Python資料分析與機器學習應用
Data Analysis and Machine Learning with Python

I. Course Information
Instructor: Cheng-Yuan Ho (何承遠、Tommy Ho)
Office: 2C, Teaching and Researching Hall
Email: tommyho@ntu.edu.tw
Office hours: Please send an email to make an appointment first, and indicate your class. DO NOT send NTU Cool's internal letter.
Curriculum Number: IM1013
Curriculum Identity Number: 705 14200
Class: 01
Required/Elective: Elective
Time: Monday 2, 3, 4 (9:10~12:10)
Classroom: B01, Building 1, College of Management
Authorization Code: None

II. TAs' Information
Name and email:

Class 01: 顏子鈞 r11942161@ntu.edu.tw
　　　　　李振勳 r11942118@ntu.edu.tw
Class 02: 顏勁賢 r11942145@ntu.edu.tw
　　　　　許宸睿 r11942085@ntu.edu.tw
Class 03: 黃冠綸 r12725046@ntu.edu.tw
　　　　　康崴　 r11725031@ntu.edu.tw
Class 04: 王心瑋 r11725033@ntu.edu.tw
　　　　　莊啟宏 r11725019@ntu.edu.tw

Office hours: Thursday Night A, B, C (18:25~21:00)
Classroom: B01, Building 1, College of Management

Note: if you send emails to TA(s), please also cc me.

TA Schedule

Week	Date (mm/dd)	TAs	
1	2/22	---	---

2	2/29	王心瑋	莊啟宏
3	3/7	王心瑋	莊啟宏
4	3/14	王心瑋	莊啟宏
5	3/21	顏子鈞	李振勳
6	3/28	黃冠綸	康崴
7 Ching Ming	4/4	---	---
8 Midterm	4/11	---	---
9	4/18	黃冠綸	李振勳
10	4/25	黃冠綸	康崴
11	5/2	許宸睿	康崴
12	5/9	顏子鈞	顏勁賢
13	5/16	顏子鈞	李振勳
14	5/23	許宸睿	顏勁賢
15	5/30	許宸睿	顏勁賢
16 Final	6/6	---	---

III. Teaching Methods

Classes in person and discussion

IV. Course Syllabus

Course Description:

This course will be your guide to learning how to use the power of Python to analyze data and use powerful machine learning algorithms, and then create beautiful visualizations for the analysis results and predictions.

This course is designed for beginners with some programming experience, the guys who already know some Python and are ready to dive deeper into using those Python skills for data analysis and Machine Learning, or experienced developers looking to make the jump to Data Science.

I want to help guide students to learning a set of skills to make them extremely hirable in today's workplace environment. Therefore, I'll teach you how to program with Python, how to create amazing data

visualizations, and how to use Machine Learning with Python! Enroll in the course and become a data scientist!

Course Objective:
      1. Use Python for Data Analysis and Machine Learning
      2. Understand how to Use Tools to Analyze Data
      3. Understand how to Use Existing Machine Learning Modules/Packages
      4. Learn Related Modules and Tools in Python, like NumPy, Pandas, Matplotlib, and SciKit-Learn

Course Prerequisites:
      Take Python programming or related course before, or understand Python

If you DO NOT learn Python before and REALLY REALLY want to take this course, PLEASE watch the following videos first.
    1. **Material in Chinese:**
       https://www.youtube.com/playlist?list=PLCGdg-V2_I4hoUsS1AHbQvjjlDN796HGK
    2. **Material in English:**
       https://www.youtube.com/playlist?list=PLCGdg-V2_I4jrLMFPBeBt8dnZ8jMSp_26
    3. **Online Practice Material in English:**
       https://www.youtube.com/playlist?list=PLCGdg-V2_I4j02BK5NqtfPvHvWxa0TS2N

