University of Sunderland

Faculty of Technology

Department of Computer Science

PROJECT PROPOSAL FORM

PROM02

MSc Dissertation

Academic Year: 2022/23

Module Leader: Neil Eliot

Email: neil.eliot@sunderland.ac.uk

Student Name:	Chu Lai Ting
Student ID:	219265092
Programme:	MSc Data Science
Mode:	PT
Modules Completed	CETM50; CETM24; CETM46; CETM47

Proposed Project Title:	A stock forecast model to assist optimal investment strategy in Hong Kong stock market
Supervisor (if known and agreed):	Marta Rudina / Ming Jiang

Aims (To be revised in the Planning Stage)

The aim of this project is:
Develop LSTM forecast model to assist optimal investment strategy in Hong Kong stock market

Objectives (To be revised in the Planning Stage)

The objectives to achieve this aim are:

- Evaluate different investment strategies
- Study modern approaches and methods used for stock forecasting
- Evaluate/ study the Hong Kong stock market landscape by performing background research and technical analysis
- Literature review to critically evaluate potential forecasting algorithms and its applications
- Discuss data collection process by Yahoo Finance
- Identification and discussion of relevant professional, ethical, social and legal issues
- Build, evaluate, deploy and visualize forecasting model(s) to assist optimal investment strategy in Hong Kong stock market

Skills from course (Include module codes)

The skill I will use from existing modules include:
CETM50: Technology Management for Organizations

- Learnt knowledge discovery in databases (KDD) process
- Developed critical understanding on professional, ethical, social and legal concerns

CETM24: Data Science Fundamentals

- Gained knowledge of supervised learning techniques, eg. Logistic regression, neural networks, etc.
- Hands on R programming for model development

CETM46: Data Science Product Development

- Exposed to various software used for data product development
- Learnt design and develop data science product to solve data problem
- Learnt systematic literature review process, project development methodologies and prototype development

CETM47: Machine Learning and Data Analytics

- Formulated the data solution among machine learning, data mining and analytics area
- Built up ability to choose possible techniques for data problem solving
- Model evaluation and selection by performance measures, eg. RMSE, confusion matrix, etc.

New skills I will need to develop

I will need to develop new skills in ….

- Python ecosystem including but not limited to numpy, pandas, matplotlib, scikit learn, keras, tensorflow
- LSTM concept and application

Practical Element

The practical element will involve…

- Overview of Hong Kong stock market and focus on Hang Seng Index (HSI) and major 3 selected stocks for model development
- Data collection and pre-processing process by Yahoo Finance
- Perform technical analysis by Python such as numpy, pandas, matplotlib, etc.
- Define a baseline model by Python such as scikit learnt
- Build, evaluate, deploy and visualize forecasting model(s) by using Python libs

To evaluate the success of the practical outcome …

- Identify insights from technical analysis
- Derive performance measure that prove the model selection process
- Test by different set of data to ensure model stability

Research Ethics

Not applicable due to the data used are publicly available and not related to personal data.

References

[1] Vijh, M., Chandola, D., Tikkiwal, V.A. and Kumar, A. (2020). *Stock Closing Price Prediction using Machine Learning Techniques. Procedia Computer Science*, 167, pp.599–606. doi:10.1016/j.procs.2020.03.326

[2] Smita, M. ed., (2021). *Logistic Regression Model - A Review*. [online] Available at: https://ijisrt.com/assets/upload/files/IJISRT21MAY1050.pdf

[3] 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

[4] 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

[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

[6] 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