University of Sunderland

Faculty of Technology

Department of Computer Science

PLANNING REVIEW

PROM02

MSc Dissertation

Academic Year: 2022/23

Student Name: Chu Lai Ting, Jackie

Student ID 219265092

Email: bi06aa@student.sunderland.ac.uk

Programme: MSc Data Science

Mode: Part Time

Supervisor: Marta Rudina

1 Terms of Reference (50%)

a. Project Title

A stock forecasting model for swing trading recommendations in the Hong Kong stock market

b. Project Overview

***Aim*:** To develop a LSTM forecasting model to assist optimization of swing trading investment strategy in the Hong Kong stock market by providing a visual result as a recommendation for further action.

Objectives:
1. Background research by evaluating different investment strategies and studying modern approaches and methods used for stock forecasting
2. Study the Hong Kong stock market landscape by performing background research and technical analysis
3. Literature review to critically evaluate available forecasting algorithms and its application
4. Discuss data collection process by Yahoo Finance
5. Identification and discussion of relevant professional, ethical, social and legal issues
6. Build, train, deploy and visualize forecasting model(s) for evaluating the one with the highest prediction accuracy and efficiency to assist optimization of swing trading in the Hong Kong stock market

Research Question:
1. What modern approaches are used in stock forecasting?
2. What are the model features to be considered?
3. What improvements for existing stock forecasting models can be done by applying modern machine learning methods and techniques?

Practical outcomes:
1. Overview of the Hong Kong stock market by focusing on Hang Seng Index (HSI) and major 3 selected stocks for model development
   It includes summary statistics and different visuals on various technical indicators to find out insights that describe the market situation.
2. Data collection and pre-processing steps by Yahoo Finance
   - Daily data on high, low, open, close, volume, etc.
   - Missing value handling
   - Derived variables management
3. Perform technical analysis by Python libraries such as numpy, pandas, matplotlib, etc.
4. Detail the experiment design as well as a baseline model by Python libraries such as scikit learnt.
   - Data to be used including features and period adopted
   - Data collection methods and pre-processing steps
   - Baseline model building, LSTM model training and it's performance metrics
5. Build and visualize forecasting model(s) based on the selected algorithms with performance measurements presentations by using Python libraries such as keras, tensorflow, matplotlib, etc. Evaluate results of developed models and select the one with highest accuracy.

c. Underpinning research with Literature Review

Citation	*Brief summary of paper*	*Relevance to your research question*	*Relevance to practical outcome of project*
[1] Manimegalai, T., Manju, J., Rubiston, M.M., Vidhyashree, B. and Prabu, R.Thandaiah. (2022). *Prediction of OPTIMIZED Stock Market Trends using Hybrid Approach Based on KNN and Bagging Classifier (KNNB)*. [online] IEEE Xplore. doi:10.1109/CSNT54456.2022.9787638	1. Stock market trends prediction using various approaches 2. Compare different algorithms and the best was a hybrid approach of KNN and bagging classifier	The paper described a hybrid approach based on the conventional KNN to KNNB. Hybrid or ensemble approaches could be a consideration of enhancing from conventional models. Moreover, it compares a few commonly use approach as a proof of success to the newly suggested model	A few commonly use algorithms will be considered to find the best model approaches
[2] Khaidem, L., Saha, S. and Dey, S. (2016). Predicting the direction of stock market prices using random forest. *Applied Mathematical Finance*, [online] 00(00), pp.1–20. Available at: https://arxiv.org/pdf/1605.00003.pdf	1. Stock trend prediction can be used for devising new trading strategies 2. How to perform random forest ensemble classification model of predicting stock price go up after n days or vice versa 3. Use of technical indicators as features of the model 4. Use of confusion matrix and ROC curve for performance measurement	The paper opined trend prediction is better as price is affected by intrinsic volatility which can minimize forecasting errors as a classification problem	Consider using different technical indicators as features of the model
[3] Smita, M. ed., (2021). *Logistic Regression Model - A Review*. [online] Available at: https://ijisrt.com/assets/upload/files/IJISRT21MAY1050.pdf	1. Stock trend prediction aimed for making better informed investment decision 2. Detailed what is logistic regression, including assumption, theory and its merit and limitation 3. LR can be used as a baseline model	The paper illustrated a thorough review on logistic regression on assumption, methodology, parameter estimates and application area	Logistic regression will be adopted as a baseline model in this project
[4] Vijh, M., Chandola, D., Tikkiwal, V.A. and Kumar, A. (2020). *Stock Closing Price Prediction using Machine Learning Techniques. Procedia Computer*	1. Quoted machine learning techniques in stock prediction have improve efficiencies by 60-86% as compared to traditional technical or basic analysis	The paper summarized the methodology used and demonstrated how ANN worked on better stock price prediction than RF. Moreover, it collected historical data from Yahoo	The paper contained the end-to-end flow of an empirical study on modeling five selected stock. It showed better result from ANN than RF. More features analysis and the

