**Techno India NJR Institute of Technology**



**Course File**

**Data Analytics (4ME2-01)**

 Dr Jitendra Shreemali

 **Department of Computer Science Engineering**



**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 |  | Identify common approaches and algorithms for basic features selection, decision trees and factor analysis. |
| 5 |  | 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** |
|  | **PO1** | **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** | 1. Why is Big Data hot now? Discuss one case study in brief.
2. Explain the Data Analytics.
 |
| **2** | 1. Write a brief essay on “Factor Analysis”.
2. Write a short note on cluster analysis.
 |
| **3** | 1. What is ANOVA?
2. Explain briefly MANOVA and MANCOVA.
 |
| **4** | 1. What is residual in multivariate analysis?
2. How would you differentiate among multiple discriminant analysis, regression analysis and logistic regression analysis?
 |
| **5** | 1. 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**

****

****

