

Homework1: Stock Price Prediction

CE6023 Artificial Intelligence (2022 Fall)

ncu-ai-2022fall-ta@googlegroups.com

Description

- **Task**

- 輸入第 1~4 天的股票交易資料，預測第 5 天的收盤價。
- **Evaluation Measure:** RMSE (Root-Mean-Square Rrror)

- **Data**

- 從臺灣證券交易所下載的台積電 (2330.TW) 每日交易資料。
- 分成 Training data 跟 Testing data
 - Training data: 2010/01~2019/12
 - Testing data: 2020/01~2022/09 (已分組並 shuffle 過)
- 請特別注意，從網站上爬出正確資料拿來做參考也將視為作弊！

Training data

- **datetime:** 日期
- **shares:** 成交總股數
- **amount:** 成交總金額
- **open:** 當日開盤價
- **high:** 當日最高價
- **low:** 當日最低價
- **close:** 當日收盤價
- **change:** 當日漲跌價差
- **turnover:** 成交總筆數

datetime	shares	amount	open	high	low	close	change	turnover
2010-01-04	39511138	2557720928	65.0	65.0	64.0	64.9	0.4	8255
2010-01-05	38394084	2464115096	65.0	65.1	63.9	64.5	-0.4	9205
2010-01-06	52734385	3390698544	64.5	64.9	63.7	64.9	0.4	12597
2010-01-07	53294614	3437221996	64.9	65.0	64.2	64.2	-0.7	11195
2010-01-08	48047497	3068341466	63.5	64.3	63.5	64.0	-0.2	9804
2010-01-11	37446686	2413758179	64.0	64.9	63.5	64.5	0.5	8066
2010-01-12	52444792	3337576396	64.4	64.4	63.3	63.6	-0.9	12865
2010-01-13	49806496	3129817596	63.0	63.1	62.6	62.8	-0.8	10609
2010-01-14	39254264	2469193916	63.6	63.6	63.0	63.2	0.4	8755
2010-01-15	50011109	3155482441	62.9	63.5	62.8	63.5	0.3	11048

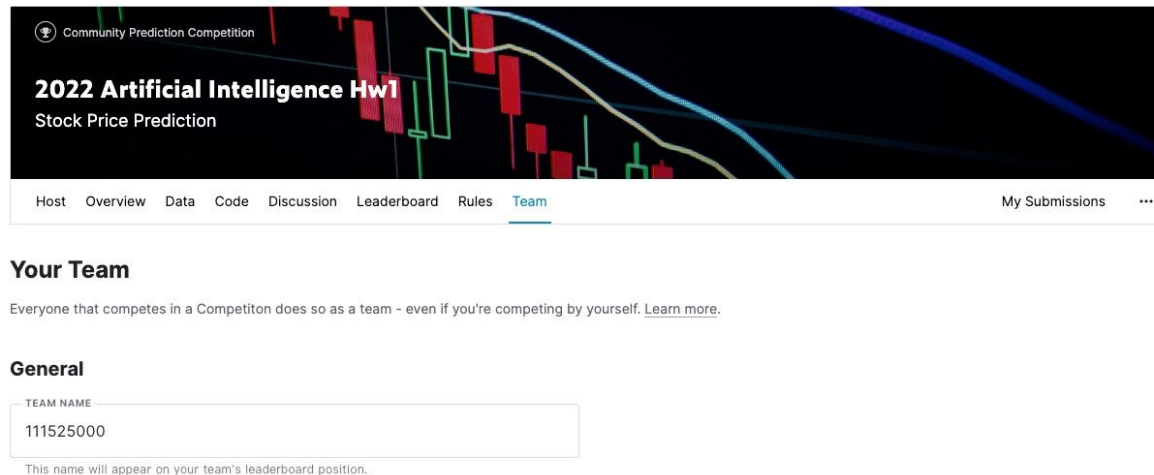
Testing data

- **id**: 測試資料集任務編號
- **shares**: 成交總股數
- **amount**: 成交總金額
- **open**: 當日開盤價
- **high**: 當日最高價
- **low**: 當日最低價
- **close**: 當日收盤價
- **change**: 當日漲跌價差
- **turnover**: 成交總筆數

id	shares	amount	open	high	low	close	change	turnover
0	60853811.0	27461867933.0	450.0	454.0	448.5	453.0	10.0	37378.0
0	67645223.0	31061721030.0	458.0	462.0	457.5	460.0	7.0	36118.0
0	46303738.0	21346346714.0	465.0	465.0	456.5	462.0	2.0	25938.0
0	52677241.0	24157138176.0	460.5	461.0	457.0	459.0	-3.0	22599.0
1	36842639.0	16140184501.0	440.0	441.0	436.0	437.0	-3.0	22692.0
1	47426963.0	20592831368.0	436.0	438.5	432.0	433.5	-3.5	27828.0
1	82510381.0	35076048246.0	425.5	429.0	423.0	423.0	-10.5	54922.0
1	41263269.0	17522224043.0	427.0	428.0	421.0	424.0	1.0	23290.0
2	51296611.0	26259941709.0	517.0	517.0	505.0	516.0	4.0	50929.0
2	30809747.0	15703026732.0	512.0	513.0	508.0	508.0	-8.0	33415.0