Science, 167, pp.599–606. doi:10.1016/j.procs.2020.03.326	2. Described ANN and Random Forest algorithms applied to stock price prediction in 5 selected stocks 3. Features and derived new variables were used for predicting next stock closing price 4. ANN showed better prediction results than random forest by RMSE, MAPE and MBE	Finance and detail the new variable derivations	model building process will be further studied in this project
[5] Chimmula, V.K.R. and Zhang, L. (2020). *Time series forecasting of COVID-19 transmission in Canada using LSTM networks. Chaos, Solitons & Fractals*, 135, p.109864. doi:10.1016/j.chaos.2020.109864	1. Covid transmission prediction by LSTM 2. Detail discussion on internal mechanism of LSTM networks 3. The model successfully predict Covid transmission with accuracy reach 93% yet Covid data was little at the time of performing this research	The paper detailed the theory of LSTM and proven successful at predicting real time transmission	As stock trend also temporal in nature and recent event affect more on the next day stock price and LSTM will be considered as the primary model to develop in the practical study
[6] Thakkar, A. and Chaudhari, K. (2021). *A Comprehensive Survey on Deep Neural Networks for Stock Market: The Need, Challenges, and Future Directions. Expert Systems with Applications*, p.114800. doi:10.1016/j.eswa.2021.114800.	1. Comprehensive review on deep neural network algorithms, indicated the urgency under the volatile stock market for the scientific methodology on deriving inherent patterns for quick investment decisioning 2. Comparison of different models has carried out and analysed 3. Discussed the challenges of DNN stock prediction	The paper detailed most of the ANN and DNN algorithms for stock prediction and stated the pros and cons	The analysis of different model results could be a good reference as the background of focusing on LSTM in this experiment
[7] Chen, K., Zhou, Y. and Dai, F. (2015). *A LSTM-based method for stock returns prediction: A case study of China stock market. 2015 IEEE International Conference on Big Data (Big Data).* [online] doi:10.1109/bigdata.2015.7364089.	1. Experimental design for LSTM on China stock market using historical daily stock data from Yahoo Finance 2. Progressive methods of adding features on demonstrating the different of accuracy in each method 3. Case study showed LSTM is powerful in stock market by adequate features	The paper described a complete structure of experimental design on testing different features used in the model training	Try to design the practical part into a few attempts of methods and compare for better model and results

