Assignment 1
Stock Price Forecast

Meng-Chien Lin

Po-Chih Kuo

Goal

- To make stock price forecasting by using the stock price of a company to predict the other in the same industry
- To implement the regression model to achieve the forecasting
- Preprocess the data for model training/testing

Basic (100%)

- Use MTK's stock prices to predict TSMC's stock prices of 20 days followed by the due day (10/14) of the homework.
- That is, to predict the prices from 10/15 to 11/11 (without holidays, totally 20 days)
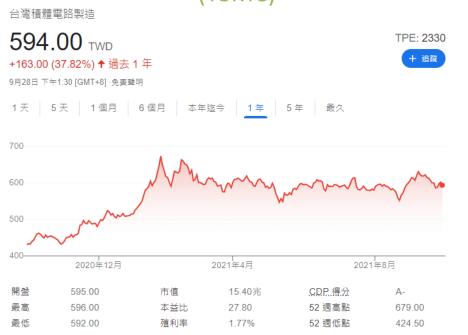
Data

- Download link:
https://www.twse.com.tw/zh/page/trading/exchange/STOCK_DAY_AVG.html

MediaTek Inc.
(MTK)

Taiwan Semiconductor Manufacturing Co., LTd
(TSMC)

How To Download

How To Download

How To Download
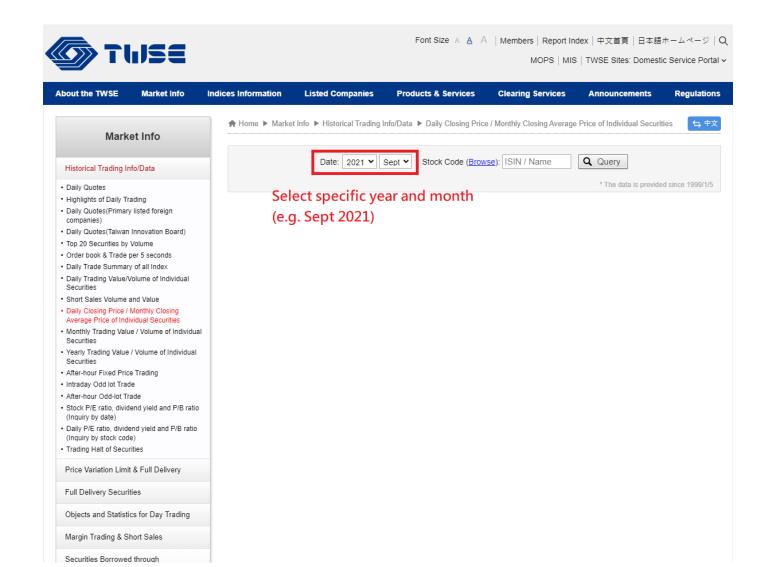
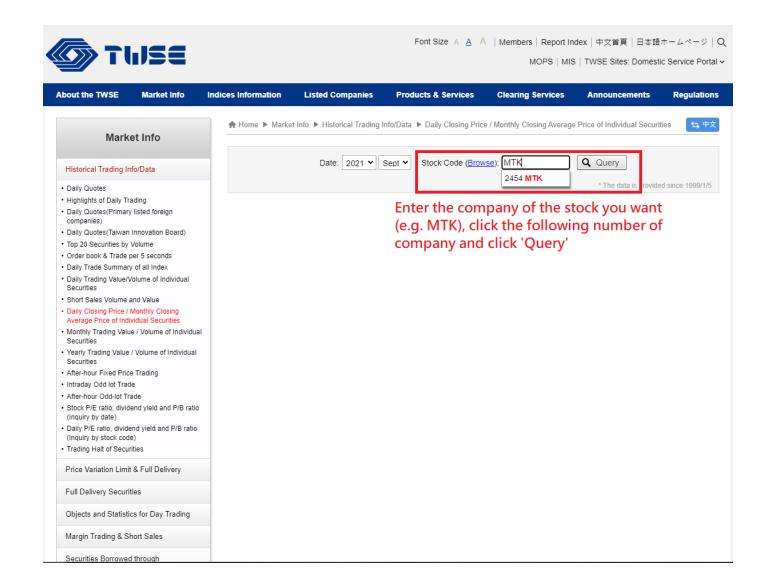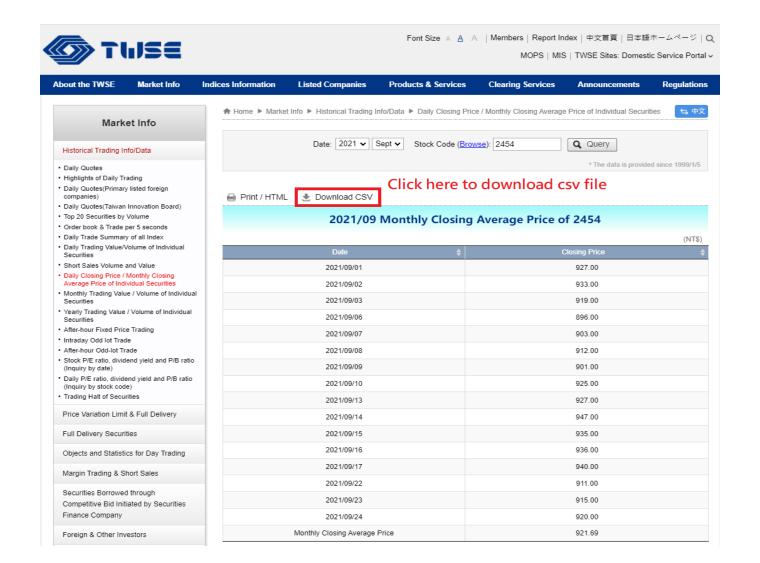# *How To Download*

