DL3

FYP Proposal

**A System for Predicting Stock Price and Offering Financial Advice**

by

Aadhar Agarwal, Jose Luis Amador Jaime, Leung Ka Wing (Michael) and Luk Yin Chai (Dragos)

**DL3**

Advised by

Prof. Dik-Lun, Lee

Submitted in partial fulfilment

of the requirements for COMP 4981 and CPEG 4901

in the

Department of Computer Science and Engineering

and the

Computer Engineering Program

The Hong Kong University of Science and Technology

2020-2021

Date of submission: September 18, 2020

[1 Introduction 5](#_Toc61510770)

[1.1 Overview 5](#_Toc61510771)

[1.2 Objectives 6](#_Toc61510772)

[1.3 Literature Survey 8](#_Toc61510773)

[1.3.1 Recurrent Neural Network 8](#_Toc61510774)

[1.3.2 Convolutional Neural Networks 9](#_Toc61510775)

[1.3.3 Time Series Forecasting 10](#_Toc61510776)

[1.3.4 Summary 11](#_Toc61510777)

[2 Methodology 12](#_Toc61510778)

[2.1 Design 12](#_Toc61510779)

[2.1.1 Data Crawler Design 12](#_Toc61510780)

[2.1.2 Database Design 13](#_Toc61510781)

[2.1.3 Model Selection 14](#_Toc61510782)

[2.1.4 Investment Algorithm Design 14](#_Toc61510783)

[2.1.5 Backend Server Design 14](#_Toc61510784)

[2.1.6 Frontend Design 14](#_Toc61510785)

[2.2 Implementation 15](#_Toc61510786)

[2.2.1 Data Crawler Development 15](#_Toc61510787)

[2.2.2 Database Implementation 15](#_Toc61510788)

[2.2.3 Model Implementation 15](#_Toc61510789)

[2.2.4 Investment Algorithm Development 15](#_Toc61510790)

[2.2.5 Backend Server Implementation 16](#_Toc61510791)

[2.2.6 Frontend Implementation 16](#_Toc61510792)

[2.3 Testing 16](#_Toc61510793)

[2.4 Evaluation 16](#_Toc61510794)

[3 Project Planning 18](#_Toc61510795)

[3.1 Distribution of Work 18](#_Toc61510796)

[3.2 GANTT Chart 19](#_Toc61510797)

[4 Required Hardware & Software 20](#_Toc61510798)

[4.1 Hardware 20](#_Toc61510799)

[4.2 Software 20](#_Toc61510800)

[5 References 21](#_Toc61510801)

[6 Appendix A: Meeting Minutes 24](#_Toc61510802)

[6.1 Minutes of the 1st Project Meeting 24](#_Toc61510803)

[6.2 Minutes of the 2nd Project Meeting 26](#_Toc61510804)

[6.3 Minutes of the 3rd Project Meeting 28](#_Toc61510805)

[6.4 Minutes of the 4th Project Meeting 30](#_Toc61510806)

[6.5 Minutes of the 5th Project Meeting 32](#_Toc61510807)

[6.6 Minutes of the 6th Project Meeting 34](#_Toc61510808)

[6.7 Minutes of the 7th Project Meeting 36](#_Toc61510809)

[6.8 Minutes of the 8th Project Meeting 38](#_Toc61510810)

[6.9 Minutes of the 9th Project Meeting 40](#_Toc61510811)

[6.10 Minutes of the 10th Project Meeting 42](#_Toc61510812)

# Introduction

## Overview

Investing has always been a crucial element for any person who is looking to grow their personal wealth. The returns gained from investing properly can often outweigh the salary one earns from their job. However, for most of human history, there hasn’t been any method to effectively know where and when to invest that is available to everyone. Most times, the average person will trust their assets with a third party that has some level of insight or experience with investing (like a bank) and will only see the result. However, if effective stock prediction model could be created and used to offer easily understandable and useable financial advice, investing can become accessible to a much larger population. This FYP will aim to create such a system for stock prediction and offering financial advice,

## Objectives

Despite recent improvements in algorithms that allow for better stock price prediction, it is still difficult to find systems that use these predictions effectively to create investment strategies. In addition, most existing commercial products do a poor job of explaining how the strategy was formulated in a manner that is easily comprehensible by the layman. Overcoming these issues would allow us to create an investment and advisory platform that is much more accessible than existing solutions allowing a much larger user base to enjoy the benefits of using modern machine learning techniques to invest more efficiently.

