MANGALORE UNIVERSITY

DEPARTMENT OF COMPUTER SCIENCE

MASTER OF COMPUTER APPLICATIONS (MCA) PROGRAMME

MCAH501: DATA MINING TECHNIQUES		
Hours/Week: 4 **Credits : 4**		**I.A. Marks: 30** **Exam. Marks: 70**

Course Outcomes:

CO1: Understand data mining principles and techniques: Introduce DM as a cutting edge business intelligence method and acquaint the students with the DM techniques for building competitive advantage through proactive analysis, predictive modelling, and identifying new trends and behaviors. Course Objectives include:

CO2: Building basic terminology.

CO3: Learning how to gather and analyze large sets of data to gain useful business understanding.

CO4: Learning how to produce a quantitative analysis report/memo with the necessary information to make decisions.

CO5: Describing and demonstrating basic data mining algorithms, methods, and tools 6. Identifying business applications of data mining

CO6: Overview of the developing areas - web mining, text mining, and ethical aspects of data mining

	UNIT-I	**12 Hours**

Introduction: Motivations, Data Mining Databases-Relational Databases, Data warehouse, Transactional Databases, Advanced Database systems and advanced Database applications. Data Mining Functionalities-Concept/Class Discrimination; characterizations and Discrimination, Association Analysis, Classification and Prediction, Cluster Analysis, Outlier Analysis and Evolution Analysis. Classifications of Data Mining Systems, Major issues in Data Mining. **Data Pre-processing**: Data Cleaning, Data Integration and Transformation, Data Reduction, Discrimination and Concepts Hierarchy Generation.

Data Warehouse and OLAP technology for data Mining: Definition of data warehouse, A Multidimensional Data Model, Data warehouse architecture, Data warehouse implementation, Further development of data cubes technology, From data warehousing to data Mining.

		UNIT-II	**12 Hours**

Mining Primitives, Languages and Systems Architectures: Data Mining Primitives, Data Mining Query Languages, Designing Graphical User Interfaces Based on Data Mining Query Languages and Architecture of Data Mining systems.

Concept Description: Characterization and Comparison: Concept Description, Data Generalization and Summarization-based Characterization, Analytical Characterization: Analysis of Attributes Relevance, Missing Class comparisons: Discriminating Between Different classes, Mining Descriptive Statistical Measures in Large Databases.

		UNIT-III	**12 Hours**

Mining Association Rules in Large Database: Association Rule Mining, Mining Single-Dimensional Association Rules From Transactional Databases, Mining Multi-Association Rules From Transaction Databases, Mining Multi-dimensional Association Rules from Relational Databases and Data Warehouses, From Association Mining Correlation Analysis, Constraint-Based Association Mining.

		UNIT-IV	**12 Hours**

Classification and Prediction: Definition of Classification, issues regarding classification and Prediction, Classification by decision tree induction, Bayesian Classification, Classification by Back propagation, Classification based on concepts from association rules mining, other classification methods, prediction, classification accuracy. **Cluster Analysis:** Definition of Cluster, Types of data in cluster analysis, A categorization of major cluster Methods, Partitioning methods, Hierarchical methods, Density-Base Methods, Grid-based methods, Model based Methods, Outlier analysis.

REFERENCE BOOKS
1. Jaiawei Han and MichelineKamber, Data Mining Concepts and Techniques,3rd Edition, Morgan Kaufmann/Elsevier Science publisher, Reprint published by Harcourt (INDIA) Private Limited.
2. David L. Olson,DursunDelen, Advanced Data Mining Techniques,Springer publishers.