MECHANICAL

DEPARTMENT

2022-23

DA

PREPARED BY

Dr. Kalpana F.

Techno India NJR Institute of Technology

Course File

Data Analytics (4ME2-01)

Dr. Kalpana F.
Department of Basic Science

RAJASTHAN TECHNICAL UNIVERSITY, KOTA

Syllabus

2nd Year - IV Semester: B.Tech. : Mechanical Engineering

4ME2-01: DATA ANALYTICS

Credit: 2 **Max. Marks: 100 (IA:20, ETE:80)**

2L+0T+0P **End Term Exam: 3 Hours**

SN	Contents	Hours
1	**Introduction:** Objective, scope and outcome of the course.	1
2	Introduction to Multivariate Statistics-Degree of Relationship among Variables-Review of Univariate and Bivariate Statistics-Screening Data Prior to Analysis-Missing Data, Outliers, Normality, Linearity, and Homoscedasticity.	4
3	Multiple Regression- Linear and Nonlinear techniques- Backward Forward-Stepwise- Hierarchical regression-Testing interactions (2way interaction) - Analysis of Variance and Covariance (ANOVA & ANCOVA) - Multivariate Analysis of Variance and Covariance (MANOVA & MANCOVA).	6
4	Logistic regression: Regression with binary dependent variable - Simple Discriminant Analysis- Multiple Discriminant analysis Assessing classification accuracy- Conjoint analysis (Full profile method).	5
5	Principal Component Analysis -Factor Analysis- Orthogonal and Oblique Rotation-Factor Score Estimation-Multidimensional Scaling-Perceptual Map-Cluster Analysis (Hierarchical Vs Nonhierarchical Clustering).	5
6	Latent Variable Models an Introduction to Factor, Path, and Structural Equation Analysis- Time series data analysis (ARIMA model) – Decision tree analysis (CHAID, CART) - Introduction to Big Data Management.	5
	TOTAL	26

Course Overview: The course has certain outcomes by virtue of which the students will get an idea of the subject Data Analytic.

Course Outcomes:

CO No	Cognitive Level	Course Outcome
1		Describe Data Analytics and the skill sets need for a data analyst.
2	Comprehension	Explain statistical inference and probability distribution commonly used as foundation for statistical modelling.
3	Synthesis	Apply basic data analytics techniques: ANOVA, MANOVA, ANCOVA, MANCOVA, liner regression.
4	Synthesis	Identify common approaches and algorithms for basic features selection, decision trees and factor analysis.
5	Synthesis	Apply common approaches and algorithms used for Cluster analysis and Time series model.

Prerequisites:

1. Fundamentals of Database Management System.
2. Students should know about MS Excel.
3. Students should be able to implement the Data Analytics algorithm with Excel.
4. Students should be able to work various Spreadsheets tools.

Course Outcome Mapping with Program Outcome:

Course Outcome	Program Outcome											
CO No.	Domain-Specific					Domain-Independent						
	PO 1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12
CO 1	2	0	0	0	0	0	0	0	0	0	0	0
CO 2	2	2	0	0	0	0	0	0	0	0	0	0
CO 3	2	2	0	3	0	0	0	0	0	0	0	0
CO 4	2	3	0	3	0	0	0	0	0	0	0	0
CO 5	0	3	0	2	0	0	0	0	0	0	0	0
1: Slight (Low), 2: Moderate (Medium), 3: Substantial (high)												

Lecture plan based on Unit 1(Introduction - Outcome)

Lecture No.	Topic	Unit Mapping
1	Objective and Concept of Data Analytics, and how are they different from traditional Database Management systems with Real-time application use of Data Analytics	1

Lecture plan based on Unit 2 (Introduction to Data Analytics)

Lecture No.	Topic	Unit Mapping
2	Introduction to Multivariate Statistics Degree of Relationship	2
3	Introduction to Outliers and Normality	2
4	Linearity and Homoscedasticity	2
5	Variable Review of Univariate and Bivariate Statistics Screening Data	2

Lecture plan based on Unit 3 (Multiple Regression)

Lecture No.	Topic	Unit Mapping
6	Understanding Linear and Nonlinear Techniques	3
7	Basic about Backward Forward Stepwise Hierarchical Regression	3
8	Testing Interactions (2 way interaction)	3
9	Analysis of Variance and Covariance (ANOVA & ANCOVA)	3
10	Multivariate Analysis of Variance	3
11	Multivariate Analysis of Covariance (MANOVA & MANCOVA)	3

Lecture plan based on Unit 4 (Logistic Regression)

Lecture No.	Topic	Unit Mapping
12	Regression with binary dependent variable	4
13	Simple Discriminant Analysis	4
14	Multiple Discriminant Analysis	4
15	Assessing Classification Accuracy	4
16	Conjoint Analysis (Full Profile Method)	4

Lecture plan based on Unit 5 (Principal Component Analysis)

Lecture No.	Topic	Unit Mapping
17	Component Analysis	5
18	Factor Analysis	5
19	Orthogonal and Oblique Rotation Factor Score Method	5
20	Multidimensional Scaling Perceptual Map	5
21	Cluster Analysis (Hierarchical VS Non-hierarchical Clustering)	5

Lecture plan based on Unit 6 (Latent Variable Models)

