Data Science-CS4048
Fall 2024

Course Outline

Instructor: Arooj Khalil
Email: saif.islam@lhr.nu.edu.pk
Office location: 47
Office Timings: Tues/Thu: 11 am- 12.45pm------4pm to 5pm

Course Information

Program: BS (CS)
Credit hours: 3
Type: Elective
Course website: Google Classroom
Class Venue: NB-310/311
Pre-requisites:

Program Learning Outcomes (PLOs)

This course covers the following PLOs:

PLO#	PLO Name	PLO description
PLO 3	Problem Analysis	Identify, formulate, research literature, and solve complex computing problems reaching substantiated conclusions using fundamental principles of mathematics, computing sciences, and relevant domain disciplines
PLO 4	Design/ Development of Solutions	Design and evaluate solutions for complex computing problems, and design and evaluate systems, components, or processes that meet specified needs with appropriate consideration for public health and safety, cultural, societal, and environmental considerations
PLO 5	Modern Tool Usage	Create, select, adapt and apply appropriate techniques, resources, and modern computing tools to complex computing activities, with an understanding of the limitations

Course Learning Outcomes

By the end of the course, students should be able to achieve the following CLOs:

CLO#	CLO description	BT Domain/ BT Level	PLO #
CLO 1	Extract, clean, and transform data for analysis	**C2 (Understanding)**	**PLO 3**
CLO 2	Apply tools for performing exploratory data analysis and visualization.	**C3 (Applying)**	**PLO 5**

CLO 3	Build and evaluate machine learning models.	**C3 (Applying)**	**PLO 4**

Text Book

1. Doing Data Science by Oreilly

Reference book

2. Python Data Science Handbook: Essential Tools for Working with Data Book by Jake VanderPlas

Course Outline

Week	Tentative Course Contents/Topics
1	Introduction to Data Science Data science vs. data analysis vs. data engineering Data science Lifecycle and Process Types of data analytics
2-3	Types of Data Data Quality Data Collection (Web Scrapping) Data preparation/ Data Wrangling Data integration Data cleaning Noise removal, outlier removal, handling missing and inconsistency in data Data transformation Normalization, aggregation Encoding techniques
4-5	Probability and statistics Exploratory Data Analysis ● Statistical measures for analysis ● Descriptive Statistics ● Frequency and mode ● Percentile ● Measure of location: Mean and median ● Measure of spread: Range and variance Distribution of Data Gaussian Distribution Skewness Kurtosis Correlation Data visualization Histograms

	Box Plots Scatter Plots Heatmaps
6-7	Feature Selection: Overview of Feature Selection approaches ● Filter-based ● Wrapper-based Feature Reduction: Dimensionality Reduction using Principal Component Analysis
8	Textual Data preprocessing Image Data Preprocessing
9-12	Introduction to Machine Learning: Supervised and Unsupervised learning Intro to Linear Regression: Model Representation for single variable Linear Regression Multiple Variable Linear Regression, Normal equation Linear Regression (Regression Problems) Gradient Descent Gradient Descent with momentum, SGD Logistic Regression (Classification Problems)
12-14	CNN (Architecture, Applications) Evaluation Measures: Accuracy, Recall, Specificity, Precision False positive rate, ROC, Cross validation
14-16	Recommendation engines, Association Analysis

Evaluation (Subject to change)

Assignments	(3 to 4)	10%
Quizzes	(3 to 4)	10%
Mid Exams	(1)	30%
Final Exam	(1)	40%
Project	(1)	10%
Total:		**100 %**

Grading Policy

- **Relative**