[8] Istiake Sunny, Md.A., Maswood, M.M.S. and Alharbi, A.G. (2020). *Deep Learning-Based Stock Price Prediction Using LSTM and Bi-Directional LSTM Model*. 2020 2nd Novel Intelligent and Leading Emerging Sciences Conference (NILES). doi:10.1109/niles50944.2020.9257950 .	1. Experimental design for LSTM on Google stock for LSTM and BiLSTM methods 2. Illustrated the concept of BiLSTM and its improvements from LSTM by utilizing all previous information from both directions 3. The paper demonstrated the tunning of hyperparameters highly affect the accuracy of the result 4. Result showed BiLSTM model had higher accuracy than LSTM model however LSTM model took lesser time to predict the data	The paper focused on LSTM and its variant of BiLSTM for stock prediction. Expanding the experiment could be by tuning hyperparameters and changing on algorithms	To be considered the experimental design could be on the aspects of tuning hyperparameters and similar algorithms
[9] Li, X., Wu, P. and Wang, W. (2020). *Incorporating stock prices and news sentiments for stock market prediction: A case of Hong Kong. Information Processing & Management,* p.102212. doi:10.1016/j.ipm.2020.102212	1. Using news sentiment as a complimentary to numerical stock price prediction 2. Short-term prediction method attempted for one-step forecast with more encouraging result 3. Long-term prediction method attempted for multi-step forecast that is more challenging with growing uncertainties 4. Combining NLP and LSTM models to form NBA with the best result among the tested model	The paper inspired the consideration of combining news sentiment analysis in literature review	To be considered if the practical work will combine with news sentiment analysis after literature review
[10] Mageswaran, G., Nagappan, S.D., Hamzah, N. and Brohi, S.N. (2018). *Machine Learning: An Ethical, Social & Political Perspective.* 2018 Fourth International Conference on Advances in Computing, Communication & Automation (ICACCA). doi:10.1109/icaccaf.2018.8776702.	1. The paper detailed machine learning technology use case in various aspects including healthcare, banking and political areas 2. Discussed on the impact to social, political and ethical considerations during machine learning application	The paper reinforced the importance of social, political and ethical issues in the machine learning application process	The experiment will well consider the social, political and ethical issues during development

2. Project Schedule (20%)

2.1 Table 1: Effort

Task Id	Task Name	Start	Deadline	Hours	Deliverable
1.0 Background research by evaluating different investment strategies and studying modern approaches and methods used for stock forecasting					
1.1	Research on different investment strategies	17Dec2022	31Dec2022	20	Understand big categories of investment strategies
1.2	Research on modern approaches and methods used for stock forecasting	17Dec2022	31Dec2022	20	Research pros and cons for different approaches and methods
1.3	Summarization	28Dec2022	8Jan2023	5	Identify the most suitable modern forecasting approaches and area for improvement
1.4	Dissertation write up	3Jan2023	8Jan2023	5	Write up on background and introduction
2.0 Study the Hong Kong stock market landscape by performing background research and technical analysis					
2.1	Identify data scope to be collected	9Jan2023	15Jan2023	10	Find out the big waves of Hong Kong stock market and summarize the analysis from longest available period to zoom in to recent 2 years market situation
2.2	Perform data analysis	16Jan2023	4Feb2023	20	Present the data analysis result of long, medium, short term that reflect the market situation
2.3	Summarize analytics finding	30Jan2023	5Feb2023	10	Consolidate the analytics finding and support the model building objectives
2.4	Dissertation write up	4Feb2023	5Feb2023	5	High level describe the Hong Kong stock market and summarize the exploratory data analysis finding
3.0 Literature review to critically evaluate potential forecasting algorithms and its application					
3.1	Literature review on logistics regression	17Dec2022	31Dec2022	10	Critically evaluate different usage by regression and its application
3.2	Literature review on neural networks and LSTM	1Jan2023	15Jan2023	10	Critically evaluate different usage by neural networks, LSTM and their application
3.3	Literature review on how LSTM variants to solve forecasting problems	16Jan2023	31Jan2023	50	Understand the theory of LSTM and its variants specific on forecasting problems
3.4	Summarization	30Jan2023	5Feb2023	10	Define the experimental design of the practical model building
3.5	Dissertation write up	5Feb2023	11Feb2023	10	Dissertation on literature review
4.0 Discuss data collection process by Yahoo Finance					
4.1	Research on available data collection method	6Feb2023	11Feb2023	10	Identify the available data collection method and corresponding advantages and disadvantages
4.2	Dissertation write up	12Feb2023	12Feb2023	5	State the available data collection method and reason of adoption

