

SCHEME OF COURSE WORK

Course Details:

Course Title	: DATA MINING		
Course Code	: 22CD1103	L T P C	3 0 0 3
Program:	: B.Tech.		
Specialization	: Data Science		
Semester	: 4th Semester		
Prerequisites	:		
Courses to which it is a prerequisite	:		

Course Outcomes (COs):

1	Interpret the given data statistically.
2	Classify data using various Classification techniques.
3	Apply Association rule mining to find frequent patterns
4	Build models using unsupervised learning techniques.
5	Apply different techniques for Anomaly Detection.

Program Outcomes (POs):

A graduate of data mining Specialization will be able to

1	Graduates will be able to apply the knowledge of Mathematics, Science, Engineering Fundamentals, Principles of Computer Science and Engineering and Data Science to solve complex problems in different domains.
2	Graduates can identify a problem, conduct data analysis experiments to interpret data, and formulate appropriate strategies to evolve solution using the basic principles of science and engineering.
3	Graduates will have the ability to design a component or a product applying all the relevant standards and with realistic constraints, including public health, safety, culture, society and environment.
4	Graduates will be able interpret data, extract meaningful information, and assess findings as per the prevailing research methods.
5	Graduates will be able to choose and apply new tools and innovative methodologies necessary for engineering practice to solve data-driven problems.
6	Graduates will be able to analyze the impact of data analytic systems on individuals, organizations, society and design appropriate solutions related to society health and safety.
7	Graduates will have adaptive thinking and adaptability in relation to environmental context and sustainable development
8	Graduates will be able to have clear understanding of professional thinking and innovation to provide more economical and effective solutions.
9	Graduates will be able to have cross cultural competency exhibited by working as a member or in teams identifying the significance of each individual's contribution.
10	Graduates will be able to have a good working knowledge of communicating in English – communication with engineering community and society
11	Graduates will be able to have good cognitive load management skills and manage data science projects adhering to financially viable options
12	Graduates will engage themselves in independent and continuous learning in the broad context of data

science and other computer Science related domains.

Course Outcome versus Program Outcomes:

COs	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12
CO-1	S	M		M								S
CO-2	S	M		M	M							S
CO-3	S	S	M	S	M						S	S
CO-4	S	S	M	S	S						S	S
CO-5	M	M	M	M	S	S					M	M

S - Strongly correlated, M - Moderately correlated, Blank - No correlation

Assessment Methods:

Assignment / Quiz / Seminar / Case Study / Mid-Test / End
Exam

Teaching-Learning and Evaluation

Week	TOPIC / CONTENTS	Course Outcomes	Sample questions	TEACHING-LEARNING STRATEGY	Assessment Method & Schedule
1	Introduction To Data Mining: Data Mining, Data Mining tasks, Types of Data, Data Quality	CO1	1. Define Data Science 2. Explain Datafication 3. Explain roles of data Scientist	Lecture / PPT /Group Discussion	Assignment-1, Test- 1 Quiz-1
2	Data Processing, Measures of Similarity and Dissimilarity	CO1	1. What is Statistical Inference 2. Define population and Sample 3. Describe Big data	Lecture / PPT	Assignment-1, Test- 1 Quiz-1
3	Exploring Data: Data Set, Summary Statistics, Visualization, OLAP and multidimensional data Analysis.	CO1	1. Describe in detail data science process with a neat diagram 2. Explain role of Data scientist in Data science process	Lecture / PPT	Assignment-1, Test- 1 Quiz-1
4	Classification: Basic Concepts, Decision Trees, and model evaluation: General approach for solving a classification problem, Decision Tree induction	CO2	1. Explain Correlation analysis and correlation coefficient 2. Explain descriptive statistics	Lecture / PPT	Assignment-1, Test- 1 Quiz-1
5	Model overfitting: Due to presence of noise, due to lack of representation samples, Evaluating the performance of classifier	CO2	3.		
6	Rule based classifier, Nearest Neighborhood classifier, Bayesian Classifier	CO2	4.		
7	Association analysis: Problem Definition, Frequent Item-set generation, Rule generation, compact representation of frequent item sets	CO2	1. Explain different categories of data 2. Write a note on languages for data science 3. Explain the methods to clean the data	Lecture / PPT	Assignment-1, Test- 1 Quiz-1
8	FP-Growth Algorithm, Handling categorical and continuous attributes, concept hierarchy	CO2	1. What is Z-scores Normalization 2. Explain Advanced Ranking Techniques	Lecture / PPT	Assignment-1, Test- 1 Quiz-1
9	MID TEST-1				

10	sequential patterns: Problem formulation, sequential pattern discovery, Apriori-like method, Candidate generation, Candidate pruning, support counting	CO2	1. Briefly explain statistical distributions 2. Describe p-test	Lecture / PPT	Assignment-1, Test- 1 Quiz-1
11	Problem formulation, sequential pattern discovery, Apriori	CO2	1. Explain different chart types for visualization 2. What is Exploratory Data Analysis	Lecture / PPT	Assignment-1, Test- 1 Quiz-1
12	Clustering: Overview, Types of Clustering, Types of clusters, K-means, Basic Agglomerative Hierarchical clustering algorithm, DBSCAN	CO4	1. Explain taxonomy of models 2. How are models developed are evaluated	Lecture / PPT	Assignment-2, Test- 2 Quiz-2
13	Cluster evaluation: Supervised and unsupervised cluster evaluation.	CO4	1. Explain Linear Regression with example 2. What is Gradient Descent	Lecture / PPT	Assignment-2, Test- 2 Quiz-2
14	Anomaly detection: Preliminaries: Causes of Anomalies, Approaches of Anomaly Detection, Issues.	CO5	3.		
15	Statistical approaches: Detecting outliers in a univariate Normal Distribution, Outliers in a Multivariate Normal Distribution, A mixture model approach for anomaly Detection, Proximitybased outlier detection, Density-based outlier detection, Clustering-based outlier detection	CO5	1. Explain Classification 2. Describe Logistic Regression with example 3. How are decision Trees constructed	Lecture / PPT	Assignment-2, Test- 2 Quiz-2
16	MID TEST-2				
17	END EXAM				