American University of Beirut

Faculty of Engineering and Architecture

Electrical and Computer Engineering Department

EECE 502 - Final Year Project

Group & Disciplines

Rami Awar *Optimization, Software*
Tarek Tohme *Statistics, Finance*
Nader Al Awar *Databases, Software*

Group Final Report

Spring 2018-2019

QuantInvest: Automated and Accessible Investment Management

Advisor: Dr. Fadi Zaraket - Electrical and Computer Engineering Department

Glossary

Alternative Data: Any kind of structured or unstructured data expected to yield information about market movements. Our usage of the term 'alternative data' includes financial news, liability networks between organizations, GPS data, social media, among others.

Algorithmic Trading: Trading that involves automation at some or all stages of the trading procedure, for example at the portfolio generation stage or the trade execution stage.

Asset: A financial security that has a monetary value and can be traded on an exchange, or over-the-counter (without any intermediary).

Backtesting: Executing a strategy on past market data to assess performance.

Book Value: A company's equity divided by the total number of shares.

Cross-validated test error: Statistical models are trained on a certain dataset and tested on a different dataset, which is the main reason why they are useful, namely, they can produce predictions on data they have never encountered during training. In practice, it is sometimes difficult to obtain a different dataset to test a model against, so the training data set is split into two sections: a train section and a test section. The test error (ratio of correct predictions) can be obtained from the test section. The train and test sections are then altered and testing is done again. Repeating this process and averaging over the resulting test errors gives us the cross-validated test error. In most cases, it provides a good estimate of the real test error. **Expected Return:** The expected profit produced by an investment strategy.

Fundamental Analysis: Analysis of an organization's fundamental data: economic data that pertains to its financial or economic well-being. Fundamentals are also used in the valuation of a company.

Interpretable Trading Strategy: An automated trading strategy where the buying and selling choices done by the algorithm are comprehensible and justifiable to a human being. This is important for the investor to trust the suggestions provided by the software.

Investment/Trading Strategy: A set of trading actions (buy, sell or hold) linked with a set of possible future events in the market that fulfill a given constraint on risk, time horizon, or both. For example, the sentence "if company X's stock price falls below D dollars within the next T weeks, buy M stocks of that company" expresses an investment strategy.
Portfolio: A set of assets that compete for resources and generate value to an individual or organization. A portfolio has three characteristics: risk level, expected return and time horizon.

Quantitative Analysis: Any technique that seeks to model and understand real world phenomena using computational, mathematical or statistical tools.

Quantitative Trading: Trading that involves Quantitative Analysis of financial data.

Aggressiveness (of a strategy): Amount of risk and return a strategy is expected to yield. For example, an aggressive strategy can produce high return, but with high risk. **Technical Analysis:** Quantitative Analysis of historical market data, like stock prices, volume and volatility. Technical analysis is used to find trends that give an edge on the market to the investor.

Time Horizon: The desired time after which an investor wishes to receive a return on their investment, in other words to sell the assets of their portfolio. **Trading Strategy:** A set of instructions that specify when to buy and when to sell assets. For example, Simple Moving Average is a trading strategy that instructs to sell an asset when its price drops below a moving average value over a set time period, and to buy when it goes above it.

Volatility: The risk associated with an asset. Standard deviation of a stock's price can be an estimate for its volatility.

1. Executive Summary

In the early 20th century, the word ‘computer’ used to refer to a human who performs mathematical calculations. It seems as though the term ‘trader’ will meet the same semantic shift in the 21st century. Quantitative trading models - models used for trading financial assets that use computational, mathematical or statistical methods - are able to produce unprecedented returns on investment and have become more accessible than ever. The proposed platform aims to make use of such methods to generate a combination of stocks (or **portfolio**) that fits the requirements of each user in terms of risk, time horizon and expected return on their investment, while showing financial data (or **fundamentals**) and clearly explaining the reasoning behind the selection of each stock. In short, it can best be described as an automated financial portfolio manager or consultant that emphasizes beginner-friendliness.

By peeling away the complexity behind making investment decisions, the suggested platform attempts make stock market trading simpler and more accessible to people with little investment experience. The user’s risk, return, and time horizon preferences, collectively referred to as an **investment profile**, are learned through an interactive setup phase analogous to a first meeting with a human asset manager: the user is asked a series of questions that are used to infer an investment profile. After this step, the model generates a portfolio customized for each user. The platform enables the user to explore new stocks, but does not take action on their behalf as in Algorithmic Trading platforms. The platform can also provide basic justifications if a user wishes to know the reasoning behind a suggestion, by showing them fundamental data that motivates a certain decision, and also allowing the user to modify the portfolio and **backtest** it instantaneously.

The market consists of a very large diversity of **financial instruments** that could be used to meet different investment objectives in terms of risk, return on investment and time horizon. Each instrument can require complex and distinct trading techniques, and investors usually specialize in a specific **asset class** (e.g. Bonds, Futures). Since we emphasize on accessibility to beginners in investing, we’ll only be concerned with stocks from the NASDAQ and NYSE stock exchanges, and will ignore other, more sophisticated financial instruments. The model only has access to publicly available market data, which limits the range of historically available stock prices and fundamental data.

The proposed platform is divided into three main components, the first of which is an interactive questionnaire which serves as a **profiler** that guides the client through finding acceptable risk and return values for their investment. The second component is a **portfolio optimizer** that uses historical market data to generate a portfolio with the required expected return and risk values. The third component is a **portfolio explorer** that allows the user to view the fundamental data that supports the inclusion of each stock in the portfolio, while also allowing the user to modify the portfolio and backtest it immediately.

In the present report we define the platform at a detailed level, describe the design decisions and iterations it went through, and present the finished product, which includes the three main components of the platform: the **Profiler**, the **Optimizer**, and the **Explorer**. These components are supplemented by a **data fetcher** that fetches both market and fundamental data and a **Backtester** that tests the portfolio on the historical market data. Since the proposed platform is composed of independent functional components, our progress will be organized in steps corresponding to each component, and each new block building on previous blocks.

2. Acknowledgements

We are thankful to Dr. Fadi Zaraket for helpful discussion and comments and Mr. Marc Tohme for good advice about investing and general knowledge in finance.

Contents

List of Figures

List of Tables

3. Introduction

In the early 20th century, the word ‘computer’ used to refer to a human who performs mathematical calculations. It seems as though the term ‘trader’ will meet the same semantic shift in the 21st century. Quantitative trading models - models used for trading financial assets that use computational, mathematical or statistical methods - are able to produce unprecedented returns on investment and have become more accessible than ever. The proposed platform aims to make use of such methods to generate a combination of stocks (or **portfolio**) that fits the requirements of each user in terms of risk, time horizon and expected return on their investment, while showing financial data (or **fundamentals**) and clearly explaining the reasoning behind the selection of each stock. In short, it can best be described as an automated financial portfolio manager or consultant that emphasizes beginner-friendliness.

3.1. Motivation

Quantitative finance is gaining a lot of momentum in the age of abundant data and accessible trading software. One example is Quantopian, a new type of online crowd-sourced hedge fund where programmers can come up with trading strategies and receive funding for profitable ones. According to Quantopian’s website, their user base has been nearly doubling every year for the last four years [21]. However, this platform does not help non-technical people looking to invest, as its usage requires programming knowledge. It does help people who want to test strategies they have already come up with against historical prices. This motivated the creation of a simple and accessible quantitative finance platform that anyone can use to gain insights into the market and test portfolio performance on historical data, without any programming experience. Our platform attempts to make sense out of the torrents of data that increasingly permeate every sector of the industry through the portfolio optimizer. It also provides a portfolio explorer that clients can use to better understand investment choices, and allow the platform to suggest portfolios whose performance matches clients’ expectations.
A good example of a widely available platform that tries to solve some of the problems discussed is Portfolio Visualizer [2]. This website allows users to visualize portfolio performance on historical stock data. Another similar website is Bear N Bull [3]. However, this website allows even less customization in the portfolio optimization process. On the other hand, our platform allows the user the ability to configure the optimizer. For example, one optimization procedure could be minimizing risk while maximizing reward, while another might be solely maximizing reward, or solely minimizing risk, or optimizing some other portfolio property other than risk or reward.
Another popular online platform that addresses some of the same problems is Fidelity [38]. The problem with Fidelity is the complexity that the user faces when he or she attempts to create an account. They are expected to have extensive financial knowledge in order to use the platform properly, namely knowing how to pick stocks and interpret financial metrics. To cater for this, our platform includes the Stock Explorer, an interactive tool to explore and learn about the financial health of companies in an accessible way.

3.2. Desired Needs

Most modern Quantitative trading software performs technical analysis, and builds a portfolio based only on historical market data, which is known to yield uncertain results at best [10],[16]. Also, only a small fraction of a company's total shares is ever traded on the stock market at any point in time. For example, AAPL's 50-day average daily volume was 26,463,819 as of this writing, and it has 4,715,280,000 shares outstanding. This means that only 0.5 percent of shareholders agreed with the market price enough to trade in the past 50 days. The minority of shareholders who do trade are what determines the market price of a stock. Market price may therefore be a poor indicator of a company's real worth, so technical analysis alone is not enough to make informed investment decisions. A company's financial promise, hence its attractiveness to an investor, can better be gauged using its fundamental information. A vigilant investor would need to know things about a company like how it manages debt or whether its earnings and equity are consistent and increasing across time. QuantInvest performs a combination of risk modeling, portfolio optimization, and fundamental analysis, combining both technical data (e.g. market price, volatility) and fundamental data (e.g. debt, cash flow) to provide a client with better insights about companies' financial promise.

Another desired feature for the platform is to provide suggestions and explanations that can serve as guidelines for an investor. Investors expect the trading strategies to be meaningful, or else will reconsider the investment advice given.

4. Requirements and Deliverables

QuantInvest can best be described as an automated financial portfolio manager that focuses on the beginner user-experience. It will not serve as a platform to perform trades, such as InteractiveBrokers, nor will it be an Algorithmic Trading platform, such as AlgoTrader, since it does not involve automation in the execution of trade signals, in other words, the decisions to sell or buy a stock are not executed automatically, and are left to the user. The next section describes the platform's high-level functions. Each high level function relies on a set of components, indicated between parentheses. The low-level functions are detailed in the specifications section.

4.1. Requirements

1. To provide investment suggestions on a given portfolio. A user should be able to input an existing portfolio of theirs and have the platform suggest modifications to improve one of its three parameters (risk level, expected return and time horizon) keeping one or two others fixed. (A and B)

2. To construct a suitable portfolio for a user given a set of constraints or preferences. For example, a user may require that their initial investment of $10,000 produce a return of 50% after five years, and that their money be invested only in real estate and utilities. In that case the return level is 50%, the time horizon is five years, and the risk level is conservative since real estate and utilities are non-volatile sectors of the market,

and such an expected return is comparable to that of a government bond. In general the user cannot immediately state the risk they are prepared to take as a number, so QuantInvest's Profiler component infers their risk aptitude and other parameters from an interactive questionnaire. (A, B and C)

3. To simulate likely future outcomes of a portfolio on an interactive and easy to understand graphical interface. This serves two purposes: to assist the user in determining their preferences for a strategy upon initialization, and to forecast the performance of a portfolio, and visualize it on a graph enabling the user to select or modify a strategy. (A and C)

4. To suggest companies to add or remove from a portfolio, give understandable explanations, and show the data that supports it. It will be able to show a user the logic behind a trade decision, pointing the user to the fundamental quantities that support it. The data that supports a trade decision could be displayed in a clean presentable form upon the user's request. (A and C)

5. To enable a user to browse and examine a company's financial situation on a clear and beginner-friendly interface without any technical jargon and long lists of numbers which are difficult to interpret. Important parameters and statistics about a company are plotted cleanly for a user as an overview, with the option to look into the cash flow statement, income statement or balance sheet of a company for more investigation.

The following subsections contain the detailed specifications of the platform.

4.2. Specifications

A **Backtester**
Description: This component serves two functions: it tests a strategy generated by component C on historical market data to see how well it would have performed in the past . It provides an important proof of validity of portfolio performance, which has to go through backtesting and perform well before being suggested to the user. The user could get a piece of information such as "if you had followed the suggested strategy a year ago you would have twice the amount of money you do now." thanks to the backtesting component.
Languages: Python
Input ports: Relevant dataset from component E, set of portfolios to test.
Output ports: Testing error (when used to verify the validity of a generated statistical model), investment return estimate
Connectors: Is called by component D to visualize how a current strategy would have performed in the past, called by component A to obtain cross-validated test error on a certain model.

B **Portfolio Optimizer/Strategy Generator**
Description: This component constructs a set of stocks that match the user's requirements in terms of risk level, expected return and time horizon. It does so by running

an optimization algorithm which is constrained by the desired inputs of the user. For example, a user could request an existing portfolio's risk to be minimized while keeping expected return and time horizon fixed. It can also generate portfolios given a certain number of preferences set by the user. For example, here is a portfolio: AAPL 30%, GOOG 50%, AMZN 20%. It uses a subcomponent to estimate risk in order to fulfill the risk requirement (or minimize it).
Languages: Python
Input ports: a portfolio, a set of constraints
Output ports: a portfolio
Connectors: Is called by component D when a user clicks on the "Optimize portfolio" button.
Sub-components: Risk estimator component, optimization algorithm.

C **Interactive Graphs and Displays**
Description: An interface consisting of interactive graphs, pop-up windows and data visualization windows. Graphs are clickable and provide forecasts for the user's desired strategy or for a suggested strategy.
Languages: JavaScript, HTML, CSS
Input ports: User clicks
Output ports: GUI
Connectors: Calls component A to request a forecast, calls component B to display a backtested portfolio.

D **Databases**
Description: Conain market datasets, as well as user profile and authentication data.
Database: MongoDB
Input ports: pandas DataFrame object
Output ports: values
Connectors: is called by component B to backtest, is called by component A to train models.

4.3. Deliverables

A website fulfilling the above requirements and specifications.

5. Technical and Non-Technical Constraints

5.1. Technical Constraints

5.1.1. Limited Computing Power

Given the computational resources available, generating daily predictions and portfolios for a large number of clients becomes infeasible at some point. Hence, the kinds of securities the platform can trade have been limited to stocks only, excluding options, bonds, foreign exchange and other commodities.

5.1.2. Free Market and News Data

The proposed software uses publicly available market data and news data. The market data is made up of around 3000 U.S. stocks. More details provided in the progress description, section 9 of this report.

5.1.3. No Trading Signal Generation

Since the proposed platform focuses on collecting stock prices and news data, and optimizing a portfolio of assets, which requires substantial amounts of computing, this excludes it from the category of high frequency trading, where the strategies make use of statistical arbitrage and trend following to make millisecond profits. Therefore, the software does not offer the option of executing the trade signals automatically, since speed is not an immediate concern. The task of sending the trade order to the broker is left to the user.

5.2. Non-Technical Constraints

- The provided investment software should make clear to users the risks involved in trading, and that the developers are not responsible for any losses that might result from using it. It must also advise the clients to refer to private financial advisors before making any investment decisions. This would clear the developers of any liabilities.
- The terms and licensing of off-the-shelf software used in the project, most likely creative commons, will be respected and honored, and all used resources will be mentioned in the report.

6. Literature Review

Quantitative Trading (QT) combines elements of statistical analysis, econometrics, constrained optimization, and software engineering, among other topics. Because of the breadth of the concepts involved, there is no coherent body of literature concerned with QT systems as a conceptual unit. However, there are well-defined and distinct components within a QT system, each of which will be treated in a separate subsection.

