Semester Lecture Plan

Name of the college: Government College of Arts, Science and Commerce, Sanquelim - Goa

Name of Faculty: Shubham Naik Subject: Computer Science Division:

Paper code: CSD102-Data Mining **Program/Course: T.Y.B.Sc.** **Total Lectures: 45**

Academic year: 2024- 2025 **Semester: V**

Course Objectives:

To get an understanding of the general properties of data in large databases Understand a variety of real-world applications that require mining

To introduce the basic concepts of Data Warehouse and Data Mining techniques.

Examine the types of the data to be mined and apply pre-processing methods on raw data using data mining software.

Become familiarized with association analysis, classification and cluster analysis of data objects.

To discover interesting patterns, analyze and estimate the accuracy of popular data mining algorithms using different data sets.

Get introduced to the challenges in mining complex data types.

To develop skills of using data mining software for solving practical data mining problems.

Course Outcome:

Upon completion of the course students should be able to::

Design a DatawarehouseSchema

Use Classification and prediction methods to solve problems

Identify suitable clustering methods for different applications

Student Learning Outcomes
Understand the basic concept of Data mining and it uses. Understand the task involved in data mining and various benefits or advantages to the society
Understand the task involved in data mining and various benefits or advantages to the society
Introduction to knowledge discovery Understand different technologies used in data mining. Introduce the data mining system architecture,
Mining Algorithm with the example of market basket Anaylsis Explain the apriori algorithm and finding itemsets using candidate generation Solved the problem
Understand the task involved in data mining and various benefits or advantages to the society
Describe the types such as temporal, streaming , text , Spatial, web mining and multimedia using the examples

Month	**Lectures**		**No. of lectures allotted**	**Topic, Subtopic to be covered**	**Experiment/Assignment**	**ICT Tools**	**Reference books**	
	From:	**To:**						
Jul	14/07/24	20/07/24	3	Unit 1 Evolution of Database Technology, What is Data Mining, Scope of Data Mining, Task of Data mining, Which Kind of Applications are Targeted-Business Intelligence, Web Search Engines, Common Data Mining Application Domains, Benefits of Data Mining, Data Mining and Society	Assignment	Microsoft PowerPoint Presentation	Jiawei Han, Micheline Kamber, Jian Pei,"Data mining concepts and Techniques" 3rd Edition	
	22/07/24	27/07/24	3	Unit 2 Data Objects and Attribute Types, Data Pre-processing – Data Quality: a Reason to Pre-process the Data, Major Task in Data Pre-processing-Data Cleaning-Missing Data; Noisy Data; Inconsistent Data, Need of Data Integration-Issues in Data Integration,	Assignment	Microsoft PowerPoint Presentation	Jiawei Han, Micheline Kamber, Jian Pei,"Data mining concepts and Techniques" 3rd Edition	

	29/07/24	03/08/24	3	Data Transformation, Need of Data Reduction, Data Visualization Unit 3 Introduction to Data Warehouse, Understanding a Data Warehouse, Data Warehouse Schema- Star, Snowflake, Fact Constellations, Data Warehouse Modeling: OLAP and Data Cube-What is OLAP, what is Data Cube, Data Cube as a Multidimensional Data Model, Dimensions: The Role of Concept Hierarchies,	Assignment	Microsoft PowerPoint Presentation	Jiawei Han, Micheline Kamber, Jian Pei,“Data mining concepts and Techniques” 3rd Edition		
Aug	05/08/24	10/08/24	3	OLAP in Data Warehouse, OLAP Vs OLTP, Types of OLAP, Data Usage in Data Warehouse, Data Warehousing, A Three Tier Data Warehouse Architecture, Data Warehouse Design Process, Data Warehousing to Data Mining, Data Warehouse Applications	Assignment	Microsoft PowerPoint Presentation	Jiawei Han, Micheline Kamber, Jian Pei,“Data mining concepts and Techniques” 3rd Edition		
	12/08/02	17/08/24	3	Unit 4 Data Mining and Knowledge Discovery, What kind of Data can be Mined, Technologies used in Data Mining-Statistics;	Assignment	Microsoft PowerPoint Presentation	Jiawei Han, Micheline Kamber, Jian Pei,“Data mining concepts and Techniques” 3rd Edition		
	19/08/24	24/08/24	3	Database and Data Warehouse Systems; Information Retrieval; Machine Learning; Pattern Recognition, Data Mining System Architecture, Data Mining Techniques,	Assignment	Microsoft PowerPoint Presentation	Jiawei Han, Micheline Kamber, Jian Pei,“Data mining concepts and Techniques” 3rd Edition		

