RO1

- Good report!
- Update the verb tenses and GANTT chart.
- Update your title on p. 2
  and on page headers.
- Mention ensemble learning

FYP Proposal

Personalized Algorithmic Trading using Machine Learning / Deep Learning models

OR: "Personalize Algo-trading Using Ensemble Learning"

by

MUN Sun Bin, KWOK Ue Nam, KONG Kin Cheung and HO Yat Man Peter

RO1

Advised by

Prof. David Rossiter

Submitted in partial fulfillment of the requirements for COMP 4981

in the

Department of Computer Science

The Hong Kong University of Science and Technology

2021-2022

1. Introduction

1.1 Overview

1.1.1 Stock Market Bubble Growth

Due to recent advances and convenience of ~~Being benefited by the technology enhancement on~~ mobile technologies, investors nowadays can easily ~~are much easier to~~ enter the stock market through mobile applications developed by ~~the~~ security brokers. Due to the economic slowdown caused by governments' COVID-19 policies, ~~With a coincidence of the COVID-19 pandemic, Governments are undergoing quantitative easing to support their economies~~ many central banks are rapidly debasing their currencies to hold off imminent economic collapse and maintain the status quo. As a result, a large amount of capital has flowed into ~~the~~ stock markets as more ~~people stepped into it to keep their assets valuable over~~ and more people fear that the value of their hard-earned savings will continue to be eroded by inflation [18]. However, many of them are not experienced in the stock or securities markets, ~~in which~~ so they have inadequate knowledge ~~and resources (Investment Capital or Time)~~ and time to manage their investments. Most are also unaware that the global elite have planned for decades to make the whole global financial system implode so that they can have a "Great Reset" and bring in a new blockchain-based system in which proof of work is done through human brain activity [20, 21].

1.1.2 Machine Learning and Deep Learning in Trading

Machine Learning (ML) is a study of constructing a model operating under complex algorithms which will improve over the iterative learning process without programmed instructions.

Further stretching, Deep Learning (DL) is an extension of Machine Learning that makes use of multiple artificial neural networks to return higher-level features of the data. The field studies constructing algorithms to perform like the neural network in the human brain.

In recent years, both Machine Learning and Deep Learning models have been ~~are~~ applied to a vast range of industries, including the financial industry. Some popular models used by quantitative analysts and data scientists use ~~used~~ for stock prediction include:

1. Convolutional Neural Network (CNN)
2. Long ~~Short Term~~ Short-term Memory (LSTM)
3. Recurrent Neural Network (RNN)
4. Reinforcement Learning (RL)
5. Q-Learning (reinforcement learning ~~model~~ without ~~the~~ an agent)

1.1.3 Algorithmic Trading

Among all the practical applications of ML and DL in the financial industry, such as fraud detection, market insights, and financial advisory, algorithmic trading (or electronic trading) plays a great role. According to Coherent Market Insights, the equity market ~~will~~ is expected to contribute $8.61 billion ~~in~~ of the ~~A~~ a lgo-trading market share by 2027 [14]. Moreover, Coalition Greenwich reported that the top 12 investment banks profited around USD 2 billion from ~~the~~ portfolio and algorithmic trading in 2020 [15].

Because the ~~trading market is a~~ financial markets are vibrant, fast-paced, and hard to predict ~~environment~~, algorithmic trading is of great use with its higher entry speed, concrete execution protocol, and lower degree of bias and emotion compared to human-initiated trading, ~~and less prone to human bias~~. Based on the input data, ~~the~~ models ~~will be~~ are trained to accurately predict ~~the~~ prices and can be applied to ~~the~~ real markets to form a profitable portfolios.

With the diverse pool of strategies, we ~~propose~~ are developing a comprehensive stock trading ~~pipeline~~ system with ~~the~~ an ~~selection~~ ensemble of algorithms/strategies and suggesting what ~~type of~~ strategies that are ~~the most~~ appropriate for each user. Our team plans to apply ~~the~~ ensemble learning, i.e., using a unique 'voting system' by first filtering out the outlier predictions and rearranging the contribution of different models in the final price decision. The predicted value of each stock will then determine how to distribute ~~the~~ stocks in a portfolio and determine whether ~~it~~ the portfolio has an ~~apt~~ appropriate risk level and offers maximum profit.

1.2 Objectives

~~1.2.1 General goal overview~~

The ~~objective~~ main goal of our project is to develop a user-friendly algorithmic trading ~~pipeline~~ application with a diverse choice of models for stock price prediction and portfolio distribution that can beat the market by more than 5%. Moreover, when comparing our project to other existing ~~R~~robo-advisors or algorithmic trading platforms, we hope that our final product can be better in terms of two ~~prospects~~ aspects. Firstly, with the help of ~~comparing~~ ensemble learning, i.e., using multiple machine learning models, and different trading strategies, we would like to achieve higher accuracy in stock price prediction ~~when comparing the~~ compared to existing ~~R~~robo-advisors in the market. Secondly, we would like to develop a truly automated platform that does not require users to make any judgments or choices during investing. To achieve ~~the above two goals, we have designed~~ this, we have been working on the following objectives:

1. **Database Implementation**

   The database stores the training data, test data, and machine learning models.

