Course Syllabus

1.	Course Number:	2110531
2.	Course Credit:	3 (3-0-6)
3.	Course Title:	Data Science and Data Engineering Tools
4.	Department	Computer Engineering
5.	Semester	First Semester
6.	Academic Year	2024
7.	Instructor	Peerapon Vateekul, Ph.D. Natawut Nupairoj, Ph.D. Veera Muangsin, Ph.D.
8.	Condition	-
9.	Status	Elective
10.	Curriculum	Computer Engineering
11.	Degree	M.Eng., M.Sc.
12.	Hours/Week	3-Hour Lecture & Lab

13. Course Description

Data Science is the study of the discovery of knowledge from data. Being a data scientist requires an integrated skill set spanning mathematics, statistics, machine learning, databases, and other branches of computer science along with a good understanding of the craft of problem. Data Engineering is the study of how to engineer or process data, i.e., data cleansing, data storing, etc. There are three main parts in this course:

- Data engineering: Data exploration & preparation
- Data analysis: Machine Learning techniques
- Data visualization: Storytelling via data

14. Course Outline

14.1. Learning Objectives

- Describe what Data Science and Data Engineering are and the skill sets needed
- To be able to explore and understand collected data
- To be able to analyze data by apply traditional machine learning techniques
- To be able to visualize data in relation to spatial and temporal points of views

14.2. Learning Contents

- Section1 on Tue 1PM-4PM, Building ENG3 Room 417
- Section5 on Sat 1PM-4PM, Building ENG4 Room 18-16
- *** Please bring your laptop to the class ***
- Students need to attend the class on-site at least 80% (at least 12 weeks) as a mandatory criterion to "pass" this course.

#	Tue 1PM-4PM	Sat 1PM-4PM	Contents	Instructor	Module
1	06-Aug-24	10-Aug-24	Introduction, Pandas, Data Prep	Aj.Peerapon	DS1
2	13-Aug-24	17-Aug-24 (no class)	Traditional ML (1) Make-up class for 17-Aug: TBA	Aj.Peerapon	DS2
3	20-Aug-24	24-Aug-24 Ext.5:30PM	Traditional ML (2)	Aj.Peerapon	DS3
4	27-Aug-24	31-Aug-24 Ext.5:30PM	Deep Learning (1); CNN, RNN (LSTM, GRU)	Aj.Peerapon	DS4
5	03-Sep-24	07-Sep-24	Deep Learning (2); Transformer	Aj.Peerapon	DS5
	10-Sep-24	14-Sep-24	No class		
6	17-Sep-24	21-Sep-24	Advanced topics (Generative AI) + Model monitoring (MLflow)	Aj.Peerapon	DS6
7	*28-Sep-24*	28-Sep-24	Midterm Exam Week (23 - 27 Sep 2024) Guest speakers	Aj.Peerapon	DS7
8	01-Oct-24	05-Oct-24	Big data architecture + data storage (Graduation Week: October 2 – 4, 2024)	Aj.Natawut	DE1
9	08-Oct-24	12-Oct-24	Web scraping	Aj.Natawut	DE2
10	15-Oct-24	19-Oct-24	Data ingestion	Aj.Natawut	DE3
11	22-Oct-24	26-Oct-24	Big data processing (Spark)	Aj.Natawut	DE4
12	29-Oct-24	02-Nov-24	MLOps: Orchestration (Airflow) and serving (FastAPI, Seldon Core)	Aj.Natawut	DE5
13	05-Nov-24	09-Nov-24	Data visualization	Aj.Veera	VIZ1
14	12-Nov-24	16-Nov-24	Python visualization	Aj.Veera	VIZ2
15	19-Nov-24	23-Nov-24	Graph analysis & spatial analysis	Aj.Veera	VIZ3
	26-Nov-24	30-Nov-24	Final Exam Week (25 Nov - 9 Dec 2024) *** Final Exam on Sat 30 Nov 2024		

* There will be up to two guest speakers in the class.

14.3. Method: Lecture and Lab
14.4. Learning Media: PowerPoint presentation, Zoom
14.5. Evaluation

- M odule1 Assignment (data analytics) 15%
- Module2 Assignment (data engineering) 15%
- Module3 Assignment (data visualization) 10%
- Midterm Exam (Kaggle) 15%
- Project 10%
- Attendance 5%
- Final Exam 30% (Lab Test)

15. Reading List
    15.1. Required Text: N/A
    15.2. Electronic Media or Websites:

16. LMS
    16.1. CourseVille: Sec1: “SeaLLM”, Sec5: “Gemini”
    16.2. Discord: https://discord.gg/DhM2cpXG79
    16.3. GitHub: https://github.com/pvateekul/2110531_DSDE_2024s1