How To Download

- Then you can get one month's stock data！

- Combine several months of data to create the final input file

Report

- Briefly describe how to implement regression
- Summarize your work
- No more than two pages

Basic Grading Policy

- Model – 50%

- Performance – 40%

- Report – 10%

Bonus (extra 20%)

- Use more than two companies to predict the stock price of TSMC in the next month (10/15~11/11 without holidays, totally 20 days)

Bonus Grading Policy

- Model -- 5%

- Performance – 10%

- Report – 5% (You should write basic and bonus part of the report together in three pages)

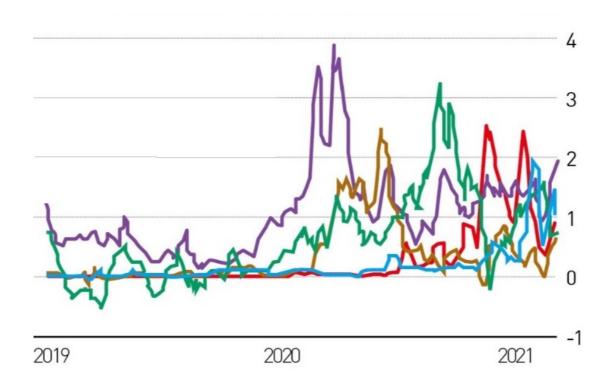# *Template*

- Basic part must follow this format (There is no restriction for the format of bonus part)

- You must use the given file "template.py" to build the model

- Except for the imported packages in the template, you cannot use any other packages

```python
template.py  ✕

1  #!/usr/bin/env python
2  # coding: utf-8
3
4  #import package
5  #Note: You cannot import any other package!
6  import numpy as np
7  import matplotlib.pyplot as plt
8  import csv
9  import math
10 import random
11
12
13
14 # Global attributes
15 StudentID = ''  # TODO 1 : Fill your student ID here
16 input_dataroot = 'input.csv'  # Please name your input csv file as 'input.csv'
17 output_dataroot = StudentID + '_basic_prediction.csv'  # Output file will be named as '[StudentID]_basic_prediction.csv'
18
19 input_datalist =   []  # Initial datalist, saved as numpy array
20 output_datalist =   []  # Your Prediction, should be 20 * 2 matrix and saved as numpy array
21                          # The format of each row should be [Date, TSMC_Price_Prediction]
22                          # e.g. ['2021/10/15', 512]
23
24 # You can add your own global attributes here
25
26
27
28 # Read input csv to datalist
29 with open(input_dataroot, newline='') as csvfile:
30     input_datalist = np.array(list(csv.reader(csvfile)))
31
32 # From TODO 2 to TODO 6, you can declare your own input parameters, local attributes and return parameters
33
34 def SplitData():
35 # TODO 2: Split data, 2021/10/15 ~ 2021/11/11 for testing data, and the other for training data and validation data
36
37
38 def PreprocessData():
39 # TODO 3: Preprocess your data  e.g. split datalist to x_datalist and y_datalist
40
41
```

