

<b>L</b>	<b>P</b>	<b>C</b>
<b>3</b>	<b>3</b>	<b>5</b>

## **20MC203 DATA MINING**

### **Course Description and Objectives:**

This course aims at knowing how to make use of historical data so that high end business decisions can be taken for the growth of an organization. The main objective of this course is to design intelligent machines which can take business decision risks on behalf of the humans using data mining techniques like classification, clustering, outlier detection, and association rule mining.

### **Course Outcomes:**

Upon completion of the course, the student will be able to achieve the following outcomes:

<b>COs</b>	<b>Course Outcomes</b>	<b>POs</b>
<b>1</b>	Learn the basic concepts of database technology and also understand the need for data mining and its functionalities.	<b>1</b>
<b>2</b>	Explore/Analyze the effective and efficient maintenance of Data Warehouses.	<b>2</b>
<b>3</b>	Apply data mining functionalities like Clustering, Classification, Association Analysis to real world data.	<b>1</b>
<b>4</b>	Investigate interesting patterns and association rules from huge volume of data in making classifications and predictions.	<b>4</b>
<b>5</b>	Design and Implement applications on emerging areas like Web Mining, Text Mining, and Spatial Mining.	<b>5</b>

### **Skills:**

- Performing data cleaning for data analytics.
- Using regression techniques for predictive analytics.
- Applying classification and clustering techniques on real time applications.
- Designing a data warehouse for an organization.

### **Activities:**

- Using data mining techniques to detect similarity of diseases among patients in a hospital database.
- Predicting of weather conditions with the help of data mining techniques.
- Classifying the disease category by applying data mining techniques on crop and diseases data set.

## **Syllabus**

### **UNIT – 1**

**9 Hours**

**INTRODUCTORY BASICS OF DATA MINING:** Introduction to data mining, kinds of data that can be mined, kinds of patterns that can be mined, which technologies are used, Applications and major issues of data mining.

### **UNIT – 2**

**9 Hours**

**DATA WAREHOUSE:** Data warehouse basics, Data cube, Schemas for warehouse, Role of concept hierarchies, OLAP operations, Data objects and attributes, Statistical description of data, Measuring dispersion, Data visualization and measuring the data similarity and dissimilarity.

### **UNIT – 3**

**9 Hours**

**DATA PREPROCESSING AND ASSOCIATION MINING:** Data preprocessing, Data cleaning, Data integration, Data transformation, Mining frequent patterns, Frequent Item set mining, Pattern growth approach.

### **UNIT – 4**

**9 Hours**

**CLASSIFICATION:** Classification basics, Decision tree induction, Attribute selection measure, Bayes method, Model evaluation and selection methods. Advanced classification methods: Back propagation, Support vector machines, K-nearest neighbor.

### **UNIT – 5**

**9 Hours**

**CLUSTER ANALYSIS:** Introduction, Partitioning methods, Density-based methods, Evaluation of clustering, Probabilistic model-based clustering, Outliers and Outlier detection methods.

### **List of Experiments:**

1. Explore various commands given in PL/SQL in Oracle 8.0.
2. Execute multi-dimensional data model using SQL queries.
3. Implement various OLAP operations such as slice, dice, roll up, drill up, pivot etc.
4. Implementation of Text Mining on the data warehouse.
5. Explore the correlation-ship analysis between the data set.
6. Evaluate attribute relevance analysis on a weather data warehouse.
7. Evaluate Information Gain of an attribute in the student database.
8. Experiment to predict the class using the Bayesian classification.
9. Find out a weight & bias updating using the Back Propagation Neural Network.
10. To perform various data mining algorithms on the give data base using WEKA.

**Text Book:**

Jiawei Han, Micheline Kamber, “Data Mining: Concepts and Techniques”, 3<sup>rd</sup> Edition, Morgan Kaufmann, 2012.

**Reference Books:**

1. Pang-Ning Tan, Michael Steinbach, Vipin Kumar, “Introduction to Data Mining”, 1<sup>st</sup> Edition (Kindle), 2012.
2. Ralph Kimball, Margy Ross, “The Data Warehouse Toolkit”, 1<sup>st</sup> Edition, John Wiley and Sons Inc., 2002.
3. Ramesh Sharda, Dursun Delen, Efraim Turban, David King, “Business Intelligence”, 2<sup>nd</sup> Edition, Pearson Education, 2011.
4. Berry, Gordon S. Linoff, “Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management”, 3<sup>rd</sup> Edition, John Wiley & Sons Inc publishers, 2011.