   (a) Historical data is ~~being~~ stored for training the models and testing the models

   (b) Real-time data is ~~being~~ pushed into the database constantly through APIs, keeping the database updated.

   one or several APIs ?

   (c) ~~While the~~ After a user is ~~being~~ classified according to desired level of risk, a ~~the~~ dedicated model will be retrieved for ~~the~~ making choices according to his risk level. ~~corresponding risk.~~

2. **Machine Learning Model~~s~~ Training**

   To have ~~a better~~ good performance in trading, having ~~an~~ accurate stock price prediction is very important. One of the ways to achieve this is to use multiple models to predict different results and then ~~use~~ combine the results to produce the final predictions of ~~the~~ stock prices. The method we use here is to filter out some of the outliners/top and bottom-n results and then average ~~it~~ out, the remaining predictions. Our ~~The~~ method of averaging out can be ~~averaged~~ done with weighting. The weighting can be ~~figured out by 2~~ calculated by using two approaches:

1. Based on every prediction, higher weighting ~~would~~ will be assigned to the data in the main cluster.
2. Based on previous training history, the models with better performance ~~would~~ will be assigned with ~~the~~ higher weighting.

3. Risk Quantification

To adequately categorize users' ~~To be more familiar with the users' profile and the~~ risk-bearing levels, we ~~will~~ designed a survey to understand users' financial situations and investment preferences ~~survey for the user to obtain the information of users and then classify them into different risk levels~~. After knowing the risk level and the areas of interest of ~~the~~ users, we ~~would~~ then automatically select a few corresponding potential stocks for price and trend prediction. With the price prediction results and the users' profiles, the platform will automatically trade the selected stocks.

4. Portfolio Presentation

We are building a ~~A~~ user-friendly web dashboard ~~will be built~~ for visualizing the users's portfolio and the performance of our models compared with the US stock market. It contains the following parts:

(a) A pie ~~Pie~~ chart is used to visualize the distribution of stocks currently held in ~~holding~~ a users portfolio.

(b) Multi-line charts are used to compare the performance of ~~between~~ our strategies and the "buy and hold" strategy.

~~1.2.3 Objectives plans~~

To achieve the first objective, we ~~would get~~ scraped X years of historical data from the US stock market and then split it into two parts: the training set and the testing set. After training the models, we will filter out some of the models that return top and bottom n-results and use the ~~rest~~ results of the remaining models to calculate the average for stock price prediction.

To achieve the second objective, a virtual machine is needed for ~~building~~ the database, and ideally, it should be able to constantly push new data into the database.

To achieve the last three objectives, a web application with an intuitive UI is being ~~will be~~ built to perform most of the interaction with the users.

~~1.2.4 Potential Challenges~~

As mentioned above, the financial market is so uncertain and difficult to predict, ~~we can foresee~~ that one of the major challenges has been ~~would be~~ how to train an accurate and reliable stock price prediction model. As mentioned above, we are using ~~plan to use~~ different machine learning models to predict stock prices independently, and then we apply filtered weighting. The major challenge has been deciding ~~would be how to decide~~ which models to be dropped out and how to allocate the weighting of the fund results among the remaining ~~of the~~ models.

1.3 Literature ~~Review~~ Survey

There exist many methods of algorithmic trading and also quant firms with electronic trading execution systems. We build on two main areas of related work: Existing ~~A~~ algo-trading companies~~,~~ and the technologies currently used.

1.3.1 Existing Algo-Trading Companies

Wealthfront

Some companies use algorithmic trading to perform portfolio management. Machine Learning models deployed in each firm tend to make decisions based on low-risk/high-return rules. One example of a company using a r ~~R~~obo-advisor (a ~~A~~ digital platform providing algorithmic financial planning without any human supervision) is Wealthfront (https://www.weal...). Wealthfront users set investing goals and duration before placing the seed money, and then the platform takes care of the rest. It automatically manages portfolios while also allowing customized portfolios of single stocks, ETFs, and other funds. However, Wealthfront provides a rigid list of stocks in creating a portfolio, and users do not get to have a 'personalized' portfolio upon their will [19].

Aqumon

One of ~~Other than Wealthfront, another~~ Asia's leading trading platforms, Aqumon ~~(link:~~ (https://www.aqumon.com/en/) has rapidly ~~is in rapid~~ developed ~~development~~ recently. Aqumon looks similar to the expected outcome of this project, with four major functions ~~which are~~: Long Term Asset Allocation Portfolios, SmartStock Thematic Stock Portfolios, Stock Trading, and Cash Management Portfolios. One of the selling points for Aqumon is that they claim that their portfolios are built using the Nobel-winning Markowitz Efficient Frontier and Black-Litterman Model.

