



INSTITUTE OF AERONAUTICAL ENGINEERING

(Autonomous)

Dundigal - 500 043, Hyderabad, Telangana

COURSE CONTENT

WEB MINING								
VII Semester: IT								
Course Code	Category	Hours / Week			Credits	Maximum Marks		
AITD26	Elective	L	T	P	C	CIA	SEE	Total
		3	-	-	3	40	60	100
Contact Classes: 48	Tutorial Classes: Nil	Practical Classes: Nil			Total Classes: 48			
Prerequisite: Probabilistic Modeling and Reasoning, Foundations of Machine Learning								

I. COURSE OVERVIEW:

Web Mining is a multidisciplinary field that involves applying data mining techniques to discover useful patterns and insights from the vast amounts of data available on the web. This course introduces students to the core concepts, methodologies, and tools used in web mining, focusing on three main categories: Web Content Mining, Web Structure Mining, and Web Usage Mining. Students will gain a comprehensive understanding of how web data is collected, processed, and analyzed to extract valuable information that can drive decision-making, improve user experiences, and enhance business strategies.

II. COURSE OBJECTIVES:

The students will try to learn

- I. The relationship between web mining, data mining, and other related fields like information retrieval and machine learning.
- II. How to work with unstructured data from the web.
- III. The ethical, privacy, and legal issues surrounding web mining and data collection.

III. COURSE OUTCOMES:

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

- CO1 Understand key concepts of Web mining to discover useful information from the World-Wide Web and its usage patterns.
- CO2 Compare various methods of web data mining and its applications.
- CO3 Analyze the structure of data on web.
- CO4 Develop skills of using WEKA tool to perform preprocessing, clustering, classification on web data.
- CO5 Demonstrate various pattern discovery and analysis techniques.
- CO6 Apply web mining techniques to real-world applications, such as social media analysis, customer feedback analysis, document categorization, and knowledge.

IV. COURSE CONTENT:

MODULE – I: INTRODUCTION (10)

Introduction: World Wide Web, History of the Web and the Internet, What is Data Mining? What is Web Mining? Introduction to Association Rule Mining, Supervised Learning & Unsupervised Learning. Information Retrieval and Web Search: Basic Concepts of Information Retrieval, Information Retrieval Models, Relevance Feedback, Evaluation Measures, Text and Web Page pre-processing, Inverted Index and Its Compression, Latent Semantic Indexing, Web Search, Meta-Search: Combining Multiple Rankings, Web Spamming.

MODULE – II: INFORMATION INTEGRATION (09)

Information Integration: Introduction to Schema Matching, pre-processing for Schema Matching, Schema - Level Matching, Domain and Instance-Level Matching, Combining Similarities, 1: m Match, Integration of Web Query Interfaces, Constructing a Unified Global Query Interface. Opinion Mining and Sentiment Analysis: The Problem of Opinion Mining, Document Sentiment Classification, Sentence Subjectivity and Sentiment Classification, Opinion Lexicon Expansion, Aspect- Based Opinion Mining, Opinion Search and Retrieval, Opinion Spam Detection.

MODULE – III: TEXT CLASSIFICATION AND SENTIMENT ANALYSIS (10)

Understand how to build text classifiers for tasks like spam detection, topic categorization, and sentiment analysis.

Learn about machine learning algorithms (e.g., Naive Bayes, SVM, neural networks) used for text classification. Explore how sentiment analysis can be performed on textual data to extract opinions, emotions, or sentiments.

MODULE - IV STRUCTURED DATA EXTRACTION (10)

Structured Data Extraction: Wrapper Generation, Preliminaries, Wrapper Induction, Instance-Based Wrapper Learning, Automatic Wrapper Generation: Problems, String Matching and Tree Matching, Building DOM Trees, Extraction Based on a Single List Page, Extraction Based on Multiple Pages.

MODULE –V: WEB USAGE MINING (09)

Web Usage Mining: Data Collection and Pre-Processing, Data Modeling for Web Usage Mining, Discovery and Analysis of Web Usage Patterns, Recommender Systems and Collaborative Filtering, Query Log Mining, Computational Advertising.

V.TEXT BOOKS:

1. Mining the Web, 1st Edition Discovering Knowledge from Hypertext Data SSSAuthor: S Chakrabarti
Mprint: Morgan Kaufmann.
2. Bing Liu, “Web Data Mining Exploring Hyperlinks, Contents and Usage Data”, 2nd Edition, Springer New York, 2011.

VI. REFERENCE BOOKS:

1. Soumen Chakrabarti, “Mining the Web”, Morgan Kaufmann,2002.
2. Gordon S. Linoff, Michael J.A. Berry, Mining the Web: Transforming Customer Data into Customer Value, John Wiley & Sons.
3. Zdravko Markov, Daniel T. Larose, “Data Mining the Web: Uncovering Patterns in Web Content, Structure, and Usage”, Wiley & Sons, 2007.

VII. WEB REFERENCES:

1. <https://www.scaler.com/topics/data-mining-tutorial/web-mining/>
2. <https://www.geeksforgeeks.org/machine-learning/web-mining/>

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)