Designated reading: No
References:
      1. Introduction to Machine Learning with Python: A Guide for Data Scientists
      2. Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython
      3. Machine Learning with Python Cookbook: Practical Solutions from Preprocessing to Deep Learning
      4. Python Cookbook
      5. Hands-On Machine Learning with Scikit-Learn and TensorFlow: Concepts,Tools, and Techniques to Build Intelligent Systems

Progress:

Week	Date	Topic (The actual teaching content will be adjusted according to the class status)
Week 1	2/19	Course Overview, What is Data Analysis, and What is Machine Learning
Week 2	2/26	Steps and Flows in Data Analysis, Introduction to Pandas, Functions and Module in Pandas for Data Analysis I
Week 3	3/4	Functions and Module in Pandas for Data Analysis II
Week 4	3/11	Machine Learning: Supervised vs Unsupervised vs Reinforcement Learning Homework 1 assigned
Week 5	3/18	Algorithms and Modules in Supervised Machine Learning I
Week 6	3/25	Algorithms and Modules in Supervised Machine Learning II
Week 7	4/1	Homework 1 DUE, Presentation, and Other Implementations and Discussions
Week 8	4/8	Midterm Test
Week 9	4/15	Algorithms and Modules in Unsupervised Machine Learning Grouping
Week 10	4/22	Data Modeling, Tuning, and Explanation (by Manual)

Week 11	4/29	Data Modeling, Tuning, and Explanation (by Automatic and Functions) **Grouping DUE**
Week 12	5/6	Some Methods MAY Improve Model Performance **Homework 2 assigned**
Week 13	5/13	Case Study: 1. Energy Saving and Product Defect Detection 2. Prognostics and Health Management (PHM) 3. When Machine Learning Meets Groundwater
Week 14	5/20	
Week 15	5/27	**Homework 2 DUE, Presentation, and Other Implementations and Discussions**
Week 16	6/3	**Final Project Presentation**

Homework & Project:

Item	Description
Homework 1	**Practice how to analyze and predict data with Python and different Machine Learning modules (and you can use AI tools)**
Homework 2	**Practice how to predict data with Python Machine Learning modules and Find the best parameters, and know how to explain the results (and you can use AI tools)**
Presentation in 7th and 15th Weeks	**Show your code (what to do), result, and analysis and prediction reasons (why) Or different views with classmates Note: Everyone MUST appear on stage once at least**