5.0 Identification and discussion of relevant professional, ethical, social and legal issues					
5.1	Research on social issues related to stock forecasting model	15Feb2023	24Feb2023	2	Discuss on social issue related to stock forecasting model
5.2	Research on ethical issues related to stock forecasting model	15Feb2023	24Feb2023	2	Discuss on ethical issue related to stock forecasting model
5.3	Research on professional issues related to stock forecasting model	15Feb2023	24Feb2023	2	Discuss on professional issue related to stock forecasting model
5.4	Research on legal issues related to stock forecasting model	15Feb2023	24Feb2023	2	Discuss on legal issue related to stock forecasting model
5.5	Dissertation write up	15Feb2023	24Feb2023	2	Summarize and state the relevant social, ethical, professional and legal issue related to stock forecasting model
6.0 Build, train, deploy and visualize forecasting model(s) for evaluating the one with the highest prediction accuracy and efficiency to assist optimization of swing trading in the Hong Kong stock market					
6.1	Learning Python to code for data preprocessing	25Feb2023	8Apr2023	50	Equip skills on Python code for data importation and preprocessing
6.2	Learning Python to code for data modeling	10Apr2023	26May2023	80	Equip skills on Python code for data modeling
6.3	Sprint 1: Develop, finetune and finalize baseline model by logistic regression model	29Apr2023	18May2023	40	Develop Python code for logistic regression modeling
6.4	Sprint 2: Develop, finetune and finalize LSTM modeling	19May2023	7Jun2023	40	Develop Python code for LSTM modeling
6.5	Sprint 3: Research for different variants and finalize for the best data model	30May2023	18Jun2023	40	Summarize model performance measures tables and charts
6.6	Summarization	19Jun2023	22Jun2023	10	Identify the model with the highest prediction accuracy and efficiency to assist optimization of swing trading in the Hong Kong stock market
7.0 Wrap up work done					
7.1	Summarize the model building process	23Jun2023	25Jun2023	20	Highlight the key steps of developing the model
7.2	Wrap up and final evaluation of the model outputs	26Jun2023	30Jun2023	40	Summarize and discuss the performance measures and discuss the findings from the output
7.3	Discuss the future enhancement	1Jul2023	2Jul2023	10	Discuss the future enhancement
7.4	Dissertation write up	3Jul2023	7Jul2023	40	Wrap up the practical work, conclusion and possible future enhancement
7.5	Viva preparation	3Jul2023	7Jul2023	40	Build presentation deck

2.2 Table 2: Deliverables

Del. No.	Name	Deadline
D1.1	Understand big categories of investment strategies	31Dec2022
D1.2	Research pros and cons for different approaches and methods	31Dec2022
D1.3	Identify the most suitable modern forecasting approaches and area for improvement	8Jan2023
D1.4	Write up on background and introduction	8Jan2023
D2.1	Find out the big waves of Hong Kong stock market and summarize the analysis from longest available period to zoom in to recent 2 years market situation	15Jan2023
D2.2	Present the data analysis result of long, medium, short term that reflect the market situation	4Feb2023
D2.3	Consolidate the analytics finding and support the model building objectives	5Feb2023
D2.4	High level describe the Hong Kong stock market and summarize the exploratory data analysis finding	5Feb2023
D3.1	Critically evaluate different usage by logistic regression and its application	31Dec2022
D3.2	Critically evaluate different usage by neural networks, LSTM and their application	15Jan2023
D3.3	Understand the theory of LSTM and its variants specific on forecasting problems	31Jan2023
D3.4	Define the experimental design of the practical model building	5Feb2023
D3.5	Dissertation on literature review	11Feb2023
D4.1	Identify the available data collection method and corresponding advantages and disadvantages	11Feb2023
D4.2	State the available data collection method and reason of adoption	12Feb2023
D5.1	Discuss on social issue related to stock forecasting model	24Feb2023
D5.2	Discuss on ethical issue related to stock forecasting model	24Feb2023
D5.3	Discuss on professional issue related to stock forecasting model	24Feb2023
D5.4	Discuss on legal issue related to stock forecasting model	24Feb2023
D5.5	Summarize and state the relevant social, ethical, professional and legal issue related to stock forecasting model	24Feb2023
D6.1	Equip skills on Python code for data importation and preprocessing	8Apr2023
D6.2	Equip skills on Python code for data modeling	26May2023
D6.3	Develop Python code for logistic regression modeling	18May2023
D6.4	Develop Python code for LSTM modeling	7Jun2023
D6.5	Summarize model performance measures tables and charts	18Jun2023
D6.6	Identify the model with the highest prediction accuracy and efficiency to assist optimization of swing trading in the Hong Kong stock market	22Jun2023
D7.1	Highlight the key steps of developing the model	25Jun2023
D7.2	Summarize and discuss the performance measures and discuss the findings from the output	30Jun2023
D7.3	Discuss the future enhancement	2Jul2023
D7.4	Wrap up the practical work, conclusion and possible future enhancement	7Jul2023
D7.5	Build presentation deck	7Jul2023

