



GUJARAT TECHNOLOGICAL UNIVERSITY

Syllabus for Master of Computer Applications, 4th Semester
Subject Name: Data Mining and Data Visualization (DMDV)
Subject Code: 4649307

With effective
from academic
year 2018-19

1. Learning Objectives:

- To understand the need for Data Mining and advantages to the business world.
- To get a clear idea of various classes of Data Mining techniques, their need, scenarios (situations) and scope of their applicability.
- To learn the algorithms used for various type of Data Mining problems
- To understand how to explore and communicate data using data visualization techniques

2. Prerequisites: Knowledge of RBMS and OLTP

3. Contents:

Unit	Course Content	Weightage Percentage
Unit I	Introduction : Data Mining An Overview; What is Data Mining; Why Data Mining?, What Kind of Data can be mined?, What Kind of Patterns Can be Mined? Major Issues in Data Mining, Data Objects and Attribute Keys Data Pre-processing: An Overview of Data Cleaning methods, Data Integration, Data Reduction and Data Transformation.	10%
Unit II	Mining Frequent Patterns, Associations, and Correlations Basic Concepts: Market Basket Analysis; Frequent Item sets, Closed Item sets, and Association Rules; Frequent Pattern Mining Methods: Apriori Algorithm: Finding Frequent Itemsets Using Candidate Generation; Generating Association Rules from Frequent Item sets Which Patterns Are Interesting? —Pattern Evaluation Methods (Strong Rules Are Not Necessarily Interesting, From Association Analysis to Correlation Analysis, A Comparison of Pattern Evaluation Measures) Overview: Improving the Efficiency of Apriori	15%



GUJARAT TECHNOLOGICAL UNIVERSITY

Syllabus for Master of Computer Applications, 4th Semester
Subject Name: Data Mining and Data Visualization (DMDV)
Subject Code: 4649307

With effective
from academic
year 2018-19

Unit III	Classification & Prediction Classification : Basic Concepts Decision Tree Induction: Decision Tree Induction, Attribute Selection Measures; Tree Pruning; Scalability and Decision Tree Induction, Visual Mining for Decision tree induction Rule-based Classification: Using IF-THEN Rules for Classification; Rule Extraction from a Decision Trees; An Overview of Rule Induction Using a Sequential Covering Algorithm Bayesian Classification: Bayes' Theorem, Naive Bayesian Classification; Bayesian Belief Networks Model Evaluation and Selection: Metrics for Evaluating Classifier Performance, Holdout Methods and Random Sub sampling, Cross-validation, Bootstrap, Model Selection Using Statistical Tests of Significance, Comparing Classifiers Based on Cost–Benefit and ROC Curves Techniques to Improve Classification Accuracy: Introducing Ensemble Methods, Bagging, Boosting and Adaboost, An Overview of Random Forests, Improving Classification Accuracy of Class-imbalanced data	20%
Unit IV	Cluster Analysis Cluster Analysis (What is Cluster Analysis? Requirement for Cluster Analysis, Overview of Basic Clustering Methods); Distance Measures in Algorithmic Methods Partitioning Methods (K-Means, K-Medoids) Hierarchical Methods: Agglomerative versus Divisive Hierarchical Clustering, An Overview of BIRCH Multiphase Hierarchical Clustering using clustering feature tree, Chameleon Multiphase Hierarchical Clustering using Dynamic Modeling, Probabilistic Hierarchical Clustering, Density-Based Methods (An Overview of DBSCAN: Density-Based Clustering Based on Connected Regions with High Density) Outliers Analysis: What are Outliers? Types of Outliers, Challenges of Outliers Detection.	20%



GUJARAT TECHNOLOGICAL UNIVERSITY

Syllabus for Master of Computer Applications, 4th Semester
Subject Name: Data Mining and Data Visualization (DMDV)
Subject Code: 4649307

With effective
from academic
year 2018-19

Unit V	Introduction to Data Visualization Acquiring Data, Visualizing Data, Simultaneous Acquisition and Visualization, Applications of Data Visualization (Uses in the Public Sector, Business-to-Business and inter-Business Uses, Business-to-Consumer Uses), Control of Presentation, Faster and Better JavaScript Processing, Rise of HTML5, Lowering the Implementation Bar Exploring the Visual Data Spectrum: Charting Primitives (Data Points, Line Charts, PerfChart, Seasonality Chart, Bar Charts, Pie Charts, Area Charts), Exploring advanced Visualizations (Candlestick Charts, Bubble Charts, Surface Charts, Map Charts, Infographics). Making use of HTML5 CANVAS, Integrating SVG	20%
Unit VI	Visualizing Data Programmatically Creating HTML5 CANVAS Charts (HTML5 Canvas Basics, Linear Interpolations, A Simple Column Chart, Adding Animations), Startig with Google Charts (Google Charts API Basics, A Basic Bar Chart, A Basic Pie Chart, Working with Chart Animations)	15%

4. Text Books:

- 1) Han, J., Kamber, M., Pei, J. Data mining concepts and techniques. Morgan Kaufmann, 3rd Edition, 2011
- 2) Jon Raasch, Graham Murray, Vadim Ogievetsky, Joseph Lowery, “JavaScript and jQuery for Data Analysis and Visualization”, WROX

