ITE2006	Data Mining Techniques	L	T	P	J	C
		3	0	0	4	4
Pre-requisite	ITE1003	**Syllabus version**				
		1.0				

Course Objectives:

- To understand the fundamental data mining methodologies and the ability to formulate and solve problems.
- To comprehend the overall architecture of a data warehouse, methods for data gathering and data pre-processing
- To learn practical, efficient and statistically sound techniques, capable of solving real world issues

Expected Course Outcome:

1) Demonstrate the knowledge of fundamental elements and concepts related to data mining and its applications.

2) Analyse and understand the various data pre-processing techniques and improve the quality of data and efficiency.

3) Understand the concept of knowledge representation and visualization techniques.

4) Use and apply important methods for finding frequent item sets and association rule mining.

5) Understand the concept of data classification methods.

6) Understand the advanced classification techniques.

7) Understand the unsupervised learning techniques and the algorithm used for data clustering.

8) Design and develop a domain specific application which will address the contemporary issues.

Student Learning Outcomes (SLO):	1, 2, 14
[1]	Having an ability to apply knowledge of mathematics, science, and engineering
[2]	Having a clear understanding of the subject related concepts and of contemporary issues
[14]	An ability to design and conduct experiments, as well as to analyze and interpret data

Module:1	Introduction	6 hours

Data Mining – Stages of the Data Mining Process – Data Mining Knowledge Representation - Technologies – Major Issues in Data Mining- Data Warehousing- Multidimensional data – OLAP Vs OLTP

Module:2	Data Preprocessing	6 hours

Data cleaning - Data reduction - Data Integration - Data Transformation – Feature Selection – Dimensionality Reduction- Discretization and generating concept hierarchies

Module:3	Data mining knowledge representation	6 hours
Task relevant data -Interestingness measures - Representing input data and output knowledge - Visualization techniques		
Module:4	Mining Frequent Patterns, Associations and Correlations	6 hours
Market Basket Analysis – Frequent Item Set Mining methods- Apriori algorithm –Generating Association Rules- A Pattern Growth Approach – Association Analysis to Correlation Analysis		
Module:5	Data Mining Algorithms : Classification	6 hours
Basic concepts – Bayesian Classification Methods -Decision Tree Induction – Rule based Classification -Experiments with Weka.		
Module:6	Advanced Classification Methods	6 hours
Bayesian Belief Networks- Classification by Back propagation- Lazy Learners- Genetic Algorithm – Rough Set Approach.		
Module:7	Clustering	6 hours
Basic issues in clustering - Partitioning methods- K-means, K-Medoids - Hierarchical methods: distance-based agglomerative and divisible clustering- Density Based Methods		
Module:8	Contemporary issues:	3 hours
	Total Lecture hours:	45 hours

Text Book(s)	
1.	J. Han and M. Kamber, Data Mining: Concepts and Techniques, Third Edition, Morgan Kaufman, 2013.
Reference Books	
1.	Charu C. Aggarwal, Data Mining: The Textbook, Springer, 2015.
2.	Zaki and Meira, Data Mining and Analysis Fundamental Concepts and Algorithms, 2014
3.	G. K. Gupta, Introduction to Data Mining with Case Studies, Easter Economy Edition, Prentice Hall of India, 2014.

Recommended by Board of Studies	05-03-2016		
Approved by Academic Council	No. 40	Date	18-03-2016