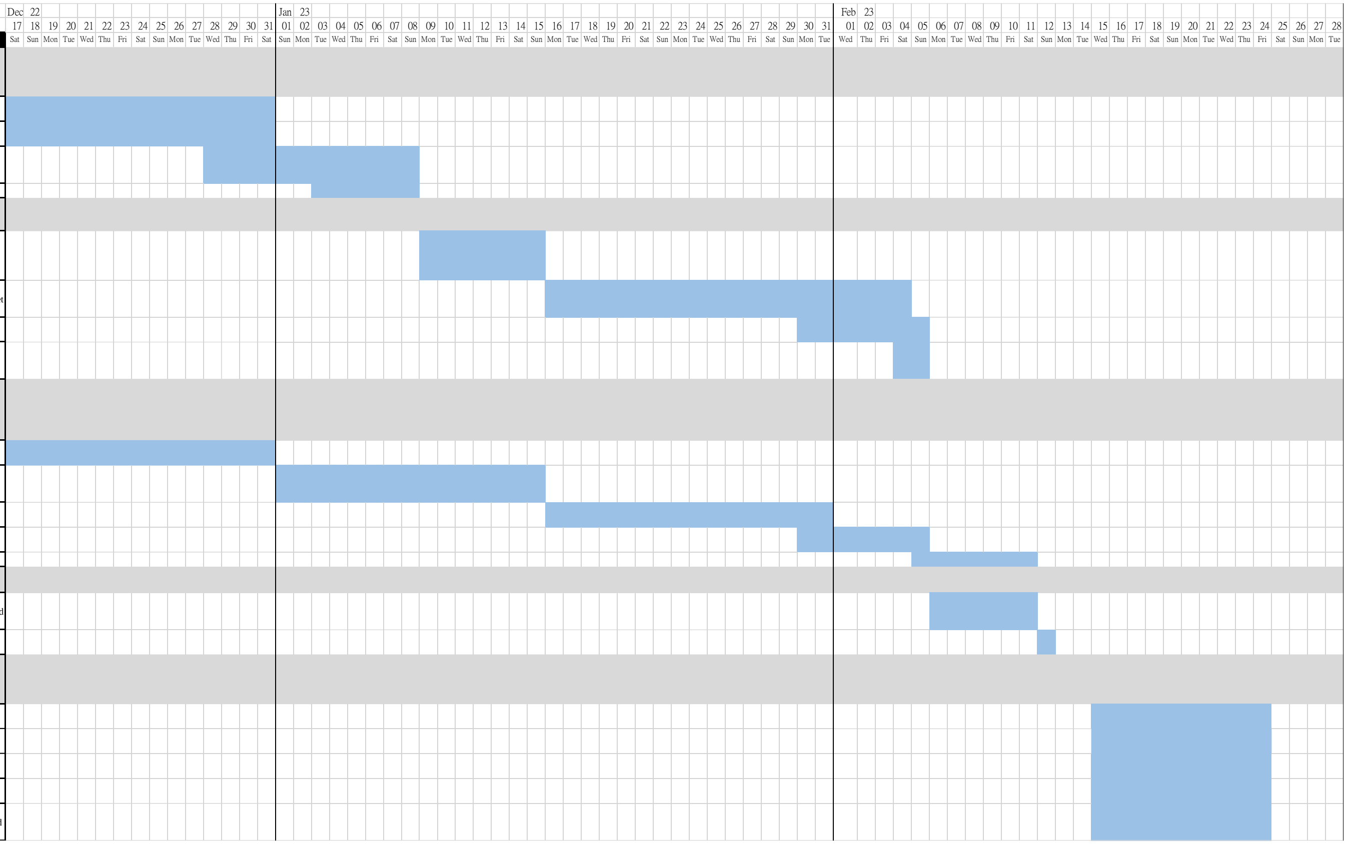
2.3 Table 3: Milestones

Milestone	Name	Deadline	Evidence
M1	Background research by evaluating different investment strategies and studying modern approaches and methods used for stock forecasting	8Jan2023	List, describe and discuss at least three investment strategies and modern approaches that can assist for decisioning
M2	Study the Hong Kong stock market landscape by performing background research and technical analysis	5Feb2023	Use of summary statistics and various charts for technical analysis that able to articulate the market situation in 2022
M3	Literature review to critically evaluate potential forecasting algorithms and its application	11Feb2023	Summarize literature review findings and lay down experimental design
M4	Discuss data collection process by Yahoo Finance	12Feb2023	Document the process flow with corresponding consideration
M5	Identification and discussion of relevant professional, ethical, social and legal issues	24Feb2023	Document the study and discussion of relevant professional, ethical, social and legal issues
M6	Build, train, deploy and visualize forecasting model(s) for evaluating the one with the highest prediction accuracy and efficiency to assist optimization of swing trading in the Hong Kong stock market	22Jun2023	Develop Python coding of the model building and evaluation process
M7	Wrap up work done	7Jul2023	Wrap up dissertation and build presentation deck