One of the advantages of Aqumon is that they provide various combinations of ETFs for the user to choose from. The users are given advice based on their risk-bearing level, but the users can make their own choices. However, financial-market newcomers might not know how to make ~~a~~ rational decisions. Therefore, our project aims to provide a fully automated trading platform that does all the trading decisions in the back end without asking for users' decisions.

1.3.2 ~~Literature Review on~~ Stock Price Prediction Models

Research on predicting stock prices was conducted by implementing two major techniques: 1) Machine Learning Methods and 2) Deep Learning Techniques. This is because these techniques can help us to predict the price movement in the future. The accuracy may not be high, but it's better than a wild guess.

Machine Learning Methods

There are several common techniques in the machine learning field that are used for stock price prediction, such as Support Vector Machine (SVM) [2,4,12], Random Forest (RF) [7,13], and Multiple Linear Regression (MLR) [10,12]. SVM and RF are practical and useful methods for classification in making trading actions (Buy/Sell/Hold), but they are not able to give a numerical prediction on the stock price. MLR is effective for finding the relationship between the historical data and the future price, but it is limited to linear relationships. ~~which~~ Deep Learning can overcome these limitations. Hybrid models [7,13] combined with data mining methods and machine learning methods are also used to enhance the improvement.

In [9], three machine learning methods ~~are being~~ were implemented, including Extended Hill Climbing (EHC), Grid Method (GC), and Differential Evolution Method (DEM).
These models are compared based on an annualized rate of return from strategy and two risk measures, annualized standard deviation, and the maximum drawdown.

Deep Learning Methods

Other than machine learning methods, deep learning has drawn our attention as well. Artificial Neural Networks (~~ANN~~ ANNs)[13], Convolutional Neural Networks (~~CNN~~ CNNs) [1,3,5,11] and Recurrent Neural Networks (RNNs) [3,5,6,8] are the generally used deep learning methods to tackle price prediction problems. The major advantage of a neural network is that it can solve more complicated problems, as the hidden layers in the neural network can extract more features from the data. [1] and [11] used a CNN to predict stock price movement by using 1-D price data as input data. A CNN is capable of capturing features from the data in the form of both 1-D and 2-D, which can be used to capture ~~the~~ patterns in the stock price data. Other than CNN, as stock price data is time-series data, an RNN may be useful as well for time series data. ~~Long Short Term~~ Long Short-term Memory (LSTM) [3,5,6,8], an improved version of RNN, ~~could~~ can also be a good choice to handle time-series data. However, the design of LSTM is more focused on historical data, which may not be very effective for predicting future events. Moreover, deep learning methods are quite sensitive to noise, which exists a lot in stock data. Therefore, hybrid models [3,5] were ~~being~~ applied in deep learning methods as well. These models are generally better than standalone models.

The existing research and products ~~always~~ usually apply one single model on the prediction of price trends. They focus on optimizing the models or making comparisons between models. However, they usually do not use ensemble learning, i.e., they do ~~it does~~ not consist of any fusion or integration of the results from different models. Thus, ~~which is included~~ in our project ~~and it should be able to spread~~ we are spreading out the prediction error from different models ~~such that~~ so as to lowering the risk.

As mentioned above, algorithmic trading ~~is the new generation of trading business.~~ now dominates stock market trading. On top of the current systems and algorithms implemented, we are ~~focus on~~ developing ~~the pipeline to~~ an application that generates ~~the~~ an optimal combination of multiple algorithms and studying whether such procedures will yield a higher return in the trading market.

2. Methodology

- Try to match the verb tenses to the current status of each part.

2.1 Design

Good chart!

