

20MCA241 – Data Science Lab

Lab Report Submitted By

ALFIYA P.S

AJC22MCA-2011

In Partial Fulfilment for the Award of the Degree Of

**MASTER OF COMPUTER APPLICATIONS
(MCA TWO YEAR)
[Accredited by NBA]**

APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY



**AMAL JYOTHI COLLEGE OF ENGINEERING
KANJIRAPPALLY**

[Affiliated to APJ Abdul Kalam Technological University, Kerala. Approved by AICTE,
Accredited by NAAC. Koovappally, Kanjirappally, Kottayam, Kerala – 686518]

2023-2024

DEPARTMENT OF COMPUTER APPLICATIONS
AMAL JYOTHI COLLEGE OF ENGINEERING
KANJIRAPPALLY



CERTIFICATE

This is to certify that the lab report, “**20MCA241 DATA SCIENCE LAB**” is the bonafide work of **ALFIYA P.S (AJC22MCA-2011)** in partial fulfilment of the requirements for the award of the Degree of Master of Computer Applications under APJ Abdul Kalam Technological University during the year **2023-24**.

Ms. Jetty Benjamin

Lab In- Charge

Rev. Fr. Dr. Rubin Thottupurathu Jose

Head of the Department

Internal Examiner

External Examiner

Course Code	Course Name	Syllabus Year	L-T-P-C
20MCA241	Data Science Lab	2020	0-1-3-2

VISION

To promote an academic and research environment conducive for innovation centric technical education.

MISSION

- MS1 - Provide foundations and advanced technical education in both theoretical and applied Computer Applications in-line with Industry demands.
- MS2 - Create highly skilled computer professionals capable of designing and innovating real life solutions.
- MS3 - Sustain an academic environment conducive to research and teaching focused to generate up-skilled professionals with ethical values.
- MS4 - Promote entrepreneurial initiatives and innovations capable of bridging and contributing with sustainable, socially relevant technology solutions.

COURSE OUTCOME

CO	Outcome	Target
CO1	Use different python packages to perform numerical calculations, statistical computations and data visualization.	60.2
CO2	Use different packages and frameworks to implement regression and classification algorithms.	60.2
CO3	Use different packages and frameworks to implement text classification using SVM and clustering using K-means.	60.2
CO4	Implement convolutional neural network algorithm using Keras framework.	60.2
CO5	Implement programs for web data mining and natural language processing using NLTK.	60.2

COURSE END SURVEY

CO	Survey Question	Answer Format
CO1	To what extent you are able to use different python packages to perform numerical calculations, statistical computations and data visualization?	Excellent/Very Good/Good/Satisfactory/Poor
CO2	To what extent you are able to use different packages and frameworks to implement regression and classification algorithms?	Excellent/Very Good/Good/Satisfactory/Poor
CO3	To what extent you are able to use different packages and frameworks to implement text classification using SVM and clustering using K-means?	Excellent/Very Good/Good/Satisfactory/Poor
CO4	To what extent you are able to implement convolutional neural network algorithm using Keras framework?	Excellent/Very Good/Good/Satisfactory/Poor
CO5	To what extent you are able to implement programs for web data mining and natural language processing using NLTK?	Excellent/Very Good/Good/Satisfactory/Poor

CONTENT

SL. NO.	LIST OF LAB EXPERIMENTS/EXERCISES	DATE	CO	PAGE NO
1	Program to perform matrix operations. Use numpy as the python library and perform the operation using built in functions.	25-09-23	CO1	1
2	Program to perform single value decomposition using numpy.	29-09-23	CO1	3
3	Program to perform data visualization using python library matplotlib.	29-09-23	CO1	5
4	Program to implement KNN classification using any standard dataset available in the public domain and find the accuracy of algorithm (Iris Dataset)	10-10-23	CO2	6
5	Program to implement KNN classification using any standard dataset available in the public domain and find the accuracy of algorithm (Load Digits)	10-10-23	CO2	7
6	Program to implement Naïve Bayes Algorithm using any standard dataset available in the public domain and find the accuracy of algorithm (Iris Dataset)	31-10-23	CO2	9
7	Program to implement Naïve Bayes Algorithm using any standard dataset available in the public domain and find the accuracy of algorithm (Breast Cancer Dataset)	31-10-23	CO2	10
8	Give one dimensional dataset represented with numpy array. Write a program to calculate slope and intercept	10-11-23	CO2	11
9	Program to implement simple linear regression using any standard dataset available in the public domain and find r2 score.	07-11-23	CO2	13
10	Program to implement linear and multiple regression techniques using any standard dataset available in the public domain and evaluate its performance	10-11-23	CO2	15
11	Program to implement decision trees using any standard dataset available in the public domain and find the accuracy of the algorithm (Iris Dataset)	03-11-23	CO3	16
12	Program to implement decision trees using any standard dataset available in the public domain and find the accuracy of the algorithm (Breast Cancer Dataset)	03-11-23	CO3	18
13	Program to implement k-means clustering technique using any standard dataset available in the public domain (Iris Dataset)	21-11-23	CO3	20
14	Program to implement k-means clustering technique using any standard dataset available in the public domain (Breast Cancer Dataset)	21-11-23	CO3	22
15	Program to implement text classification using support vector machine.	30-11-23	CO3	24
16	Program on artificial neural network to classify images from any standard dataset in the public domain using Keras framework.	01-12-23	CO4	27
17	Program to implement a simple web crawler using requests library	06-12-23	CO5	29

18	Program to implement a simple web crawler and parse the content using BeautifulSoup.	06-12-23	CO5	30
19	Implement problems on natural language processing – Part of Speech tagging, N-gram & smoothening and Chunking using NLTK	07-12-23	CO5	32