2.4 Table 4: Outline Schedule / Gantt chart

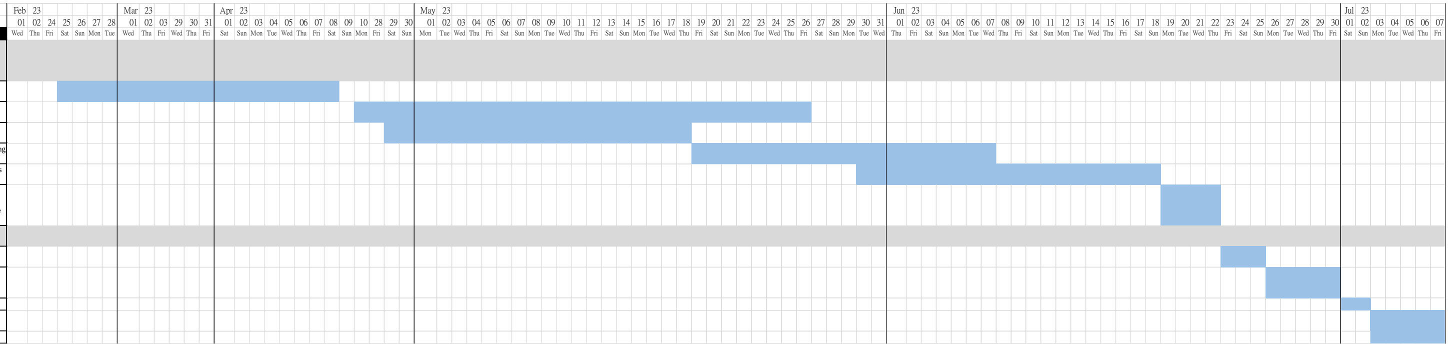
Detail Gantt chart refer to appendix

Task Id	Task Name	Start	Deadline	Hours	Deliverable
1.0 Background research by evaluating different investment strategies and studying modern approaches and methods used for stock forecasting					
1.1	Research on different investment strategies	17Dec2022	31Dec2022	20	Understand big categories of investment strategies
1.2	Research on modern approaches and methods used for stock forecasting	17Dec2022	31Dec2022	20	Research pros and cons for different approaches and methods
1.3	Summarization	28Dec2022	8Jan2023	5	Identify the most suitable modern forecasting approaches and area for improvement
1.4	Dissertation write up	3Jan2023	8Jan2023	5	Write up on background and introduction
2.0 Study the Hong Kong stock market landscape by performing background research and technical analysis					
2.1	Identify data scope to be collected	9Jan2023	15Jan2023	10	Find out the big waves of Hong Kong stock market and summarize the analysis from longest available period to zoom in to recent 2 years market situation
2.2	Perform data analysis	16Jan2023	4Feb2023	20	Present the data analysis result of long, medium, short term that reflect the market situation
2.3	Summarize analytics finding	30Jan2023	5Feb2023	10	Consolidate the analytics finding and support the model building objectives
2.4	Dissertation write up	4Feb2023	5Feb2023	5	High level describe the Hong Kong stock market and summarize the exploratory data analysis finding
3.0 Literature review to critically evaluate potential forecasting algorithms and its application					
3.1	Literature review on logistics regression	17Dec2022	31Dec2022	10	Critically evaluate different usage by logistic regression and its application
3.2	Literature review on neural networks and LSTM	1Jan2023	15Jan2023	10	Critically evaluate different usage by neural networks, LSTM and their application
3.3	Literature review on how LSTM variants to solve forecasting problems	16Jan2023	31Jan2023	50	Understand the theory of LSTM and its variants specific on forecasting problems
3.4	Summarization	30Jan2023	5Feb2023	10	Define the experiment design of the practical model building
3.5	Dissertation write up	5Feb2023	11Feb2023	10	Dissertation on literature review
4.0 Discuss data collection process by Yahoo Finance					
4.1	Research on available data collection method	6Feb2023	11Feb2023	10	Identify the available data collection method and corresponding advantages and disadvantages
4.2	Dissertation write up	12Feb2023	12Feb2023	5	State the available data collection method and reason of adoption