Input File Format

- Basic part must follow this format (There is no restriction for the format of bonus part)
- Named as "input.csv" and contains a n * 3 matrix, n means number of dates
- Each row represents "Date, MTK Price, TSMC Price"
- The part to be predicted (TSMC Price of 10/15 ~ 11/11) is filled with 0
- You can refer to the given file "sample_input.csv"
- We will use this format of csv file to test your model with n = 209 (1/4 ~ 11/11)
- Please make sure your model can be correctly input into this format of csv file

Date	MTK	TSMC
2021/1/4	861	624
2021/1/5	591	840
2021/1/6	677	748
2021/1/7	542	866
2021/1/8	512	766
2021/1/11	915	852
2021/1/12	589	997
2021/1/13	552	508
2021/1/14	947	743
2021/1/15	707	784
2021/1/18	680	539
2021/1/19	679	700
2021/1/20	669	899
2021/1/21	511	753
2021/1/22	747	840
2021/1/25	820	746
2021/1/26	823	639
2021/1/27	776	795
2021/1/28	589	705
2021/1/29	724	781
2021/2/1	927	781
2021/2/2	967	780
2021/2/3	735	545
2021/2/4	941	505
2021/2/5	838	592
2021/2/17	554	697
2021/2/18	638	512
2021/2/19	865	527
2021/2/22	705	954
2021/2/23	870	682
2021/2/24	614	959
2021/2/25	696	847
2021/2/26	670	916
2021/3/2	922	543

Date	MTK	TSMC
2021/9/27	504	863
2021/9/28	704	713
2021/9/29	833	995
2021/9/30	929	604
2021/10/1	952	793
2021/10/4	883	809
2021/10/5	695	723
2021/10/6	703	783
2021/10/7	691	657
2021/10/8	858	652
2021/10/12	928	876
2021/10/13	536	843
2021/10/14	922	836
2021/10/15	519	0
2021/10/18	565	0
2021/10/19	625	0
2021/10/20	975	0
2021/10/21	804	0
2021/10/22	579	0
2021/10/25	822	0
2021/10/26	734	0
2021/10/27	729	0
2021/10/28	730	0
2021/10/29	906	0
2021/11/1	926	0
2021/11/2	957	0
2021/11/3	880	0
2021/11/4	712	0
2021/11/5	684	0
2021/11/8	938	0
2021/11/9	789	0
2021/11/10	929	0
2021/11/11	766	0

Output File Format

- Both basic and bonus parts must follow this format
- The prediction you turned in must also follow this format
- Named as "[StudentID]_basic(bonus)_prediction.csv" and contains a 20 * 2 matrix
- Each row represents "Date, TSMC Price(Prediction)"
- You can refer to the given file "sample_output.csv"
- Please make sure your model can correctly output this format of csv file

2021/10/15	564
2021/10/18	423
2021/10/19	526
2021/10/20	628
2021/10/21	730
2021/10/22	799
2021/10/23	710
2021/10/26	654
2021/10/27	622
2021/10/28	543
2021/10/29	510
2021/11/1	499
2021/11/2	526
2021/11/3	589
2021/11/4	614
2021/11/5	688
2021/11/8	624
2021/11/9	635
2021/11/10	611
2021/11/11	584

HMI

Assignment 1 Requirement

- Do it individually! Not as a team! (team is for final project)
- Announce date: 2021/9/30
- Deadline: 2021/10/14 23:59 (Late submission is not allowed!)
- Hand in your files in the following format
  - [StudentID]_basic.py
  - [StudentID]_basic_prediction.csv
  - [StudentID]_bonus.py (optional)
  - [StudentID]_bonus_prediction.csv (optional)
  - [StudentID]_report.pdf
  - Compress all files into [StudentID]_HW1.zip

The Evaluation Metric

- MAPE (Mean absolute percentage error):

$$\text{MAPE} = \frac{100\%}{N} \sum_{i=1}^{N} \left| \frac{y_i - \hat{y}_i}{y_i} \right|$$

- For example:
- The value you predicted:
- $\hat{y}$ = [592, 486, 538, 689, 752, 841, 491]
- Ground Truth :
- y = [491, 584, 541, 599, 615, 741, 512]

- MAPE = 1/7 *0.928 = 0.1326 = 13.26%.

Penalty

- 0 points if any of the following conditions

- Plagiarism

- Late submission

- Not using template or import any other packages in basic part

- Incorrect input/output format

- Incorrect submission format

Questions?

- TA: Meng-Chien Lin (charlie2039667@gapp.nthu.edu.tw)