



Department of Computer Science (New Campus) University of Engineering & Technology, Lahore

Subject: **Introduction to Data Science**

(6th Semester, 2021 Session, SE)

Total Marks: 10

Due Date: **28-04-2024 (11:55 PM)**

Semester Project

CLOs	Questions	Marks				
2	<table><tr><td>A</td><td>https://archive.ics.uci.edu/dataset/602/dry+bean+dataset https://archive.ics.uci.edu/dataset/887/national+health+and+nutrition+health+survey+2013-2014+(nhanes)+age+prediction+subset</td></tr><tr><td>B</td><td>https://archive.ics.uci.edu/dataset/1/abalone https://archive.ics.uci.edu/dataset/78/page+blocks+classification</td></tr></table> <p>Instructions:</p> <ol style="list-style-type: none">1. If your registration number is ODD, then you will work on A datasets otherwise on B datasets.2. First, you need to perform Exploratory Data Analysis (EDA) on the given datasets. You should draw 4 plots and explain those plots in detail. While drawing 4 plots, you will draw <u>2 different plots in each Bivariate and Multivariate analysis</u>.3. Second, apply kNN and Naïve Bayes on these datasets for classification and draw confusion matrix, calculate accuracy, precision, recall and F1-score.4. You need to write a report to explain 4 plots and the results of classification. In the report, you will add all the plots as Figures (e.g., Figure 1, Figure 2 etc.) and then explain the Figure / plots. Add the results of classification as a table and figures of confusion matrix and explain the results.5. Your <u>report should not exceed 7 pages</u> including the title page. On the title page, you must mention your Name and Registration number.6. You can use MS Word for writing the report and then convert it to a PDF file. The name of the PDF file should be your registration number in the format 2021SExxx. (If you fail to follow this format, you will lose marks).	A	https://archive.ics.uci.edu/dataset/602/dry+bean+dataset https://archive.ics.uci.edu/dataset/887/national+health+and+nutrition+health+survey+2013-2014+(nhanes)+age+prediction+subset	B	https://archive.ics.uci.edu/dataset/1/abalone https://archive.ics.uci.edu/dataset/78/page+blocks+classification	10
A	https://archive.ics.uci.edu/dataset/602/dry+bean+dataset https://archive.ics.uci.edu/dataset/887/national+health+and+nutrition+health+survey+2013-2014+(nhanes)+age+prediction+subset					
B	https://archive.ics.uci.edu/dataset/1/abalone https://archive.ics.uci.edu/dataset/78/page+blocks+classification					

	<p>7. All the plots should have <u>your registration number as a watermark</u>. Similarly, all the pages of the report should be watermarked with your registration number.</p> <p>8. Before submitting, make sure that the PDF file can be opened. You can save plots as PNG or JPEG files and then insert in the Word Document.</p>	
--	---	--