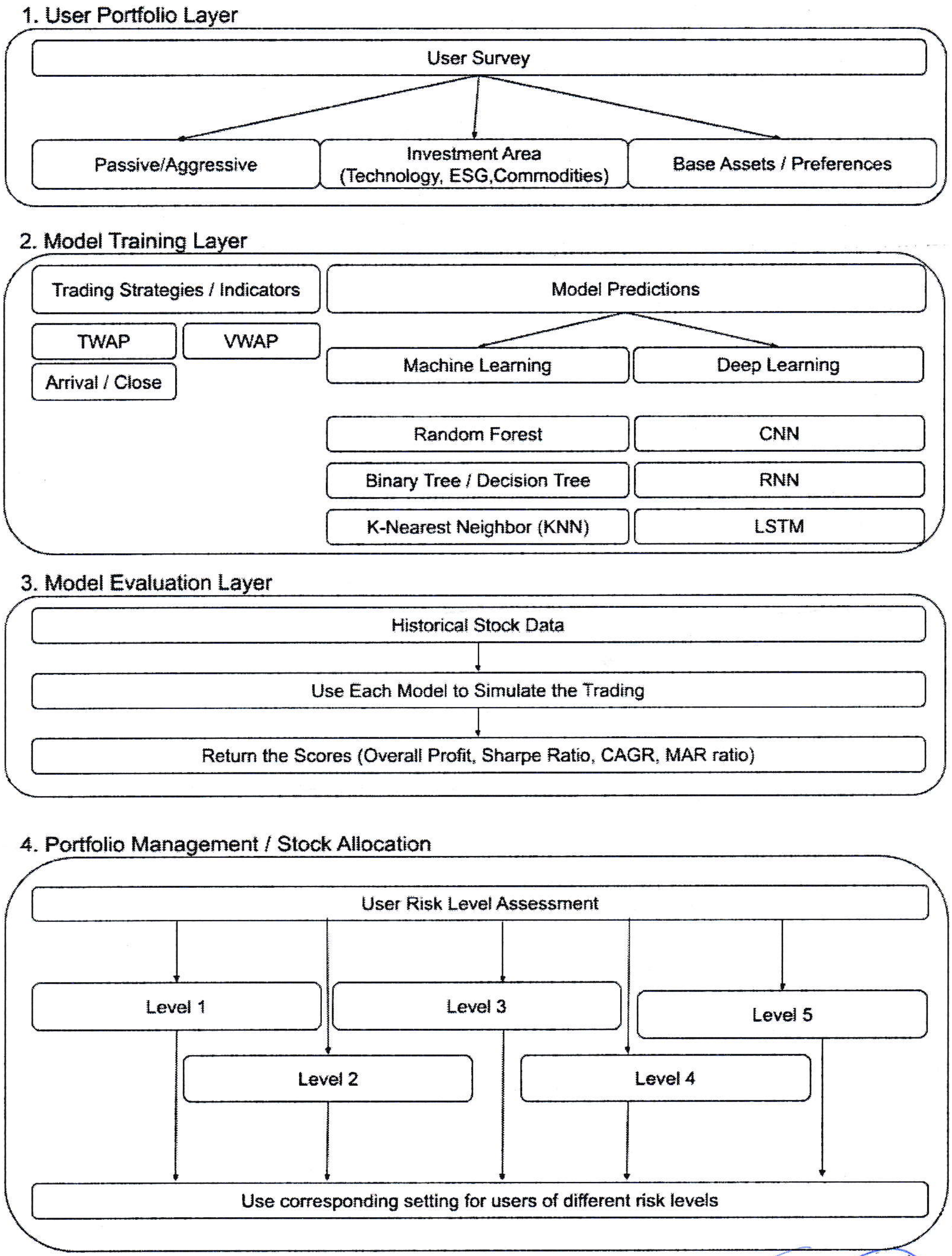

Figure 1. The initial design of the Customized portfolio pipeline ***"name"***

?

Use a smaller font size

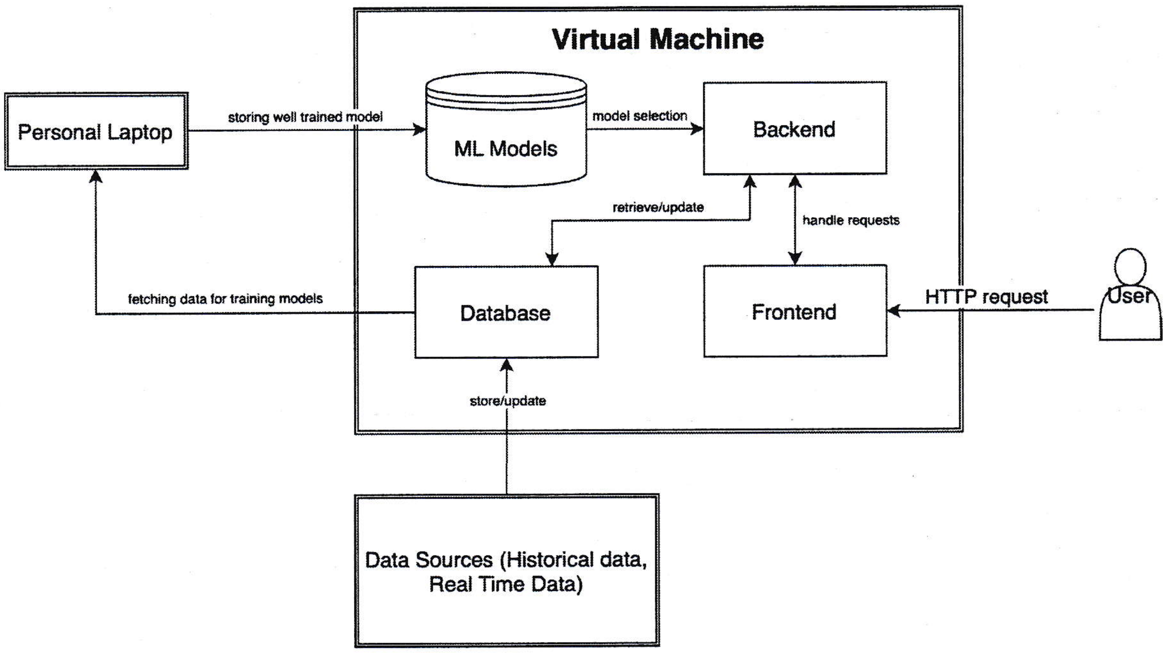

Figure 2. Initial Design of System Architecture for our platform } smaller font

The pipeline our team designed goes through the following steps:

2.1.1. User preference survey

Users will first be given a series of questionnaires related to their preference in specific fields and style of investment. All the responses will be recorded in the database and will classify users on a scale of 1 to 5 (1 being extremely passive and 5 being extremely active/risky).