This project aims to build a web application that allows users to get predictions on stock prices as well as advice on how to maximize their return on investment. The outcome of the project shall be achieved by reaching the following goals:

1. Achieving a satisfactory level of accuracy on the predictions.
2. Using the predictions to offer investment strategies- including buy, sell, hold, and diversify- that result in a return on investment that is higher than the market return.
3. Ensuring the web application provides an intuitive and personalized interface

The technical challenges involved in reaching the first goal are mainly related to the selection of the algorithms used to make the predictions as well as the training of optimization of the various machine learning models to maximize accuracy and precision across all stocks. This is because it is unfeasible to create a new model for every different stock the user chooses at it would take far too long and be very unfriendly to the end user. In particular, choosing the ideal hyperparameters will be a difficult task.

The second goal is more dependent on our knowledge of investment strategies and how effectively we can choose the correct strategies for the various scenarios that can occur when investing. The reason for this is that it would be very inefficient to have the program generate financial advice through machine learning as it would be very resource intensive and may not provide the right advice in all cases.

The key challenge for the third goal is ensuring that the results and predictions are provided quickly and smoothly as it is imperative that the frontend and the backend are able to communicate easily.

## Literature Survey

### Recurrent Neural Network

A Recurrent Neural Network is a class of neural networks with its units being recurrently connected. This enables them to have an internal memory to process a sequence of inputs. This allows them to process the inputs and recognize their long-term dependencies in the time-series data of stock prices. A shortcoming is that RNN are not able to hold this memory for an extended period of time, as the Vanishing Gradient descent problem causes that after every iteration in the neural net, the memory and data it has gets vanished going deeper. A way to overcome this is to use Long Short-Term Memory cells instead of the typical neuron-like cells, this to learn more accurately the long-term dependencies in the data.

In [1] it was found that RNN is better at capturing non-linear relationships in stock price data as it was compared against the linear of model ARIMA, and the RNN error percentage was much lower at 3.90% compared to 31.91%. Moreover, it was also found that LSTM performed slightly better due to its ability to hold its memory better. However, the paper also found that CNN performed better as it was able to capture better dynamic changes in trends in the stock price data.

Furthermore, in [2] found that Recurrent Neural Networks (RNN) with Long Short-Term Memory cells performed better than traditional Machine Learning Algorithms, such as regression, support vector machine, random forest, feed forward neural network and backpropagation as RNN-LSTM had much less prediction error when evaluating the models.

### Convolutional Neural Networks

Convolutional neural network (CNN) is another class of deep neural network that is being studied for stock price prediction in recent years. Although CNNs are more commonly used to solve for problems related to computer vision, they are recently found to be performing well for time series forecasting problems too. In fact, [3] found that RNNs, which are generally associated with sequential data, could be outperformed by a simple CNN for sequential modeling.

As for the application of CNN to stock price prediction, several studies have been done to examine the performance of CNN, or an ensemble of CNN and other models, against other methods. These studies generally found that CNNs offer a novel direction in tackling the problem of stock price prediction. For example, in [4], a three-dimensional tensor is used to represent the data of the different features of several markets across a period of time, and input into a CNN. Similarly, [5] found that a 1D CNN can predict stock prices as accurately as some LSTM architectures, while only a fraction of the time is needed to train the CNN model.

An alternate category of studies related to applying CNN to the prediction of stock prices are [6] and [7]. Both studies took stock chart images as input to the CNN; for example, [6] experimented with using different kinds of charts, including a candlestick chart, a line chart, a bar chart, or even a fusion of the different charts as images to be input into the model.

In addition to using CNN as a single model, [8] examined the possibility of combining CNN with a bi-directional LSTM. The CNN is used to extract high level features before passing in data to the LSTM; the rationale behind this was that by performing convolutions on the data, the CNN will be able to exploit the time structure of the input data.

### Time Series Forecasting

Aside from deep-learning based approaches, traditional regression models have shown great effectiveness in forecasting values for time series data such as stock prices [9]. Traditionally, the method used for times series data regression is the Autoregressive Integrated Moving Average (ARIMA) model [9].This model involves using the forecast outcome and the sum of the forecast errors to provide forecasts that should become more accurate the more data there is available.

However, recent interest in the field has led to the development of a new model for making forecasts using regression called Prophet. Prophet is a time series forecasting model made by Facebook and compiled into open source libraries for use in both Python and R [10]. Prophet is an additive regression model with 4 key components; A linear or logistic growth curve trend, a yearly seasonal component modeled using Fourier series, a weekly seasonal component modeled with dummy variables, and the user provided list of important holidays. The benefit of such a model over a traditional regression model is that it leverages both the objectivity of the model as well as “insights” and “hunches” that industry professionals may have observed in their experiences in their industry.