	26/08/24	31/08/24	3	Issues in Data Mining- Mining Methodology and User Interaction Issues; Performance Issues; Diverse Data Type Issues	Assignment	Microsoft PowerPoint Presentation	Jiawei Han, Micheline Kamber, Jian Pei,“Data mining concepts and Techniques” 3rd Edition		
Sept	02/09/24	07/09/24	3	Unit 5 Introduction to Association Analysis, Frequent Patterns, Market Basket Analysis, Association Rule Mining- Problem Definition, Important Concepts;	Assignment	Microsoft PowerPoint Presentation	Jiawei Han, Micheline Kamber, Jian Pei,“Data mining concepts and Techniques” 3rd Edition		
	09/09/24	14/09/24	3	The Apriori Algorithm: Finding Frequent Itemsets Using Candidate Generation; Pseudocode for Apriori; Example of Apriori for Generating Frequent Itemsets ;	Assignment	Microsoft PowerPoint Presentation	Jiawei Han, Micheline Kamber, Jian Pei,“Data mining concepts and Techniques” 3rd Edition		
	19/09/24	21/09/24	3	Example of Apriori for Generating Association Rules, Mining Multilevel Association Rules, Mining Multidimensional Association Rules, Other Applications of Association Rule Mining	Assignment	Microsoft PowerPoint Presentation	Jiawei Han, Micheline Kamber, Jian Pei,“Data mining concepts and Techniques” 3rd Edition		

	23/09/24	28/09/24	3	Unit 6 Introduction, Classification and Prediction Techniques, How Does Classification work, Building the Classifier;	Assignment	Microsoft PowerPoint Presentation	Jiawei Han, Micheline Kamber, Jian Pei,“Data mining concepts and Techniques” 3rd Edition		
Oct	30/09/24	05/10/24	3	Using Classifier for Classification, General Approach to Classification, Classification and Prediction Issues.,Classifier Accuracy-Confusion Matrix: Accuracy; Recall; Precision; F-Measure, Type-I, Type-II errors.	Assignment	Microsoft PowerPoint Presentation	Jiawei Han, Micheline Kamber, Jian Pei,“Data mining concepts and Techniques” 3rd Edition		
	07/10/24	12/10/24	3	**Unit 7** **Introduction to Cluster Analysis, What is Clustering, Clustering Applications, Requirements for Clustering Algorithms,**	Assignment	Microsoft PowerPoint Presentation	Jiawei Han, Micheline Kamber, Jian Pei,“Data mining concepts and Techniques” 3rd Edition		
	14/10/24	19/10/24	3	Major Clustering Methods- Partitioning Methods, Hierarchical Methods, Density Based Methods	Assignment	Microsoft PowerPoint Presentation	Jiawei Han, Micheline Kamber, Jian Pei,“Data mining concepts and Techniques” 3rd Edition		

	21/10/24	26/10/24	3	unit 8 Introduction, Mining Complex Data Types-Temporal Data Mining, Streaming Data Mining, Spatial Data Mining, Text Mining and Multimedia Data Mining, Web Mining-Categories of Web Mining	Assignment	Microsoft PowerPoint Presentation	Jiawei Han, Micheline Kamber, Jian Pei,“Data mining concepts and Techniques” 3rd Edition		
Month	**Practicals**		**No. of practicals allotted**	**Experiments/Topics to be covered**	**Experiment/Assignment**	**ICT Tools**	**Reference books**		
	From:	**To:**							
Jul	14/07/24	20/07/24	1	Create employee.arff dataset with attributes gender, age, salary, performance in the following format. Also create employee.csv dataset. View both the files.	Practicals	Microsoft PowerPoint Presentation	Jiawei Han, Micheline Kamber, Jian Pei,“Data mining concepts and Techniques” 3rd Edition		
	22/07/24	27/07/24	1	pply preprocess on employee.arff and employee.csv dataset.	Practicals	Microsoft PowerPoint Presentation	Jiawei Han, Micheline Kamber, Jian Pei,“Data mining concepts and Techniques” 3rd Edition		

