

SEMESTER S5

DATA MINING

(Common to CS/CD/CM/CA/AM)

Course Code	PECST525	CIE Marks	40
Teaching Hours/Week (L: T:P: R)	3:0:0:0	ESE Marks	60
Credits	5/3	Exam Hours	2 Hrs. 30 Min.
Prerequisites (if any)	None	Course Type	Theory

Course Objectives:

1. To provide a thorough understanding of the key processes and concepts involved in data mining and data warehousing within application domains
2. To enable students to understand the different data pre-processing techniques, fundamentals and advanced concepts of classification, clustering, association rule mining, text mining and web mining, and apply these techniques in real-world scenarios

SYLLABUS

Module No.	Syllabus Description	Contact Hours
1	Data Mining Fundamentals :- Data Mining - concepts and applications, Knowledge Discovery in Database Vs Data mining, Architecture of typical data mining system, Data Mining Functionalities Data warehouse - Differences between Operational Database Systems and Data Warehouses, Multidimensional data model- Warehouse schema, OLAP Operations, Data Warehouse Architecture	8
2	Data Preprocessing :- Data Preprocessing - Need of data preprocessing, Data Cleaning- Missing values, Noisy data, Data Integration and Transformation	9

	Data Reduction - Data cube aggregation, Attribute subset selection, Dimensionality reduction, Numerosity reduction, Discretization and concept hierarchy generation.	
3	<p>Classification And Clustering :-</p> <p>Classification - Introduction, Decision tree construction principle, Information Gain, Gini index, Decision tree construction algorithm - ID3, Neural networks, back propagation, Evaluation measures - accuracy, precision, recall, F1 score</p> <p>Clustering - Introduction to clustering, distance measures, Clustering Paradigms, Partitioning Algorithm - k means, Hierarchical Clustering, DBSCAN</p>	9
4	<p>Association Rule Analysis And Advanced Data Mining :-</p> <p>Association Rule Mining - Concepts, Apriori algorithm, FP Growth Algorithm</p> <p>Web Mining - Web Content Mining, Web Structure Mining- Page Rank, Web Usage Mining- Preprocessing, Data structures, Pattern Discovery, Pattern Analysis</p> <p>Text Mining - Text Data Analysis and information Retrieval, Basic measures for Text retrieval, Text Retrieval methods, Text Indexing Technique</p>	10

Criteria for Evaluation(Evaluate and Analyse): 20 marks

Students must be asked to identify problems involving large datasets and identify the right solution from the concepts already learned. A comparison of the results with a similar approach also need to be performed to assess the Knowledge Level 5.

End Semester Examination Marks (ESE):

In Part A, all questions need to be answered and in Part B, each student can choose any one full question out of two questions

Part A	Part B	Total
<ul style="list-style-type: none">• 2 Questions from each module.• Total of 8 Questions, each carrying 3 marks (8x3 =24 marks)	2 questions will be given from each module, out of which 1 question should be answered. Each question can have a maximum of 3 subdivisions. Each question carries 9 marks. (4x9 = 36 marks)	60

Course Outcomes (COs)

At the end of the course students should be able to:

Course Outcome		Bloom's Knowledge Level (KL)
CO1	Understand the key process of data mining and data warehousing concepts in application domains.	K2
CO2	Apply appropriate pre-processing techniques to convert raw data into suitable format for practical data mining tasks	K3
CO3	Illustrate the use of classification and clustering algorithms in various application domains	K3
CO4	Comprehend the use of association rule mining techniques	K3
CO5	Explain advanced data mining concepts and their applications in emerging domains	K2

Note: K1- Remember, K2- Understand, K3- Apply, K4- Analyse, K5- Evaluate, K6- Create

CO-PO Mapping Table:

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12
CO1	2	2										2
CO2	3	3	3	3	2							2
CO3	3	3	3	3	2							2
CO4	3	3	3	3	2							2
CO5	2	2										2

Text Books				
Sl. No	Title of the Book	Name of the Author/s	Name of the Publisher	Edition and Year
1	Data Mining Concepts and Techniques	Jaiwei Han, Micheline Kamber	Elsevier	3/e, 2006
2	Data Mining: Introductory and Advanced Topics	Dunham M H	Pearson Education	1/e, 2006

Reference Books				
Sl. No	Title of the Book	Name of the Author/s	Name of the Publisher	Edition and Year
1	Introduction to Data Mining	Pang-Ning Tan, Michael Steinbach	Addison Wesley	1/e, 2014
2	Data Mining: Concepts, Models, Methods, and Algorithms	Mehmed Kantardzic	Wiley	2/e, 2019

Video Links (NPTEL, SWAYAM...)	
Module No.	Link ID
1	https://youtu.be/ykZ-_UGcYWg?si=qiqynQyjI1sNNiHE
2	https://youtu.be/NSxEiohAH5o?si=ZIJHMiRvpFcNQNMA
3	https://youtu.be/VsYKqOokgaE?si=rgndBZqpzB29LUGg
4	https://youtu.be/N_whCVtfL9M?si=VPMH9NP4vdAaiuPe