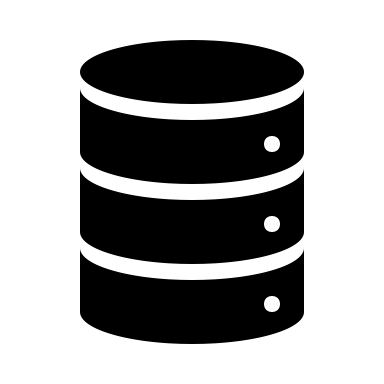
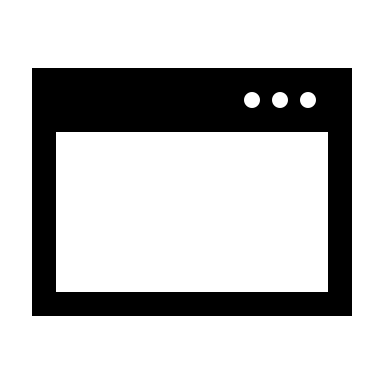
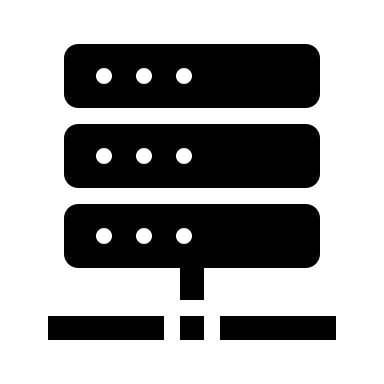
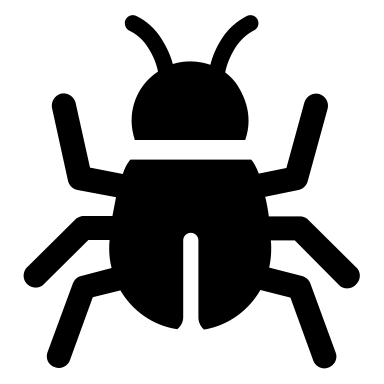
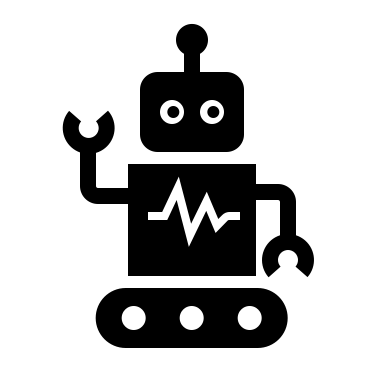
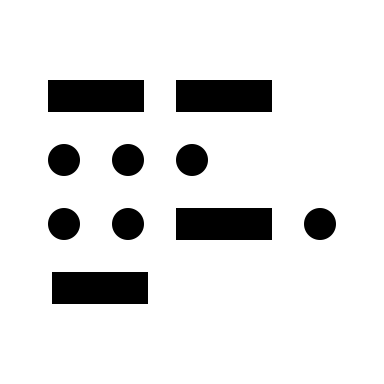
However, the accuracy and precision of the forecasts created by Prophet is lower than the results obtained from a LSTM model. This results in lower return on investment (ROI) than LSTM models [11]. Despite this, Prophet models can still be useful as they can be used alongside LSTM models as a part of an ensemble model to reduce errors. In addition, the Prophet model can also be used as an input layer for the LSTM model in order to reduce the impact of random event of the predicted outcomes [12]. For these reasons we will be using a linear regression model – specifically Prophet – alongside our deep learning models.

### Summary

In summary, this FYP will aim to use methods from both Deep Learning as well as traditional regression to further extend of existing models and generate accurate predictions. In addition to the above-mentioned methods, we shall also use Natural Language Processing (NLP) to perform sentiment analysis which will then be used to make predictions.

# Methodology

## Design

The Design Phase of the project started in July, and we will continue working on the following aspects:

Data Crawler

Frontend

Investment Algorithm

ML Models

Database

Server

### Data Crawler Design

We will design a data crawler that periodically crawls data and stores it in a database for each stock on the market, which includes: 1) market indicators, 2) financial statements, and 3) related news.

Stocks traded in the Hong Kong Exchange shall be included.

Market indicators shall include the following daily data: opening price, closing price, highest price, lowest price, trade volume, turnover, relative strength index (RSI), and moving average convergence divergence (MACD), etc.

Financial statements shall include the following yearly data: profit margin, return on assets (ROA), current ratio, etc.

Related news shall include the news title, news content, date of publishment, and positive vs negative votes.

### Database Design

The data shall be preprocessed before storing in the database. Each of the set of data crawled shall be stored in a separate table, i.e. the market indicators, financial statements, and related news are stored in separate tables.

There should also be a user database, where information for a specific user is stored. The table should contain the following columns: username, password, and risk aversion level.

For the user to select which stocks to include in their portfolio, another table shall be used to store the selected stocks. This table should have these columns: stock symbol, number of shares, and average purchase price.

### Model Selection

We will conduct experiments to select a model or an ensemble of models for stock price prediction. All the data will be used after preprocessing and normalization. Models that we will test include RNN (LSTM or GRU), CNN, Random Forest, and SVM. Regression on the RSI values as well as classification on buy-sell-hold labels will be performed. We will also perform feature selection.