6.1. Financial Indicator Based Movement Predictor Models

The question of correlating technical market indicators with stock price movements has been the subject of numerous research papers. This section summarizes a relevant subset of these papers. Zheng et al. [22] proposed a model to predict short-term price movements using time series data of stock prices only. Their attempt was based on using financial technical indicators alone such as relative strength index (RSI) values, bollingerbands (BBands) values, CCI (commodity channel index) values, and 14 others as inputs in their machine learning models. However the paper's interpretation of the results was questionable, as only a single stock (microsoft MSFT) was considered, and prediction accuracy of 92.3% for price changes of greater than 2$ was claimed, which is not something that can be generalized. Financial models usually operate on percent changes instead of actual prices since they could then be generalized to any stock. As further elaboration, if given a smaller company with a stock price of 10$, a 2$ change would be highly improbable, and using such a model would yield very inaccurate results, whereas a model trained on percentage changes would do much better. The proposed software's stock price prediction model will be using artificial neural networks combined with market data and financial indicators to produce an explainable high prediction accuracy model. By explainable model, it is meant that for every prediction, a backpropagation of the neuron activations results in a weight assigned to every input. In this way, the input contribution to the prediction can be visualized and thus the prediction is 'explained'. In the presence of financial indicators as inputs, assigning weights to indicators will make sense to investors since the indicators are human interpretable, and not machine learned features.

6.2. News Feature Based Movement Predictor Models

In this section, the focus will be on engineered news based features used in machine learning models to help predict stock price movements. Hagenau et al. [9] engineered news momentum related features to predict weekly stock price movements. The first feature was a news tone aggregate based feature, which was comprised of the sum of the "tonality" values of single articles concerning news of a certain stock. Here "tonality" is a sentiment metric developed by Michael Hagenau et al. in [12]. The second feature was a news proportion aggregate, which measured the ratio of positive tonality news to all news concerning some stock. The main model in that paper was a rule-based decision model that predicted price increases if the current week tonality aggregate was greater than that of the previous week, and a price decrease otherwise. This yielded a maximum of 71.5%,which is substantial relative to other papers reviewed. Kranti et. al [20] combined data mining and time series analysis to predict the stock price movements. The model developed in this paper relies on historical stock prices and market news, which are fed into a neural network. The prices, risk-adjusted prices, volumes, and booleans representing the presence of some key words in the text are fed in at the input layer, and the output layer attempts to predict the trading day's closing price for a certain stock. The proposed software's stock price movement prediction model will integrate news sentiments in the form that Hagenau et al. did, as well as add a stock specific feature that indicates the news information propagation delay to the market. By that it is meant that some news affect the market directly, while others take longer to do so. An attempt to estimate this delay will yield better stock price movement predictions. This idea is inspired from behavioral finance, where it is common knowledge that market reactions to different news can differ substantially between companies, even if the news is very similar in content.

6.3. Portfolio Optimization

Portfolio optimization forms one of the most researched fields of finance, as it poses a very complex problem: trying to integrate various factors such as investor views of the market with minimizing portfolio risk and maximizing portfolio return. The first major breakthrough in the field was made by Harry Markowitz, after whom Markowitz portfolios were named. Markowitz modelled the rate of return of assets as random variables, and then chose the portfolio asset allocations in such a way as to maximize return and minimize risk. This model considered the optimal set of weights to be the set that minimizes volatility while achieving the expected rate of return. Formally, for n assets, let w_i be the weight of asset i in our portfolio, such that $\sum_{i=1}^{n} w_i = 1$, and r_i be the random variable representing the rate of return of asset i, and $\mu_i = E(r_i)$. Now let

$$\mathbf{r} = \begin{pmatrix} r_1 \\ \vdots \\ r_n \end{pmatrix}, \mu = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_n \end{pmatrix}, Cov(r) = \Sigma$$

Defining portfolio return as a random variable r_p, portfolio expected return $E(r_p)$, and portfolio volatility $\sigma^2_{r_p}$ then:

$$r_p = r_1 w_1 + r_2 w_2 + ... + r_n w_n = \sum_{i=1}^{n} r_i w_i$$

$$E(r_p) = E(r_1 w_1) + ... + E(r_n w_n) = \sum_{i=1}^{n} E(r_i w_i) = E(r)^T w = \mu^T w$$

$$\sigma^2_{r_p} = E[(r_p - E(r_p))^2] = w^T \Sigma w$$

The Markowitz theory proposes minimizing the portfolio variance, while meeting the minimum expected reward (set by the investor). This optimization problem can be formalized using lagrangian multipliers in the following way, and solved using any optimizer:

$$\begin{aligned} \min_{w} \quad & \frac{1}{2}\sigma^2_{r_p} \\ \text{s.t.} \quad & \mu^T w \geq \mu_{p_{min}}, \\ & w^T 1 = 1 \end{aligned} \tag{1}$$

Despite the theoretical success of Markowitz's mean-variance model, it is not frequently used in practice. The major drawback of using this mean-variance Markowitz optimization method is that it usually leads to undiversified portfolios. As put by Michaud [19], "Although Markowitz efficiency is a convenient and useful theoretical framework for portfolio optimality, in practice it is an error prone procedure that often results in error-maximized and investment irrelevant portfolios." This error obviously comes from the random variable statistics themselves, which have been derived from historical data. In addition, investor views of the market or alternative news source influence cannot be integrated into such a model. In some cases, investors have external information about some stock future movements, and would obviously want to integrate this information into the model dynamically. This cannot be done if a simple Markowitz mean-variance portfolio optimization model is used. These issues were solved in more advanced models such as the Black-Litterman model [18], [8], which is now widely available in mathematical packages. Other artificial intelligence search algorithms fit the problem very well, such as covariance matrix adaptation evolution strategies [7], and genetic algorithms.

The PyPortfolioOpt [33] package is based on Markowitz's mean-variance portfolio optimization. This type of optimization relies on a risk model, a way of quantifying the assets' risk. A simple risk model is the covariance matrix Σ, which contains information about each variable's variance and the covariance of every pair of variables. PyPortfolioOpt enables the user to choose from different risk models other than sample covariance matrix. We will not discuss all the possible ways to model risk, as it is a very broad and complex endeavor and is beyond the scope of our project. In this project, the covariance matrix is used as a risk model.

6.4. Other QT software on the Market

Multiple software solutions exist for all the needs of a quantitative analyst. Here is a non-comprehensive list of some popular software:

	Free Data	Alternative Data	Portfolio Optimization	Data Analysis
AlgoTrader	YES	NO	YES	NO
Quantopian	NO	YES	YES	NO
Tradestation	YES	NO	NO	NO
NinjaTrader	NO	NO	NO	NO
Multicharts	NO	NO	NO	NO
Google Finance	YES	YES	NO	NO
Yahoo! Finance	YES	YES	NO	YES
Finviz	YES	YES	NO	YES

Table 1: Comparison of different algorithmic trading platforms. Sources: [23] [24] [25] [26] [27]

7. Applicable Standards

The proposed platform (QuantInvest) will be adopting the MIT License, stated in Appendix B, and the deployed platform contains the legal information included in Appendix A.

8. Proposed Solution Methodology and Initial Design

8.1. Evaluate the best stock predictor model

The first step to build the stock prediction model is to gain experience on Quantopian. The tools and documentation provided there are enough to cover the basics of technical analysis and to become flexible with new types of data thanks to the versatile pandas package. Figure 1 is a screen capture of a piece of code and graph from a Quantopian tutorial depicting the simple 20-day and 50-day moving average of an AAPL stock from 2013 to 2015. This simple strategy instructs to buy when the 20-day average moves above the 50 day average and sell when it moves below. This is the simplest type of trading strategy. The easy-to-use backtesting framework on Quantopian will give accurate feedback on the performance of our strategies.

In [18]:

```python
# Research environment functions
from quantopian.research import prices, symbols

import pandas as pd

aapl_close = prices(
    assets=symbols('AAPL'),
    start='2013-01-01',
    end='2016-01-01',
)

# Compute 20 and 50 day moving averages on
# AAPL's pricing data
aapl_sma20 = aapl_close.rolling(20).mean()
aapl_sma50 = aapl_close.rolling(50).mean()

# Combine results into a pandas DataFrame and plot
pd.DataFrame({
    'AAPL': aapl_close,
    'SMA20': aapl_sma20,
    'SMA50': aapl_sma50
}).plot(
    title='AAPL Close Price / SMA Crossover'
);
```

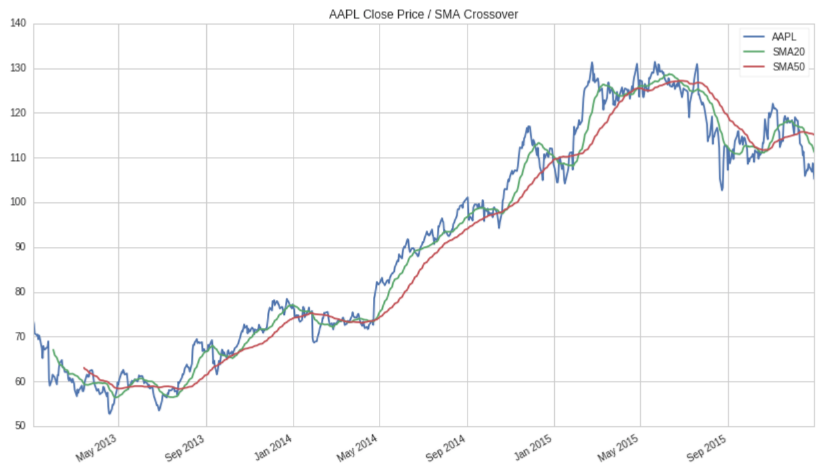


Figure 1: Screen capture from a Quantopian tutorial

8.2. Research on stock prediction models

Literature suggests many available models to predict stock returns, including but not limited to AR (Autoregressive) models, non-linear models such as neural networks, decision-tree based models, support vector machines, and other statistical models.

8.3. Exploring Alternative Datasets

This will be a purely exploratory exercise which we will work on continuously throughout the year.

Source	t Fetch	Data range	Constraints	Offline availability
Quandl	2.02	1996 - 2017	50000 calls per day	available
iex	1.7	5 years past - present	over 100 per second	not available
robinhood	0.829	1 year past - present	undeclared	not available

Table 2: Comparison of different data sources. Sources: Tests performed, and the following [13] [14] [15]

9. Progress Description

The progress on the software so far has been distributed across all main components: the stock price movement predictor using financial and news data, the portfolio optimization model, and the graphical user interface for interacting with clients. Progress has been going on smoothly and the platform is ahead of schedule with all components currently in progress. For each component, the progress made will be described, along with design decisions made and alternatives that were not chosen and for what reason.

9.1. Preliminary Design Alternatives

9.1.1. Data Sources

First of all, for the market data, several data sources have been tested, and the results can be summarized in the following table:

Notice that even though Yahoo! Finance and Google Finance are popular market data sources, due to recent API changes, they have to be manually scraped to extract data. The effort required to perform that would render parts of the platform infeasible in the given time span, and the difference is immeasurable between different APIs given the software's needs. The InvestorsExchange (IEX) seems to be the only source providing present data for free, and hence it has been chosen as the data provider for the proposed platform.

9.1.2. Stock Movement Predictor

Basic trading strategy review has been done, and basic strategies such as OLS regression predictor based strategies, simple moving average crossover, and double moving average crossovers have been tested to make sure the data available is what was expected, and everything is running smoothly.

9.1.3. Graphical User Interface

After comparing several GUI libraries, a python based GUI seemed well suited given the ease of communication between components written in the same language. This would allow for faster and more efficient development. The python GUI Tkinter library seemed to be the best fit, as it has been used for stock visualization programs previously, coupled with matplotlib for charting and data visualization. Alternatives included Dash, a python GUI library designed by plotly, but was rejected as an alternative due to the low amount of control given to developers. Another alternative was Kivy, and that was rejected due to bad documentation and no prior publicly available usage of graphs in Kivy GUIs. In the end, charting in python seemed too simple, not allowing for any interactivity.

The solution chosen was using the Electron framework built on the chrome V8 engine to design the GUI. The python models will interface with the Electron GUI as binaries, accessible through a well-designed API. This allows for better visualizations at a negligible speed cost.

9.1.4. Portfolio Optimization

Several models have been investigated, such as Markowitz mean-variance based models, Black-Litterman models, and evolutionary strategy based models. The plan is to implement each and compare the results in the final phase of the project.

9.2. Preliminary Implementation

Further research has been done into portfolio optimization models, and several quantitative finance books and courses are being covered by the developers to further enhance the proposed software. Working python notebooks have been developed to act as a base for future explorations and model developments, with some extracts available in the preliminary testing section.

9.3. Preliminary Testing

The first component worked on was the stock price predictor component. This had to use the data sources and software libraries, hence ensuring the base of the platform is built. In Figure 1, it can be seen how using percent changes instead of closing prices can result in near normal distributions, as discussed in the literature review section concerning Zheng et al.'s results.

```python
# Using percentage changes to get a near normal distribution
aapl['10d_close_pct'] = aapl['Adj. Close'].pct_change(10)
aapl['10d_close_pct'].plot()
plt.show()

aapl['10d_close_pct'].hist(bins=100)
plt.show()
```

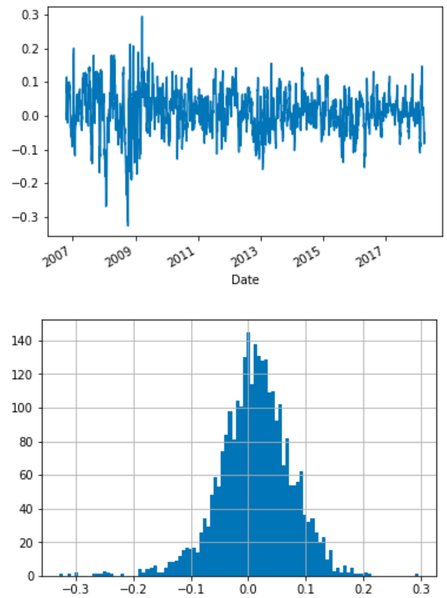


Figure 2: Visualizing distribution of close prices vs. percent changes in close prices

Performing the Shapiro-Wilk test for normality however, it is found that the distribution is very far from normal.

```python
from scipy import stats

print(stats.shapiro(aapl['10d_close_pct'].dropna()))
```

```
(0.980026125907898, 1.0543773353945954e-19)
```

Figure 3: Shapiro-Wilk test for normality

Further exploring stock data, it is tested for correlation between historical close prices and future close prices, seperated by 10 days. What is found is seemingly great, but deceiving.

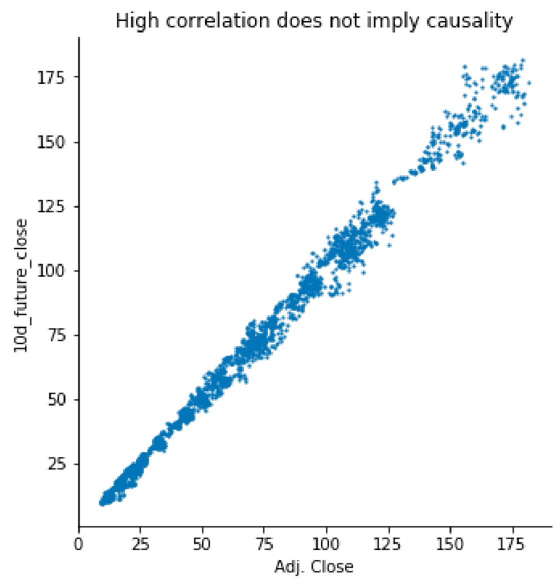


Figure 4: Correlation between close price and 10 day future close price

This correlation is large because the raw magnitude of the prices can't change that fast in 10 days. However, this is deceiving. The range of future prices based on the current price is too large to be useful, and hence this is meaningless. For this interpreting data correctly is very important when working on the proposed model. Correctly evaluating our results is as big a step as getting to the results.

