INSTITUTE OF AERONAUTICAL ENGINEERING
(Autonomous)
Dundigal - 500 043, Hyderabad, Telangana

COURSE CONTENT

DATA MINING AND WAREHOUSING								
V Semester: CSE(DS)								
Course Code	**Category**	**Hours / Week**			**Credits**	**Maximum Marks**		
		L	**T**	**P**	**C**	**CIA**	**SEE**	**Total**
ACDD06	Core	3	0	0	3	40	60	100
Contact Classes: 45	**Tutorial Classes: Nil**	**Practical Classes: Nil**				**Total Classes: 45**		
Prerequisite: There are no prerequisites to take this course.								

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 valuable 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 financial analysis, telecommunication industry, intrusion detection, and complex data mining applications in stream, web, text, spatial and other scientific applications.

II COURSE OBJECTIVES:

The students will try to learn

I. The scope and essentiality of data mining and warehousing.
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:

At the end of the course, students should be able to:

CO1 Relate knowledge discovery in databases (KDD) process with the help of data warehouse fundamentals and data mining functionalities

CO2 Summarize data mining concepts and preprocessing techniques for knowledge discovery

CO3 Apply Apriori and FP growth methods on transaction data for frequent pattern mining.

CO4 Choose classification or clustering algorithm for building a classification or prediction model

CO5 Infer complex data models with respect to multimedia, streams, spatial and web mining.

CO6 Examine data mining algorithms for solving real world problems

IV. COURSE CONTENT:

MODULE-I: DATAWAREHOUSING (08)
Introduction to Data warehouse, Differences between OLAP and OLTP, A Multi-dimensional data model- Star, Snow flake and Fact constellation schemas, Measures, Concept hierarchy, OLAP operations in Multi dimensional data model, Data warehouse architecture- A three tier Data ware house architecture, types of OLAP servers, Data warehouse Implementation, Data warehouse models

MODULE-II: DATA MINING (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 objects. Data Preprocessing: Data cleaning, Data integration and transformation, Data reduction, Data discretization and Concept hierarchy.

MODULE-III: ASSOCIATION RULE MINING (10)
Association Rules: Problem Definition, Frequent item set generation, The APRIORI Principle, support and confidence measures, association rule generation; APRIORI algorithm.

FP-Growth Algorithm, Compact Representation of Frequent item Set-Maximal Frequent item set, close frequent itemset.

MODULE-IV: CLASSIFICATION AND PREDICTION (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

MODULE-V: CLUSTERING (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: Multidimensional 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 WideWeb.

V. TEXT BOOKS:

1. JiaweiHan, Jian Pei and Hanghang Tong "Data Mining-Concepts and Techniques", Morgan Kaufmann Publishers, Elsevier, 4th Edition, 2023.
2. Dr. Jugnesh Kumar "Data Warehouse and Data Mining: Concepts, Techniques, and Real-Life Applications", BPB Publications, 1st Edition, 2024.

VI REFERENCE BOOKS:

1. Jiawei Han, Jian Pei, and Hanghang Tong, "Data Mining: Concepts and Techniques", Elsevier (Morgan Kaufmann Publishers), 4th Edition,2022.
2. Parteek Bhatia, "Data Mining and Data Warehousing: Principles and Practical Techniques", Cambridge University Press, 1st Edition ,2019
3. RalphKimball, "The DataWarehouse Life Cycle Toolkit", Wiley, Student Edition, 2006.
4. VikramPudi, PRadha Krishna,""Data Mining", Oxford University, 1st Edition, 2007.

VII. ELECTRONICS RESOURCES:

1. http://www.anderson.ucla.edu
2. https://www.smartzworld.com
3. http://iiscs.wssu.edu
4. https://www.cisco.com/application/pdf/en/us/guest/products/ps2011/c2001/ccmigration_09186a00802342c f.pdfhttps://www.jntubook.com
5. http://ftp.utcluj.ro/pub/users/cemil/dwdm/dwdm_Intro/0_5311707.pdf.

VIII.MATERIALS ONLINE

1. Course template
2. Tutorial question bank
3. Tech-talk topics

4. Open-ended experiments
5. Definitions and terminology
6. Assignments
7. Model question paper – I
8. Model question paper – II
9. Lecture notes
10. PowerPoint presentation
11. E-Learning Readiness Videos (ELRV)