### Investment Algorithm Design

The algorithm is used for suggesting how a user can rebalance their portfolio according to their target risk level. Based on the user’s preference for short-term, medium-term, or long-term investment, the algorithm should be able to suggest different stocks that the user can add to or remove from their investment portfolio. For example, the algorithm can adjust the overall standard deviation of a portfolio according to a user’s preference. Also, buy utilizing the prediction results from the models, the algorithm can suggest stocks, with a certain confidence level, for the user to invest in.

### Backend Server Design

The backend consists of the data crawler, the prediction models, the investment algorithm, the logic in handling user data, and also databases.

### Frontend Design

The frontend will be a web application that can be accessed through a web browser. The application shall contain the following pages:

1. A page for stock price prediction that can be accessed by the public
2. A page for the user to login/register and to select their risk aversion level
3. A page for the user to view their stock portfolio and to view suggested investment strategies including whether to diversify

## Implementation

The Implementation Phase will include the following aspects:

### Data Crawler Development

Based on our design, we will use Python to crawl for data from AASTOCKS and Yahoo Finance.

### Database Implementation

Based on our design, Apache Parquet may be used…

### Model Implementation

Based on our design, we will evaluate each model and select features and search for hyperparameters in order to optimize the models and the prediction results.

### Investment Algorithm Development

Based on our design, the algorithm will need to take in different inputs, including a user’s risk aversion level and short/medium/long-term preference, the standard deviation of their overall investment portfolio, as well as the predicted stock prices.

### Backend Server Implementation

Based on our design, we will use either Flask or Django to develop the backend server.

### Frontend Implementation

Based on our design, we will use React as a framework to develop the user interface.

## Testing

The testing phase will be carried out when some milestones such as big functions are coded, a subsystem is developed, etc. are achieved. For the frontend and backend testing, we apply white box testing by making several test cases such that most if not every independent execution path is tested, and the output is as expected. During the development of the frontend and backend system, when we face bugs, we will record the test cases, and apply regression testing in the later stage of the project. When we integrate the frontend and the backend to one system, we will also check whether the frontend receives the same output as the result that is computed in the backend system. When we build the whole advisory system, we will apply black box testing by entering the stock number to the application and check whether the application can provide advices about the stock.

## Evaluation

In the evaluation phase, it is likely that we will build several machine learning models for predicting the stock price. We will plot graphs based on an array of metrics such as model training time, r2 score, f1 score, accuracy, etc. to evaluate and compare the performance of the models. We will try to investigate and make hypothesis of why a model is better than the other one by observing and comparing those graphs of different models.

We will also evaluate the profitability of our advisory system by stimulating the process of buying and selling stocks based on the advices the system provided.

We will also invite users to test the system and let them to evaluate the user-friendliness of the system.