Moving on to more significant results, below is an example extract from implementing a simple moving average crossover strategy:

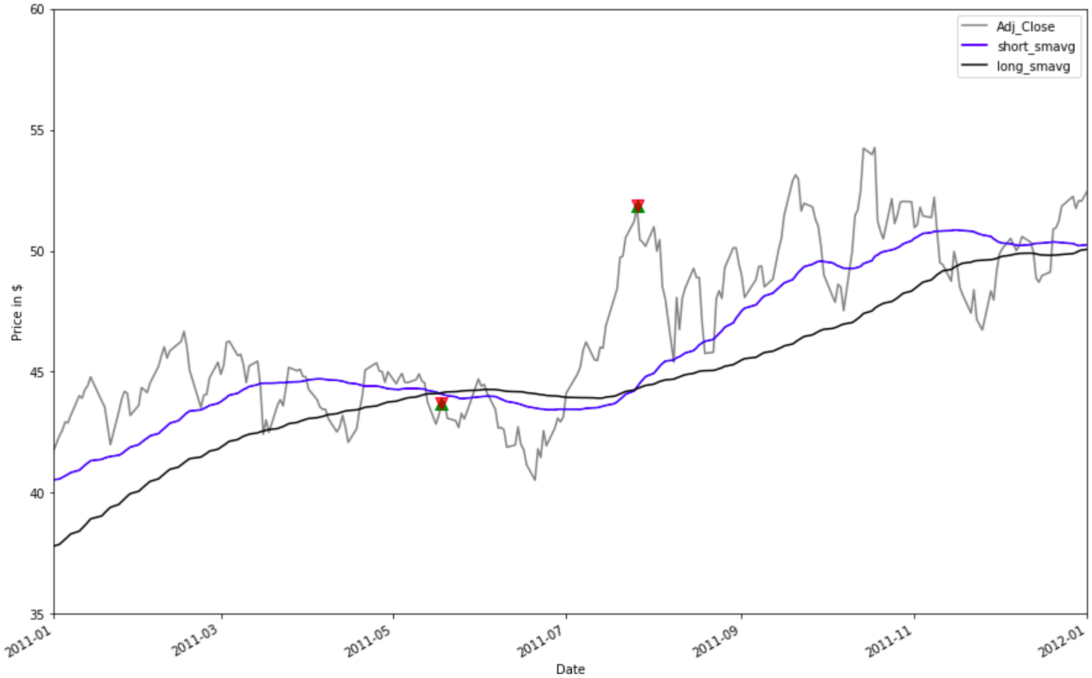


Figure 5: Simple moving average crossover strategy with 2 signals visualized

It has also been attempted to engineer basic features such as weekdays, and checking their importance in predicting stock price movements. The following Figure 5 shows the correlation matrix obtained:

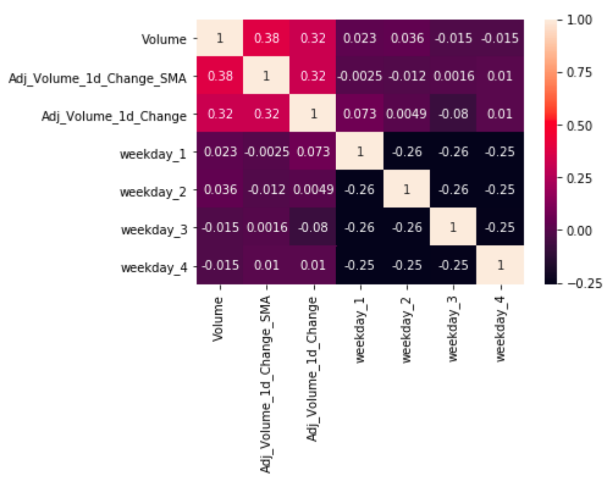


Figure 6: Correlation between different features and engineered features

Even though the correlations are weak, the engineered features can still help by interacting with other features in predictions.

9.3.1. Autoregressive Models

Basic AR models take only one lagged value into account, so the following formulation is formed. Let P_t be the time series of some stock. An AR(1) model is simply:

$$P_t = \alpha + \beta P_{t-1} + \epsilon_t$$

where ϵ_t is some noise.

From this formulation, it is seen that beta was positive, then the lagged series is positively correlated with the non-lagged series, which suggests that the time series is trend following, as the current value is highly correlated with the previous value. On the other hand, if beta was negative, then this would suggest that the series is mean reverting. Integrating several lagged values into the model, a model that could give us insights about the future behavior of a stock is formed. Below 9.3.1 is an example AR(1) model forecasting interest rates:

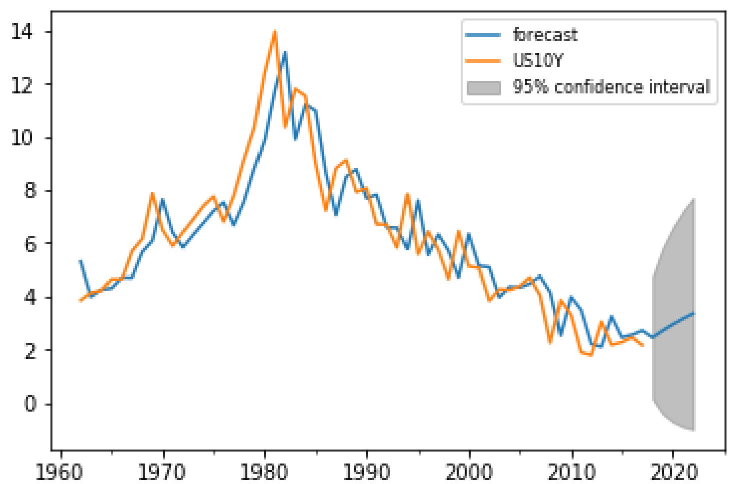


Figure 7: Autoregressive Model with Lag 1 - Forecasting Interest Rates. Author: Rami Awar

Taking this further, more lags can be incorporated, depending on how statistically significant their correlation to the current value is, and more advanced models can be created by integrating moving average models (MA) to get ARMA and ARIMA models.

A more detailed review will be provided as an appendix in the final version of this report.

9.4. Backtester

Another component that we have started implementing is the backtesting framework. It will be built using Python and the Zipline library, which is what Quantopian uses to do their backtesting [28]. Using Python would also give us access to a large amount of statistics and machine learning libraries.

9.5. Final Design and Implementation

9.5.1. System Design

Before discussing the individual components in detail, a general overview of the system must be revealed to make things clearer. The following diagram represents an eagle-eye view of QuantInvest's system architecture.

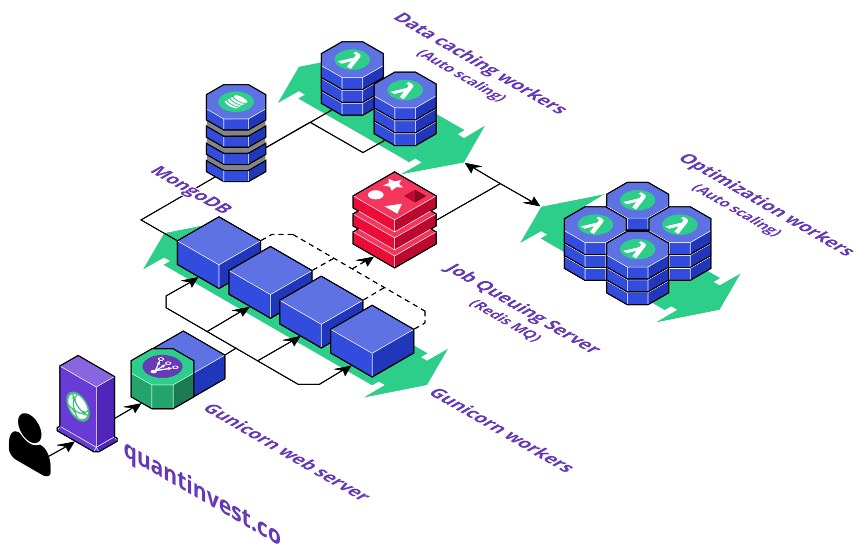


Figure 8: QuantInvest System Architecture Overview

Beginning with the web server, QuantInvest uses Gunicorn, a python-based production-grade web server. This server gives the option to customize the number of worker threads that can process incoming network requests at a time, hence making our web-server quickly scalable as the number of users increase. Moving on to the server backend, QuantInvest uses the Flask python microframework. This encompasses routing, rendering html templates to the user, authentication, etc.

Using Flask allows the usage of Flask's Blueprints structure [40], which breaks down the server code into very modular components. For example, QuantInvest uses an 'api' blueprint, which includes all routes and handlers relating to api calls. This also allows for url prefixing all routes within a certain Blueprint, allowing for better organization even from a user's perspective. A call to fetch data would be "https://quantinvest.co/api/stockfetcher/AAPL". As another example, all authentication routes are encompassed under the '/auth' prefix, and all authentication code is inside the 'auth' blueprint.

As a new user visits quantinvest.co and signs up, a hashed and salted version of their password is saved in a MongoDB database, along with their username and email. QuantInvest uses mongoengine as a means of communicating with MongoDB, and this is in order to enforce some structure to the data being saved in the database. This circumvents the issue of migration, and allows for more flexibility when in need to save some unstructured data. This database is also used to cache stock price data after market close every single day. Caching will be discussed in more detail later on.

As for the optimization, fetching data, performing the optimization to generate portfolios, and backtesting these portfolios takes some time. These calls cannot be blocking, as the browser will probably mistake the long wait for no response and terminate the connection. Also, the user should be able to see some feedback about progress so as to understand that his call was registered and will be done soon. For this reason, QuantInvest uses a Redis server as a message queuing server to enqueue jobs that need to be performed [34]. Hence, whenever a user requests some optimization, an optimization job is enqueued for workers to perform. Then, whenever a worker is free, a job is taken out of the queue and completed, and the result returned to the queue. This implementation allows for horizontal scalability, since as the number of users increase, the number of workers could be increased to account for the increased number of calls. This elegant solution is a standard one for such scenarios. This job queuing architecture is also used for our data-caching tasks, which will be mentioned later.

This redis message queuing server is shown as the red stack in the diagram above. The optimization workers are the workers that handle these optimization jobs after popping them from the job queue. The data caching workers handle the caching jobs. The Gunicorn workers are created by the Gunicorn web server to handle multiple simultaneous incoming requests from several users.

9.5.2. Risk Profiling

Just like a traditional asset manager, QuantInvest begins by getting to know the client. The purpose of Risk Profiling is to understand a user's aptitude for risk, their tolerance for risk and their investment objectives. Typically, a user will not be able to give a number to quantify their risk aptitude, so the Risk Profiler infers this information through an interactive questionnaire. The questionnaire is based on actual documents used by asset management firms, and is included in the Appendix.

The goal of this questionnaire is to obtain three numbers using the responses of the user: their aptitude for risk (measured on a scale from 1 to 6), the time after which they would like to access their returns (in years), and the minimal amount of return they wish to make (as a percentage of their initial investment). Fundamentally all three numbers are related to each other and it is possible to rank users using a single compound variable, without losing too much information. For our purposes we need three distinct numbers to construct the portfolio accordingly.

Also, there is a difference between willingness to take financial risks and the ability to do so. A user whose source of income is not very secure cannot opt for an aggressive risk strategy, and a user who owns many assets and has secure income may want to opt for capital-preserving investments even if they are able to make risky investments. We refer to the ability to take risks as "risk tolerance". If we find that their willingness to take a risk exceeds their risk tolerance, we could issue a warning that says something like: "based on the information you provided, it seems you aren't financially secure enough to undertake such a risky investment strategy. Please reconsider the plausibility of your investment objectives."

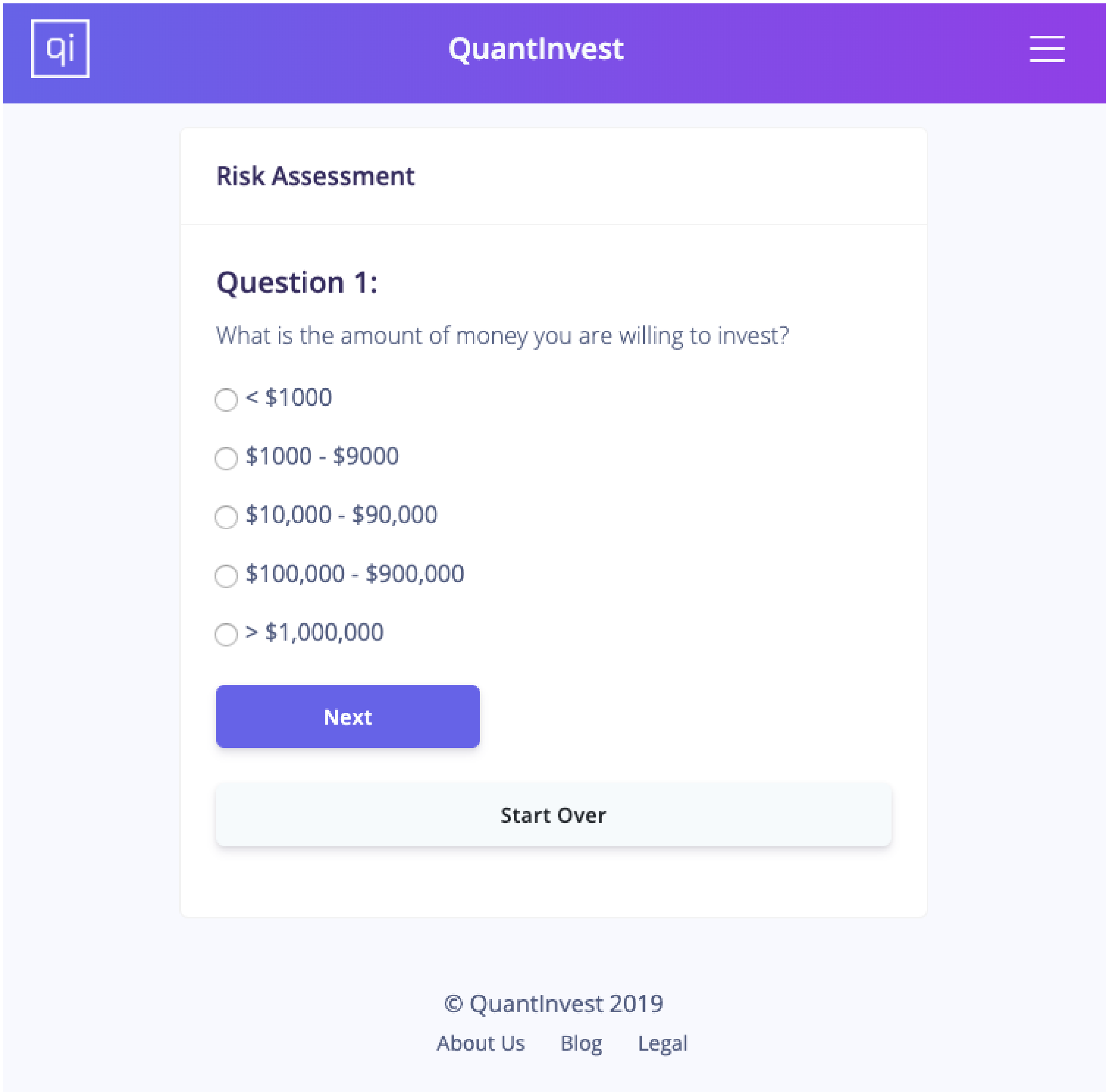


Figure 9: QuantInvest Risk Profiling Page

9.5.3. User Interface Design

For the user interface, the website is designed purely using vanilla HTML, CSS, and Javascript. The Argon design library was used for CSS and Javascript components and styles [41].

