Faculty of Engineering & Technology
Department of Computer Science
COMP4381 - Data Science and Analytics
Course outline – Summer 2023/2024

Course Description:

Provide students with a comprehensive introduction to the principles, techniques, and applications of data science. Students will explore the entire data science workflow, from data collection, cleaning, analysis, visualization, and modeling. The course emphasizes hands-on experience using and "Python" programming language with "Notebook" tools, enabling students to develop practical skills in data manipulation, statistical analysis, and machine learning.

Course Objectives:

- **Understand the foundational concepts of data science:** Introduce students to the fundamental principles, terminology, and methodologies used in data science.
- **Master data collection and preprocessing techniques:** Teach students how to acquire, clean, and preprocess various types of data to make it suitable for analysis.
- **Develop proficiency in data analysis and visualization:** Enable students to analyze datasets using statistical techniques and visualize findings effectively to gain insights.
- **Gain practical experience in Python programming:** Familiarize students with the Python programming language and its libraries for data manipulation, analysis, and visualization.
- **Learn foundations of machine learning concepts and techniques:** Provide students with a fundamental understanding of machine learning algorithms and their applications in solving real-world problems.

Prerequisites: Object Oriented Programming

Instructor:

- Hussein Soboh
- hsoboh@birzeit.edu

Office hours:

Please check your instructor's office hours on Ritaj. If you want to meet your instructor outside their public office hours please request an appointment through email/Ritaj.

Course Material:

Text books:

- **Python for Data Analysis 3rd edition** - Wes McKinney – O'RIELLY: chapters (2-10)
- **Python data science handbook 2nd edition** - Jake VanderPlas – O'RIELLY: chapters (37-40)
- **Statistics unplugged 4th edition** – Sally Cardwell - Wadsworth: chapters (1,2)

Schedule:

# of lectures	Content	Reference	Sections
1	Outline and introduction to data science		
1	Data types, formats, process	Slides	
3	Python for data science	Python for Data Analysis	2, 3
2	Numpy arrays	Python for Data Analysis	4
2	Statistics for data Science	Statistics unplugged	1, 2
2	Pandas Dataframes	Python for Data Analysis	5, 6
3	Data wrangling and aggregating	Python for Data Analysis	8, 10
Midterm Exam			
2	Data cleaning and preparation	Python for Data Analysis	7
2	Data visualization and plotting	Python for Data Analysis	9
1	Introduction to machine learning	Python data science handbook	37
2	Training regression models	Python data science handbook	38
2	Training classification models	Python data science handbook	38
2	Model validation and tuning	Python data science handbook	39
2	Feature engineering	Python data science handbook	40
1	Big Data overview	Slides	
2	Project presentations		
Final Exam			

Evaluation:

- Assignments and quizzes 20%
- Project 15%
- Midterm exam 30%
- Final exam 35%

Student responsibilities:

- **Class participation and independent work.** Students are expected to actively participate in all classes and perform independent work.
- **Attendance.** Attendance is mandatory. University regulations regarding this matter will be strictly enforced.

- **Academic Honesty.** Individual work must be each student's own work. Plagiarism or cheating will result in official University disciplinary review.
- **Missed Exams.** There are no makeup exams.
- **Class Etiquette.** Please keep all cell phones and other electronic devices turned off during class. If your activities during class are deemed disruptive, you will be asked to leave.