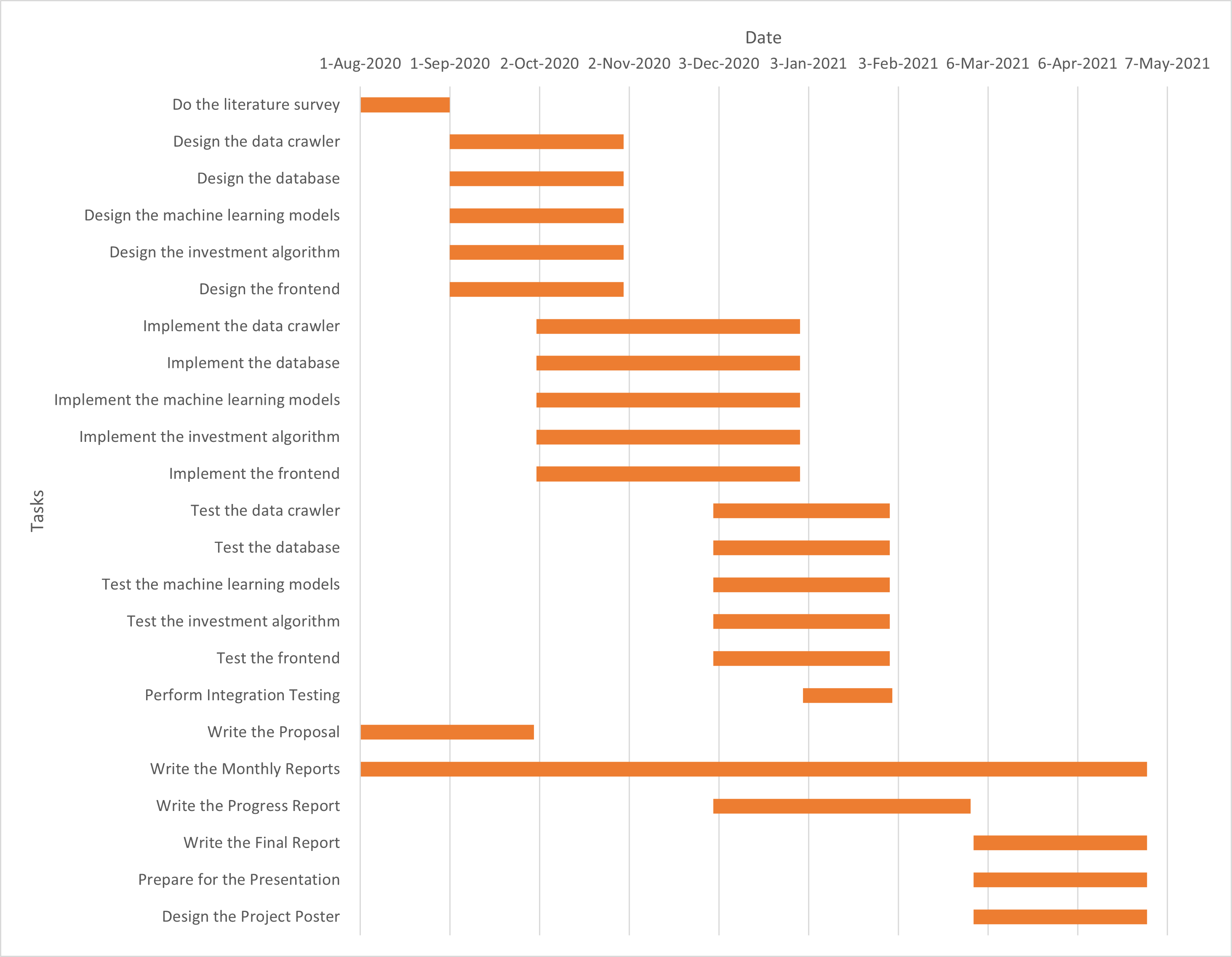
# Project Planning

## Distribution of Work

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Task | Aadhar | Dragos | Jose | Michael |
| Do the literature survey | L | A | A | A |
| Design the data crawler | A | A | L | A |
| Design the database | A | A | A | L |
| Design the machine learning models | A | L | A | A |
| Design the investment algorithm | L | A | A | A |
| Design the frontend | A | A | A | L |
| Implement the data crawler | A | A | L | A |
| Implement the database | A | A | A | L |
| Implement the machine learning models | A | L | A | A |
| Implement the investment algorithm | L | A | A | A |
| Implement the frontend | A | A | A | L |
| Test the data crawler | A | A | L | A |
| Test the database | A | A | A | L |
| Test the machine learning models | A | L | A | A |
| Test the investment algorithm | L | A | A | A |
| Test the frontend | A | A | A | L |
| Perform Integration Testing | A | A | A | L |
| Write the Proposal | A | A | A | L |
| Write the Monthly Reports | L | A | A | A |
| Write the Progress Report | L | A | A | A |
| Write the Final Report | A | A | A | L |
| Prepare for the Presentation | L | A | A | A |
| Design the Project Poster | A | L | A | A |

L = Leader, A = Assistant

## GANTT Chart



# Required Hardware & Software

## Hardware

The hardware requirements of our project would be comprised of the following:

1. Web server on cloud to handle the front end and some basic back end functions
2. SQL database on the cloud to store our financial data
3. GPU servers on cloud or on-premise to train the machine learning models
4. Personal development environment of each team members

## Software

The software requirements of the project would be comprised of the following:

1. SQL for the database and SQL Server Management Studio for its development
2. Django and Flask as web development frameworks
3. Python for development and data processing and analytics
4. Anaconda framework for data cleaning, processing and analytics
5. Visual Studio, Visual Studio Code as IDEs for full stack development
6. Jupyter lab and Jupyter notebook for data cleaning, processing and analytics

