



INSTITUTE OF AERONAUTICAL ENGINEERING

(Autonomous)

Dundigal - 500 043, Hyderabad, Telangana

COURSE CONTENT

ADVANCED DATA MINING								
VII Semester: CSE(DS)								
Course Code	Category	Hours / Week			Credits	Maximum Marks		
ACDD26	Elective	L	T	P	C	CIA	SEE	Total
		3	0	0	3	40	60	100
Contact Classes: 48	Tutorial Classes: Nil	Practical Classes: Nil			Total Classes: 48			
Prerequisites: Data mining and warehousing								

I. COURSE OVERVIEW:

Advanced Data Mining is the art and science of discovering patterns, relationships, and insights from complex and large-scale datasets. It is a subset of analytics that focuses on extracting meaningful information from data, often using sophisticated algorithms and techniques. According to INFORMS, analytics involves transforming data into insights for better decision-making.

II. COURSE OBJECTIVES:

The students will try to learn:

- I. The fundamental concepts and techniques of data mining provide the foundation for understanding how to extract valuable insights from vast amounts of data.
- II. The principles and methods of data preprocessing and feature selection for mining large datasets
- III. The advanced models and algorithms used in data mining, including supervised, unsupervised, and semi-supervised learning approaches, enable the development of intelligent data-driven systems.
- IV. The ability to discover interesting patterns, analyze results, and evaluate the performance and accuracy of various mining algorithms is crucial for effective decision-making and knowledge discovery.

III. COURSE OUTCOMES:

- CO1 **Understand** data mining concepts and techniques, including preprocessing, feature selection, and learning approaches.
- CO2 **Apply** classification models, including overfitting and under fitting analysis, performance evaluation, rule-based and nearest neighbor classifiers, and ensemble methods for effective decision-making.
- CO3 **Apply** association rule mining concepts and techniques to identify infrequent patterns in large datasets.
- CO4 **Apply** clustering techniques and dimensionality reduction methods to analyze high-dimensional data.
- CO5 **Analyze** dynamic and time-series data to classify evolving data streams using advanced methodologies.
- CO6 **Apply** Natural Language Processing (NLP) techniques for semantic discovery and document summarization.

IV. COURSE CONTENT:

MODULE – I: INTRODUCTION TO DATA MINING (09)

Motivation and significance of data mining, data mining functionalities – clustering, classification, association rule mining, data pre-processing, major issues in data mining.

MODULE – II: ADVANCED CLASSIFICATION (10)

Definition and Basics, Model over-fitting and under-fitting, performance evaluation and comparisons, rule-based classifier – direct and indirect methods, Nearest neighbor classifiers, Ensemble methods.

MODULE – III: ASSOCIATION RULE MINING (10)

Definition and basic concepts, handling categorical attributes, sequential patterns, mining patterns from graphs and sub-graphs, techniques for infrequent pattern mining.

Case Studies: Some commercial Data mining software and functionalities

MODULE – IV: CLUSTER ANALYSIS (09)

Definition and basic concepts, characteristics of clustering algorithms, fuzzy clustering, grid-based clustering, subspace clustering, constraint-based clustering, clustering high dimensional data.

Mining time series data, sequences, and data streams: Basic concepts, trend analysis, similarity search in time series data, sequential pattern mining, methods for stream data processing, frequent pattern mining from data streams, classification of dynamic data streams and evolving data stream concepts

MODULE – V: ADVANCED TOPICS IN SPATIAL AND WEB DATA MINING (10)

Definition and basic concepts, finding spatial associations and co-location patterns, spatial classification, spatial trend analysis, mining webpages, discovering semantics based on natural language processing and document summarization.

V. TEXTBOOKS:

1. J. Han, M. Kamber, and J. Pei, Data Mining: Concepts and Techniques, 3rd ed., Burlington, MA, USA: Morgan Kaufmann, 2011.
2. P.-N. Tan, M. Steinbach, and V. Kumar, Introduction to Data Mining, 2nd ed., Boston, MA, USA: Pearson, 2018.
3. C. C. Aggarwal, Data Mining: The Textbook, 1st ed., Cham, Switzerland: Springer, 2015.

VI. REFERENCE BOOKS:

1. Jiawei Han, Jian Pei, and Hanghang Tong, “Data Mining: Concepts and Techniques”, Elsevier (Morgan Kaufmann Publishers), 4 th Edition, 2022.
2. Parteek Bhatia, “Data Mining and Data Warehousing: Principles and Practical Techniques”, Cambridge University Press, 1st Edition, 2019
3. Ralph Kimball, “The Data Warehouse Life Cycle Toolkit”, Wiley, Student Edition, 2006.
4. Vikram Pudi, P Radha Krishna, “Data Mining”, Oxford University, 1 st Edition, 2007.

VII. ELECTRONIC RESOURCES

1. <http://www.anderson.ucla.edu>
2. <https://www.smartworld.com>
3. <http://iiscs.wssu.edu>

VIII. MATERIALS ONLINE

1. Course template
2. Tutorial question bank
3. Tech talk topics
4. Definitions and terminology
5. Assignments
6. Model question paper – I
7. Model question paper – II
8. Lecture notes
9. PowerPoint presentation
10. E-Learning Readiness Videos (ELRV)