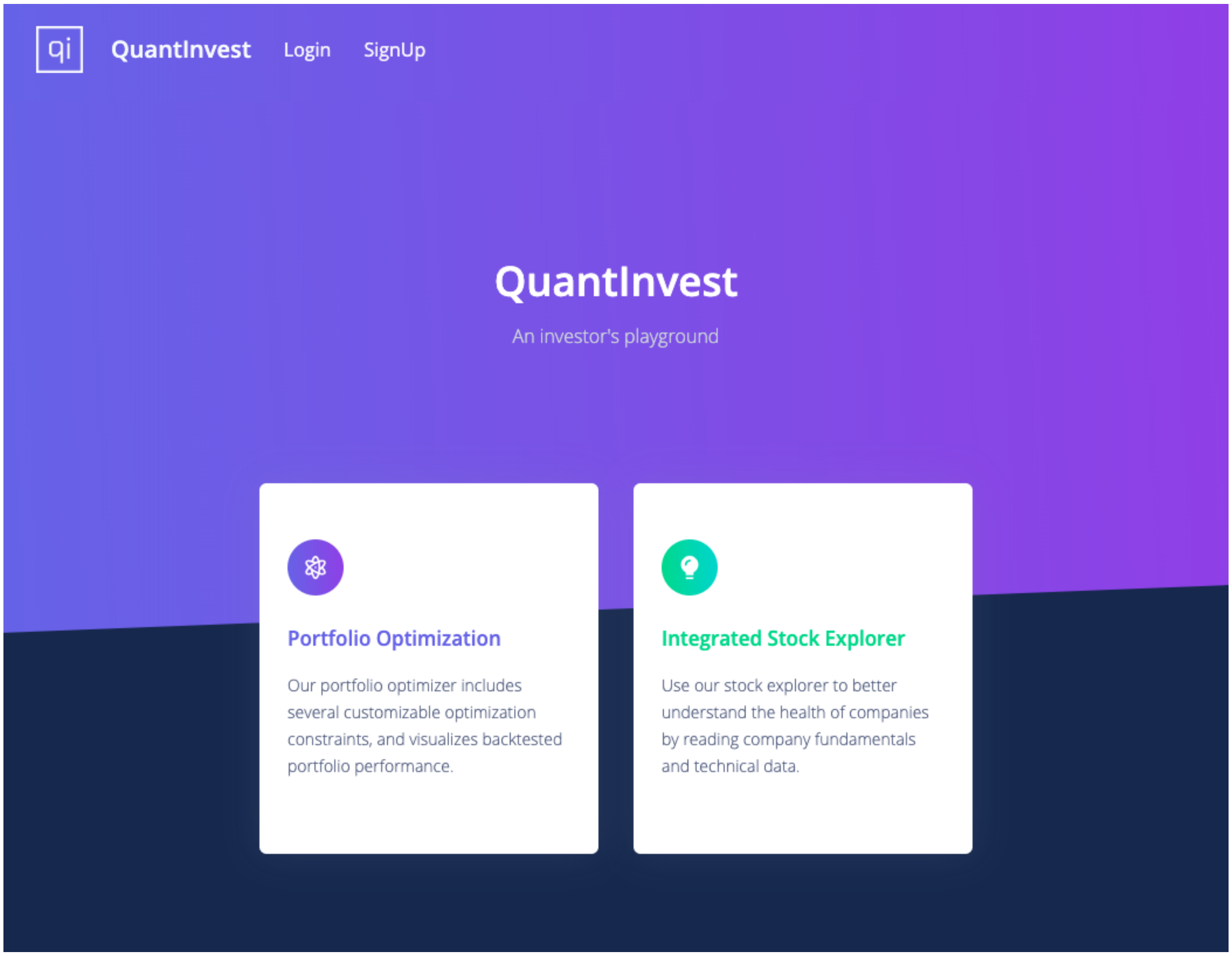


Figure 10: QuantInvest Landing Page

The interface includes a landing page, that then redirects to a signup/login page. The standard new user flow when starting with QuantInvest is signing up, visiting the dashboard page, and choosing whether or not the user is a beginner or advanced investor.

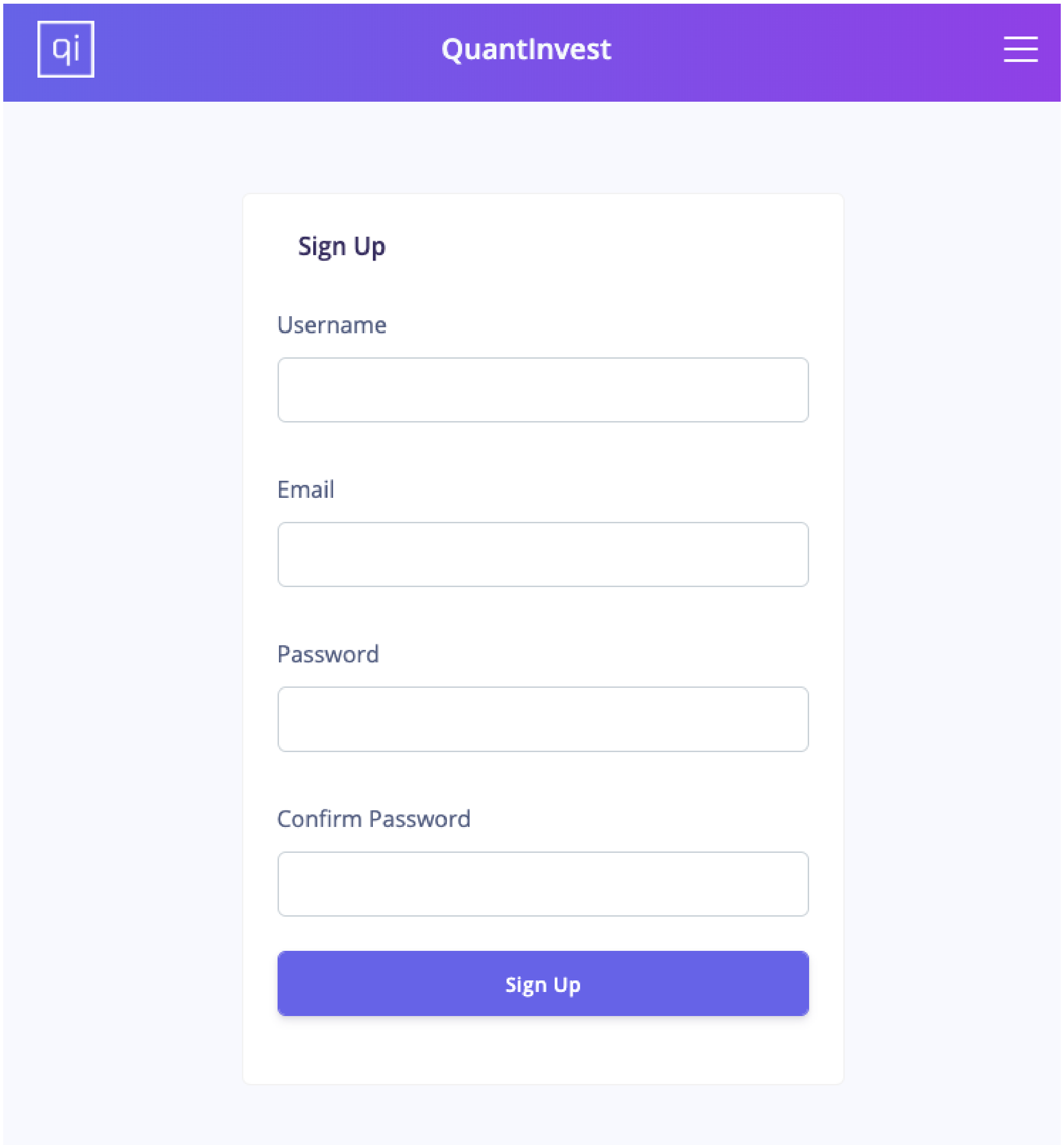


Figure 11: QuantInvest Registration Page

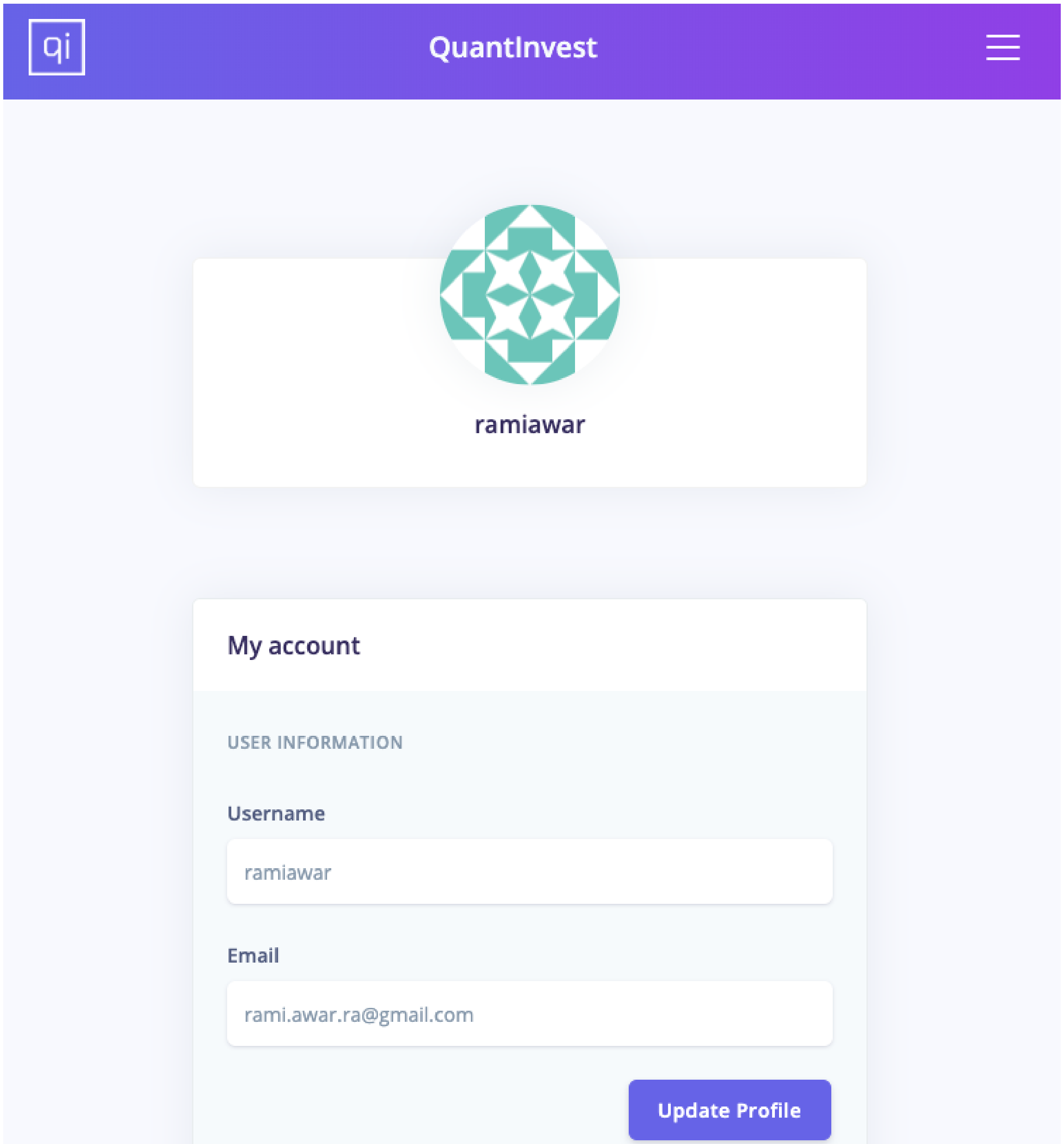


Figure 12: QuantInvest Profile Page

In case the user is a beginner user, they are directed to QuantInvest's profile extractor quiz, which generates a risk score after asking the user several questions that act as a risk assesser (mentioned in risk profiling section).

After that, the user is redirected to the basic optimization page, where they can further explore the effects of different expected return and volatility values in generating portfolios. The basic and custom optimization pages offer a helpful text at the beginning of the page, explaining to the users some of the concepts and inner-workings of both optimizers.

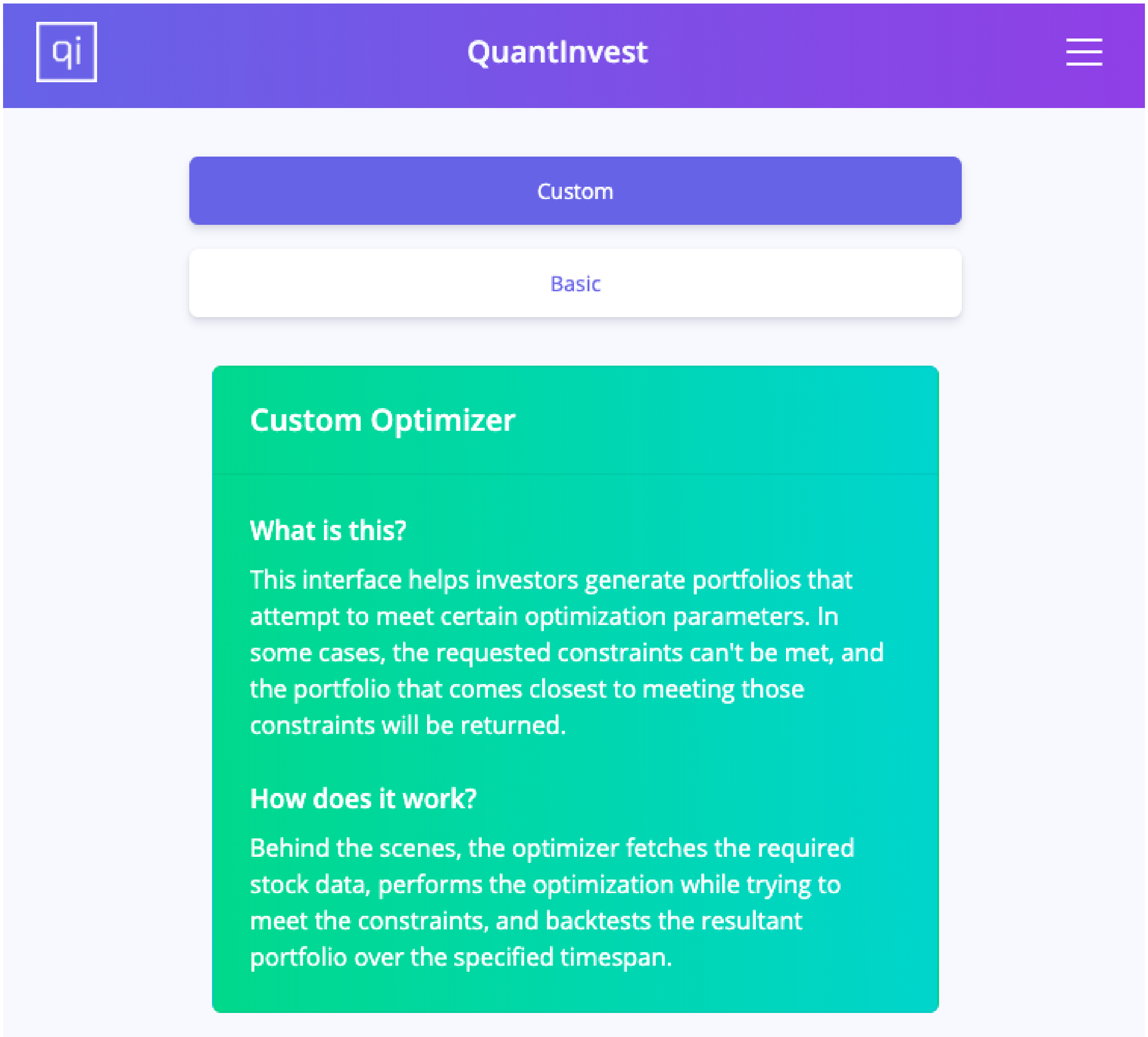


Figure 13: QuantInvest Custom Optimization Page - 01

At this point, the new users will have generated their first optimized portfolio, and can choose to save the portfolio into their portflios page, or explore the portfolio further. If they choose to explore the portfolio, they are redirected to the portfolio explorer page, which provides more details about the individual stocks, and allows the users to explore what happens if they modify the weights themselves and perform another backtest.