# References

|  |  |
| --- | --- |
| [1] | S. Selvin, G. E. Vinayakumar R, V. K. Menon and S. K.P, "A System for Predicting Stock Price and Offering Financial Advice". |
| [2] | K. Pawar, R. S. Jalem and V. Tiwari, "Stock market price prediction using LSTM RNN". |
| [3] | S. Bai, J. Z. Kolter and V. Koltun, "An Empirical Evaluation of Generic Convolutional and Recurrent Networks for Sequence Modeling," 2018. |
| [4] | E. Hoseinzade and S. Haratizadeh, "CNNpred: CNN-based stock market prediction using a diverse set of variables," *Expert Systems With Applications,* pp. 273-285, 2019. |
| [5] | C. StoeanI, W. Paja, R. Stoean and A. Sandita, "Deep architectures for long-term stock price prediction with a heuristic-based strategy for trading simulations," *PLoS ONE,* 2019. |
| [6] | T. Kim and H. Y. Kim, "Forecasting stock prices with a feature fusion LSTM-CNN model using different representations of the same data," *PLoS ONE,* 2019. |
| [7] | M. U. Gudelek, S. A. Boluk and A. M. Ozbayoglu, "A Deep Learning based Stock Trading Model with 2-D CNN Trend Detection," in *2017 IEEE SSCI*, Honolulu, 2017. |
| [8] | J. Eapen, A. Verma and D. Bein, "Novel Deep Learning Model with CNN and Bi-Directional LSTM for Improved Stock Market Index Prediction," in *2019 IEEE 9th Annual CCWC*, Las Vegas, 2019. |
| [9] | N. R., "Introduction to ARIMA: Nonseasonal Models," 2018. [Online]. Available: https://people.duke.edu/~rnau/411arim.htm. [Accessed 18 09 2020]. |
| [10] | L. B. and J. S. Taylor, "Prophet: Forecasting At Scale - Facebook Research," Facebook Research, 23 02 2017. [Online]. Available: https://research.fb.com/prophet-forecasting-at-scale/. [Accessed 18 09 2020]. |
| [11] | A. G. Mata, "A comparison between LSTM and Facebook Prophet models: a financial forecasting case study," Hdl.handle.net, 01 2020. [Online]. Available: http://hdl.handle.net/10609/107367. [Accessed 18 09 2020]. |
| [12] | P. L. W. L. C. H. a. Y. W. W.Fang, "Combine Facebook Prophet and LSTM with BPNN Forecasting financial markets: the Morgan Taiwan Index," pp. 1-2International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS), 2019. |

# Appendix A: Meeting Minutes

## Minutes of the 1st Project Meeting

Date: July 19, 2020

Time: 10:00 pm

Place: Home, via Zoom meeting

Present: Aadhar, Jose, Michael, Dragos

Absent: None

Recorder: Michael

1. Approval of minutes

This was the first formal group meeting, so there were no minutes to approve.

1. Report on progress
   1. All team members have read the instructions of the Final Year Project online and have done research for the topic.
   2. All team members have read the information provided by Prof. Lee.
2. Discussion items
   1. The goal of the project is to build a system for predicting stock prices and offering financial advice.
   2. Instead of just doing the prediction, the system should be able to offer personalized suggestion to users on how they should their investment portfolio based on their different risk aversions. Users should be able to make decision on buy, sell, hold, or diversify.
   3. The project plan needs to include a list of the main tasks, who will work on each task, and a GANTT chart.
   4. React should be used in the frontend and perhaps flask will be used for backend.
3. Goals for the coming week
   1. All team members will continue to research on the topic.
   2. All team members will read the past FYP reports provided by Prof. Lee.
4. Meeting adjournment and next meeting

The meeting was adjourned at 11:00 pm.

The date and time of the next meeting will be sent through the WhatsApp group.

## Minutes of the 2nd Project Meeting

Date: Sep 4, 2020

Time: 10:00 pm

Place: Home, via Zoom meeting

Present: Aadhar, Jose, Michael, Dragos

Absent: None

Recorder: Michael

1. Approval of minutes

The minutes of the last meeting were approved without amendment.

1. Report on progress
   1. All team members have read about some papers about the topic.
   2. All team members have read past FYP papers.
2. Discussion items
   1. We divided the tasks and each person is responsible for different subsections for the FYP report.
   2. We have decided to ask Prof. Lee on our next meeting after we have some preliminary thoughts on the design and implementation.
3. Goals for the coming week
   1. Each member should complete their parts for the proposal and send questions to the group as soon as possible; also, we will add to the references section whenever we find something useful.
   2. Aadhar will work on the introduction, including the overview, objectives, and the literature survey.
   3. Michael will start working on the design and implementation subsections of the methodology part.
   4. Dragos will start the testing and evaluation parts.
   5. Jose will do the required hardware and software part and also help with Aadhar on the introduction.
   6. Michael will email Prof. Lee to schedule our next meeting.
4. Meeting adjournment and next meeting

The meeting was adjourned at 11:00 pm.

The date and time of the next meeting will be set later by email.

## Minutes of the 3rd Project Meeting

Date: Sep 11, 2020

Time: 10:00 am

Place: Home, via Zoom meeting

Present: Prof Lee, Aadhar, Jose, Michael, Dragos

Absent: None

Recorder: Michael

1. Approval of minutes

The minutes of the last meeting were approved without amendment.

