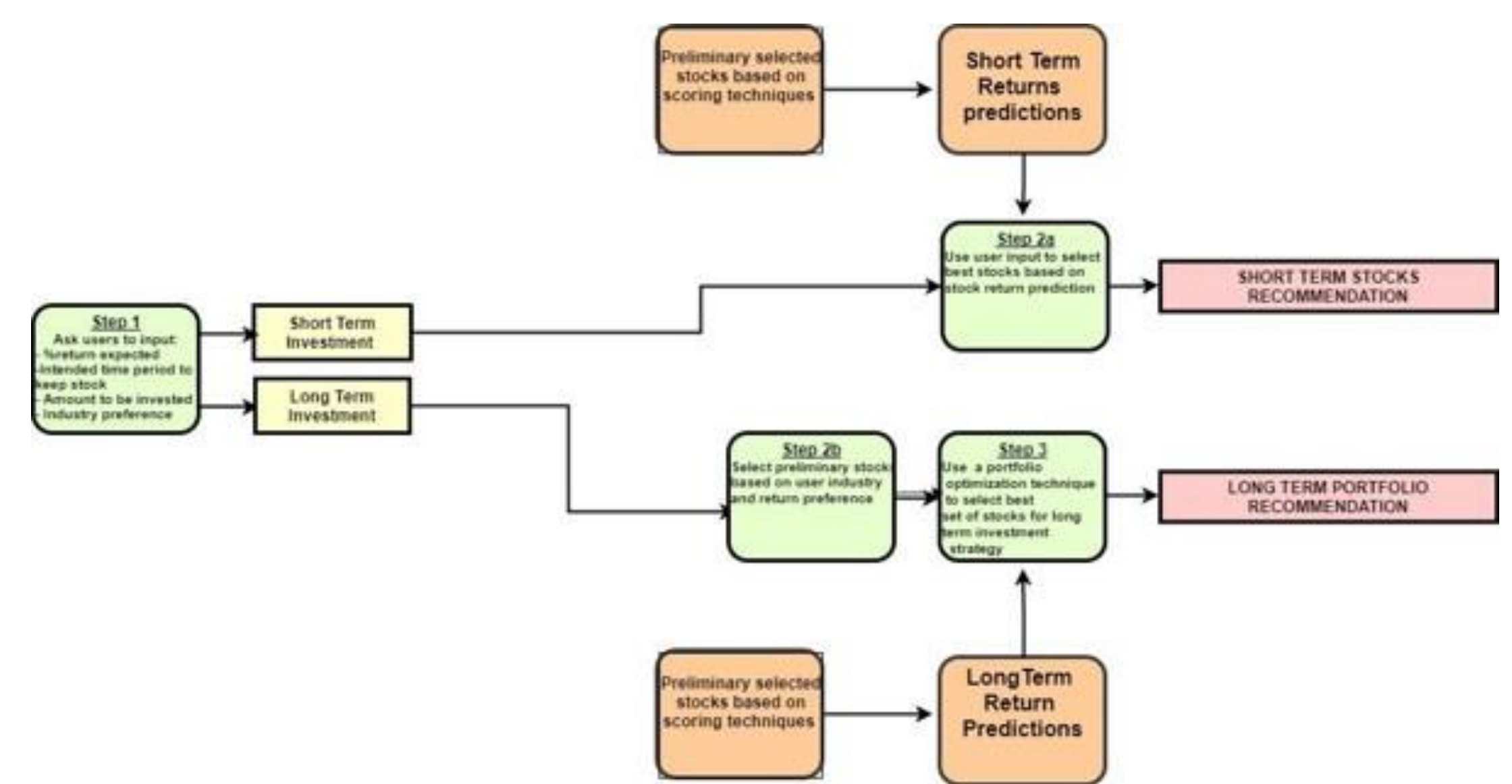
**Payout:** It is a fraction of profits paid out to shareholders. It can be seen as a measure of shareholder friendliness.

$$\text{Quality} = z (\text{Profitability} + \text{Growth} + \text{Safety} + \text{Payout})$$

## 6. OPTIMIZATION

We used an optimization procedure to determine the percentages of investment capital (portfolio weights) that should be invested in each stock selected in through the quality scoring. This procedure, called Mean-Variance Optimization, is done by finding the vector of weights that minimizes the objective function (total portfolio variance), subject to constraints that the weights must sum to one and that the portfolio return must equal a given target return. To ensure that the user will not be overwhelmed by having too many stocks to invest in, we added an additional constraint limiting the number of the stocks in the portfolio to ten or less.

## FLOW-CHART OF METHODOLOGY



## 8. WEB BASED UI TOOL

**Optimal Portfolio Calculator**

Enter investment amount:  
100000

Enter desired return (percentage):  
15

Time Horizon:  
1 year

**Calculate**

We created an interactive, web-based UI that allows users to enter their portfolio preferences and then see what our system recommends for their optimal portfolio. On the first page, the user enters the amount they want to invest, their desired return on investment, and the time horizon over which they hope to generate that return. They are then taken to a page displaying the stocks and allocations they should invest in to create an optimal portfolio, as well as a visualization for what our models predict for this portfolio's performance over time.

