National University of Computer and Emerging Sciences, Lahore Campus

Course:	Introduction to Data Science	Course Code:	DS-2001
Program:	BS(Data Science)	Semester:	Fall 2023
Duration:	-	Total Marks:	45
Due Date:	02-Dec-23	Weight	
Section:	A	Page(s):	3
Exam:	Project Part 2	Roll No.	

Instruction/Notes:

- Read the assignment carefully. Make sure you understand the requirements and expectations of the assignment.
- Ensure that you have all the necessary files and documents ready for submission in the CORRECT format.
- Only group leader should submit the files.
- The assignment must be submitted before the announced DEADLINE. One mark will be deducted for each day of late submission.

Machine Learning Modeling and Evaluation

Congratulations on completing the second stage of your data science project! You have collected, cleaned, transformed, explore and visualize the dataset. In this assignment, you will use various machine learning techniques to build a model on your dataset. Machine learning algorithm takes the training data as input, learns patterns from it and then predicts the unseen data based on its experience. Evaluate the performance of the model based on its prediction and then retrain it if needed.

Instructions:

Develop a predictive model to make predictions about the target variable based on the other variables in the data set. You can apply regression, classification, or clustering according to your problem.

Evaluate the performance of your model and refine it as needed. Evaluation metrics depend on the nature of your problem domain. Follow the below given chart to evaluate your model.

Regression- R2 Score, MAE, MSE, RMSE

Classification- Accuracy, Loss, Confusion Matrix, Precision, Recall, F1 Score, ROC/AUC

Clustering Extrinsic Measures: Mutual Information, Rand Index, Adjusted Rand Index

Clustering Intrinsic Measures: Silhouette Score Davies-Bouldin Index

Deliverables:

1- Complete Notebook

A Jupyter Notebook containing all the phases of your project including data collection, wrangling, exploration, visualization, transformation, machine learning modeling and evaluation.

2- Project Report

Write a detailed report explaining the problem, background, literature review, methodology, implementation and results section. Format is attached with this document. Report must be of latest 3 pages (excluding references).

Evaluation Criteria:

· Machine learning modeling (20 marks)
· Model evaluation and improvement (10 marks)
· Presentation (15 marks) [Schedule will be communicated soon]
. Report (15 marks)