2309 CS482/SE482 Course Outline

Subject Code	:	CS482/SE482
Subject Title	:	DATA SCIENCE
Course Type	:	Elective
Level	:	3
Credits	:	3
Teaching Activity	:	Lecture 36 hours Presentation 9 hours
Prerequisites	:	CS110 Computer Programming
Class Schedule	:	

Class	Week	Time	Classroom	Date
D2	Thu	15:30-18:20	C408	04/09/2023 - 17/12/2023

Instructor	:	Tian Jin Yu
Contact Number	:	(853)6580 0305
E-mail Address	:	jytian@must.edu.mo
Office	:	A306b
Office Hour	:	Monday (12:00-14:00) Tuesday (17:30-19:30) Wednesday (13:30-17:30) Thursday (13:30-15:30)

I. COURSE DESCRIPTION

The objective of this course is to introduce the basic syntax and control structure of Python programming language, as well as the commonly used modules in Python such as Numpy, Pandas, Matplotlib, Scipy, Sqlite, Scikit-learn etc. Finally, it will also cover common operations in data analysis, such as crawling network data, regular expressions, storing and accessing data, Monte Carlo simulation, cluster analysis, principal component analysis, and prediction.

II. PROGRAM LEVEL OUTCOME

Upon successful completion of the program students will have:

(a) Understand the basic concepts of Data Science and the basic skills needed to be a Data Scientist.

(b) Carry out Exploratory Data Analysis using Python and modules, such as Numpy, Pandas, Matplotlib, Scipy, etc

(c) Familiar with HMTL and Use Python to crawl web data.

(d) Apply basic Machine Learning Techniques (Linear Regression, K-NN, SVM, Naïve Bayes, CNN, etc) for predictive modeling.

III. INTENDED LEARNING OUTCOMES

Upon successful completion of this subject, students will be able to:

(a) Understand the foundational concepts of data science, its role in various industries, and its importance in decision-making.

(b) Learn techniques for collecting, storing, and managing data from various sources, including databases, APIs, and web scraping.

(c) Learn statistical techniques such as hypothesis testing, confidence intervals, and ANOVA to make informed decisions based on data analysis.

(d) Understand the concepts of supervised and unsupervised learning, overfitting, underfitting, bias-variance trade-off, and basic algorithms like linear regression and k-means clustering.

(e) Learn Python libraries that provide tools and functions for data manipulation, analysis, visualization, and modelling.

IV. WEEKLY SCHEDULE

Week	Topic	Hours	Teaching Method
1	Introduction, Python Installation	3	Lecture
2	Python Grammer and OOP in Python	3	Lecture / Lab
3	Python Library	3	Lecture / Lab
4	Visualization using Matplotlib, Seaborn	3	Lecture / Lab
5	Regular Expression, HMTL, Web Crawler	3	Lecture / Lab
6	Statistical Methods for Data Science	3	Lecture / Lab
7	Supervised Machine Learning: Traditional Method	3	Lecture / Lab
8	Supervised Machine Learning: Deep Learning Method	3	Lecture / Lab
9	Unsupervised Machine Learning: Traditional Method	3	Lecture / Lab
10	Unsupervised Machine Learning: Deep Learning Method	3	Lecture / Lab
11	Artificial Intelligence Generated Content	3	Lecture / Lab
12	Multimodal Data and Model.	3	Lecture / Lab
13	Security and Ethic for Data Science	3	Lecture / Lab
14-15	Final Project	3	presentation

V. TEACHING AND LEARNING APPROACH

The teaching and learning approach of this subject is to help students in understanding the basic concepts and techniques in data science. Application-oriented thinking is strongly suggested to promote study interest, help student develop right program for real problem-solving.

The teaching activities are conducted by means of explanation and codes demo through problem solving and plenty of examples of real applications as well. Students are encouraged to respond to lecturer's questions through in-class discussion.

VI. ASSESSMENT APPROACH

No.	Approach	Weighting	COs to be addressed	Description of Assessment Tasks
1	Class Participation	10%		
2	Midterm	30%	2-4	Midterm is used to measure how far students have achieved their course outcomes. Questions will primarily be analysis and calculations based to assess the student's ability in the design and analysis of Data Science.
3	Final Project	60%	1-5	Final exam is used to measure how far students have achieved their course outcomes. Questions will primarily be analysis and calculations based to assess the student's ability in the design and analysis of Data Science.

Notes:

Students will be assessed on several assessment items (i.e. Class participation, Assignment, Case Study and Final Project.).

- The class participation evaluates the student's participation of discussion in class.
- The case study evaluate the student's performance in speech.
- The final project evaluates the student's comprehensive ability in writing report and understanding data science.

VII. GUIDELINE FOR LETTER GRADE

Marks	Grade
93-100	A+
88-92	A
83-87	A-
78-82	B+
72-77	B
68-71	B-
63-67	C+
58-62	C
53-57	C-
50-52	D
0-49	F

VIII. TEXTBOOK AND REFERENCES

No recommended textbook, but the learning materials will be provided to students during the classes.

References:

Book name：Data Analysis with Open Source Tools
Author/Editor：Philipp K. Janert
Edition：1
ISBN： 978-0-596-80235-6.
Publisher：O'REILLY
Date：2011

Book name：Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython
Author/Editor：Wes McKinney
Edition：2
ISBN： 978-1491957660.
Publisher：O'REILLY
Date：2017

Revised on 2023/09/01