1. Report on progress
   1. All team members have written most of their parts for the proposal report.
   2. Aadhar invited the other team members to help with the literature survey as he thought it was an important part of the proposal report.
2. Discussion items
   1. Prof Lee said the literature survey should focus on machine learning methods instead of traditional methods on stock analysis; and focus on what has been done in deep learning in the past few years, whether the researches were successful.
   2. For financial advice, Prof Lee noted we should think from the user’s point of view, as they won’t trust the prediction blindly but want advice to help them make good investment.
   3. There is no need for a feasibility section, but the literature review should highlight why the problem is difficult.
   4. Prof Lee reminded us to approach from the bottom up, start building from a smaller scale to ensure everything works before fine tuning the models.
   5. We should take about 1 to 2 months to think about the algorithm and should relate back to the literature survey.
   6. Prof Lee suggested that we should minimize programming, so we should probably choose Django instead of Flask for the backend framework.
   7. For testing and evaluation, Prof Lee said we should ask real users to test it and also run simulations, with the target that the algorithm should at least not lose money.
   8. Prof Lee said we should have a functional prototype by January and the project should be 80% complete by then; the goal would be that the system is technically reliable, functionally rich, and has good UI and back testing.
3. Goals for the coming week
   1. We will complete the remaining parts of the proposal report.
   2. We will each help with a subsection of the literature survey.
4. Meeting adjournment and next meeting

The meeting was adjourned at 10:45 am.

The date and time of the next meeting will be set later by email.

## Minutes of the 4th Project Meeting

Date: Oct 2, 2020

Time: 10:00 pm

Place: Home, via Zoom meeting

Present: Aadhar, Jose, Michael, Dragos

Absent: None

Recorder: Michael

1. Approval of minutes

The minutes of the last meeting were approved without amendment.

1. Report on progress
   1. The proposal report was submitted on schedule and awaiting feedback.
   2. Michael set up a Jira board to track tasks, progress and deadlines.
2. Discussion items
   1. We agreed that for the coming month the priority should be the data and backend setup.
   2. The target is to have a basic functional prototype by the end of semester.
3. Goals for the coming week
   1. Aadhar will research on the data stream.
   2. Jose will setup a server and host the project on Azure.
   3. Dragos will research on using Django as a backend framework.
   4. Michael will setup a repository and start setting up a database to store user information.
4. Meeting adjournment and next meeting

The meeting was adjourned at 10:45 pm.

The date and time of the next meeting will be set later by email.

## Minutes of the 5th Project Meeting

Date: Oct 16, 2020

Time: 10:00 pm

Place: Home, via Zoom meeting

Present: Aadhar, Jose, Michael, Dragos

Absent: None

Recorder: Michael

1. Approval of minutes

The minutes of the last meeting were approved without amendment.

1. Report on progress
   1. Dragos and Jose started setting up some code with Django framework and prototyped with it.
   2. Dragos created a web application with Django and supported user profile creation (including registration, login, and changing profile details).
   3. Michael started a frontend application with Create React App and connected it to the Django backend set up by Dragos. He used the Django Rest Framework library to create APIs that support authentication.
2. Discussion items
   1. Aadhar will research on different data sources for the financial and stock data, the current target is to use the Yahoo Finance API.
   2. Concluded that if we want to include news sentiment analysis then we can add that input to the model.
3. Goals for the coming week
   1. Continue with the infrastructure setup.
   2. Write the monthly report and ask for Prof Lee’s signature.
4. Meeting adjournment and next meeting

The meeting was adjourned at 10:45 pm.

The date and time of the next meeting will be set later by email.

## Minutes of the 6th Project Meeting

Date: Nov 16, 2020

Time: 10:00 pm

Place: Home, via Zoom meeting

Present: Jose, Michael, Dragos

Absent: Aadhar

Recorder: Michael

1. Approval of minutes

The minutes of the last meeting were approved without amendment.

1. Report on progress
   1. Michael continued with the integration between the Django backend and the React frontend and chose PostgreSQL as the database in place of the default SQLite provided by Django.
   2. Michael created a “portfolio” page for the web application where users can add their previous stocks transactions so that they can track their stocks portfolios.
   3. Jose prototyped with Django to create an application where a user can search for a stock symbol and then get the daily indicators of that stock.
2. Discussion items
   1. Agreed that a data crawler may not be necessary at this stage as we could find an API to get and store the data we want.
   2. We should create a prototype application using simple models first to illustrate the whole pipeline, before increasing performance and accuracy.
3. Goals for the coming week
   1. Jose will select an appropriate API to fetch financial data and then store into the database regularly.
   2. Dragos will start experimenting with different machine learning packages to create a pipeline where data fetched and stored in the database can be pre-processed and fed into a simple model, then the model can be trained, validated, and tested.
   3. Michael will continue with the frontend/backend integration.
4. Meeting adjournment and next meeting

The meeting was adjourned at 10:45 pm.

The date and time of the next meeting will be set later by email.