Lecture No.	Topic	Unit Mapping
22	Introduction to Factor, Path and Structural Equation Analysis	6
23	Time Series data Analysis (ARIMA Model)	6
24	Decision Tree Analysis (CHAID, CART)	6
25	Introduction to Big Data Management	6
26	Case Studies of Big Data Management	6

Textbook – Dr Anil K Maheshwari, "Data Analytics Made Accessible:2020 Edition"

Reference Sessions – https://youtu.be/igmWbLprb_A

MOOC (Coursera) -https://www.coursera.org/articles/what-is-data-analysis-with-examples

Course Level Problems (Test Items):

CO No	Problem Description
1	A. Why is Big Data hot now? Discuss one case study in brief. B. Explain the Data Analytics.
2	A. Write a brief essay on "Factor Analysis". B. Write a short note on cluster analysis.
3	A. What is ANOVA? **B.** Explain briefly MANOVA and MANCOVA.
4	A. What is residual in multivariate analysis? B. How would you differentiate among multiple discriminant analysis, regression analysis and logistic regression analysis?
5	**A.** What do you understand by ARIMA model in time series data analysis?

Assessment Methodology:

1. Online quiz/poll question after every unit.
2. Practical exam in the lab where they have to implement their skills to manage the data for the given problem statement.
3. Midterm subjective paper where they have to write algorithms to perform different operations.
4. Final paper (subjective paper) at the end of the semester.

Teaching and Learning Resources Unit-Wise

Unit 2 (Introduction to Data Analytics)

Video Tutorial: https://youtu.be/igmWbLprb_A
Theory Concepts: https://nptel.ac.in/courses/110/106/110106072/

Unit 3(Multiple Regression)

Video Tutorial: https://www.youtube.com/watch?v=PrNslXgJNP8&ab_channel=IITRoorkeeJuly2018
Theory Concepts: https://www.investopedia.com/terms/m/mlr.asp

Unit 4(Logistic Regression)

Video Tutorial: https://www.youtube.com/watch?v=L_xBe7MbPwk
Theory Concepts: https://towardsdatascience.com/logistic-regression-detailed-overview-46c4da4303bc

Unit 5 (Principal Component Analysis)

Video Tutorial: https://www.youtube.com/watch?v=FgakZw6K1QQ&ab_channel=StatQuestwithJoshStarmer
Theory Concepts: https://en.wikipedia.org/wiki/Principal_component_analysis

Unit 6 (Latent Variable Models)

Video Tutorial: https://www.youtube.com/results?search_query=latent+variable+models&ab_channel=STUDYSQADACADEMY
Theory Concepts: https://ermongroup.github.io/cs228-notes/learning/latent/

Previous Year Question Paper

4E1231

Roll No. Total No of Pages: 3

4E1231

B. Tech. IV - Sem. (Main) Exam., May - 2019
BSC Automobile Engineering
4AE2 – 01 Data Analytics
AE, ME

Time: 2 Hours **Maximum Marks: 80**

Instructions to Candidates:

Attempt all five questions from Part A, four questions out of six questions from Part B and two questions out of three from Part C.

Schematic diagrams must be shown wherever necessary. Any data you feel missing may suitably be assumed and stated clearly. Units of quantities used /calculated must be stated clearly.

Use of following supporting material is permitted during examination. (Mentioned in form No. 205)

1. NIL 2. NIL

PART – A

(Answer should be given up to 25 words only) [5×2=10]

All questions are compulsory

Q.1 Define Homoscedasticity and Heteroscedasticity. [2]

Q.2 What is residual in multivariate analysis? [2]

Q.3 What is ANOVA? [2]

Q.4 Write a short note on cluster analysis. [2]

Q.5 "The analysis of time series is done to understand the dynamic conditions for achieving the short term and long term goals of business firms." Discuss. [2]

PART – B

(Analytical/Problem solving questions) **[4×10=40]**

Attempt any four questions

Q.1 Discuss the objectives and scope of data analytics in the industry. [10]

Q.2 How would you differentiate among multiple discriminant analysis, regression analysis and logistic regression analysis? [10]

Q.3 Find the multiple linear regression equation of X_1, X_2 and X_3 from data relating of three variables given below - [10]

X_1	4	6	7	9	13	15
X_2	15	12	8	6	4	3
X_3	30	24	20	14	10	4

Q.4 Write a brief essay on "Factor analysis". [10]

Q.5 What do you understand by ARIMA model in time series data analysis? [10]

Q.6 Explain briefly MANOVA and MANCOVA. [10]

PART – C

(Descriptive/Analytical/Problem Solving/Design Questions) **[2×15=30]**

Attempt any two questions

Q.1 Write a short note on (any three) –

(a) Missing data [5]

(b) Rotation [5]

(c) Conjoint Analysis [5]

(d) Outliers [5]

Q.2 Differentiate between CHAID and CART. How CHAID is better than CART? [10+5=15]

Q.3 The following table gives the average monthly sale of 4 salesman in three different types of territories: [15]

Territory	Salesman				Total
	A	B	C	D	
X	5	4	4	7	20
Y	7	8	5	4	24
Z	9	6	6	7	28
Total	21	18	15	18	72

Carry out a two way analysis of variance table from the above facts and interpret the result.

The 5% value of F for (3, 6) and (2, 6) degree of freedom are 4.76 and 5.14 respectively.