For returning users, those would come back to either check their existing portfolios performance, generate new portfolios, or perform analysis on stocks and portfolios using QuantInvest's stock and portfolio exploration tools.

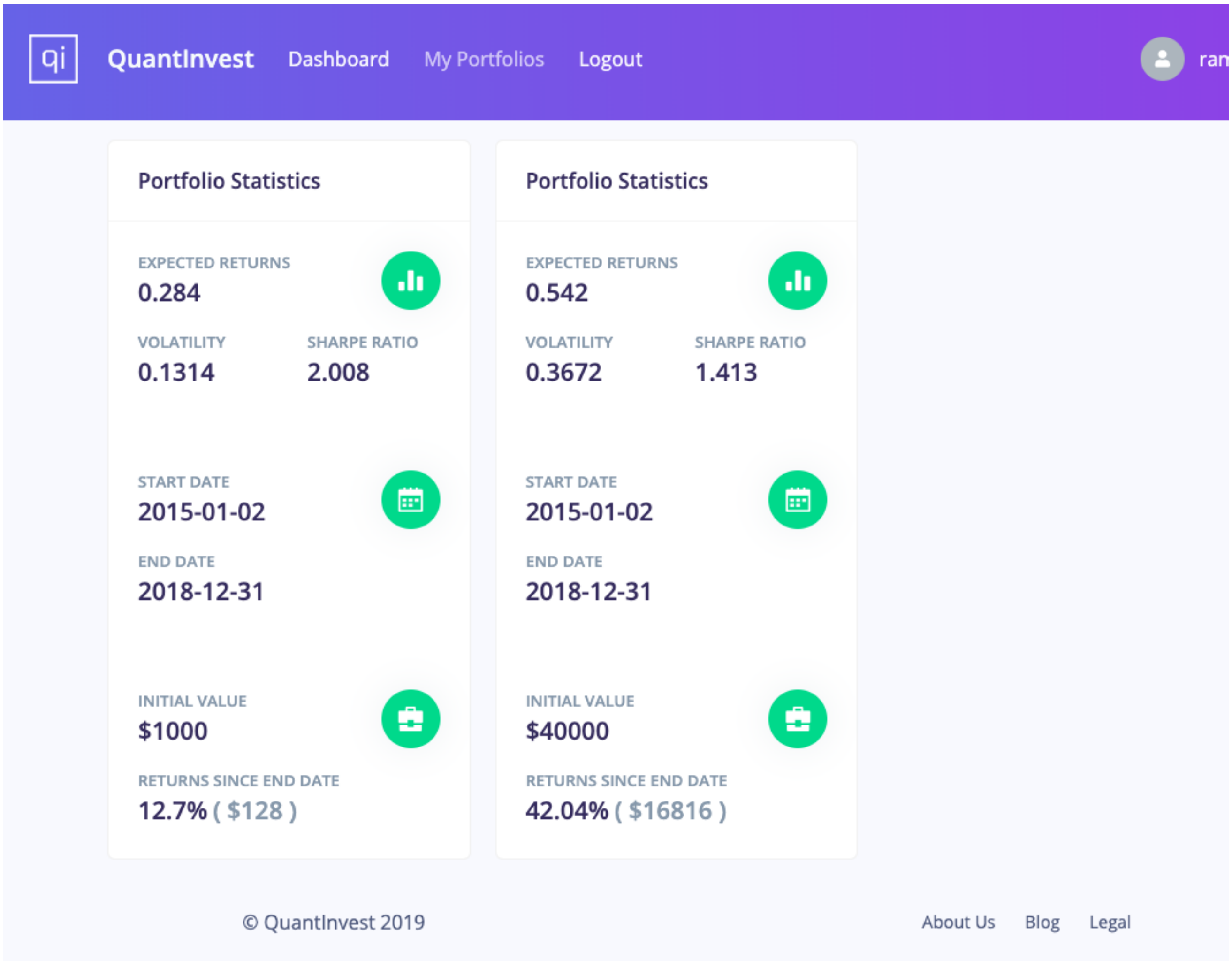


Figure 14: QuantInvest My Portfolios Page

9.5.4. Portfolio Optimizer

The portfolio optimizer is a major component in the first design iteration of QuantInvest. The functionality is divided across two pages in the user interface; the first page is titled "Basic Optimizer", while the second offers more advanced optimization options and is titled "Custom Optimizer". The third subsection of this section gets into the details of the efficient frontier optimization, and how it is implemented in python.

Basic Optimizer

The basic optimizer is designed for beginner users that still haven't fully understood the relationship between portfolio volatility and expected return. It offers two sliders which allow the modification of these two variables, and then visualizing the results of this modification after performing the optimization.

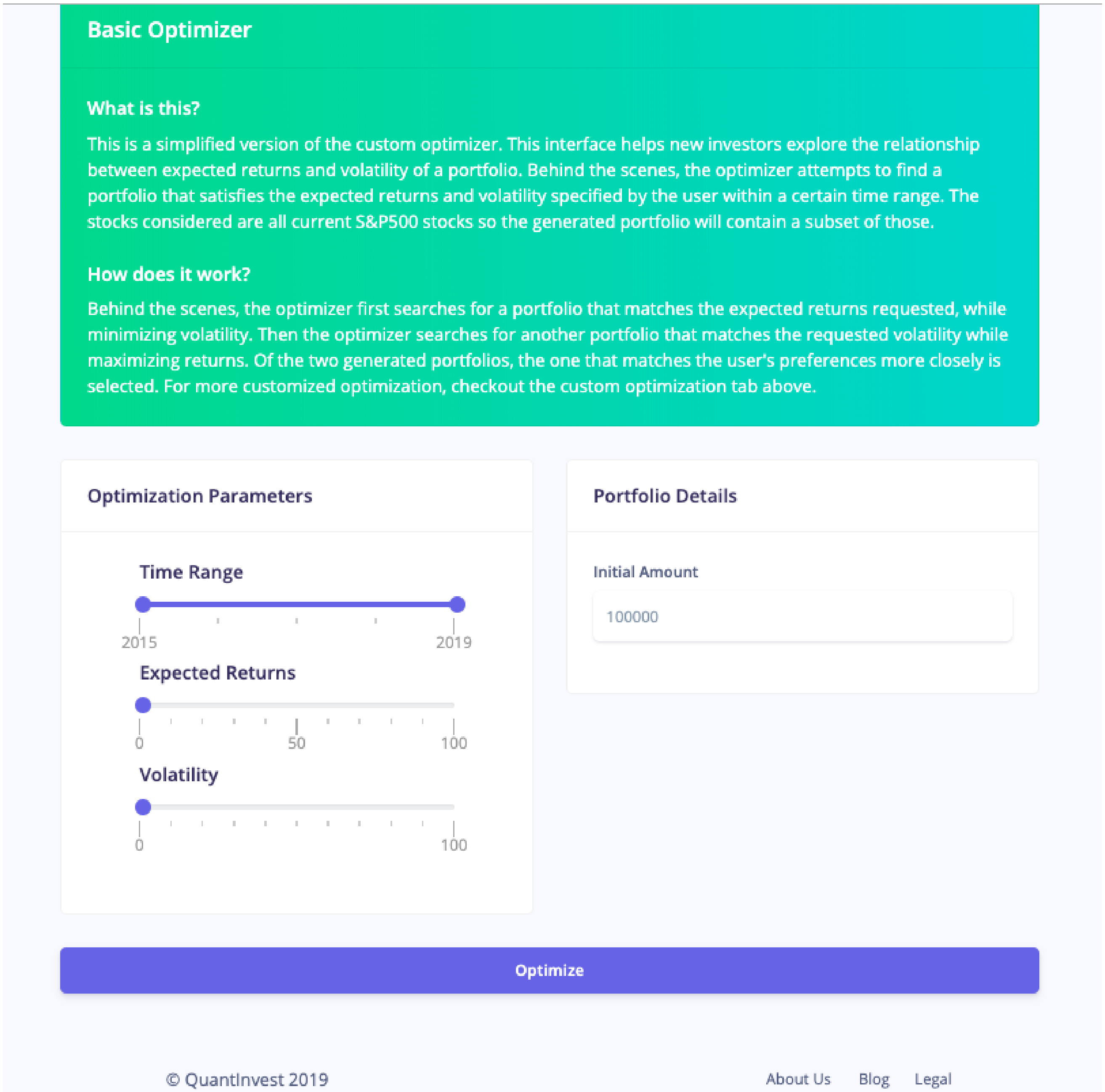


Figure 15: QuantInvest Basic Optimization Page

This interface helps new investors explore the relationship between expected returns and volatility of a portfolio. Behind the scenes, the optimizer attempts to find a portfolio that satisfies the expected returns and volatility specified by the user within a certain time range. The stocks considered are all current SP500 stocks so the generated portfolio will contain a subset of those. This is a key difference between this optimizer and the custom optimizer, where the stocks considered for optimization can be manually selected. In this case we simply select a preset index, which contains a long list of stocks, and use that to generate the optimal portfolio given the specified constraints.

The optimization method used in this basic optimization page is the efficient frontier optimization method proposed by Harry Markowitz, which was discussed previously. When a user selects a target volatility and expected return, two optimization calls take place in reality; the first attempts to find a portfolio that minimizes volatility while trying to meet the constraint on expected return, and the other maximizes expected return while trying to meet the constraint on volatility. Whichever results in a portfolio that is closer to the one the user requested is returned to the user, and visualized in the charts underneath the optimization options interface. Other inputs to this optimizer include the initial portfolio value, or the initial amount the user plans to invest, and the time range over which the backtesting should be performed.

The charts show the performance of the portfolio over the selected time range, while also showing an upper and lower envelope around the portfolio values. This envelope is what is referred to as the Bollinger band in finance, or Bollinger bands if talking about the upper and lower limit individually. A Bollinger Band® is a tool used in technical analysis of stocks, which is formed by calculating the simple moving average of a stock, and adding/subtracting to that line two standard deviations. This results in a band around the stock price. When looking at the whole portfolio, it gives us a sense of the volatility of the prices visualized by looking at the thickness of the band. This is one of the educational tips the interface offers the user when the help icon is clicked, emphasizing the role of the platform as an educational one.

Custom Optimizer

This interface helps slightly more experienced investors generate portfolios with customizable optimization constraints. The optimization method here can be changed. The methods offered so far include maximizing Sharpe ratio, minimizing volatility, maximizing expected returns under the constraint of not exceeding some target volatility, and finally minimizing volatility while trying to meet the constraint of reaching some target return. The other key difference between this optimizer and the basic optimizer is the fact that stocks can be manually chosen here. The user can choose any number of stocks starting from at least 2, and then choose an optimization method and enter the constraints, then choose a time range and initial portfolio value and optimize. This allows the user to explore portfolios they are considering for investments, or just visualize how some portfolios would have performed in certain time ranges.

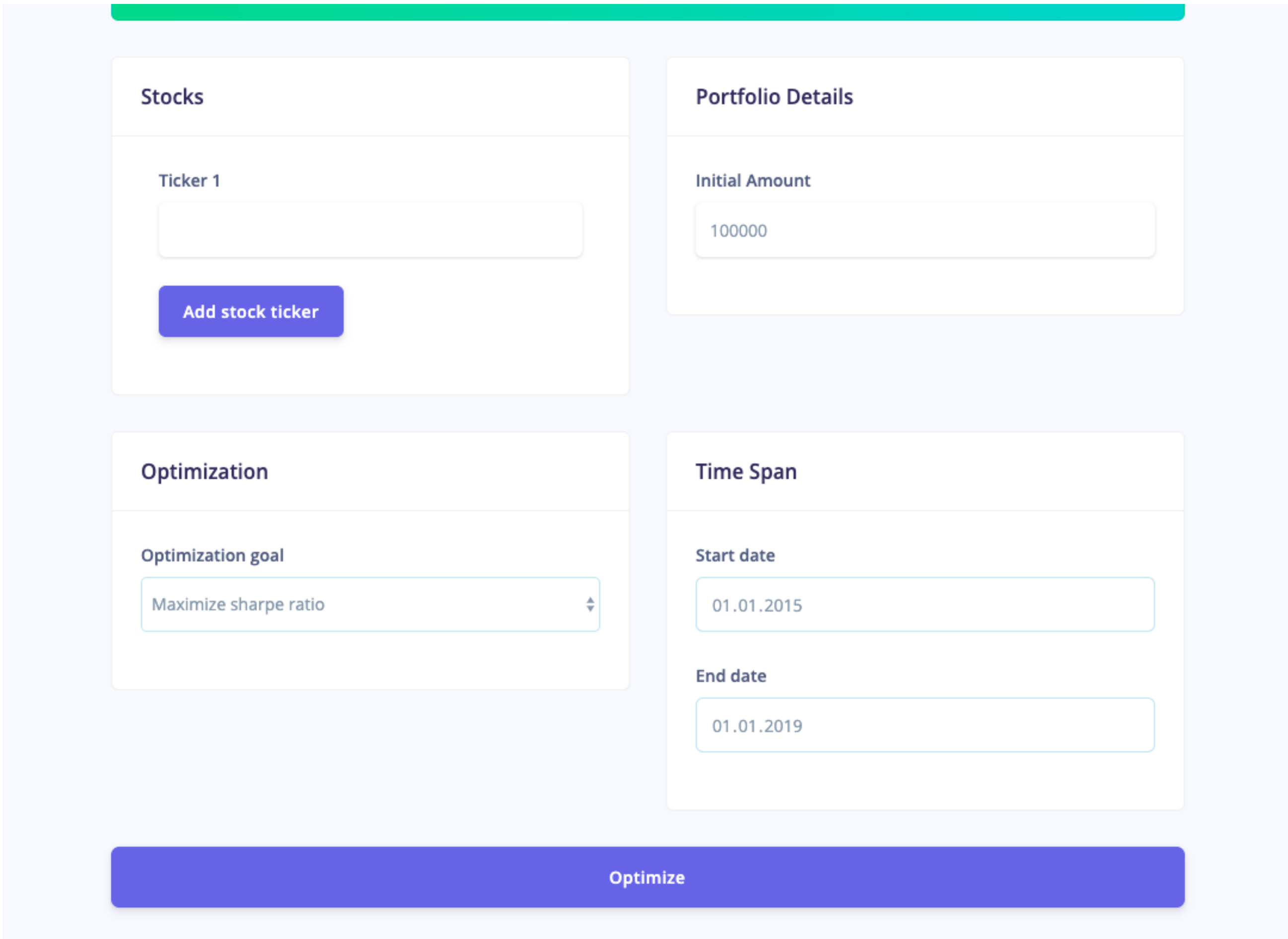


Figure 16: QuantInvest Custom Optimization Page - 02

9.5.5. Optimizer Backend

This section applies to both optimizers, as they basically perform the same behavior, but send different parameters with their requests to the backend servers. Both interfaces perform input validation checks before submitting the results to the API optimization endpoint as an AJAX post request. When this request is submitted, further optimization is prevented by disabling all the website inputs until a response is received from the server. At this point the server would have forwarded the job to the message queuing server, and a worker may have picked it up, if any were free. The number of workers can be changed by changing the heroku configuration file, hence allowing out of the box scalability as the number of users increases in the future.

At this point, the Javascript running on the webpage regularly checks whether or not the job has been completed by sending GET requests using AJAX to the job status endpoint every 200 milliseconds to see whether or not the queued job has been completed. Once the job is done, its result is fetched, the charts and tables are updated with the portfolio performance and statistics, the loading animation stops, and the inputs are enabled again.

9.5.6. Portfolio Explorer

The portfolio explorer is opened in a separate window, and shows a set of graphs and statistics that express the overall financial health of a company, as well as tools for closer investigation. The following sections describe the functions performed by this component.

Intrinsic value calculation

Estimating a company's intrinsic value is a subjective and complex process. Warren Buffett defines a company's intrinsic value as "the discounted value of the cash that can be taken out of a business during its remaining life" [43]. Since intrinsic value depends, among other things, on future cash flows, calculating it involves estimation and prediction. This is where most of the difficulty comes from. Our component relies on some assumptions about a company's cash flows in order to make a reasonable estimate.