## Minutes of the 7th Project Meeting

Date: Dec 23, 2020

Time: 10:00 am

Place: Home, via Zoom meeting

Present: Jose, Michael, Dragos

Absent: Aadhar

Recorder: Dragos

1. Approval of minutes

The minutes of the last meeting were approved without amendment.

1. Report on progress
   1. Michael fixed a bug where users could sell stocks that he doesn’t own
   2. Michael further developed the frontend; stocks data can now be shown in the frontend.
2. Discussion items
   1. Agreed that the data fetched from current API is not comprehensive enough
   2. We should next find a way to display prediction results at the frontend
      1. Prediction results from database?
      2. Running ML models at the backend?
3. Goals for the coming week
   1. Jose will find an appropriate API that can fetch comprehensive stock data.
   2. Michael will continue to work on the frontend, trying to show last 5 months close price of the stocks.
   3. Dragos will create ml model with stock data in current database.
4. Meeting adjournment and next meeting

The meeting was adjourned at 10:45 am.

The date and time of the next meeting will be set later by email.

## Minutes of the 8th Project Meeting

Date: Jan 3, 2021

Time: 10:00 am

Place: Home, via Zoom meeting

Present: Aadhar, Jose, Michael, Dragos

Absent: None

Recorder: Michael

1. Approval of minutes

The minutes of the last meeting were approved without amendment.

1. Report on progress
   1. Aadhar emailed the communication tutor for the feedback on the proposal report but has yet to receive a reply.
   2. Jose analyzed several APIs and found that Alpha Vantage is the most viable one even though it has a limit on the number of calls per minute.
   3. Michael implemented the frontend to show graphs and also a user’s portfolio holdings.
   4. Dragos built a simple model by fetching data from the database.
2. Discussion items
   1. Aadhar suggested that we can use the alphas and betas to calculate the risk of a stock and also a user’s portfolio.
   2. Michael suggested to use the day difference in stock prices as the input to the machine learning models.
   3. Jose discussed the advantages and disadvantages of Alpha Vantage and we concluded that we should use that API.
   4. We agreed to use a risk assessment survey to determine the risk aversion level of a user.
3. Goals for the coming week
   1. Aadhar will try to set up a meeting with Prof. Lee on the coming Friday.
   2. Aadhar will calculate the risks of a stock and of a portfolio in the frontend.
   3. Jose will use Alpha Vantage to store some data into the database.
   4. Dragos will continue with the model and try using the day difference in stock prices as input.
   5. Michael will try displaying prediction results on the frontend.
   6. Michael will implement the survey in the frontend.
4. Meeting adjournment and next meeting

The meeting was adjourned at 10:45 am.

The date and time of the next meeting will be set later by email.

## Minutes of the 9th Project Meeting

Date: Jan 10, 2021

Time: 9:00 pm

Place: Home, via Zoom meeting

Present: Prof Lee, Aadhar, Jose, Michael, Dragos

Absent: None

Recorder: Michael

1. Approval of minutes

The minutes of the last meeting were approved without amendment.

1. Report on progress
   1. Jose implemented the functions using the API from Alpha Vantage that can store stock data into the database.
   2. Michael found a risk tolerance survey online and implemented it into the frontend.
   3. Michael used Drago’s model to display some prediction results on the frontend, including the predicted price for the next 30 days.
2. Discussion items
   1. Prof Lee suggested that we can have separate models for different sectors, as stocks in the same sector may have similar trends.
   2. We should decide how many stocks and how many years of data to play with.
   3. For our prototype, we should focus on completing the whole pipeline to make sure it works, including the risk and portfolio weighting recommendation, and data collection, storage, training, prediction, and visualization.
   4. Start with small amount of data and then finetune the accuracy at a later stage, then sample different models to see which is more accurate.
   5. After the prototype is tested and works, add more other features such as sentiment analysis, building more sophisticated models, and advisory for investors.
3. Goals for the coming week
   1. Provide the link to the prototype for Prof Lee to try it out.
   2. Give Prof Lee an overall plan in the monthly report and include the features we plan to implement and how we plan to integrate individual components together.
   3. Try to assign weighting in a portfolio for investment.
   4. Calculate the risk of a stock and a portfolio to balance return vs risk.
4. Meeting adjournment and next meeting

The meeting was adjourned at 9:45 pm.

The date and time of the next meeting will be set later by email.

## Minutes of the 10th Project Meeting

Date: , 2021

Time: 10:00 pm

Place: Home, via Zoom meeting

Present: Aadhar, Jose, Michael, Dragos

Absent: None

Recorder: Michael

1. Approval of minutes

The minutes of the last meeting were approved without amendment.

1. Report on progress
2. Discussion items
3. Goals for the coming week
4. Meeting adjournment and next meeting

The meeting was adjourned at 10:45 pm.

The date and time of the next meeting will be set later by email.