Handbook of B.Tech. Programmes offered by USICT at Affiliated Institutions of the University.

24c/C

Data Science	L	P	C
	3		3

Discipline(s) / EAE / OAE	Semester	Group	Sub-group	Paper Code
CSE/IT/CST/ITE	7	PCE	PCE-4	CIE-405T
MAE	7	OAE-MAE	OAE-1	MAO-417T

Marking Scheme:
1. Teachers Continuous Evaluation: 25 marks
2. Term end Theory Examinations: 75 marks

Instructions for paper setter:
1. There should be 9 questions in the term end examinations question paper.
2. The first (1st) question should be compulsory and cover the entire syllabus. This question should be objective, single line answers or short answer type question of total 15 marks.
3. Apart from question 1 which is compulsory, rest of the paper shall consist of 4 units as per the syllabus. Every unit shall have two questions covering the corresponding unit of the syllabus. However, the student shall be asked to attempt only one of the two questions in the unit. Individual questions may contain upto 5 sub-parts / sub-questions. Each Unit shall have a marks weightage of 15.
4. The questions are to be framed keeping in view the learning outcomes of the course / paper. The standard / level of the questions to be asked should be at the level of the prescribed textbook.
5. The requirement of (scientific) calculators / log-tables / data – tables may be specified if required.

Course Objectives :

1.	To introduce the students about the knowledge and overview of R or Octave statistical package, data transformation and merging, data visualization and illustration of techniques through R or Octave.
2.	To understand statistical techniques like regression analysis and structural equation modelling.
3.	To promote deeper understanding of forecasting, time series data analysis and auto regression models.
4.	To provide overview of support vector machine, linear discriminant analysis and clustering techniques.

Course Outcomes (CO)

CO 1	Develop relevant programming abilities.
CO 2	Demonstrate proficiency with statistical analysis of data.
CO 3	Develop the ability to build and assess data-based models.
CO 4	Execute statistical analyses with professional statistical software.

Course Outcomes (CO) to Programme Outcomes (PO) mapping (scale 1: low, 2: Medium, 3: High)

	PO01	PO02	PO03	PO04	PO05	PO06	PO07	PO08	PO09	PO10	PO11	PO12
CO 1	3	2	3	3	-	2	-	-	-	-	-	2
CO 2	3	2	2	3	-	3	-	-	-	-	-	2
CO 3	2	2	3	3	-	3	-	-	-	-	-	2
CO 4	3	3	2	3	-	3	-	-	--	-	-	2

UNIT- I

Introduction to Data Science, Difference among AI, Machine Learning and Data Science.
Basic introduction of python, Google Colab and their features
Popular Dataset Repositories along with discussion on some datasets
Data Pre-processing, Data Scales, Similarity and Dissimilarity measures, sampling and quantization of data, filtering, Data transformation and merging, Data visualization, PCA, Correlation, Chi-Square test. Illustration of these techniques through Python.

UNIT- II

Regression Analysis, linear, generalized, regularized regression, Cross-validation, Training and Testing data set, Overview of nonlinear regression, Overview of Ridge regression, Latent variables, Structure Equation

modelling. Illustration of these techniques through Python.

UNIT- III

Forecasting, time series data analysis, Stationarity, Seasonality, recurrent models, autoregressive models. Illustration of these techniques through Python.

UNIT- IV

Classification, Linear discriminant analysis, overview of support vector machine, Decision trees, Clustering, Clustering techniques. Illustration of these techniques through Python.

Text Books:
1. Runkler, Thomas A., "Data Analytics: Models and Algorithms for Intelligent Data Analysis", Springer, 2012.
2. Friedman, Jerome, Trevor Hastie, and Robert Tibshirani, "The elements of statistical learning". Vol. 1. New York: Springer Series in Statistics, 2001.
3. Thareja, R., "Data Science and Machine Learning using Python". McGraw Hill, 2022.

References Books:
1. Kroese, D. P., Botev, Z., Taimre, T., &Vaisman, R., "Data Science and Machine Learning: Mathematical and Statistical Methods." CRC Press, 2019.
2. Grus, J. "Data Science from Scratch: First Principles with Python." O'Reilly Media, 2019.

24a/C

Data Science Lab			L	P	C
				2	1

Discipline(s) / EAE / OAE	Semester	Group	Sub-group	Paper Code
CSE/IT/CST/ITE	7	PCE	PCE-4	CIE-405P
MAE	7	OAE-MAE	OAE-1	MAO-417P

Marking Scheme:
1. Teachers Continuous Evaluation: 40 marks
2. Term end Theory Examinations: 60 marks

Instructions:
1. The course objectives and course outcomes are identical to that of (Data Science) as this is the practical component of the corresponding theory paper.
2. The practical list shall be notified by the teacher in the first week of the class commencement under intimation to the office of the Head of Department / Institution in which the paper is being offered from the list of practicals below. Atleast 10 experiments must be performed by the students, they may be asked to do more. Atleast 5 experiments must be from the given list.

Preform following experiments using Python:

1. Describing data, viewing and manipulating data.

2. To plot the probability distribution curve.

3. To perform chi square test on various data sets.

4. To use Python as a programming tool for the analysis of data structures.

5. To perform various operations such as data storage, analysis and visualization.

6. To perform descriptive statistics analysis and data visualization.

7. To perform Principal Component Analysis on datasets.

8. To perform linear regression on datasets.

9. To perform Data Aggregation and GroupWise Operations.