The first assumption is that the change in a company's intrinsic value is likely to be close to the change in its **book value** [43]. The second assumption is that the yearly **dividends** of a company are constant over time (In a healthy company, dividends tend to increase over time. This will lead to a conservative valuation of the company, which is acceptable for our purposes.) The third assumption is that a healthy company's book value will keep increasing steadily if it has done so in the past 4 years. Note that this is not as ambitious an assumption as with market values, which fluctuate quasi randomly in the short term. Building on the approach outlined by Buffett in [43], we compare a company's future returns with those of a **federal note**, which is taken to be a zero-risk investment. For this reason, we assume that the federal interest rate will remain constant.

To estimate a company's intrinsic value, we set the returns of our company equal to those of a **federal note** over a fixed number of years in the future. The price which would guarantee that return, taking into account the dividends taken out of a company and the time value of money, is the intrinsic value of that company today. In other words, if one was able to buy a share of a company at its intrinsic value, one could guarantee a risk free return on investment equal to a federal reserve note in a fixed number of years in the future. This is expressed in equations (2) and (3):

$$IV\,(1+i)^n = BV\,(1+\Delta BV)^n + nD \tag{2}$$

$$IV = \frac{BV\,(1+\Delta BV)^n + nD}{(1+i)^n} \tag{3}$$

where IV is the intrinsic value, BV is the book value, ΔBV the average change in book value over a past interval of years, i the interest rate, D the yearly dividends, and n the number of years into the future we are taking into account to make the estimation.

Using this number, a user can get an idea as to whether a stock is over-valued or under-valued. The method for producing the estimation is also displayed in an information window for the user to understand the meaning of this estimate.

3D plot of company health

Three important ratios that encompass a company's financial health are the price to book value ratio (PBV), the debt to equity ratio (DE) and the current ratio (CR). PBV gives a rough idea of how far the market value is from the company's book value, and serves as an indicator of over-valuation. DE is the ratio of a company's total liabilities to its equity, and gives an idea of how well it is managing its debt. CR is equal to assets divided by liabilities, and gives investors an idea of a company's ability to pay short-term obligations (due within one year). A point is represented on a 3D plot with these three numbers as coordinates, for a given year. QuantInvest can hold data up to 4 years in the past, so a company's "trajectory" across this plot gives a visual impression of its financial health across time.

The user has the option to show the company's peers-competing companies-on the same plot. The "safe region" on the plot is based on common wisdom among value investors about the acceptable values of each of these indicators. Users get a message when the company moves away from the safe region, explaining what is going wrong (e.g. debt too high, company too overvalued).

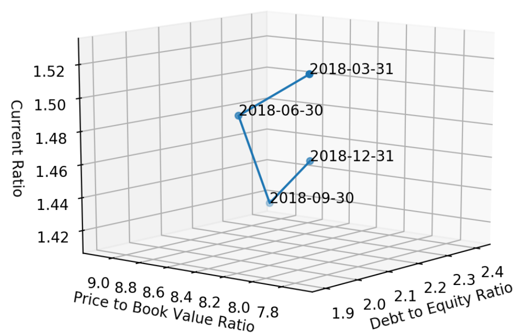


Figure 17:

Cash flow history

A company's cash flow is a financial report describing its operating, investing and financing activities. Operating activities are the money generated by the business, investing activities represents money invested by the company for things like purchasing assets, and financing activities include things like paying off debt to banks or investors [42]. To an inexperienced user, reading a cash flow statement can be intimidating, let alone interpreting the numbers and making conclusions about a company's potential. This feature plots a company's cash flow and debt over the past four years, performs linear regression on the scatterplot and gives informative comments to a user about interpreting the plot. For example, the following message is displayed if the linear regression's r-squared metric is less than 0.9: "Cash flow is not stable enough over the last 4 years. This may be a risky company to invest in." and the following message is displayed if debt is increasing consistently: "Debt is increasing.

Remember that increasing debt means decreasing income in the future.”

9.5.7. Data Fetcher

There are two types of data to be fetched: market data and fundamental data. Both of these are fetched from IEX Trading [14]. The market data of a specific list of stocks is fetched and stored on server startup, effectively caching that data. This is accomplished using a Redis task queue and workers that fetch the data in the background [34][35]. Once the data has been cached, this component can start receiving requests. If the data of a certain stock is not available in the cache, it is fetched directly from IEX. The fundamental data is also fetched from IEX, but is not cached.

The Data fetcher also uses another source of market data: Alpha Vantage [36]. This allows us to fall back to another source of data should the main one fail, and increases the reliability and availability of our platform.

9.5.8. Backtester

The backtester that QuantInvest uses was built from scratch, instead of using Zipline as planned, because the functionality that was needed was too simple, and Zipline had too much overhead. The built-in backtester receives a portfolio as input, along with a time range, and tests the performance of that portfolio over that period. Specifically, it calculates the value of the portfolio over the specified range of time, as well as other helpful statistics such as the moving average and the moving standard deviation. This backtester is connected to the optimizer, and used to visualize the portfolios generated by the optimizer.

9.5.9. Databases

The database used was a NoSQL database, MongoDB. NoSQL offers many advantages over SQL databases, such as scalability, and the flexibility to change the database models without needing to perform any migrations. MongoDB specifically is one of the most popular NoSQL databases [37], and therefore has a large user base and is continuously being updated. Additionally, MongoDB allows SQL-like queries, and therefore has the positives of all the SQL databases without any of the negatives [31].

10. Experiment Design and Testing

The experiment designed to test QuantInvest's performance is the monitoring of a generated optimal portfolio using some stock index. Simply put, some portfolios were generated, such as to maximize their sharpe ratios, and the stocks universe chosen was the whole SP500 stocks list. The following was the resultant portfolio: This initial test portfolio was generated by a $1000 investment on 01-01-2015 and optimized up until 01-01-2019, such as to maximize its sharpe ratio. The following allocations were generated.

Stock Ticker	Percentage
AMD	3.48%
AMZN	8.38%
AWK	6.83%
ANTM	2.19%
CHD	0.55%
CME	18.43%
DRI	0.23%
DLR	0.97%
HUM	1.01%
MKC	5.46%
MCD	7.09%
NFLX	2.61%
NEM	8.75%
NEE	14.62%
NVDA	7.8%
PGR	6.11%
RMD	0.72%
UNH	4.76%

Table 3: Test portfolio 1

For the test, it is assumed that on 01-01-2019, this portfolio was purchased with the initial investment of $1000. The generated returns up until 25-04-2019 were 12.7% (or $128 in profit). This means that after the initial portfolio generation, if the user had bought the portfolio at the end date specified, they would have gained 12.7% of their initial investment of $1000 by 25-04-2019.

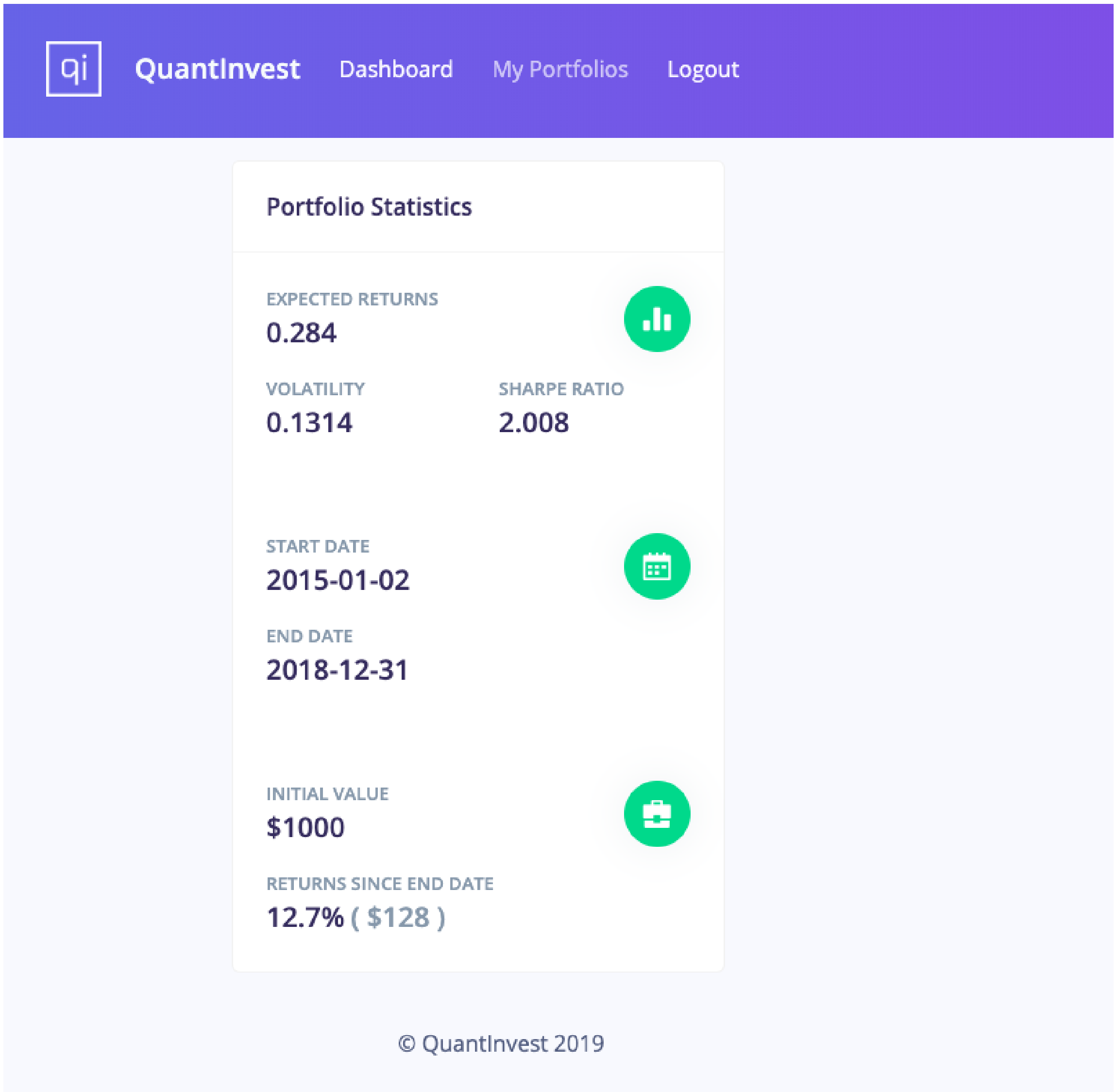


Figure 18: Test portfolio 1 on QuantInvest

The other tests performed were done using the custom optimizer, specifying manually the stock universe, and the optimization method. The following stocks were selected:

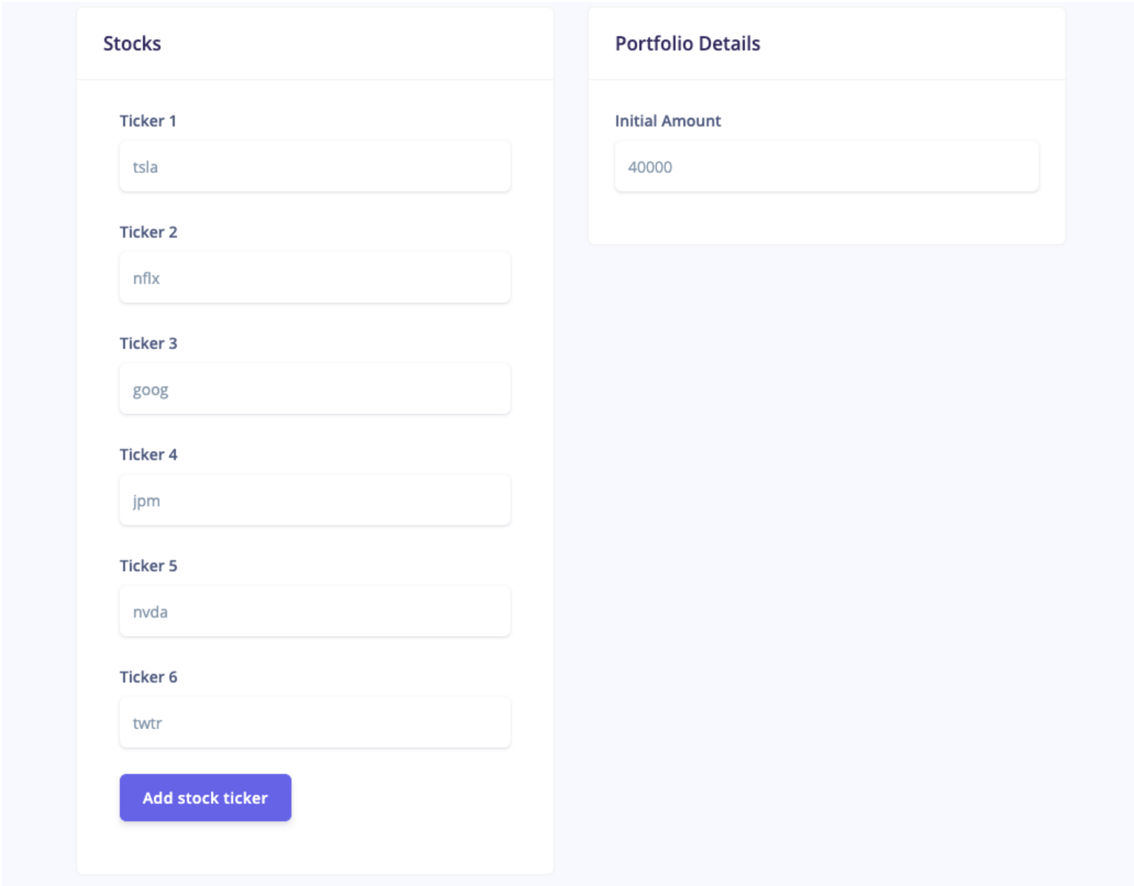


Figure 19: Test portfolio 2 stocks list on QuantInvest

The optimization results were the following:

Figure 20: Test portfolio 2 backtest results on QuantInvest

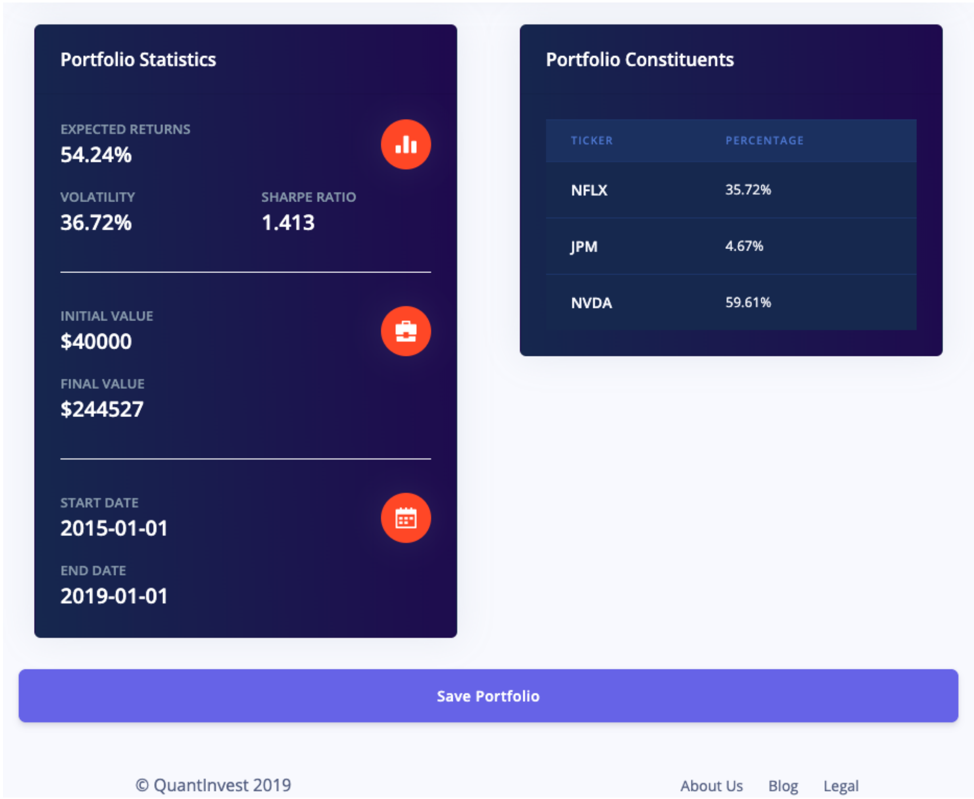


Figure 21: Test portfolio 2 statistics on QuantInvest

After saving the portfolio, the user can see how the portfolio would be performing if they had bought it on the end date specified in the optimization (in this case 01-01-2019).

