Bansilal Ramnath Agarwal Charitable Trust's
Vishwakarma Institute of Information Technology, Pune-48
(An Autonomous Institute affiliated to Savitribai Phule Pune University)
Department of Computer Science and Engineering (Artificial Intelligence)

LIST OF LAB ASSIGNMENTS

ACADEMIC YEAR: 2023-24 DATE : 08-01-2024
DEPARTMENT: CSE-AI
CLASS: SY-A SEMESTER: II

SUBJECT: CAUA22201: Data Science and Machine Learning

Sr. No.	Titles of Lab Assignment	CO	BL
1	Perform the following operations using R/Python on suitable data sets: a) read data from different formats (like csv, xls) b) indexing and selecting data, sort data, c) describe attributes of data, checking data types of each column, d) counting unique values of data, format of each column, converting variable data type (e.g. from long to short, vice versa), e) identifying missing values and fill in the missing values	CO2	L1
2	Perform the following operations using R/Python on the data sets: a) Compute and display summary statistics for each feature available in the dataset. (e.g. minimum value, maximum value, mean, range, standard deviation, variance and percentiles b) Data Visualization-Create a histogram for each feature in the dataset to illustrate the feature distributions. c) Data cleaning, Data integration, Data transformation, Data model building (e.g. Classification)	CO2	L1
3	Apply appropriate ML algorithm on a dataset collected in a cosmetics shop showing details of customers to predict customer response for special offers.	CO3	L2
4	Write a program to do following: We have given a collection of 8 points. P1=[0.1,0.6] P2=[0.15,0.71] P3=[0.08,0.9] P4=[0.16, 0.85] P5=[0.2,0.3] P6=[0.25,0.5] P7=[0.24,0.1] P8=[0.3,0.2]. Perform the k-mean clustering with initial centroids as m1=P1=Cluster#1=C1 and m2=P8=cluster#2=C2. Answer the following: a) Which cluster does P6 belong to? b) What is the population of a cluster around m2? c) What is the updated value of m1 and m2?	CO4	L2
5	Visualize the data using R/Python by plotting the graphs for assignment no. 1 and 2. Consider a suitable data set. a) Use Scatter plot, bar plot, Box plot and Histogram OR b) Perform the data visualization operations using Tableau for the given dataset.	CO6	L2

6	Assignment on Regression technique. Download temperature data from the link below. https://www.kaggle.com/venky73/temperaturesof-india?select=temperatures.csv This data consists of temperatures of INDIA averaging the temperatures of all places month wise. Temperatures values are recorded in CELSIUS a) Apply Linear Regression using a suitable library function and predict the Month-wise temperature. b) Assess the performance of regression models using MSE, MAE and R-Square metrics c) Visualize a simple regression model.	CO5	L3
7	Assignment on Classification technique Every year many students give the GRE exam to get admission in foreign Universities. The data set contains GRE Scores (out of 340), TOEFL Scores (out of 120), University Rating (out of 5), Statement of Purpose strength (out of 5), Letter of Recommendation strength (out of 5), Undergraduate GPA (out of 10), Research Experience (0=no, 1=yes), Admitted (0=no, 1=yes). Admitted is the target variable. Data Set: https://www.kaggle.com/mohansacharya/graduate-admissions The counselor of the firm is supposed to check whether the student will get an admission or not based on his/her GRE score and Academic Score. So to help the counselor to take appropriate decisions, build a machine learning model classifier using a Decision tree to predict whether a student will get admission or not. a) Apply Data pre-processing (Label Encoding, Data Transformation….) techniques if necessary. b) Perform data-preparation (Train-Test Split) c) Apply Machine Learning Algorithm d) Evaluate Model.	CO5	L3

Mini project is to be performed in a group of 3 to 4 students.

Develop a mini project in a group using different predictive models techniques to solve any real life problem.

Rubrics for Progressive Assessment -

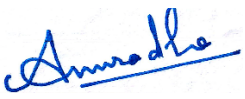
Assignment / File submission (10 Marks)	Mock Exam / Test Marks (10 Marks)	Attendance (5 Marks)	Total (25 Marks)

Rubrics for Assignment Submission -

Timely Performance & submission (4 Marks)	Understanding (3 Marks)	Presentation (3 Marks)	Total (10 Marks)

Rubrics for Mock Exam -

Implementation (5 Marks)	Oral (5 Marks)	Total (10 Marks)



Dr. Anuradha Yenkikar

Subject Coordinator

Dr. Nilesh Sable

HOD IT