5.0 Identification and discussion of relevant professional, ethical, social and legal issues					
5.1	Research on social issues related to stock forecasting model	15Feb2023	24Feb2023	2	Discuss on social issue related to stock forecasting model
5.2	Research on ethical issues related to stock forecasting model	15Feb2023	24Feb2023	2	Discuss on ethical issue related to stock forecasting model
5.3	Research on professional issues related to stock forecasting model	15Feb2023	24Feb2023	2	Discuss on professional issue related to stock forecasting model
5.4	Research on legal issues related to stock forecasting model	15Feb2023	24Feb2023	2	Discuss on legal issue related to stock forecasting model
5.5	Dissertation write up	15Feb2023	24Feb2023	2	Summarize and state the relevant social, ethical, professional and legal issue related to stock forecasting model



PLANNING REVIEW

Task Id	Task Name	Start	Deadline	Hours	Deliverable
6.0 Build, train, deploy and visualize forecasting model(s) for evaluating the one with the highest prediction accuracy and efficiency to assist optimization of swing trading in the Hong Kong stock market					
6.1	Learning Python to code for data preprocessing	25Feb2023	8Apr2023	50	Equip skills on Python code for data importation and preprocessing
6.2	Learning Python to code for data modeling	10Apr2023	26May2023	80	Equip skills on Python code for data modeling
6.3	Sprint 1: Develop, finetune and finalize baseline model by logistic regression model	29Apr2023	18May2023	40	Develop Python code for logistic regression modeling
6.4	Sprint 2: Develop, finetune and finalize LSTM modeling	19May2023	7Jun2023	40	Develop Python code for LSTM modeling
6.5	Sprint 3: Research for different variants and finalize for the best data model	30May2023	18Jun2023	40	Summarize model performance measures tables and charts
6.6	Summarization	19Jun2023	22Jun2023	10	Identify the model with the highest prediction accuracy and efficiency to assist optimization of swing trading in the Hong Kong stock market
7.0 Wrap up work done					
7.1	Summarize the model building process	23Jun2023	25Jun2023	20	Highlight the key steps of developing the model
7.2	Wrap up and final evaluation of the model outputs	26Jun2023	30Jun2023	30	Summarize and discuss the performance measures and discuss the findings from the output
7.3	Discuss the future enhancement	1Jul2023	2Jul2023	10	Discuss the future enhancement
7.4	Dissertation write up	3Jul2023	7Jul2023	40	Wrap up the practical work, conclusion and possible future enhancement
7.5	Viva preparation	3Jul2023	7Jul2023	30	Build presentation deck