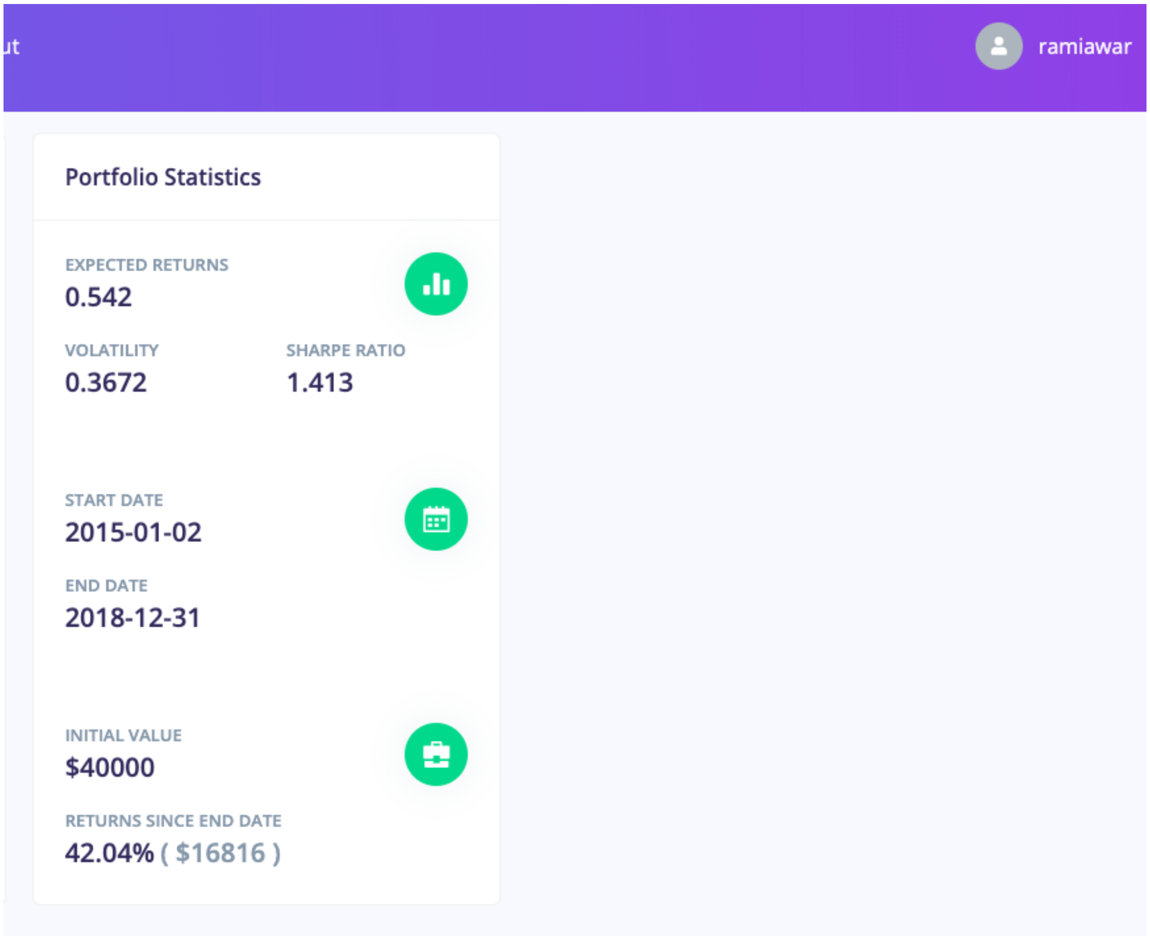


Figure 22: Test portfolio 2 test results on QuantInvest

The preceding statistics show the returns of the portfolio if it had been bought at that date, hence testing the actual effectiveness of generating a portfolio depending on past data only, to be used in the future. It is found that the portfolio yields a 42.04% return in 4 months only. As can be seen, most portfolios yield positive returns, with low volatility. It is clear that Markowitz portfolio optimization is still a very valid portfolio construction technique to this day. Of course, this is dependent on the stocks themselves. In the second test portfolio, relatively stable stocks were picked, as NFLX, JPM, and NVDA are safe choices for an investor under normal circumstances.

11. Broad Impact

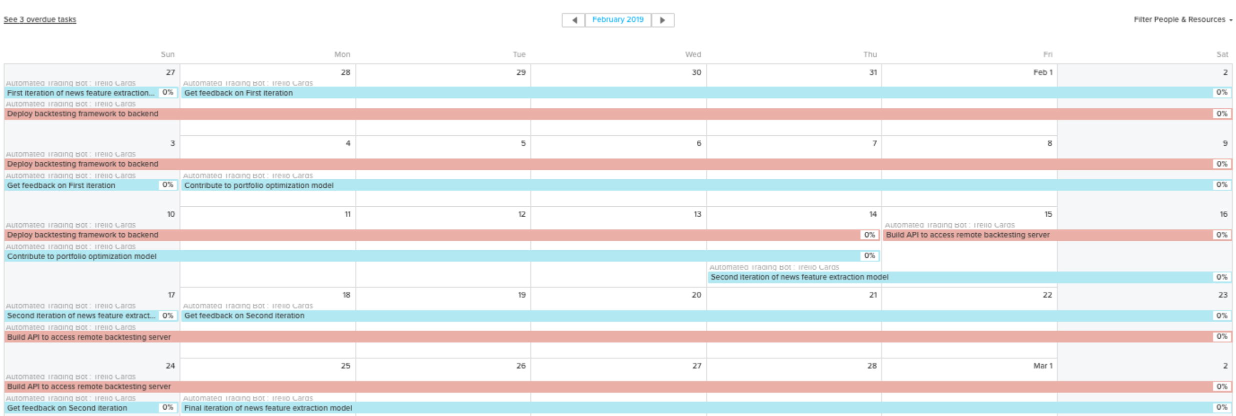
The impact of our platform will be felt most in the finance education sector, while also benefiting beginner to intermediate-level investors. As stated previously, our platform includes a profiler that helps the user find out what sort of portfolio suits them best. This allows beginners to understand what sort of questions they need to be asking themselves before investing. It also gives finance students the intuition needed for financial decision making, something that no other tool currently provides.

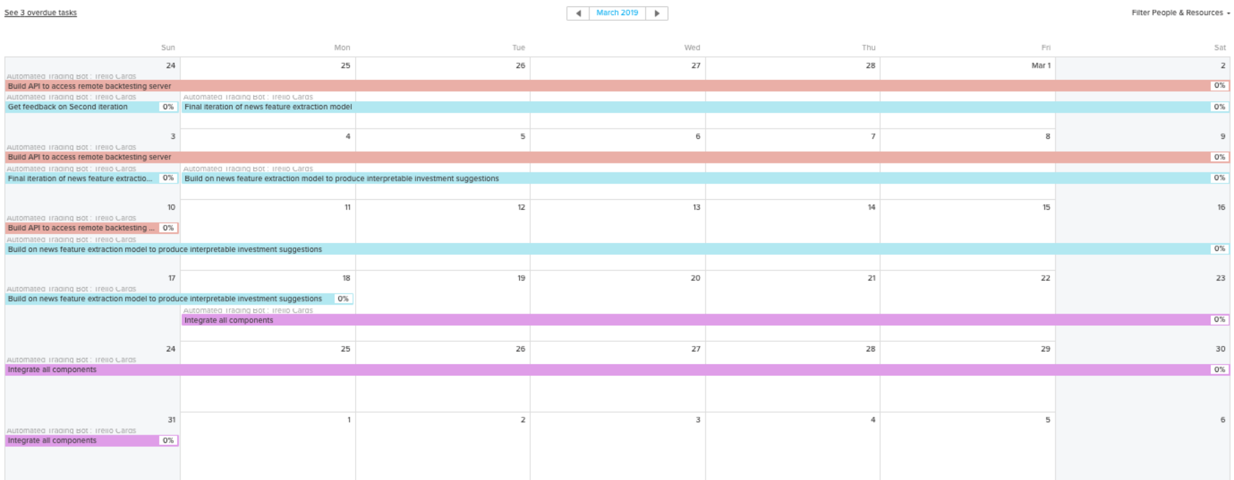
12. List of Resources and Engineering Tools Needed

- Python data science libraries : Pandas - SciKit Learn - Pytorch : Must learn
- Electron GUI Framework : C++, Javascript, HTML, CSS
- AWS Compute Machines for running optimization algorithms : Must learn
- Python 3
- Zipline library : Must learn

13. Detailed Project Schedule

The following schedule is divided by color, where Rami Awar handles green tasks, Nader Al Awar handles red tasks, and Tarek Tohme handles blue tasks. Purple tasks are a combined effort. The tasks are written inside the colored bars.





See 3 overdue tasks
March 2019
Filter People & Resources
Sun
Mon
Tue
Wed
Thu
Fri
Sat
Automated Trading Bot : Trello Cards
Build API to access remote backtesting server
Get feedback on Second iteration
Final iteration of news feature extraction model
Final iteration of news feature extractio...
Build on news feature extraction model to produce interpretable investment suggestions
Build API to access remote backtesting ...
Integrate all components
0%

References

[1] T. Li Im, P. Wai San, C. Kim On, R. Alfred and P. Anthony, "Analysing market sentiment in financial news using lexical approach," 2013 IEEE Conference on Open Systems (ICOS), Kuching, 2013, pp. 145-149.

[2] https://www.portfoliovisualizer.com/optimize-portfolio

[3] https://www.bearnbull.com

[4] Kolanovic, M. and T. Krishnamachari, R. (2018). Big Data and AI Strategies. [online] Valuesimplex.com. Available at: http://valuesimplex.com/articles/JPM.pdf

[5] Salimans, T., Ho, J. and Chen, X. (2018). Evolution Strategies as a Scalable Alternative to Reinforcement Learning. [online] Arxiv.org. Available at: https://arxiv.org/pdf/1703.03864.pdf

[6] Mankert, C. (2018). The Black-Litterman Model : Towards its use in practice. [online] DIVA. Available at: http://kth.diva-portal.org/smash/record.jsf?pid=diva2:372812 .

[7] Hansen, N. (2018). The CMA Evolution Strategy: A Tutorial. [online] Arxiv.org. Available at: https://arxiv.org/abs/1604.00772 .

[8] Glensk, B. and Madlener, R. (2018). Review of Selected Methods for Portfolio Optimization. [online] Cejsh.icm.edu.pl. Available at: http://cejsh.icm.edu.pl/cejsh/element/bwmeta1.element.cejsh-383a07f5-e0ef-4977-9be3-1f8654d81f88/c/02.pdf .

[9] M. Hagenau, M. Hauser, M. Liebmann and D. Neumann, "Reading All the News at the Same Time: Predicting Mid-term Stock Price Developments Based on News Momentum," 2013 46th Hawaii International Conference on System Sciences, Wailea, Maui, HI, 2013, pp. 1279-1288.

[10] Fang, Jiali, Ben Jacobsen, and Yafeng Qin (2014). Predictability of the simple technical trading rules: An out[U+2010]of[U+2010]sample test. Review of Financial Economics 23.1:30-45.

[11] Treleaven, Philip, Michal Galas, and Vidhi Lalchand. (2013) Algorithmic trading review. Communications of the ACM 56.11:76-85.

[12] Liebmann, M. and Hagenau, M. (2018). Effects behind Words: Quantifying Qualitative Information in Corporate Announcements .

[13] https://www.quandl.com/databases/WIKIP

[14] https://iextrading.com/developer/docs/

[15] https://github.com/sanko/Robinhood

[16] Park, Cheol[U+2010]Ho, and Scott H. Irwin.(2007) What do we know about the profitability of technical analysis? Journal of Economic Surveys 21.4 : 786-826.

[17] Friedman, Jerome, Trevor Hastie, and Robert Tibshirani. (2001) The elements of statistical learning. Vol. 1. No. 10. New York, NY, USA:: Springer series in statistics.

[18] Walters, J. (2018). The Black-Litterman Model in Detail.

[19] Michaud, R. and Michaud, R. (2008). Efficient asset management. New York: Oxford University Press.

[20] Jaybhay, Kranti M. et al. "Stock Market Prediction Model by Combining Numeric and News Textual Mining." (2012).

[21] https://www.quantopian.com/about Accessed on Nov. 2 2018

[22] Zheng, A. and Jin, J. (n.d.). Using AI to Make Predictions on Stock Market. [online] Cs229.stanford.edu. Available at: http://cs229.stanford.edu/proj2017/final-reports/5212256.pdf .

[23] https://www.algotrader.com/

[24] https://www.quantopian.com/get-funded

[25] https://www.tradestation.com/

[26] https://ninjatrader.com/

[27] https://www.multicharts.com/

[28] https://www.zipline.io/ Accessed on Nov. 15 2018

[29] https://addisonlynch.github.io/iexfinance/stable/

[30] http://flask.pocoo.org/

[31] https://www.mongodb.com/

[32] https://www.heroku.com/

[33] https://pyportfolioopt.readthedocs.io/en/latest/UserGuide.htmlefficient-frontier-optimisation

[34] https://redislabs.com/ebook/part-2-core-concepts/chapter-6-application-components-in-redis/6-4-task-queues/

[35] https://python-rq.org/

[36] https://www.alphavantage.co/

[37] https://db-engines.com/en/ranking

[38] https://www.fidelity.com/

[39] https://www.harringtoninvestments.com/legal-disclaimer/

[40] http://flask.pocoo.org/docs/1.0/blueprints/

[41] https://demos.creative-tim.com/argon-design-system/index.html

[42] https://www.investopedia.com/investing/what-is-a-cash-flow-statement/

[43] Buffett, Warren (1996). An owner's manual. Omaha, NE: Berkshire Hathaway.

14. Appendix A

The following legal information was compiled from investment websites such as [39], and seems to fit the purpose of QuantInvest's services.

```
TERMS AND CONDITIONS
QuantInvest and its members, officers, directors, owners,
   employees, agents, representatives, suppliers and service
   providers (collectively "QuantInvest") provides this website
    (the "Site") for informational purposes only. Use of and
   access to the Site and the information, materials, services,
    and other content available on or through the Site ("
   Content") are subject to these terms of use and all
   applicable laws.

NO INVESTMENT ADVICE
The Content is for informational purposes only, you should not
   construe any such information or other material as legal,
   tax, investment, financial, or other advice. Nothing
   contained on our Site constitutes a solicitation,
   recommendation, endorsement, or offer by QuantInvest or any
   third party service provider to buy or sell any securities
   or other financial instruments in this or in in any other
   jurisdiction in which such solicitation or offer would be
   unlawful under the securities laws of such jurisdiction.

All Content on this site is information of a general nature and
    does not address the circumstances of any particular
   individual or entity. Nothing in the Site constitutes
   professional and/or financial advice, nor does any
   information on the Site constitute a comprehensive or
   complete statement of the matters discussed or the law
   relating thereto. QuantInvest is not a fiduciary by virtue
   of any person's use of or access to the Site or Content. You
    alone assume the sole responsibility of evaluating the
   merits and risks associated with the use of any information
   or other Content on the Site before making any decisions
   based on such information or other Content. In exchange for
   using the Site, you agree not to hold QuantInvest, its
   affiliates or any third party service provider liable for
   any possible claim for damages arising from any decision you
    make based on information or other Content made available
   to you through the Site.

INVESTMENT RISKS
```

There are risks associated with investing in securities. Investing in stocks, bonds, exchange traded funds, mutual funds, and money market funds involve risk of loss. Loss of principal is possible. Some high risk investments may use leverage, which will accentuate gains & losses. Foreign investing involves special risks, including a greater volatility and political, economic and currency risks and differences in accounting methods. A security's or a firm's past investment performance is not a guarantee or predictor of future investment performance. For this reason backtests generated by our Optimizer should not be misinterpreted as guarantees of future profit or loss.