	29/07/24	03/08/24	1	Carry out filter operation on employee dataset to do the following-1)Add attributes qualification of nominal type and experience of numeric type 2) Remove attribute confirmed 3) Normalize the dataset. Perform preprocess on the modified dataset.	Practicals	Microsoft PowerPoint Presentation	Jiawei Han, Micheline Kamber, Jian Pei,“Data mining concepts and Techniques” 3rd Edition		
Aug	05/08/24	10/08/24	1	Create studentbuyspc.arff dataset in the following format. Also create studentbuyspc.csv dataset. Perform preprocess on the datasets. (Add 20 more record instances of your choice and get the results).	Practicals	Microsoft PowerPoint Presentation	Jiawei Han, Micheline Kamber, Jian Pei,“Data mining concepts and Techniques” 3rd Edition		
	12/08/02	17/08/24	1	Create PhoneBuyDecision.arff dataset in the following format. Perform preprocess on the dataset. Identify noisy and missing data. (Add 10 more record instances of your choice and get results). Visualize the results.	Practicals	Microsoft PowerPoint Presentation	Jiawei Han, Micheline Kamber, Jian Pei,“Data mining concepts and Techniques” 3rd Edition		
	19/08/24	24/08/24	1	Create the following five different datasets with 25 or more record instances in .arff and .csv file formats. Perform preprocess on these datasets. Analyze the results	Practicals	Microsoft PowerPoint Presentation	Jiawei Han, Micheline Kamber, Jian Pei,“Data mining concepts and Techniques” 3rd Edition		

	26/08/24	31/08/24	1	Consider a real life data mining system of Stock Market/Airline/Retail Industry/Financial Organisation/Telecommunication/Social Media or any other of your choice. Collect the data, Create the dataset and perform preprocess on the dataset. Visualize the results.	Practicals	Microsoft PowerPoint Presentation	Jiawei Han, Micheline Kamber, Jian Pei,“Data mining concepts and Techniques” 3rd Edition		
Sept	02/09/24	07/09/24	1	Use studentbuyspc.arff dataset created with the attributes age ,income, student, credit-rating, buyspc. Perform classify (using J48 classifier (C4.5 algorithm)) on the dataset and carry out performance evaluation of the classifier	Practicals	Microsoft PowerPoint Presentation	Jiawei Han, Micheline Kamber, Jian Pei,“Data mining concepts and Techniques” 3rd Edition		
	09/09/24	14/09/24	1	Create twodcluster.arff dataset with the following data. Perform Cluster (using SimpleKmeans) on the dataset. Visualize cluster assignments.	Practicals	Microsoft PowerPoint Presentation	Jiawei Han, Micheline Kamber, Jian Pei,“Data mining concepts and Techniques” 3rd Edition		
	19/09/24	21/09/24	1	Create customer.arff dataset in the following format. Perform Preprocess, Classify and Cluster on the dataset.	Practicals	Microsoft PowerPoint Presentation	Jiawei Han, Micheline Kamber, Jian Pei,“Data mining concepts and Techniques” 3rd Edition		

	23/09/24	28/09/24	1	create shopping.arff dataset in the following format. Perform Association Rule Mining with Apriori (Associate) and visualize the results.	Practicals	Microsoft PowerPoint Presentation	Jiawei Han, Micheline Kamber, Jian Pei,“Data mining concepts and Techniques” 3rd Edition		
Oct	30/09/24	05/10/24	1	Convert the following shopping data to the necessary format and carry out Association Rule Mining with Apriori on the dataset. (Add 10 more record instances of your choice and get the results)	Practicals	Microsoft PowerPoint Presentation	Jiawei Han, Micheline Kamber, Jian Pei,“Data mining concepts and Techniques” 3rd Edition		
	07/10/24	12/10/24	1	Create weather.arff dataset using the following relation. Perform preprocess, classify, cluster and associate on the dataset. Convert temperature and humidity attributes to numeric type, make necessary changes to the respective data. Visualize the results	Practicals	Microsoft PowerPoint Presentation	Jiawei Han, Micheline Kamber, Jian Pei,“Data mining concepts and Techniques” 3rd Edition		
	14/10/24	19/10/24	1	Use five created datasets: 1.Patient 2.Recruitment 3.Applicant 4.Project 5.Bank . Perform preprocess, classify, cluster and associate on the datasets. Analyze the results.	Practicals	Microsoft PowerPoint Presentation	Jiawei Han, Micheline Kamber, Jian Pei,“Data mining concepts and Techniques” 3rd Edition		

	21/10/24	26/10/24	1	Consider the following details of a Movie. Make your assumptions and create a dataset movie.arff. Apply preprocess, classify, cluster and associate on the datasets. Analyze the results Year;Length;Title;Subject;Actor;Actress;Director;Popularity;Awards Identify two data sets from UCI machine learning repository or https://sites.google.com/site/labitis462/lab-tutorials or the default datasets available with the software installation or create datasets of your choice. Perform preprocess, classify, cluster and associate on these dataset	Practicals	Microsoft PowerPoint Presentation	Jiawei Han, Micheline Kamber, Jian Pei,“Data mining concepts and Techniques” 3rd Edition		
Assessment Rubrics	Component	Max Marks							
	ISA 1	8							
	ISA 2	7							
	ISA 3	NA							
	Practical	25							
	Project	NA							
	SEE	60							