5. Reference Books:

- 1) Field Cady, 'The Data Science Handbook The Data Science Handbook', Wiley Publication ISBN-13: 978-1119092940
- 2) Pieter Adriaans & Dolf Zentinge, “Data Mining”, Addison-Wesley, Pearson (2000)
- 3) Vikram Pudi & P. Radhakrishnan, “Data Mining”, Oxford University Press (2009)
- 4) G. K. Gupta, “Introduction to Data Mining with Case Studies”, EEE, PHI (2006)
- 5) A Julie Steele and Noah Iliinsky, Designing Data Visualizations: Representing Informational Relationships, O’Relly
- 6) Scott Murray, Interactive Data Visualization for Web, O’Relly

6. Chapter wise Coverage from Main Book(s):

Unit#	Book#	Topics
I	1	Chapter 1 (1.1,1.2,1.3,1.4,1.7), Chapter 2-2.1,2.2, Chapter 3- 3.1,3.2,3.3,3.4.1, 3.5
II	1	Chapter 6 (6.1, 6.2.1-6.2.3,6.3)



GUJARAT TECHNOLOGICAL UNIVERSITY

Syllabus for Master of Computer Applications, 4th Semester
Subject Name: Data Mining and Data Visualization (DMDV)
Subject Code: 4649307

With effective
from academic
year 2018-19

III	1	Chapter 8, 9.1
IV	1	Chapter 10.1,10.2,10.3,10.4.1, 12.1
V	2	Chapter 1,3
VI	2	Chapter 9

7. Accomplishments of the student after completing the course:

- Student will understand fundamentals of Data Mining.
- Ability to apply Apriori algorithm for Association Mining
- Ability to apply Decision Tree and Bayesian algorithms for Classification
- Ability to differentiate between Classification & Clustering, and similarly between Supervised Learning & Unsupervised Learning
- Ability to understand Data visualization and visualize data programmatically.



Practical List

PART-I Data Mining Practical List

Tools: Data Mining using 'R' Programming / Python

1. Create your own data set (like customer, weather, agriculture etc.), load it and apply any pre-processing technique and clean the data, show results.
 - a. Clean missing values
 - b. Remove Data
 - c. Data Smoothing using Binning
2. Pre-process and classify any data set like customer, agriculture, weather etc.
3. Implement Decision Tree algorithm by taking appropriate data set and predict the result. Calculate entropy and information gain.
4. Implement Association mining algorithm by taking appropriate data set and find support and confidence. Also show confusion matrix.
5. Implement Naive Bayesian algorithm taking any dataset of your choice and predict the result.
6. Implement K-means clustering algorithm by taking appropriate data set and predict the result.
7. Implement K-medoids clustering algorithm by taking appropriate data set and predict the result.
8. Implement CART algorithm by taking appropriate data set and predict the result.
9. Implement DBSCAN clustering algorithm by taking appropriate data set and predict the result.
10. Implement K-Nearest Neighbor algorithm by taking appropriate data set and predict the result.

Suggested Data set URLs:

- i) www.kdnuggets.com/datasets/
- ii) <http://archive.ics.uci.edu/ml/>

PART-II Data Visualization Practical List

Tools: HTML5(Canvas and SVG tags), D3.js, Google API, Canvas.js

A) Installation

1. Setup Environment for All the Tools

B) Develop the following Program Using HTML5 CANVAS and SVG TAG

- 1) Develop the Different basic Graphical Shapes using HTML5 CANVAS
- 2) Develop the Different Advanced Graphical Shapes using HTML5 CANVAS
- 3) Develop the Different basic Graphical Shapes using HTML5 SVG
- 4) Develop the Different Advanced Graphical Shapes using HTML5 SVG

C) Develop Following Program Using HTML5 and JavaScript



- 5) Develop the simple bar chart using TML5 CANVAS
- 6) Read the data .txt file and draw Data Table
- 7) Read the data .txt file and draw Simple Bar Chart
- 8) Read the data .csv file and draw Data Table
- 9) Read the data .csv file and draw Column Bar Chart
- 10) Read the data XML file and draw Data Table
- 11) Read the data XML file and draw Simple Chart
- 12) Read JSON Data and draw Data Table
- 13) Read JSON Data and draw Simple Chart

D) Develop Following Program Using HTML5 and D3.js and Canvas.js

- 14) Showing the data as a column chart (simple)
- 15) Showing the data as a stacked column chart
- 16) Showing the Data as a column chart for four age group
- 17) Showing the data as a Line chart (single, fewer and multiple lines)
- 18) Showing the data as a Pie Chart (single and multiple pie)
- 19) Showing the data as a Bar Chart (Simple and multiple)

E) Develop Following Program Using HTML5 and Google Charts API and Map API

- 20) Using Google Charts API Basics draw charts like a Bar chart
- 21) Using Google Charts API Basics draw charts like a Line chart
- 22) Using Google Charts API Basics drawPieChart.
- 23) Using Google Charts API Basics draw Donut Chart.
- 24) Using Google Charts API Basics draw Candle Chart.
- 25) Using Google Charts API Basics draw other types of Chart.
- 26) Using Google API read JSON file and create Google Map.

F) Develop the Dashboard using all types of Graphics charts using any one graphical Tools (desirable)

- 27) Building interconnected Dashboard

Note: Some of the practicals from the above practical list may have seemingly similar definitions. For better learning and good practice, it is advised that students do maximum number of practicals. In the practical examination, the definition asked need not have the same wordings as given in the practical list. However, the definitions asked in the exams will be similar to the ones given in the practical list.