Name of the Programme: MSc Integrated
Course Code: IMC-711
Title of the Course: Data Warehousing and Data Mining
Number of Credits: 4(4L-0T-0P)
Effective from AY: 2022-23

Prerequisites for the course	Probability and Statistics	
Objectives	Data warehousing and data mining are the essential components of decision support systems for the modern day industry and business. These techniques enable the knowledge worker (analyst, manager, executive) to make better and faster decisions. The objective of this course is to introduce the student to various Data Warehousing and Data Mining concepts and techniques. A database perspective has to be used throughout the course to introduce principles, algorithms, architecture, design and implementation of data mining and data warehousing techniques.	
Content	**Introduction and Background:** Introduction to the multidisciplinary field of data mining. Discussion on the evolution of database technology that has led to the need for data warehousing and data mining. Stress on importance of its application potential. Introduction to the different key words and techniques.	6 hours
	Data Warehousing And OLAP: Insight of data warehouse and on-line analytical processing, AggregationOperations, models for data Warehousing, star schema, fact and dimension tables Conceptualization of data warehouse and multidimensional databases. Life cycle of data warehouse development. Relationship between data warehouse and data mining.	6 hours
	Data Mining Primitives: Data preprocessing including data cleaning, data integration, data transformation. Definition and Specification of a generic data mining task. Description of Data mining query language with few example queries.	12 hours
	Association Analysis: Different methods(algorithms) for mining association rules in transaction based databases. Illustration of confidence and support. Multidimensional and multilevel association rules. Classification of association rules. Discussion on few association rule algorithms e.g. Apriori, frequent pattern growth etc.	12 hours
	Classification and Predictions: Different Classification algorithm, including C4.5, CART etc., use of genie index, decision tree induction, Bayesian classification, neural network technique of back propagation, fuzzy set theory and genetic algorithms.	12 hours
	Clustering: Partition based clustering, Hierarchical clustering, model based clustering for continuous and discrete data. Discussion on scalability of clustering algorithms. Parallel approaches for clustering.	
	Web Mining: Web usage mining, web content mining, web log attributes. Use of web mining in efficient surfing and personalization	
	Mining Complex Type of Data: Data mining issues in object oriented databases, spatial databases and multimedia databases, time series databases, and text databases.	
	Applications of Data Warehousing And Data Mining: Exploration of websites on data warehousing and data mining applications including bibliography databases, Corporate Houses and Research labs.	

Pedagogy	Lectures/ Tutorials/Hands-on assignments/Self-study

References/ Readings	**Main Reading:** 1. Jiawei Han and MichelineKamber, "Data Mining Concepts and Techniques," 1st Edition Indian Reprint 2001, Harcourt India Private Limited, ISBN 1-55860-489-8. 2. Margaret Dunham, "Data Mining: Introductory and Advanced Topics," 1st Edition, 2003, Prentice Hall (Pearson Publication), ISBN 0-13-088892-3. 3. Arun K Pujari, "Data Mining Techniques". University Press, 2001. **Supplementary Reading** 1. T. Mitchell, "Machine Learning", 1997, McGraw Hill. 2. S.M. Weiss and N. Indurkhya, "Predictive Data Mining", 1998, Morgan Kaufmann. 3. M. Jarke, M. Lenzerni, Y. Vassiliou, and P. Vassiladis, "Fundamentals of Data Warehouses", 2000, Springer Verlag, Isbn 3-540-65365-1.
Course Outcomes	1. Understand Data Warehousing And OLAP 2. Understand Data Mining Primitives, Association Analysis 3. Understand Classification and Predictions, Clustering 4. Web Mining, Mining Complex Type of Data and applications of data mining and data warehousing.