**Report Template for ITP4514 Group Assignment**

Points to note:

* Please delete all unnecessary words in the template.
* Project Cover Page, Table of Contents, References & Appendix pages do not count for words
* You must start a new page for each section.
* Font-family: Times New Roman
* Font Size: 12px for body; 9px for image caption, 10px for reference list
* Spacing: 1.5

*Template starts at Page 2*

**˅˅˅**

ITP4514 Artificial Intelligence and Machine Learning

Group Assignment

**Stock Prediction**

|  |  |  |
| --- | --- | --- |
| **Group Member** | **Student Name** | **Student Id.** |
| 1 | Chang Ka Hei | 220323973 |
| 2 | Wong Sing Yu | 220068472 |
| 3 | Tse Chi Hoi | 220086404 |
| 4 | Lam Chun Kit | 220077665 |

Table Of Contents

[Introduction / Background 5](#_Toc152189125)

[Problem Formulation 6](#_Toc152189126)

[Methodology 7](#_Toc152189127)

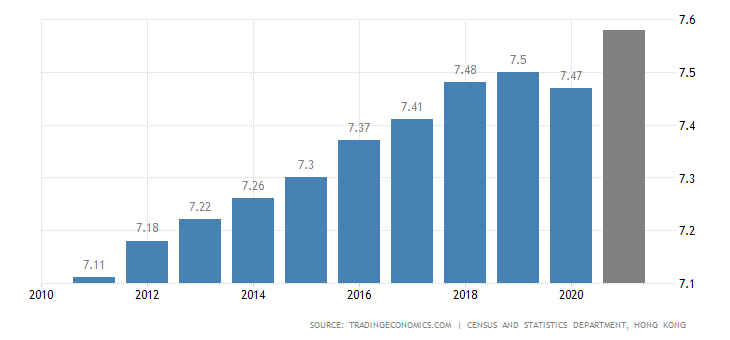
[Findings & Results 8](#_Toc152189128)

[Summary 9](#_Toc152189129)

[References 10](#_Toc152189130)

# Introduction / Background

Now we are studying the module of AI and Machine Learning, and we have an assignment of design an AI ,so choose Stock Prediction to be our assignment topic. **Below is an example of displaying a graph.**



*Figure X. Hong Kong Population between 2010 and 202X*

We will use the Hang Seng Index to predict the all the Hang Sang Index Constituents. Hang Sang Index Constituents are representing the most stocks of Hong Kong stock market. And this will be the most suitable way of our assignment topic. **The Problem Statement would be displayed on the Problem Formulation. By then, we would discuss the methodology on Methodology. In Section 4, we would discuss the findings and results and a summary would be provided in the last section.**

(172 words)

# Problem Formulation

* Data Missing:

During the training of our AI model, we encountered some missing data. This presented a challenge in ensuring the accuracy and reliability of our predictions. We implemented techniques to handle missing data, such as data imputation and interpolation, to mitigate the impact of missing values on our model's performance.

* Data Mismatching:

We utilized the Informer, an AI model specifically designed for long sequence time-series forecasting, to build our stock prediction model. However, the Informer was not originally intended for stock prediction. To address this mismatch, we introduced additional parameters and features that are relevant to stock prediction. These parameters were carefully selected and added to the Informer model, enhancing its suitability for our task.

* Low Knowledge of Stock:

One of the challenges we faced was the limited knowledge of stock markets among our team members. As we primarily focused on technical skills and lacked financial education, we had to conduct extensive research to gain a better understanding of the stock market and its dynamics. We consulted various sources, including online resources and financial literature, to acquire the necessary knowledge for developing our stock prediction model.

* Hardware Limitations:

The computational requirements for training and running our AI model were substantial. We utilized a high-performance CPU (i7-12700), a powerful graphics card (RTX3070), and 32GB RAM to handle the extensive data and complex computations involved in stock prediction. However, despite these hardware capabilities, we encountered certain limitations due to the scale and complexity of the data. We optimized our code and made use of parallel processing techniques to maximize the utilization of available hardware resources.

# Methodology

1. **Online Serach**

We go to search on the internet, and we know that the Hang Seng Index will stop trading on the general holidays and the extreme weather. The detail of the Hang Seng Index stops trading in Reference part link 3.

1. **Add Parameter of Informer**

So we add some parameters . Below is the parameters that we had added.