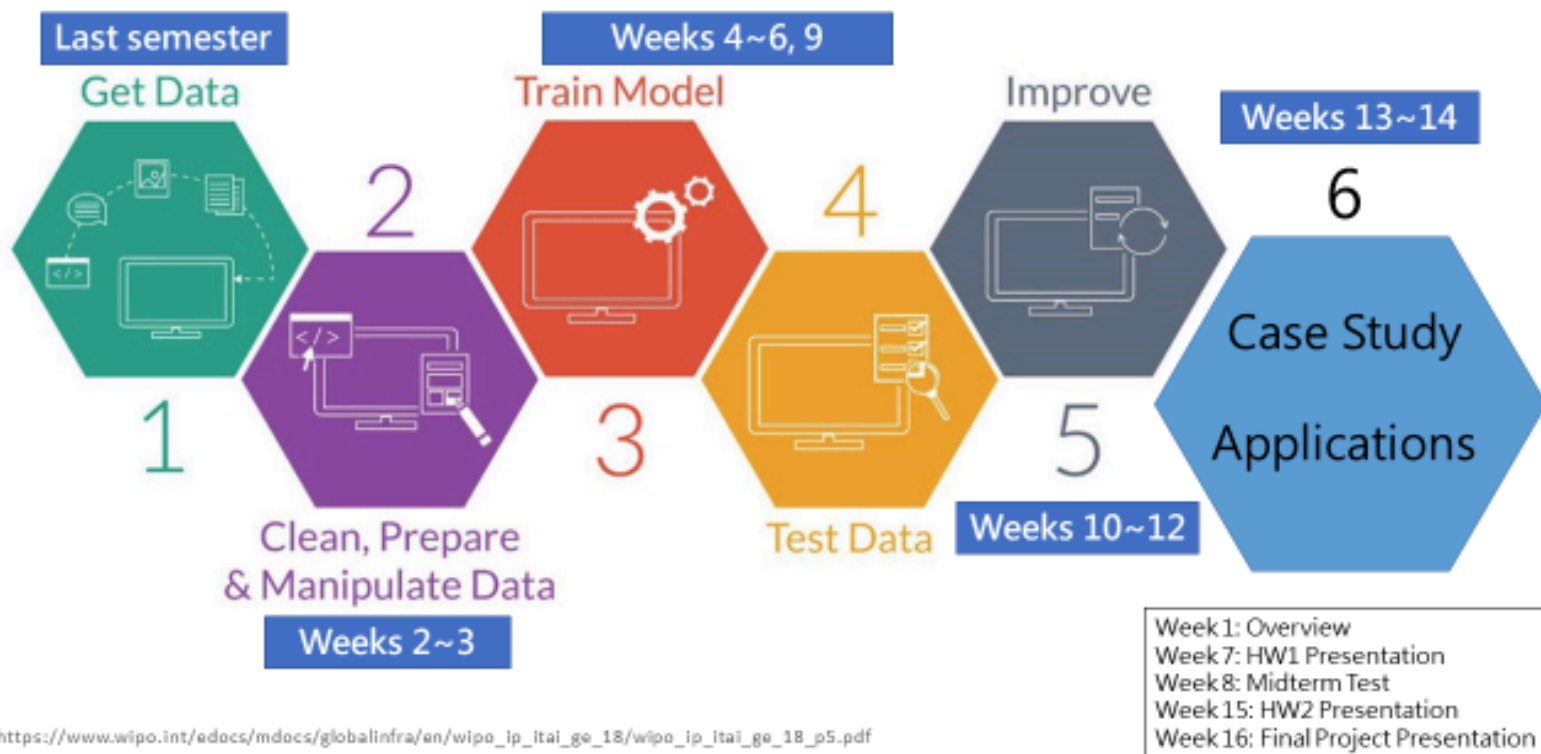| Midterm Test	Midterm Test
Final Project	**Final Project,** **including but not limited to the following items:** 1. **Target: what problem in which field** 2. **Dataset: where it comes from, or collected by web crawler** 3. **Packages and tools** 4. **Machine Learning modules** 5. **Analysis and prediction results** 6. **Team work assignments** **Note:** 1. **DO NOT** **copy problem and solution from the Internet, but can use the existing dataset** 2. **Upload the Python source code and report to NTU Cool** 3. **Demo in the class**

Score Rules:

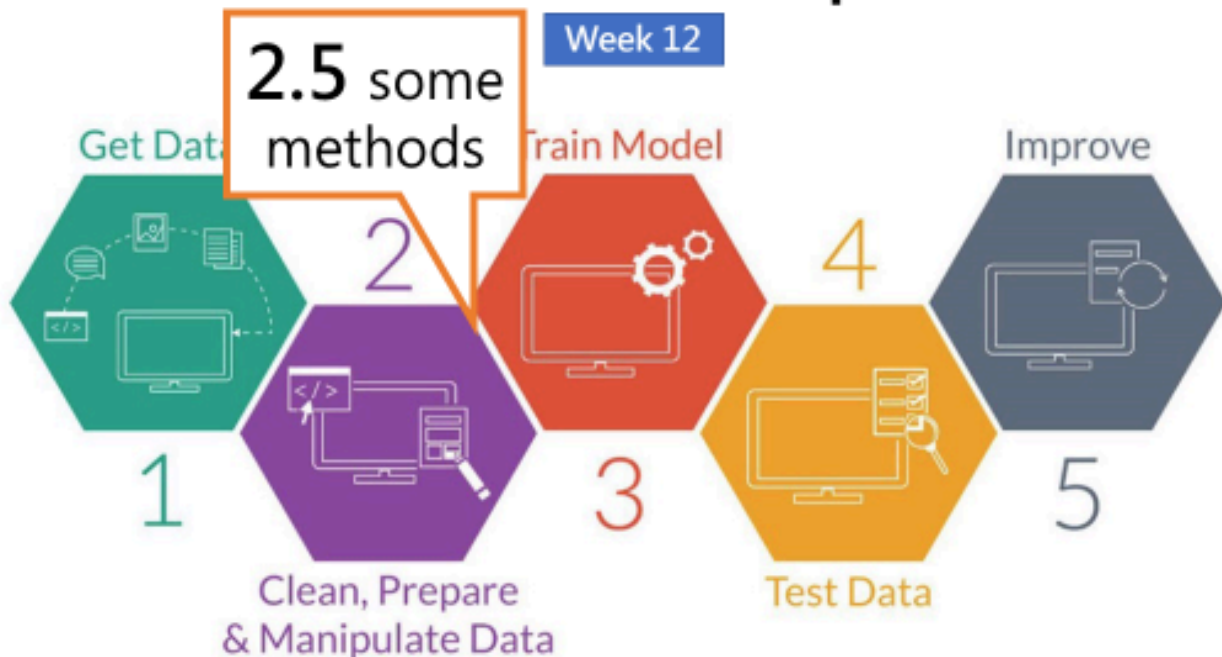| 30%	**Homework 1, 2, & presentation (12%, 12%, 6%)** **(There won't be any late submission issues, as the NTU Cool system will automatically close the submission window)**
30%	**Midterm test**
40%	**Final project (including intra- and inter-group evaluations)**
Extra Points	**1. Interaction in class (earn it every week)** **2. Participate in program language-related competitions, set the mentor to me, and provide supporting materials, you will earn extra 2 points to the course grade. The maximum times is up to 3**

Others:

1. During class, if there is something you don't understand, please ask questions immediately or after class or TAs' hours, TAs and I will try our best to let you understand it.

2. If I ask you or your group to think, talk, discuss, share and express, please try to participate in the activity.

3. About online, e-learning, live streaming, and recorded video:
   There are no online classes, e-learning, and live streaming.

Course Module Arrangement

Last semester
Get Data
1

2
Clean, Prepare
& Manipulate Data
Weeks 2~3

Weeks 4~6, 9
Train Model
3

4
Test Data
Weeks 10~12

Improve
5

Weeks 13~14
6
Case Study
Applications

Week 1: Overview
Week 7: HW1 Presentation
Week 8: Midterm Test
Week 15: HW2 Presentation
Week 16: Final Project Presentation

Some Methods MAY Improve Model

2.5 some methods

Week 12

Get Data
1

2
Clean, Prepare
& Manipulate Data

Train Model
3

4
Test Data

Improve
5

What is Data Analysis?   (source: wikipedia)

Data analysis is a process of inspecting, cleansing, transforming, and modeling data with the goal of discovering useful information, informing conclusions, and supporting decision-making.

Data analysis has multiple facets and approaches, encompassing diverse techniques under a variety of names, and is used in different business, science, and social science domains.

There are several phases that can be distinguished, described below.

- Data requirements
- Data collection
- Data processing
- Data cleaning
- Exploratory data analysis
- Modeling and algorithms
- Data and information visualization