THIRD PARTY LINKED SITES

As a convenience to you, QuantInvest may provide hyperlinks to web sites operated by third parties. When you select these hyperlinks you will be leaving the QuantInvest site. Because QuantInvest has no control over such sites or their content, QuantInvest is not responsible for the availability of such external sites or their content, and QuantInvest does not adopt, endorse or nor is responsible or liable for any such sites or content, including advertising, products or other materials, on or available through such sites or resources. Other web sites may provide links to the Site or Content with or without our authorization. QuantInvest does not endorse such sites and shall not be responsible or liable for any links from those sites to the Site or Content, or for any content, advertising, products or other materials available on or through such other sites, or any loss or damages incurred in connection therewith. QuantInvest may, in its sole discretion, block links to the Site and Content without prior notice.

YOUR USE OF THIRD PARTY WEB SITES AND CONTENT, INCLUDING WITHOUT LIMITATION, YOUR USE OF ANY INFORMATION, DATA, ADVERTISING, PRODUCTS, OR OTHER MATERIALS ON OR AVAILABLE THROUGH SUCH WEB SITES, IS AT YOUR OWN RISK AND IS SUBJECT TO THEIR TERMS OF USE.

USE OF COOKIES

QuantInvest's website utilizes different technologies to collect, store, and aggregate data about website usage. We may use electronic tags called "cookies" to help us understand and analyze use of our site. This work is either performed directly by us or by a third party we've hired to assist us. We collect information about which pages have been accessed and for how long, the country the user accesses the site from, and certain technical information regarding the user's computer and operating systems, such as user Internet protocol address, domain name and browser, etc.

Certain sections of QuantInvest's site require cookies to be enabled to enhance site performance. For example, cookies provide a secure way for us to verify user identity during a session and any return visits, they enable us to personalize a user's experience on our sites, and they help enhance site navigation. Cookies also help us to understand how people use our sites so we can improve site functionality.

When a user comes to the website, our server sends a cookie to the user's computer. Standing alone, cookies do not identify the user personally; they merely recognize the user's browser. Generally, personally identifiable information is obtained by us only when a user decides to provide it, such as when requesting additional information via email or providing personal information.

We use two types of cookies on our sites, temporary cookies and persistent cookies. Temporary cookies are used to store information during a browser session and will expire shortly after concluding a visit to one of our sites. Persistent cookies are used to store information between visits to one of our sites and are stored permanently or for a specified length of time. Persistent cookies are used to facilitate easier navigation within our sites and provide a higher level of convenience for the user.

A user can choose to have their computer issue a warning each time a cookie is being sent, or a user can choose to turn off all cookies. The management of cookies generally is handled through the user's browser settings (e.g., Internet Explorer). To obtain more information about managing cookies, visit www.aboutcookies.org.

SITE AND CONTENT NOT WARRANTED

THE SITE AND CONTENT, ARE PROVIDED "AS IS" AND WITHOUT WARRANTIES OF ANY KIND. YOU BEAR ALL RISKS ASSOCIATED WITH THE USE OF THE SITE AND CONTENT, INCLUDING WITHOUT

LIMITATION, ANY RELIANCE ON THE ACCURACY, COMPLETENESS OR USEFULNESS OF ANY CONTENT AVAILABLE ON THE SITE. QuantInvest AND ITS EMPLOYEES, OFFICERS, DIRECTORS, , PARTNERS, AGENTS, REPRESENTATIVES, SUPPLIERS AND SERVICE PROVIDERS, DISCLAIM ALL WARRANTIES, EXPRESS OR IMPLIED, INCLUDING, WITHOUT LIMITATION, ALL WARRANTIES OF TITLE, NON-INFRINGEMENT, ACCURACY, COMPLETENESS, USEFULNESS, MERCHANTABILITY, AND FITNESS FOR A PARTICULAR USE, AND WARRANTIES THAT MAY ARISE FROM COURSE OF DEALING/PERFORMANCE OR USAGE OF TRADE.

LIMITATION OF LIABILITY

YOUR EXCLUSIVE REMEDY FOR DISSATISFACTION WITH THE SITE AND CONTENT IS TO STOP USING THE SITE AND CONTENT. QuantInvest IS NOT LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, CONSEQUENTIAL, SPECIAL OR PUNITIVE DAMAGES, UNDER ANY THEORY OF LIABILITY, INCLUDING WITHOUT LIMITATION, DAMAGES FOR LOSS OF PROFITS, USE, DATA, OR LOSS OF OTHER INTANGIBLES. IN PARTICULAR, AND WITHOUT LIMITATION, QuantInvest WILL NOT BE LIABLE FOR DAMAGES OF ANY KIND RESULTING FROM YOUR USE OF OR INABILITY TO USE THE SITE OR CONTENT.

While we try to maintain the integrity and security of the Site and the servers from which the Site is operated, we do not guarantee that the Site or Content is or remains secure, complete or correct, or that access to the Site or Content will be uninterrupted or error free. The Site and Content may include inaccuracies, errors and materials that violate or conflict with these Terms. Additionally, third parties may make unauthorized alterations to the Site or Content. If you become aware of any unauthorized third party alteration to the Site or Content, contact us with a description of the material(s) at issue and the URL.

NOTICES, COMMUNICATIONS, AND ELECTRONIC SIGNATURES

You agree to be bound by any affirmation, assent or agreement that you transmit on or through the Site or any other aspect of QuantInvest's services that you access by computer or other electronic device, including internet, telephonic and wireless devices, including but not limited to any consent you give to receive communications from us solely through electronic transmission. You agree that, when in the future you click on a button worded "I agree" with your mouse, keystroke or other device, your agreement or consent will be legally binding and enforceable and the legal equivalent of your handwritten signature.

LIMITED RIGHT OF USE/OWNERSHIP OF CONTENT

You are permitted to use the Site and Content for your personal , non-commercial use only. The Site and Content are and shall remain the property of QuantInvest and is protected by copyright, trademark, patent, and/or other intellectual property, proprietary, work product rights and laws. You may use the Site and Content for your personal, noncommercial use, provided that you keep intact all copyright, trademark, patent and other proprietary notices. Except as expressly authorized in advance by QuantInvest in writing, you agree not to reproduce, modify or create derivative works based on , rent, lease, loan, sell, distribute, publish, publicly perform or display, reverse engineer, de-compile or dissemble, all or any part of the Site or Content.

Trade names, trademarks and service marks of QuantInvest include, without limitation, QuantInvest and any associated logos. All trademarks and service marks on the Site not owned by QuantInvest are the property of their respective owners. Nothing contained on the Site should be construed as granting, by implication, estoppel or otherwise, any license or right to use any of QuantInvest QuantInvest's trade names, trademarks or service marks without our express prior written consent.

TERMINATION

QuantInvest, in its sole discretion, may terminate your access to or use of the Site and Content, at any time and for any reason. Your access to or use of the Site and Content may be terminated without notice. QuantInvest shall not be liable to you or any third party for any termination of your access to the Site or Content, or to any such information or files , and shall not be required to make such information or files available to you after any such termination.

RULES OF CONDUCT

Your use of the Site and Content is conditioned on your compliance with the rules of conduct set forth here. You will not:

Use the Site or Content for any fraudulent or unlawful purpose.

Interfere with or disrupt the operation of the Site or Content or the servers or networks used to make the Site and Content available; or violate any requirements, procedures,

policies or regulations of such networks.
Restrict or inhibit any other person from using the Site or Content (including without limitation by hacking or defacing any portion of the Site or Content).
Use the Site or Content to advertise or offer to sell or buy any goods or services without QuantInvest's express prior written consent.
Reproduce, duplicate, copy, sell, resell or otherwise exploit for any commercial purposes, any portion of, use of, or access to the Site or Content.
Modify, adapt, reverse engineer, de-compile/disassemble any part of the Site or Content.
Remove any copyright, trademark or other proprietary rights notice from the Site or materials originating from the Site or Content.
Frame or mirror any part of the Site or Content without QuantInvest's express prior written consent.
Create a database by systematically downloading and storing Content.
Use any robot, spider, site search/retrieval application or other manual or automatic device to retrieve, index, "scrape", "data mine" or in any way gather Content or reproduce or circumvent the navigational structure or presentation of the Site without Company's express prior written consent.

MODIFICATIONS
QuantInvest may amend the terms of use at any time in its discretion, by posting revisions on the Site.

15. Appendix B

16. Appendix C

16.1. Updated Project Description and Agreement Form

Faculty Supervisor	Fadi Zaraket
Co-Supervisor [optional]	
Sponsor [optional] *Is there industry support or funding the project?*	
Project Title *Descriptive title* *not necessarily the final title that will be adopted by the team*	Quantitative Trading Software
Project Description and Design Aspects *What is the main motivation for the project?* *Specify the desired needs that the final product is expected to meet.*	Motivation: - quantitative trading is gaining large momentum and funding, in industry, quantitative analysts are among the highest paying jobs - increasing need to extract meaning from large unstructured datasets, applications in finance are very rewarding Desired Needs: - integrating alternative (non-financial) data in quantitative finance - low-entry barrier software for non-technical people who cannot code
Expected Deliverables *Required deliverable(s) from the team at the conclusion of the design project*	1 – Desktop application with the specifications mentioned above 2 - User guide and introduction to quantitative trading documentation
Technical Constraints *A preliminary list of multiple realistic technical constraints, e.g. power, accuracy, real-time operation , ... The technical constraints included should be detailed and specific to the design project not generic.*	1- Limited computing power, consequently limited kinds of tradeable securities. Stocks only. 2- Access to free data only 3- No automation at the trading signal generation stage

Non-Technical Constraints *A preliminary list of multiple realistic non-technical constraints, e.g. cost, environmental friendliness, social acceptance, political, ethical, health and safety, etc… The non-technical constraints included should be detailed and specific to the project not generic.*	1- Financial risks of trading should be taken into consideration and well communicated to the user.
Contemporary Issues *Cite one or more recent articles pertaining to the project or project area: news articles, blog discussions, academic articles, conference topics, etc.*	https://www.bloomberg.com/news/features/2017-12-05/how-ai-will-invade-every-corner-of-wall-streetee
Resources and Engineering Tools *Identify resources and engineering tools needed and whether they are available or need to be acquired (if known), e.g. software licenses, instruments, facilities, components, …*	All components will be built, except the APIs,e which are the only required resource.
Possible Applicable Standards *List potential standards directly or indirectly used or involved in the project*	NONE
List of Disciplines *Identify at least THREE engineering disciplines (within or outside ECE)*	○ Biomedical Systems ○ Circuits and Electronics ○ Communications and Networking ○ Control, Robotics, and Instrumentation ○ Electromagnetics and RF ○ Hardware, Computer Architecture, and Digital Systems ○ HMI, Graphics, and Visualization ○x Intelligent Systems ○ Machines and Power Systems ○ Signal Processing ○ Software Engineering ○ Engineering discipline outside ECE (Specify): financial engineering Additional Non-Engineering discipline(s) (Specify): statistics
Number of Students *Please consider the number of disciplines checked above*	○ 3 students OR ○ 4 students (if 4 please provide a justification)

16.2. Minutes of All Meetings Up-to-Date

Meeting #1 :			
Date: 21 - 09 - 2018	Time: from 6pm to 7pm	Location: BECHTEL	
Meeting called by	Group		
Attendees	Rami - Nader - Tarek		
Minutes taker	Nader		
Agenda Item: How to build the model			
Discussion			
We need more experience with financial models before deciding on this.			
Conclusions			
1 - We decided to finish 3 courses we found online in 2 weeks before moving on with our brainstorming.			
Action Items		**Person Responsible**	**Deadline**
Study 3 courses asap in 2 weeks		Rami, Tarek, Nader	30 - 09 - 2018

Meeting #: 2			
Date: 28 – 09 – 2018	Time: Friday 5:00 pm to 5:30 pm	Location: Jafet	
Meeting called by	Group		
Attendees	Nader Al Awar – Rami Awar – Tarek Tohme		
Minutes taker	Nader Al Awar		
Agenda Item: Scheduling a Weekly Meeting with Advisor			
Discussion	**Finding a weekly time slot that is suitable for us and for our advisor**		
We needed two time slots, with one serving as a backup to the other. The standard time slot would be in the evening, while the alternative time slot had to be during daylight			
Conclusions			
1) Our standard weekly meeting will be every Monday form 6:30 pm to 7:30 pm 2) We could not agree on an alternative time slot			
Action Items		**Person Responsible**	**Deadline**
Find the alternative time slot		Nader Al Awar	01/10/2018
Agenda Item: Scheduling a Weekly Meeting among the Group Members			
Discussion	**Finding a weekly time slot that is suitable for us**		
We needed to find another time slot that fits all our schedules. This proved to be harder than expected			
Conclusions			
▪ We decided to have our weekly meeting every Wednesday from 11 to 12 am.			

Meeting #3 :			
Date: 01 - 10 - 2018	Time: from 6pm to 7pm		Location: Dr. Zaraket's office
Meeting called by	Group		
Attendees	Fadi Zaraket - Rami - Nader - Tarek		
Minutes taker	Tarek		

Agenda Item: Performing case studies of companies

Discussion

- Prof. Zaraket suggested that we do case studies of famous companies and learn from other analysts' work.
- We should find a way to automate the process of extracting information from company earning reports. That might involve NLP. Look at why companies don't meet their expected earnings etc.
- We were discussing how companies can react to events that seem unrelated at first glance, and how we would have to look through diverse kinds of datasets to find correlations with price changes. Someone was fired from the company, the weather changed, etc.
- We should look for fundamental analysis data, and be creative with where we find it (news, reports, other events etc.)

Conclusions

Read company reports, understand their format, figure out how to mine useful info automatically

Action Items	Person Responsible	Deadline
	Rami, Nader, Tarek	two weeks from now

Agenda Item: Obtaining data for technical analysis

Discussion

- Reuters trade room in OSB, how to get access
- Learn how to use Yahoo finance, Google finance API, and Quandl

Conclusions **Perform**

-

Action Items	Person Responsible	Deadline
x	Rami, Nader, Tarek	two weeks from now

Agenda Item: Handling the optimization problem

Discussion

- Figuring out what to do to one client alone might be different than figuring out what to do to 100 clients asking for our services in parallel, so we should take that into account when optimizing our trading strategy.
- start thinking about optimization in general

Conclusions **read what other people have done**

Action Items	Person Responsible	Deadline
xx	Rami, Nader, Tarek	two weeks from now

Meeting #5 :		
Date: 15 - 10 - 2018	Time: no meeting	Location:
Meeting called by		
Attendees		
Minutes taker		
Agenda Item: finding meaningful measures/predictors of market movement		
Discussion		
No meeting took place exceptionally because of time conflicts in everyone's schedules.		

Meeting #5 :		
Date: 15 - 10 - 2018	Time: no meeting	Location:
Meeting called by		
Attendees		
Minutes taker		
Agenda Item: finding meaningful measures/predictors of market movement		
Discussion		
No meeting took place exceptionally because of time conflicts in everyone's schedules.		