3. Evaluation Plan (10%)

Guidance: Complete the following table - one page maximum.

	Evaluation Approach	Evidence
Background research by evaluating different investment strategies and studying modern approaches and methods used for stock forecasting	Criticize on different investment strategies and modern approaches for stock forecasting. Then define the data problem and expected deliverables correspondingly	List, describe and discuss at least three investment strategies and modern approaches that can assist for decisioning
Study the Hong Kong stock market landscape by performing background research and technical analysis	Perform exploratory data analysis of long term, mid-term and short-term study in Hong Kong stock market. Describe the market situation in 2022 and set the foundation of prediction model building	Use of summary statistics and various charts for technical analysis that able to articulate the market situation in 2022
Literature review to critically evaluate potential forecasting algorithms and its application	Research relevant literatures that can provide directions on potential forecasting algorithms, such as logistic regression, neural network algorithms and their experiment results on stock predictions	Summarize literature review findings and lay down experimental design
Discuss data collection process by Yahoo Finance	Describe the data collection process with reasons on data period, collection method, pre-processing method and corresponding consideration on each step	Document the process flow with corresponding consideration
Identification and discussion of relevant professional, ethical, social and legal issues	Elaborate the consideration on relevant professional, ethical, social and legal issues during data collection, application adoption process	Document the study and discussion of relevant professional, ethical, social and legal issues
Build, train, deploy and visualize forecasting model(s) for evaluating the one with the highest prediction accuracy and efficiency to assist optimization of swing trading in the Hong Kong stock market	Iteratively develop, evaluate and fine tune a model using the selected algorithm by data collected, ie. Logistic regression and LSTM	Develop Python coding of the model building and evaluation process

4. Social, Ethical, Legal and Professional issues (20%)

4.1 Social, Ethical, Legal and Professional Issues Table

Social issues	This project aims to develop a LSTM forecasting model to assist optimization of swing trading in the Hong Kong stock market. With the machine learning technologies, decision making process will be faster and interfered by the model recommendation. In case the model will be widely adopted by a lot of investors, the investment decisions will be fed in to as the data inflow then form the cycle of looping back to the model and eventually disrupted the stock market. However, this project is an academic study and can be use as recommendation only.
Ethical issues	While the developed machine learning algorithms may affect human decision, it is important to ensure the usage maintained at fair situation in general public terms. General refers to not personal identifiable and unbiased to any individual or groups such as age, gender, religion, races, etc. This project uses data from stock market and are publicly available therefore no ethical issues in this context.
Professional issues	Professional modeling building focused on quality data and adequate methodology to be used. This project adopted data collected from Yahoo Finance which guarantee the data quality. Data cleansing and verification during exploratory analysis will be performed to ensure again the data quality. Data model will be built by iterative evaluation process to ensure the best among results will be selected and adopted.
Legal Issues	The data source of this project comes from Yahoo Finance web site. It allows data download as well as using API like Python yfinance library. There is a disclaimer clause in yfinance that "It is an open-source tool that uses Yahoo's publicly available APIs, and is intended for research and educational purposes". In this case, there is no legal concern as this project is for academic purposes only.

4.2 Ethics Approval

Guidance: If your proposal requires ethical approval attach approval (or submission if not yet approved) from the Research Ethics Committee as an appendix.

5. Appendices

The attachment also saved to Github Repository for easy access
https://github.com/JackieChu2022/ML_Project/tree/Planning-Review

PROM02_project
planning v3.xlsx