Relationship of Data, Information and Intelligence
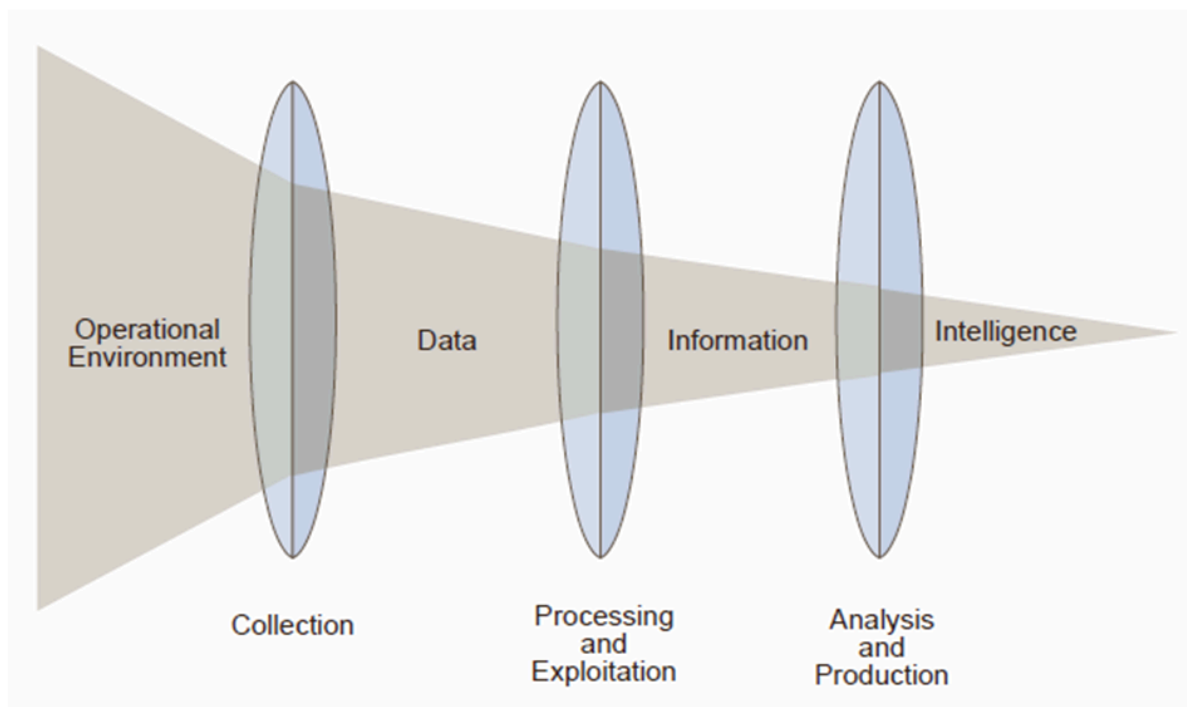
Source: Joint Intelligence / Joint Publication 2-0 (Joint Chiefs of Staff)

Practice DATA Analysis and Prediction:

1. Try to analyze and predict the rates for USD and JPY after one hour, and tell me your reasons.
   foreign-exchange-rates:
   Chinese version
   https://www.esunbank.com/zh-tw/personal/deposit/rate/forex/foreign-exchange-rates
   English version
   https://www.esunbank.com/en/personal/deposit/rate/forex/foreign-exchange-rates

2. Which covid-19 dashboard do you like? Why?
   a. WHO
   https://data.who.int/dashboards/covid19/cases?n=c
   b. Taiwan https://covid-19.nchc.org.tw/index.php
   c. Tableau
   https://www.tableau.com/covid-19-coronavirus-data-resources

Practice Problem: What is Machine Learning?

Please try to follow the steps (questions) to answer the problem, "what is machine learning?"

1. What data do you need in order to present "Machine Learning?"
2. Where do you collect data?
3. Do you need to clean or process collected data?
4. Is there any conflict among the collected data? If yes, what will you do?
5. Can you answer the problem? Please show it with data visualization if possible.

Upload your document to the following Google Drive.
Please set File name as Student ID-Name, like B12345678-Tommy
https://drive.google.com/drive/folders/1NRQhZfy9-h98VHNJeh0mUndfC4qx0w92?usp=sharing

Difference between TP and ML

Source:
https://d1m75rqqgidzqn.cloudfront.net/2019/10/What-is-Machine-Learning-Machine-learning-model-vs-traditional-model.jpg