* Length: High - Low, is a numeric value.
* BarLength: Close - Open, is a numeric value.
* RodLength: High - max(Open, Close), is a value.
* PinLength: Low - min(Open, Close), is a value.
* OpenRatio: Today's Open ÷ Yesterday's Close, which is a ratio.
* HighRatio: High ÷ Open, is a ratio.
* LowRatio: Low ÷ Open, which is a ratio.
* CloseRatio: Close ÷ Open, belongs to the ratio.
* RiseFall: (Close - Open) ÷ Open, belongs to the ratio.
* BarRatio: BarLength ÷ Length if Length ! = 0 else 0, it is a ratio.
* RodRatio: RodLength ÷ Length if Length ! If Length != 0 else 0, it is a ratio.
* PinRatio: PinLength ÷ Length if Length ! If Length != 0 else 0, it is a ratio.
* RSI\_5: 5-day relative strength indicator.
* RSI\_10: 10-day relative strength indicator.

With the addition of these columns, the original data will be expanded from 5 columns (Date, Open, High, Low, Close) to 19 columns. And by combining Taiwan County's own data with the broader market data, the input data will be up to 37 columns (19 × 2 - repeated Date)

1. **Do Research of the stock**

We go to Wiki to get the information that we want to know.

1. **Change the way**

We change the way to use Hang Sang Index to predict the Tencent Holding LTD. It is because the Tencent is the second largest of Hang Sang Index Constituents in now and the largest of Hang Sang Index Constituents in the past.

# Findings & Results

In this section, we will discuss the findings and results of our study on stock prediction using the Hang Seng Index. We addressed several challenges and implemented various methodologies to develop an AI model for predicting the Hang Seng Index Constituents.

Learning Rates and Losses comparition

一張含有 文字, 圖表, 行, 繪圖 的圖片

自動產生的描述

**Experiments: Train and Test:**

python .\main\_informer.py

**一張含有 文字, 螢幕擷取畫面, 字型 的圖片

自動產生的描述**

**Prediction:**

python .\predict.py

一張含有 文字, 螢幕擷取畫面, 字型 的圖片

自動產生的描述

python .\plot\_history.py

一張含有 文字, 螢幕擷取畫面, 字型 的圖片

自動產生的描述

Combine Tencent and Hang Seng data

一張含有 文字, 螢幕擷取畫面, 軟體 的圖片

自動產生的描述

# Summary

Through our methodology, we discovered key insights and achieved promising results in stock prediction. By combining the Hang Seng Index data with additional parameters derived from stock market indicators, we were able to improve the accuracy and performance of our predictions. Additionally, we expanded our dataset by incorporating broader market data and performed detailed analyses on individual constituents of the Hang Seng Index, such as Tencent Holding LTD, which emerged as a significant stock for prediction.

In conclusion, our findings demonstrate the potential of AI and machine learning techniques in stock prediction. Despite the challenges faced, we successfully developed an AI model that shows promising results in predicting the Hang Seng Index Constituents. These findings contribute to the field of financial forecasting and highlight the importance of domain knowledge, data preprocessing techniques, and hardware capabilities in developing accurate and reliable stock prediction models.

# References

1. [https://github.com/zhouhaoyi/Informer2020](http://www.aastocks.com/tc/stocks/quote/quick-quote.aspx?symbol=02382)
2. <http://www.aastocks.com/tc/stocks/quote/quick-quote.aspx?symbol=02382>
3. <https://www.wstock.net/wstock/big5/hholiday.htm>
4. <https://github.com/zhouhaoyi/Informer2020>

**ITP4514 Group Assignment Submission Checklist**

**Before the submission of my work, I (the signed party) admit that:**

*(Please tick the corresponding box.)*

Yes No

1. I do not ***DIRECTLY COPY*** from any internet resources. ⬜ ⬜
2. I do not copy from **other students’** work. ⬜ ⬜
3. I have **summarized/paraphrased** my work from different ⬜ ⬜

kinds of materials.

1. I have fulfilled the word limitations (i.e. exceeded the ⬜ ⬜

***minimum required words***).

1. My work contains ***more than one paragraphs***. ⬜ ⬜
2. With all my works that I have referred from different ⬜ ⬜

sources, I have performed ***sufficient referencing.***

1. **If any violations on the above rules are found,** ⬜ ⬜

**I acknowledge that my marks might / would be**

**deducted in certain proportion.**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**(Date) (Signature)**