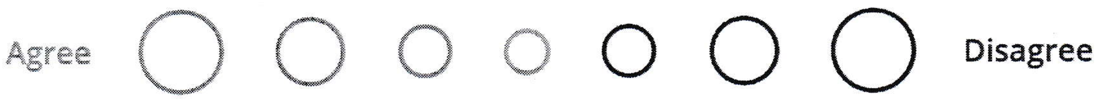

Figure 2. Example of possible question choices [17]

Example questions:

a. I am looking for the long term / short term investment
b. I prefer stocks that recently claimed IPOs / start-ups
c. I wouldn't mind higher risk if the portfolio promises a higher return (high risk - high return)
d. I would like to invest along with ETF indexes (safer)
e. I would like to invest in technology-related funds
f. I would like to invest on ESG related businesses

Around 20 questions will be asked for the user to categorize users.

you can put the details in an appendix
→ one for the survey
→ one for the results

2.1.2 User classification & Asset deposit

After the user has completed all the questionnaires, the pipeline finalizes the following regarding the user :

- Which type of portfolio the user expects from the **"name"**
- Which category (Blockchain, traveling, Semiconductors, Artificial Intelligence, Cloud, Biotechnology, Social Media, Entertainment, Retail, Franchise, Real estate, Telecommunication, Energy & Resources, Commodity, Luxury goods, etc) user prefer investing on
- How much of a risk the user is willing to take (user risk level assessment)
- Whether the user is looking for long term investment or HFT (High-Frequency Trading)

If the user wishes to adjust the level on their own, they can re-take the survey (from section 2.1.1) to re-adjust their category.

- Seed money balance

After the pipeline returns the list of information above, the user can take them into account in their trading activities.

2.2 Implementation

– Focus on How you implement what you designed

2.2.1 User survey - Input form

An input form will be included in the web application. The result will be sent back to the database and then decide the strategy/algorithm used in the backend based on the corresponding risk level the user has chosen.

2.2.2 Model training/testing/evaluation

We will use Jupyter Notebook or Google Colab for our major model training and testing tools. We will try out various models such as Random Forest (RF), Multiple Linear Regression (MLR), Long Short Term Memory (LSTM), and Convolutional Neural Network (CNN) with hyperparameters tuning.

Evaluation of the models will be compared to other research papers as we aim for better performance than other models. Evaluation will also be done on the entire platform so that we can check if our platform can outperform other existing platforms.

2.2.3 Data ~~S~~s Storage

) Be consistent on your use of CAPS in headings

A database is needed for storing the historical data of one/more stocks. It is essential to the models' training and evaluation. MySQL database is the first option since our team is familiar with it and it is suitable for time series data like stocks' price data. For historical data, our target is to find free data with around 20 years of stock prices. For real-time data, we will try to extract data from different brokers' APIs, for example, Interactive Broker, First Trade, FUTU, and then store it in the database for later use. If we could not afford the charges of real-time data, we may use free data from the internet with a little delay of the data (i.e. 1-day delay). [Provide citations for each data source you use.]

Historical data will be stored in our database as a backup in case of any modification of the data source. Before the training, necessary data preprocessing such as normalization and removing outliers will be conducted for removing the effect on the scale of different features.

2.2.4 Graphic User Interface ~~(GUI)~~

A web application will be built to present the portfolio of the user in a dashboard view. There are numerous templates available online and we can choose one and then modify it to suit our needs. There are also many functions available online so that we can simply integrate them into our application. HTML/CSS/Javascript or ReactJS are the possible options for developing the web application. Some examples are below:

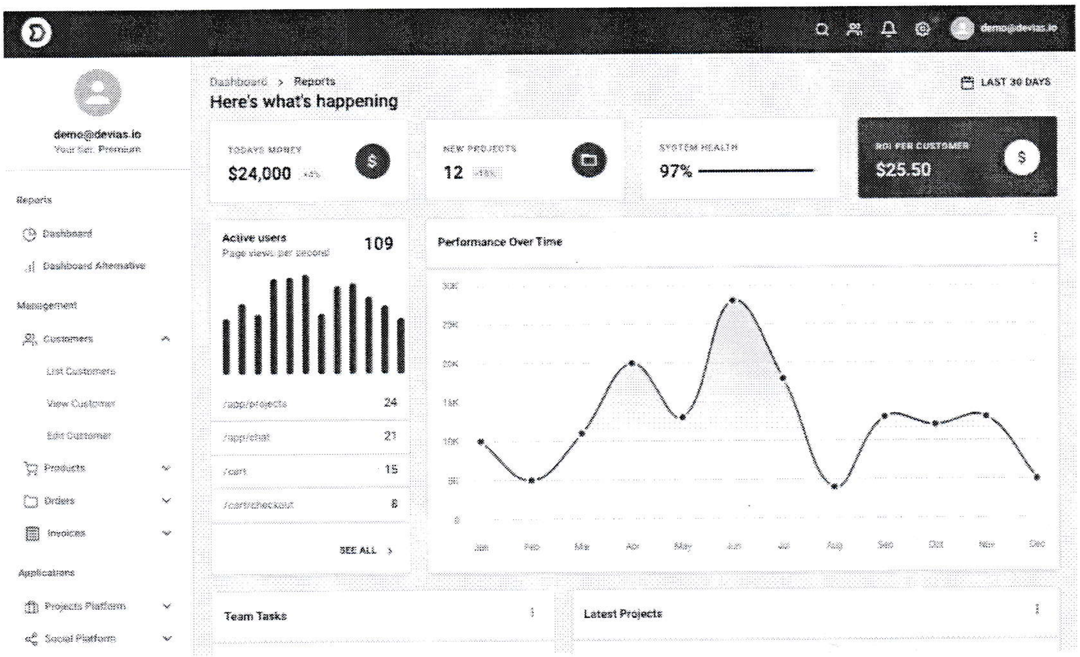

Figure 3. Dashboard Design Temple - 1

source: https://material-ui.com/store/collections/free-react-dashboard/

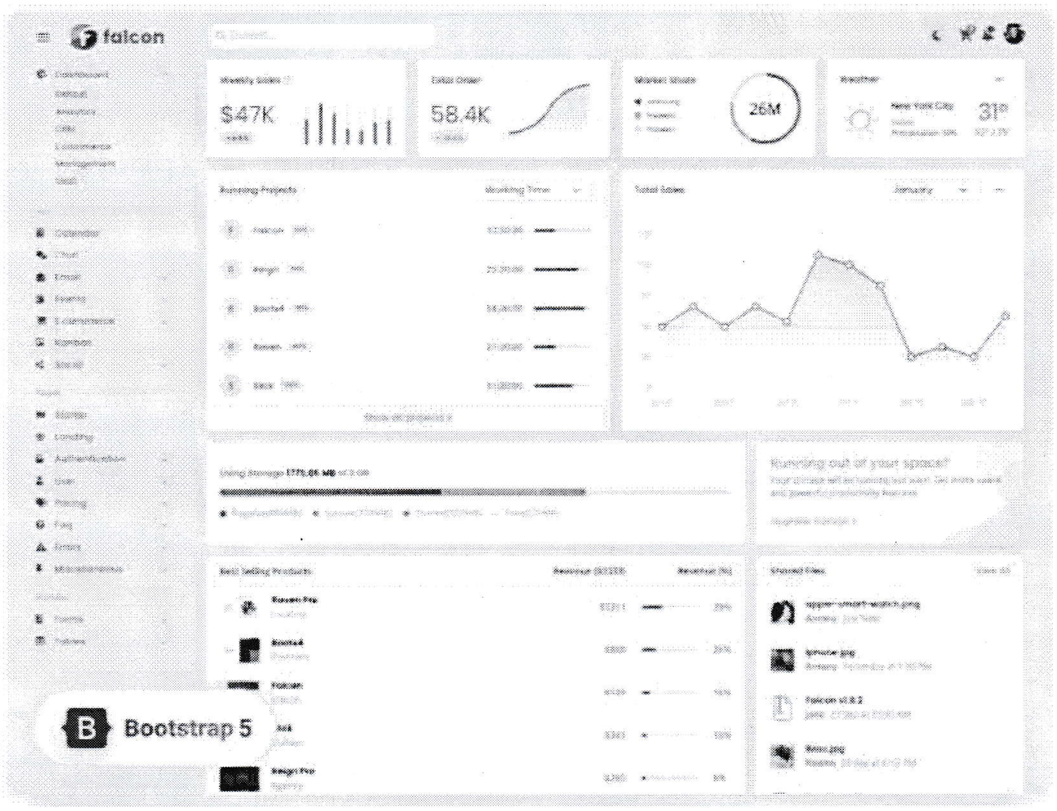

Figure 4. Dashboard Design Template - 2

source: https://themes.getbootstrap.com/product/falcon-admin-dashboard-webapp-template/

2.3 Testing ~~(Backtest)~~

Testing will be done on both the Price Prediction Model and Our Entire Platform with 3-5 years of historical data.

To evaluate the performance of the model, we will use a simulator to simulate the trading process. After we train a model, historical and unseen testing data will be fed into the model for generating testing results (list of buying and selling), like in CSV format. Then, the results will be imported into a simulator and start simulating the trading process and calculating the corresponding profit and some performance indicators mentioned above such as CAGR, Sharpe Ratio, and Annual Income.

2.3.1 Testing ~~on~~ the Price Prediction Models

For the historical data we have, we will first set a cut-off date as the date to separate the dataset. The data before the cut-off date would be used to train the model while the data later than the date will be used to test the model. The performance of the model would be assessed by the various Indicators such as Model Accuracy Score, Model Trend Score,

3. Project Planning

Very good!

But try to squeeze it onto one A4 page.

3.1 Distribution of work

Task	Peter	Anthony	Sun	Thomas
Project Designing				
1. Database Design	○	●	○	○
2. GUI Design	●	○	○	○
Model Training & Testing				
3. CNN	○	●	○	○
4. RNN	○	●	○	○
5. LSTM	○	○	●	○
6. Random Forest	○	○	●	○
7. KNN Algo	○	○	●	○
8. Decision Tree	○	○	○	●
9. SVM	○	○	○	●
Database Implementation				
10. Database set up on VM	○	○	○	●
11. Database Structure	○	○	●	○
API Data Extraction				
12. Apt Dataset Selection	○	○	●	○
13. Database Storage Automation	○	○	○	●
14. Data API & Pipeline Connectivity	○	●	○	○
GUI Implementation				
15. User portfolio	●	○	○	○
16. Pie chart distribution	●	○	○	○
17. Trend line analysis	●	○	○	○
18. Returned portfolio dashboard	●	○	○	○
Model Evaluation				
19. Backtesting - Model	○	●	●	○
20. Backtesting - User	○	●	●	●
21. Further Evaluation	●	●	●	●
Other				
22. ~~Poster Design~~ Video Trailer	●	●	●	●

- Good, but try to fit the chart on one A4 page
- Also update the chart to match the current status.

3.2 Gantt Chart

	2021				2022				
Tasks	**Sep**	**Oct**	**Nov**	**Dec**	**Jan**	**Feb**	**Mar**	**Apr**	**May**
Project Designing									
1. Database Design									
2. GUI Design									
Model Training & Testing									
3. Model Building - Different files for a Departmentdifferent model									
4. Model Testing - CNN									
5. Model Testing - RNN									
6. Model Testing - LSTM									
7. Model Testing - Random Forest									
8. Model Testing - KNN Algo									
9. Model Testing - Decision Tree									
Database Implementation									
10. Database set up on VM									
11. Database Structure									
API Data Extraction									
12. Apt Dataset Selection									
13. Database Storage Automation									
14. Data API & Pipeline Connectivity									
GUI Implementation									
15. User portfolio									
16. Pie chart distribution									
17. Trend line analysis									
18. Returned portfolio dashboard									
Model Evaluation									
19. Backtesting - Model									
20. Backtesting - User									
21. Further Evaluation									
Other									

4. Required Hardware and Software

4.1 Hardware

- 4 Laptops

4.2 Software

- Jupyter / Google Colab — For testing
- Python — Programming language
- Virtual Machine (AWS) — For hosting the DB and server
- React JS? — For building the front end
- HTML / CSS / Javascript?
- Any others?
- Git — For software version management

5. References

Very good, but the IEEE style is more popular in the CSE Dept. → see the CT website

1. Chen, S. & He, H. (2018). Stock Prediction Using Convolutional Neural Network. In *IOP Conf. Series: Materials Science and Engineering,435*. Retrieved from University of Shanghai for Science and Technology, 1 School of Optical-Electrical and Computer Engineering
   Web site: https://iopscience.iop.org/article/10.1088/1757-899X/435/1/012026/pdf
2. Henrique, B. M., Sobreiro, V. A. & Kimura, H. (September 2018). Stock price prediction using support vector regression on daily and up to the minute prices. In *The Journal of Finance and Data Science,4*(3). Retrieved from University of Brasília, Department of Economics
   Web site: https://www.sciencedirect.com/science/article/pii/S2405918818300060
3. Lu, W., Li, J. & Li, Y. (24 November 2020). A CNN-LSTM-Based Model to Forecast Stock Prices. In *Artificial Intelligence for Smart System Simulation,2020*. Retrieved from University Abdelmalek Essaadi, Morocco,
   Web site: https://www.hindawi.com/journals/complexity/2020/6622927/
4. Madge, S. (2015). *Predicting Stock Price Direction using Support Vector Machines*. Retrieved from Stanford University, Department of Computer Science
   Web site: https://www.cs.princeton.edu/sites/default/files/uploads/saahil_madge.pdf
5. Mehtab, S. & Sen, J. (8 November 2020). *Stock Price Prediction Using CNN and LSTMBased Deep Learning Models*. Retrieved from Praxis Business School, Kolkata, INDIA, Department of Data Science
   Web site: https://arxiv.org/pdf/2010.13891.pdf
6. Moghar, A. & Hamiche, M. (2020). Stock Market Prediction Using LSTM Recurrent Neural Network. In *Procedia Computer Science,170*. Retrieved from University Abdelmalek Essaadi, Morocco,
   Web site: https://www.sciencedirect.com/science/article/pii/S1877050920304865
7. Polamuri, S. R., Srinivas, K. & Mohan, A. K. (September 2019). Stock Market Prices Prediction using Random Forest and Extra Tree Regression. In *International Journal of Recent Technology and Engineering,8*(3). Retrieved from University Abdelmalek Essaadi, Morocco, Web site:
   https://www.researchgate.net/publication/347994783_Stock_Market_Prices_Prediction_using_Random_Forest_and_Extra_Tree_Regression
8. Qiu, J., Wang, B. & Zhou, C. (3 January 2020). *Forecasting stock prices with long-short term memory neural network based on attention mechanism*. Retrieved from University of Brasília, Department of Economics
   Web site: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0227222

