DATA WAREHOUSING AND DATA MINING

VI Semester: CSE | IT

Course Code	Category	Hours / Week			Credits	Maximum Marks		
		L	**T**	**P**	**C**	**CIA**	**SEE**	**Total**
AIT006	**Core**	3	1	-	4	30	70	100
Contact Classes: 45	**Tutorial Classes: 15**	**Practical Classes: Nil**				**Total Classes:60**		

I. COURSE OVERVIEW:

Data mining refers to extracting or mining knowledge from large amounts of data. It emphasizes various techniques and algorithms used to explore, analyze and leverage data and turn it into valu- able and actionable information. It includes data warehousing and data mining functionalities such as analytical processing, descriptive analysis, association mining, classification, clustering and outlier analysis. The techniques are used to tackle data centric applications in various domains such as finan- cial analysis, telecommunication industry, intrusion detection, and complex data mining applications in stream, web, text, spatial and other scientific applications.

II. OBJECTIVES:

The course should enable the students to:

I	The scope and essentiality of data warehousing and mining.
II	The analysis of data, choosing relevant models and algorithms for respective applications.
III	The process and mining of complex data types such as streams, spatial, web and multimedia
IV	The research perspectives towards advances in data mining

III. COURSE OUTCOMES:

After successful completion of the course, students should be able to:

CO 1	**Relate** knowledge discovery in databases (KDD) process with the help of data warehouse fundamentals and data mining functionalities	Understand
CO 2	**Select** appropriate preprocessing techniques on real time data for usage of data mining algorithms	Apply
CO 3	**Apply** Apriority and FP growth methods on transaction data for frequent pattern mining	Apply
CO 4	**Choose** classification or clustering algorithm for building a classification or prediction model.	Apply
CO 5	**Infer** complex data models with respect to multimedia, streams, spatial and web mining	Understand
CO 6	**Examine** data mining algorithms for solving real world problems	Analyze

IV. SYLLABUS:

UNIT-I	DATAWAREHOUSING	Classes: 08

Introduction to Data warehouse, A Multi dimensional data model- Star, Snow flake and Fact constellation schemas, Measures, Concept hierarchy, Data warehouse architecture- A three tier Data warehouse architecture, types of OLAP servers, Data warehouse Implementation, Data Marts, Differences between OLAT and OLTP.

UNIT-II	DATA MINING	Classes: 10

Introduction, What is Data Mining, Definition, Knowledge Discovery in Data (KDD), Kinds of data bases, Data mining functionalities, Classification of data mining systems, Data mining task primitives, Data Preprocessing: Data cleaning, Data integration and transformation, Data reduction, Data discritization and Concept hierarchy.

UNIT-III	ASSOCIATION RULE MINING	Classes: 10

Association Rules: Problem Definition, Frequent item set generation, The APRIORI Principle, support and confidence measures, association rule generation; APRIORI algorithm.

FP-Growth Algorithms, Compact Representation of Frequent item Set-Maximal Frequent item set, closed frequent item set.

UNIT-IV	CLASSIFICATION AND PRIDICTION	Classes: 10

Issues Regarding Classification and Prediction, Classification by Decision Tree Induction, Bayesian Classification, Classification by Back propagation, Classification Based on Concepts from Association Rule Mining, Other Classification Methods, Prediction, Classifier Accuracy.

UNIT-V	CLUSTERING	Classes: 07

Types of data, categorization of major clustering methods, K-means partitioning methods, hierarchical methods, density based methods, grid based methods, model based clustering methods, outlier analysis.
Mining Complex Types of Data: Multi dimensional Analysis and Descriptive Mining of Complex, Data Objects, Mining Spatial Databases, Mining Multimedia Databases, Mining Time-Series and Sequence Data, Mining Text Databases, Mining the World Wide Web.

Text Books:

1. Jiawei Han, Michelin Kamber, "Data Mining-Concepts and Techniques", Morgan Kaufmann Publishers, Elsevier, 2nd Edition, 2006.
2. Alex Berson, Stephen J.Smith, "Data Warehousing Data Mining and OLAP", Tata McGraw-Hill, 2nd Edition, 2007.

Reference Books:

1. Arun K Pujari, "Data Mining techniques", Universities Press,3rd Edition,2005
2. Pualraj Ponnaiah, "Data Warehousing Fundamentals", Wiley, Student Edition, 2004.
3. Ralph Kimball, "The Data Warehouse Life Cycle Toolkit", Wiley, Student Edition, 2006.
4. Vikram Pudi, P Radha Krishna,"Data Mining", Oxford University,1st Edition,2007.

Web References:

1. http://www.anderson.ucla.edu
2. https://www.smartzworld.com
3. http://iiscs.wssu.edu

E-Text Books:

1. https://www.cisco.com/application/pdf/en/us/guest/products/ps2011/c2001/ccmigration_09186a008 02342cf.pdfhttps://www.jntubook.com
2. http://ftp.utcluj.ro/pub/users/cemil/dwdm/dwdm_Intro/0_5311707.pdf.

MOOC Course

https://3ca1513rbm.wordpress.com

Course Home Page: