20MCA241	DATA SCIENCE LAB	CATEGORY	L	T	P	CREDIT
		LAB	0	1	3	2

Preamble: This is an introductory practical course on Data Science and student will learn how to use various scientific libraries in python to implement data mining techniques and machine learning algorithms.

Prerequisite: Fundamentals of programming, python programming fundamentals, Machine learning, fundamentals of web programming,

Course Outcomes: After the completion of the course the student will be able to

CO No.	Course Outcome (CO)	Bloom's Category Level
CO 1	Use different python packages to perform numerical calculations, statistical computations and data visualization	Level 3: Apply
CO 2	Use different packages and frameworks to implement regression and classification algorithms.	Level 3: Apply
CO 3	Use different packages and frameworks to implement text classification using SVM and clustering using k-means	Level 3: Apply
CO 4	Implement convolutional neural network algorithm using Keras framework.	Level 3: Apply
CO 5	Implement programs for web data mining and natural language processing using NLTK	Level 3: Apply

Mapping of course outcomes with program outcomes

	PO 1	PO 2	PO 3	PO 4	PO 5	PO 6	PO 7	PO 8	PO 9	PO 10	PO 11	PO 12
CO 1	3	3	3	1	3	2	3		2			
CO 2	3	3	3	2	3	2	3		2			
CO 3	3	3	3	2	3	2	3		2			
CO 4	3	3	3	2	3	2	3		2			
CO 5	3	3	3	2	3	3	3		2			

3/2/1: High/Medium/Low

Assessment Pattern

Bloom's Category	Continuous Assessment Tests		End Semester Examination
	1	2	
Remember (K1)			
Understand (K2)			
Apply (K3)	50	50	50
Analyse (K4)			
Evaluate (K5)			
Create(K6)			

Mark distribution

Total Marks	CIE	ESE	ESE Duration
100	50	50	3 hours

Continuous Internal Evaluation Pattern:

Maximum Marks: 50	
Attendance	7½
Maintenance of daily lab record and GitHub management	10
Regular class viva voce	7½
Timely completion of day-to-day tasks	10
Tests/Evaluation	15

End Semester Examination Pattern:

Maximum Marks: 50			
Verification of Daily program record and Git Repository			5 marks
Viva			10 marks
Problem solving (Based on difficulty level, one or more questions may be given)	Flowchart / Algorithm / Structured description of problem to explain how the problem can be solved / Interface Design	15%	35 marks
	Program correctness	50%	
	Code efficiency	15%	
	Formatted output	20%	

Course Level Assessment Questions

Course Outcome 1 (CO1):

- Review of python programming – Programs review the fundamentals of python (simple python programs ice breaker) – (at most one lab session)

- Matrix operations (using vectorization) and transformation using python and SVD using Python.
- Programs using matplotlib / plotly / bokeh / seaborn for data visualisation.
- Programs to handle data using pandas.

Course Outcome 2 (CO2)

- Program to implement k-NN classification using any standard dataset available in the public domain and find the accuracy of the algorithm.
- Program to implement Naïve Bayes Algorithm using any standard dataset available in the public domain and find the accuracy of the algorithm
- Program to implement linear and multiple regression techniques using any standard dataset available in the public domain and evaluate its performance.

Course Outcome 3(CO3):

- Program to implement text classification using Support vector machine.
- Program to implement decision trees using any standard dataset available in the public domain and find the accuracy of the algorithm.
- Program to implement k-means clustering technique using any standard dataset available in the public domain

Course Outcome 4 (CO4):

- Programs on feedforward network to classify any standard dataset available in the public domain.
- Programs on convolutional neural network to classify images from any standard dataset in the public domain.

*[Note] : Encourage students to refer standard neural network architectures such as LeNet5, ResNet, GoogLeNet etc. and use these as starting points for their models.

Course Outcome 5 (CO5):

Web Data Mining

- Implement a simple web crawler (ensure ethical conduct).
- Implement a program to scrap the web page of any popular website – suggested python package is scrappy (ensure ethical conduct).

Natural Language Processing

Problems may be designed for the following topics so that students can get hands on experience in using python for natural language processing:

- Part of Speech tagging
- N-gram and smoothening
- Chunking

Syllabus
Review of python programming, Matrix operations, Data Visualisation using matplotlib / plotly / bokeh / seaborn, Data handling using pandas, Classification k-NN algorithm, Naïve Bayes algorithm, Implementation of linear and multiple regression techniques, Text classification using Support vector machine, Implementation of Decision Trees, Clustering using k-means algorithm, Convolutional Neural Network to classify images using Keras framework, Web Crawler and Scrapping web pages, Implementation of NLP - Part of Speech tagging, N-gram & smoothening and Chunking using NLTK.

Reference Books

1. Christopher M Bishop, “Pattern Learning and Machine Learning”, Springer, 2006
2. E. Alpayidin, “Introduction to Machine Learning”, Prentice Hall of India (2005)
3. T. Hastie, RT Ibrashiran and J. Friedman, “The Elements of Statistical Learning”, Springer 2001
4. Toby Segaran, “Programming Collective Intelligence: Building Smart Web 2.0 Applications”, O' Reilly Media; 1 edition (16 August 2007).
5. Drew Conway, John Myles White, “Machine Learning for Hackers: Case Studies and
6. Algorithms to Get You Started”, O' Reilly Media; 1 edition (13 February 2012)
7. Simon Rogers, Mark Girolami, “A First course in Machine Learning”, CRC Press, First Indian reprint, 2015.
8. Tom Mitchell, “Machine Learning”, McGraw Hill, 1997.
9. Bing Liu, Web Data Mining - Exploring Hyperlinks, Contents and Usage Data, Second edition, Springer 2011

Course Contents and Lab Schedule

Sl No.	Topic	No. of hours
1	Review of python programming, Matrix operations, Programs using matplotlib / plotly / bokeh / seaborn for data visualisation and programs to handle data using pandas.	8
2	Program to implement k-NN classification using any standard dataset available in the public domain and find the accuracy of the algorithm	2
3	Program to implement Naïve Bayes Algorithm using any standard dataset available in the public domain and find the accuracy of the algorithm	2
4	Program to implement linear and multiple regression techniques using any standard dataset available in the public domain and evaluate its performance.	4
5	Program to implement text classification using Support vector machine.	4
6	Program to implement decision trees using any standard dataset available in the public domain and find the accuracy of the algorithm	4
7	Program to implement k-means clustering technique using any standard dataset available in the public domain	2
8	Program on convolutional neural network to classify images from any standard dataset in the public domain using Keras framework.	6
9	Program to implement a simple web crawler and scrapping web pages.	6
10	Implement problems on natural language processing - Part of Speech tagging, N-gram & smoothening and Chunking using NLTK	8