



SILVER OAK UNIVERSITY

College of Computer Application

Bachelor Computer Application

Subject Name: Data Mining

Subject Code: 2040233341

Semester: V (Model -2)

Prerequisite: Programming Concepts

Objective: This course is designed for a section level integration of data mining. It is about how to discover significant data and therefore separate important patterns from it.

Teaching and Examination Scheme:

Teaching Scheme				Evaluation Scheme				Total Marks
L	T	P	Credits	Internal		External		
				CIE (TH)	CIE (PR)	ESE (TH)	ESE (PR)	
4	0	2	5	40	20	60	30	150

Content:

Unit No.	Course Contents	Teaching Hours	Weightage %
1	Introduction to data mining (DM): <ul style="list-style-type: none"> What is data mining What kind of data is mined? Database data Data Warehouses Transactional data Other kinds of data What kind of patterns can be mined? Class/Concept Description: Characterization and Discrimination Mining Frequent patterns, Association and Co relations Classification and Regression for Predictive Analysis Cluster Analysis Outlier Analysis Are all patterns interesting? Which technologies are used? Statistics Machine Learning Database system and Data Warehouse 	10	18%

2	<p>Data Pre-processing</p> <ul style="list-style-type: none"> ● Data Preprocessing: An Overview <ul style="list-style-type: none"> ○ Data Quality: Why Preprocess the Data? ○ Major task in Data Preprocessing ● Data Cleaning <ul style="list-style-type: none"> ○ Missing Values ○ Noisy Data ○ Data Cleaning as Process <p>Data Integration</p> <ul style="list-style-type: none"> ○ Entity identification problem ○ Redundancy and correlation analysis ○ Tuple Duplication ○ Data Value conflict <ul style="list-style-type: none"> Detection and Resolution <p>Data Reduction</p> <ul style="list-style-type: none"> ○ Overview of data reduction strategies ○ Histogram ○ Sampling ○ Data Cube Aggregation 	12	21%
3	<p>Data Warehouse:</p> <ul style="list-style-type: none"> ● Data Warehouse: Basic Concepts What is a Data Warehouse? ● Difference between Operational Database Systems and Data Warehouse ● Why have a separate Data Warehouse? ● Data Warehousing: A multitiered architecture ● Data Warehouse Models: Enterprise Warehouse, Data Mart, and Virtual Warehouse ● Extraction, Transformation and Loading ● Data Warehouse Modeling: Data cube and OLAP ● Data Cube: Multidimensional Data Model ● Stars, Snowflakes, and Fact ● Constellations: Schemas for Multidimensional Data Models ● Typical OLAP Operations ● Data Warehouse Design and Usage ● A business Analysis Framework for Data Warehouse Design ● Data warehouse Design Process ● Data Warehouse USAGE FOR Information Processing from OLAP to Multidimensional data Mining 	12	21%

4	Mining Frequent Patterns, Association: Basic Concepts and Methods <ul style="list-style-type: none"> • Basic Concepts • Market Basket Analysis: A Motivating Example • Frequent itemset, Closed itemset and Association Rules • Frequent Itemset Mining Methods • Apriori Algorithm: Finding Frequent Itemset by confined candidate generation • Generating Association rules from frequent itemset • Which Patterns are Interesting? -Pattern Evaluation methods • Strong rules are not necessarily interesting From association analysis to correlation analysis 	12	21%
5	Classification and Cluster Analysis: Basic Concepts <ul style="list-style-type: none"> • Basic Concepts of Classification • What is Classification? • General approach to Classification • Decision Tree Induction • Bayes Classification method • Basic Concepts of Clustering • What is Cluster Analysis? • Requirements of Cluster Analysis • Overview of Basic Clustering Methods • Partitioning Methods • K-Means: A centroid based technique 	10	19%

Course Outcome:

Sr. no.	CO statement	Unit No
CO-1	Perform the preprocessing of data and apply mining techniques on it.	1,2
CO-2	Identify the association rules, classification, and clusters in large data sets.	4,5
CO-3	Solve real world problems in business and scientific information using data mining.	3,4,5
CO-4	Use data analysis tools for scientific applications.	3,4,5

CO-5	K-Means: A centroid-based technique and Classification	5
-------------	---	----------

List of Experiments/Tutorials:

1. Demonstration of preprocessing on dataset student.arff
2. Demonstration of preprocessing on dataset labor.arff
3. Demonstration of Association rule process on dataset contactlenses.arff using apriori algorithm
4. Demonstration of Association rule process on dataset test.arff using apriori algorithm
5. Demonstration of classification rule process on dataset student.arff using j48 Algorithm
6. Demonstration of classification rule process on dataset employee.arff using j48 algorithm
7. Demonstration of classification rule process on dataset employee.arff using id3 algorithm
8. Demonstration of classification rule process on dataset employee.arff using naïve bayes algorithm
9. Demonstration of clustering rule process on dataset iris.arff using simple k-means
10. Demonstration of clustering rule process on dataset student.arff using simple k-means

Teaching & Learning Methodology: -

- The course includes a laboratory, where students have an opportunity to build an appreciation for the concepts being taught in lectures.
- Lectures with live practical example using Projector and Computer
- Experiments shall be performed in the laboratory related to course contents

Major Equipment:

- Computer System
- Internet Connectivity
- Projector

Reference Books:

1. J. Han, M. Kamber, “Data Mining Concepts and Techniques”, Morgan Kaufmann
2. M. Kantardzic, “Data mining: Concepts, models, methods and algorithms, John Wiley & Sons Inc.
3. Ning Tan, Vipin Kumar, Michael Steinbach Pang, “Introduction to Data Mining”, Pearson Education
4. Data Warehousing, Reema Thareja Oxford University Press, 2009

Web Resources:

https://www.tutorialspoint.com/data_mining/index.htm
<https://www.tutorialspoint.com/dwh/index.htm>