Kaggle

- **Link:** <https://www.kaggle.com/t/206ba0a9a44f453e8638955f4109d111>
- **Team Name:** 學號 (e.g. 111525000)
- 你可以選擇 2 個 Submissions 來計入您的 Leaderboard 的分數。
- 你一天可以繳交 20 個 Submissions。



Community Prediction Competition

2022 Artificial Intelligence Hw1

Stock Price Prediction

Host Overview Data Code Discussion Leaderboard Rules **Team** My Submissions ...

Your Team

Everyone that competes in a Competition does so as a team - even if you're competing by yourself. [Learn more.](#)

General

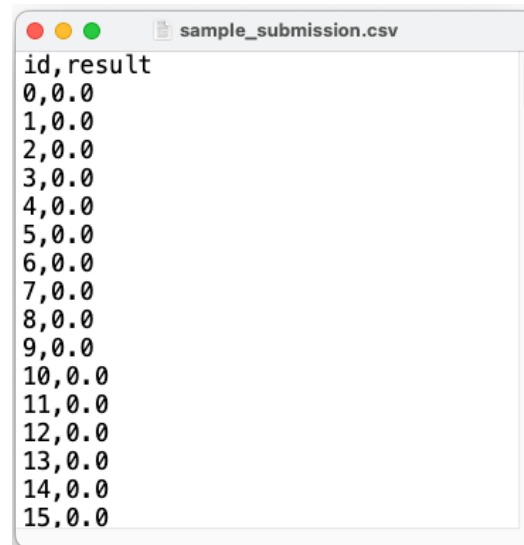
TEAM NAME

111525000

This name will appear on your team's leaderboard position.

Kaggle Submission Format

- 預測 132 筆 Testing data 中的收盤價，並將預測結果上傳至 Kaggle
 - 以 CSV 格式上傳
 - 第一行必須是 id,result
 - 第二行開始，每行分別為 id 值及預測的收盤價，以逗號隔開。



Report

- **Template Link:**

https://docs.google.com/document/d/1atkVdBoeEsqOrorj4fQadtXN_YjrpBGAPRYDs_tffNbo/edit?usp=sharing

- 請標明學號與姓名，並根據題號回答問題。
- 輸出 A4 大小的 PDF 檔，檔名為：report.pdf。

Report (cont.)

1. (5%) 使用四種不同的 **Learning Rate** 進行 **Training** (方法參數需一致) , 作圖並討論其收斂過程 (橫軸為 **Iteration** 次數 , 縱軸為 **Loss** 的大小 , 四種 **Learning Rate** 的收斂線請以不同顏色呈現在一張圖裡做比較) 。
2. (5%) 比較取前 2 天和前 4 天的資料的情況下 , 於 **Validation data** 上預測的結果 , 並說明造成的可能原因 。
3. (5%) 比較只取部分特徵和取所有特徵的情況下 , 於 **Validation data** 上預測的結果 , 並說明造成的可能原因 。
4. (5%) 比較資料在有無 **Normalization** 的情況下 , 於 **Validation data** 上預測的結果 , 並說明造成的可能原因 。
5. (10%) 請說明你超越 **Baseline** 的 **Model** (最後選擇在 **Kaggle** 上提交的) 是如何實作的 (若你有額外實作其他 **Model** , 也請分享是如何實作的) 。

Grading Policy

- **Programming (70%)**

- Top10: Get 70%
- Top11~20: Get 60%
- \geq Baseline: Get 50%
- else: Get 0%

- **Report (30%)**

- 請輸出 A4 大小的 PDF 檔，檔名為 report.pdf (**Report.pdf 是不正確的**)
- 格式錯誤 -5%

Overview

- 請寫一個 Python3 的程式，輸入第 1~4 天的股票交易資料，預測第 5 天的收盤價。
- 將預測的結果上傳到 Kaggle，並試著超過 Baseline。
- 請將檔案請包成一個 zip 壓縮檔，上傳到 ee-class 系統。
 - 在你的 學號.zip 內至少要有下列 3 類檔案
 - report.pdf 你的報告
 - main.py Training 及 Testing 的主程式
 - README.md 說明你的 Python3 版本，以及要如何執行。
- 範例程式碼：

https://colab.research.google.com/drive/1vzstbxjwRrNx2SAr8iB2JIP_v9blxYhn?usp=sharing

- 允許同學以範例程式碼為基礎進行修改。