



Kot, Bhalwal, Jammu

Model Institute of Engineering
& Technology (Autonomous)
Course Handout

COURSE HANDOUT

INTRODUCTION TO DATA SCIENCE (COM-501)

B.E-5th SEMESTER

ACADEMIC YEAR (2024-25)

Dr. Palvi Sharma

Assistant professor

Department of Computer Science & Engineering



Department of CSE

Model Institute of Engineering & Technology (Autonomous)

KotBhalwal, Jammu - 181122

www.mietjmu.in



Dr. Arun K. Gupta Teaching-Learning Centre

Version 1.1

श्रेष्ठ

श्रम

नवीनता

Please Do Not Print Unless Necessary



Kot, Bhalwal, Jammu

Course Code	Course Name	Course Type	Cd	L	T	P	Marks		
							Sessional	Final Exam	Total
COM-501	Introduction to Data Science	PCC	4	3	1	0	50	100	150

COURSE OUTCOMES

At the end of the course the student will be able to	
CO1	Describe big data concepts and emerging trends in analytics.
CO2	Apply data analysis lifecycle phases to real-world case studies.
CO3	Utilize R for descriptive statistics and exploratory data analysis.
CO4	Analyze clustering algorithms and association rules for data patterns.
CO5	Evaluate the effectiveness of regression and classification methods.

Unit-I: Introduction to Big Data Analytics: Big Data Overview: Data Structures, Analyst Perspective on Data Repositories, State of the Practice in Analytics: BI Versus Data Science, Current Analytical Architecture, Drivers of Big Data, Emerging Big Data Ecosystem and a New Approach to Analytics, Key Roles for the New Big Data Ecosystem, Examples of Big Data Analytics.

(8 Hours)

Unit-II: Data Analysis Lifecycle: Lifecycle Overview, Phase 1: Discovery, Phase 2: Data Preparation, Phase 3: Model Planning, Phase 4: Model Building, Phase 5: Communicate Results, Phase 6: Operationalize, Case Study, Global Innovation Network and Analysis (GINA).

(12 Hours)

Unit-III: Basic Data Analytics Methods using R: Introduction to R - R Graphical User Interfaces, Data import and export, Attributes and Data Types, Descriptive Statistics, Exploratory Data Analysis, Statistical Methods for Evaluation: Hypothesis Testing, Difference of Means, Wilcoxon Rank-Sum Test, Type I and Type II Errors, Power and Sample Size, ANOVA.

(8 Hours)

Unit-IV: Clustering and Association Rules: Overview of Clustering, K-means: Use Cases, Overview of the Method, Determining the Number of Clusters, Diagnostics, Association Rules overview, Apriori Algorithm, Evaluation of Candidate Rules, Applications of Association Rules, Examples, Validation and Testing.

(12 Hours)

Unit-V: Advanced Analytical Theory and Methods: Linear Regression and Logistic Regression with Use Cases and Model Description. Decision Trees, Naïve Bayes: Bayes' Theorem, Naive Bayes Classifier, Smoothing and Diagnostics, Diagnostics of Classifiers.

(8 Hours)



Text Books

S.No	Name of the Books	Name of the Author	Publisher Name	Edition (Pub.Yr.)
1	Data Science and Big Data Analytics	EMC Education Services (Editor)	Wiley	1st (2015)

Reference Books

S.No	Name of the Books	Name of the Author	Publisher Name	Edition (Pub.Yr.)
1	Introduction to Data Science: Data Analysis and Prediction Algorithms with R	Rafael A. Irizarry	Chapman and Hall/CRC	1st (2019) 2
2	Python Data Science Handbook	Jake Vanderplas	O Reilly	2nd (2022)

COURSE PLAN
Unit-I Introduction to Big Data Analytics

S.No	Topics	Recommended Books
1	Data Structures	Book 1,Ch 1
2	Analyst Perspective on Data Repositories	Book 1,Ch 1
3	State of the Practice in Analytics: BI Versus Data Science	Book 1,Ch 1
4	Current Analytical Architecture	Book 1,Ch 1
5	Emerging Big Data Ecosystem	Book 1,Ch 1
6	Key Roles for the New Big Data Ecosystem	Book 1,Ch 1
7	Examples of Big Data Analytics.	Book 1,Ch 1
Unit-II Data Analysis Lifecycle:		
8	Lifecycle Overview	Book 1,Ch 2
9	Phase 1: Discovery	Book 1,Ch 2
10	Phase 2: Data Preparation	Book 1,Ch 2
11	Phase 3:Model Planning	Book 1,Ch 2
12	Phase 4: Model Building	Book 1,Ch 2
13	Phase 5: Communicate Results	Book 1,Ch 2
14	Phase 6: Operationalize	Book 1,Ch 2
15	Case Study:Global Innovation Network and Analysis (GINA).	Book 1,Ch 2
Unit-III Basic Data Analytics Methods using R		
16	Introduction to R	Book 2,Ch 1
17	R Graphical User Interfaces	Book 1,Ch 3
18	Data import and export	Book 1,Ch 3
19	Attributes and Data Types	Book 2,Ch 1
20	Descriptive Statistics	Book 1,Ch 3
21	Exploratory Data Analysis	Book 1,Ch 3
22	Statistical Methods for Evaluation	Book 1,Ch 3- Book 2,Ch 8
23	Hypothesis Testing	Book 1,Ch 3
24	Type I and Type II Error,	Book 1,Ch 3
25	ANOVA	Book 1,Ch 3

Unit-IV Clustering and Association Rules		
26	Overview of Clustering	Book 1,Ch 4
27	K-means: Use Cases	Book 3,Ch 5
28	Determining the Number of Clusters, Diagnostics,	Book 1,Ch 4
29	Association Rules overview	Book 1,Ch 5
30	Apriori Algorithm	Book 1,Ch 5
31	Evaluation of Candidate Rules	Book 1,Ch 5
32	Applications of Association Rules	Book 1,Ch 5
33	Validation and Testing.	Book 3,Ch 5
Unit-V Advanced Analytical Theory and Method		
34	Linear Regression and Logistic Regression with Use Cases,	Book 3,Ch 5
35	Model Description. Decision Trees	Book 3,Ch 5
36	Naïve Bayes	Book 2,Ch 4
37	Bayes' Theorem	Book 1,Ch 7
38	Naive Bayes Classifier	Book 1,Ch 7
39	Smoothing and Diagnostics	Book 1,Ch 7
40	Diagnostics of Classifiers	Book 1,Ch 7

ADDITIONAL WEB RESOURCES

1.	Coursera: Foundations: Data, Data, Everywhere Link: https://www.coursera.org/programs/test-program-93zov/learn/foundations-data?source=browse
2.	Coursera: Python for Applied Data Science AI. Link : https://www.coursera.org/learn/python-for-applied-data-science-ai/home/welcome

GRADING AND ASSESSMENT

- **MST 1** : 10 marks
- **MST 2** : 10 marks
- **Assignment**: 20 marks
- **Attendance**: 10 marks
- **Final Examination**: 100 marks

COURSE POLICIES

- **Attendance**: Minimum 75% attendance is mandatory to appear in the final examination of the course.
- **Academic Integrity**: MIET's academic integrity policies apply. Plagiarism will not be tolerated.
- **Late Submissions**: Assignments and projects must be submitted by the specified timelines.

FACULTY INFORMATION

- **Office Hours**
Monday (12:05 PM - 12:55 PM)
Friday (12:05 PM - 12:55 PM)
- **Contact Information**
palvi.cse@mietjammu.in