

DATA SCIENCE LAB
(20MCA241)

LAB RECORD

Submitted in partial fulfilment of the requirements for the award of the degree
of Master of Computer Applications of A P J Abdul Kalam Technological
University, Kerala.

Submitted by:

SREYAS SATHEESH (SJC23MCA-2053)



MASTER OF COMPUTER APPLICATIONS

**ST. JOSEPH'S COLLEGE OF ENGINEERING AND
TECHNOLOGY, PALAI**

CHOONDACHERRY P.O, KOTTAYAM
KERALA

November 2024

ST. JOSEPH'S COLLEGE OF ENGINEERING AND TECHNOLOGY, PALAI

(AUTONOMOUS)

(An ISO 9001: 2015 Certified College)

CHOONDACHERRY P.O, KOTTAYAM, KERALA



CERTIFICATE

This is to certify that the Data Science Lab Record (20MCA241) submitted by **Sreyas Satheesh**, student of **Third** semester **MCA** at **ST. JOSEPH'S COLLEGE OF ENGINEERING AND TECHNOLOGY, PALAI** in partial fulfilment for the award of Master of Computer Applications is a bonafide record of the lab work carried out by him under our guidance and supervision. This record in any form has not been submitted to any other University or Institute for any purpose.

Dr. Rahul Shajan

Associate Professor
(Head, Department of
Computer Application)

Mr. Anish Augustine K

Assistant Professor,
Dept. of Computer Application
(Faculty In-Charge)

Submitted for the End Semester Examination held on _____

Examiner 1

Examiner 2

DECLARATION

I Sreyas Satheesh, do hereby declare that the Data Science Lab Record (20MCA241) is a record of work carried out under the guidance of Mr. Anish Augustine K, Asst. Professor, Department of Computer Applications, SJ CET, Palai as per the requirement of the curriculum of Master of Computer Applications Programme of A P J Abdul Kalam Technology University, Thiruvananthapuram. Further, I also declare that this record has not been submitted, full or part thereof, in any University / Institution for the award of any Degree / Diploma.

Place: Choondacherry

SREYAS SATHEESH

Date :

(SJC23MCA-2053)

DEPARTMENT OF COMPUTER APPLICATIONS

VISION

To emerge as a center of excellence in the field of computer education with distinct identity and quality in all areas of its activities and develop a new generation of computer professionals with proper leadership, commitment and moral values.

MISSION

- Provide quality education in Computer Applications and bridge the gap between the academia and industry.
- Promoting innovation research and leadership in areas relevant to the socio economic progress of the country.
- Develop intellectual curiosity and a commitment to lifelong learning in students, with societal and environmental concerns.

COURSE OUTCOMES

After the completion of the course 20MCA241 Data Science Lab the student will be able to :

CO 1	Use different python packages to perform numerical calculations, statistical computations and data visualization	K3 (Apply)
CO 2	Use different packages and frameworks to implement regression and classification algorithms.	K3 (Apply)
CO 3	Use different packages and frameworks to implement text classification using SVM and clustering using k-means	K (Apply)
CO 4	Implement convolutional neural network algorithm using Keras framework.	K3 (Apply)
CO 5	Implement programs for web data mining and natural language processing using NLTK	K3 (Apply)

Sl. No.	Program List	Pg No	Date
1	Write a program represent the days of the week using a List, Tuple, and Dictionary, and display their types.	1	
2	Write a program to find the sum of 2 matrices using nested List.	2	
3	Write a program to perform bubble sort on a given set of elements.	3	
4	Program to find the count of each vowel in a string(use dictionary)	4	
5	Create a Python program that repeatedly subtracts the sum of the digits of a positive number from the number itself until the result is positive.	5	
6	Construct a 2x3 2-dimensional array with complex numbers and print it. Additionally, show: <ul style="list-style-type: none"> a. The number of rows and columns b. The array's dimensions c. The reshaped array into a 3x2 format 	7	
7	Construct a 1D array using the arange function with 10 elements. Display <ul style="list-style-type: none"> a. First 4 elements b. Last 6 elements c. Elements from index 2 to 7 	9	
8	Create an 1D array with arrange containing first 15 even numbers as elements <ul style="list-style-type: none"> a. Elements from index 2 to 8 with step 2(also demonstrate the same using slice function) b. Last 3 elements of the array using negative index c. Alternate elements of the array d. Display the last 3 alternate elements 	10	
9	Create a 2D array with 4 rows and 4 columns. <ul style="list-style-type: none"> a. Display all elements excluding the first row b. Display all elements excluding the last column c. Display the elements of 1 st and 2 nd column in 2 nd and 3 rd row d. Display the elements of 2 nd and 3 rd column e. Display 2 nd and 3 rd element of 1 st row f. Display the elements from indices 4 to 10 in descending order(use-values) 	11	

10	Create two 2D arrays using array object and <ul style="list-style-type: none"> a. Add the 2 matrices and print it b. Subtract 2 matrices c. Multiply the individual elements of matrix d. Divide the elements of the matrices e. Perform matrix multiplication f. Display transpose of the matrix g. Sum of diagonal elements of a matrix 	13	
11	Create a square matrix with random integers using randint() and d Demonstrate the following: <ul style="list-style-type: none"> a. Inverse b. Rank of matrix c. Determinant d. Transform matrix into 1D array e. Eigen values and vectors 	15	
12	Create a matrix X with appropriate rows and columns. Demonstrate the following: <ul style="list-style-type: none"> a. Cube each element using different methods (multiply(), *, power(), **) b. Display the identity matrix of the given square matrix c. Raise each element of the matrix to different powers 	17	
13	Define two matrices: A (5x6) and B (3x3). Extract a 3x3 submatrix from A, multiply it by B, and replace the original submatrix in A with the resulting matrix.	19	
14	For a given matrix-vector equation $AX = b$, write a program to solve for X using the solve() function, with provided matrices A and b	21	
15	Write a program to perform Singular Value Decomposition (SVD) on a given matrix A using numpy.linalg.svd(). Also, reconstruct the original matrix from the three matrices obtained after performing SVD.	22	

16	Represent the following information using a line graph with following style properties.	23	
17	The table gives the daily sales of items in a shop. Use subplot function to draw the line graphs with grids(color as blue and line style dotted) for the above information as 2 separate graphs in two rows.	25	
18	Create a scatter plot for the given data in Scatter function.	27	
19	Display the data using multiline plot(3 different lines in same graph) a. Display the description of the graph in upper right corner(use legend()) b. Use different colors and line styles for 3 different lines	29	
20	Construct a bar graph to represent the survey results of 100 students' primary modes of transport to school.	30	
21	Using the heights of 30 cherry trees, create a histogram with a bin size of 5 inches.	32	

22	Using the pandas read_csv() function, read the given 'iris' data set and demonstrate the following: a. Display Shape of the data set. b. First 5 and last five rows of data set(head and tail). c. Size of dataset. d. No. Of samples available for each variety. e. Description of the data set(use describe).	33	
23	Utilize the pairplot() function in seaborn to display pairwise relationships between Attributes. Experiment with various plot types ('scatter', 'kde', 'hist', 'reg') And different markers.	34	
24	Using the 'iris dataset' familiarize with the following seaborn functions: a. Displot() b. Histplot() c. Relplot()	36	
25	Using the 'iris dataset' , implement the KNN algorithm. Take different values for the Test and training data set .Also use different values for k. Also find the accuracy level.	38	
26	Using 'blood_transfusion dataset' implement KNN algorithm.	39	
27	Implement Naive Bayes classification on the iris data set using different algorithms (i.e., Gaussian, Bernoulli). Additionally: a. Determine the accuracy level for each algorithm. b. Display the number of mislabeled classifications from the test data set. c. List the class labels of the mismatching records.	41	

28	Using a 'car_details' CSV file, implement the decision tree algorithm. Additionally: <ul style="list-style-type: none"> a. Determine the accuracy level b. Display the number of mislabeled classifications from the test data set c. List the class labels of the mismatching records 	45	
29	Implement simple linear regression using the 'student_score.csv' dataset and Multiple linear regression using the 'company_data.csv' dataset.	47	
30	Create a neural network using the 'houseprice.csv' dataset to predict whether the price of the house is above or below the median value.	52	
31	Write a program to implement a simple web crawler using Python to extract and display the content of the page from <p> tags.	54	
32	Write a program to implement a simple web crawler using Python to display all the Hyperlinks on the page.	55	
33	Write a program to implement n-grams for Natural Language Processing without using any external libraries.	57	

34	Write a program for Natural Language Processing to perform n-grams using the NLTK library.	58	
35	For a given text, perform the following Natural Language Processing tasks: <ul style="list-style-type: none"> a. Word tokenization b. Sentence tokenization c. Remove stop words from the text d. Create n-grams 	59	
36	Using k means clustering analyze the dataset 'customer_data_.csv' that creates 6 clusters of customers based on their spending pattern.	61	