IEEE style

16 Personalities
17. Free personality test. ~~(n.d.). Retrieved from~~ (n.d.)
~~Website~~: https://www.16personalities.com/free-personality-test (accessed 15 Sept, 2021).

18. Cox, J. (2021, April 9). Investors have put more money into stocks in the last 5 months than the previous 12 years combined. *CNBC,* p..
Website:https://www.cnbc.com/2021/04/09/investors-have-put-more-money-into-stocks-in-the-last-5-months-than-the-previous-12-years-combined.html

19. Smith, Christopher. "Wealthfront Review: How It Works, Pros & Cons." Clark Howard, 27 Apr. 2021,
Website:
https://clark.com/personal-finance-credit/investing-retirement/wealthfront-review/

[20] Microsoft, "Cryptocurrency System Using Body Activity Data," Int'l Patent WO 2020/060606 A1, Mar. 26, 2020.

[21] H. Sutcliffe. "COVID-19: The 4 building blocks of the Great Reset." World Economic Forum. https://www.weforum.org/agenda/2020/08/building-blocks-of-the-great-reset (accessed 6 Feb, 2022).

6. Appendix

6.1 ~~Meeting~~ Minutes of the 1st Meeting

~~1st Meeting Minutes~~

Date: 7 Aug 2021
Time: 11:00 - 11:57
Via: Discord

Participants:
KWOK, Ue Nam (20597580 unkwok)
HO, Yat Man Peter (20606082 ympho)
KONG, Kin Cheung (20595166 kckongaa)
MUN, Sun Bin (20455378 sbmun)

Meeting Discussion:

1. Contact the professor
2. Starting to prepare the proposal (Due on 17 Sept)
    - Due on 17 Sept
    - Suggested page number: 6
    - The task for this week: Overview - Sun and Thomas
    - The task for this week: Objective - Peter and Anthony
3. Discussing the direction
    - More focusing on Machine Learning
4. Schedule meeting 3-4 weeks later for discussing different methodologies
5. Regular meeting (Sunday 10 am)
6. Focusing on the pipeline (X on the GUI or application platform/server)
    - Data Extraction (Where to get stock data?)
    - Data Preprocessing
    - Model Design (Different algorithms and strategies)
    - Model Training
    - Evaluation (Evaluation Method? Simulate the trading once)

    Main focus: devising the unique algorithm for the stock prediction

Task Distribution:
1.1 Overview - Sun, Thomas
1.2 Objective - Peter, Anthony
1.3 - Altogether next week

Research Distribution:

- Sun: Algorithm for trading (VWAP, TWAP...) + some previous Python/ML implementation of these algorithms
- Thomas: Sentiment analysis/NLP
- Peter: previous FYP possible competitor
- Anthony: CNN or DNN (LSTM) for stock price data

Next Meeting: 14 Aug 2021

Minutes of the

6.2 **2nd Meeting ~~Minutes~~**

Date: 14 Aug 2021
Time: 10:30 - 11:30
Via: Discord

Participants:
KWOK, Use Nam (20597580 unkwok)
HO, Yat Man Peter (20606082 ympho)
KONG, Kin Cheung (20595166 kckongaa)
MUN, Sun Bin (20455378 sbmun)

Meeting Discussion:

1. Confirmation on using US Stock Market
2. Brainstorm on the final product
    - More focus on a trading model
    - Multiple Models for different combinations
    - Base on Users' risk level
    - Should we add Sentiment Analysis by NLP
    - Collaborative Filtering for User Classification (SVM, CF, etc...)
3. Visualization of fund results the Product and the Logic Flow
4. Email to Professor (Seek Advice and Share our Ideas)

Task Distribution:

1. FYP Report - Overview (Sun and Anthony)
2. FYP Report - Objective (Peter and Thomas)

Next Meeting: 1 Sept 2021 15:00-16:00