21CYS04	WEB MINING	L	T	P	C
		3	0	0	3

Course Objectives

- To describe web mining and understand the need for web mining
- To differentiate between Web mining and data mining
- To understand the different application areas for web mining
- To understand the different methods to introduce structure to web-based data
- To describe Web mining, its objectives, and its benefits
- To understand the methods of Web usage mining

UNIT I	INTRODUCTION TO WEB DATA MINING	9 Hours

Introduction to Web Data Mining and Data Mining Foundations, Introduction – World Wide Web (WWW), A Brief History of the Web and the Internet, Web Data Mining-Data Mining, Web Mining. Data Mining Foundations :Association Rules and Sequential Patterns – Basic Concepts of Association Rules, Apriori Algorithm- Frequent Itemset Generation, Association Rule Generation, Data Formats for Association Rule Mining, Mining with multiple minimum supports – Extended Model, Mining Algorithm, Rule Generation, Mining Class Association Rules, Basic Concepts of Sequential Patterns, Mining Sequential Patterns on GSP, Mining Sequential Patterns on PrefixSpan, Generating Rules from Sequential patterns.

UNIT II	SUPERVISED AND UNSUPERVISED LEARNING	9 Hours

Supervised Learning – Basic Concepts, Decision Tree Induction – Learning Algorithm, Impurity Function, Handling of Continuous Attributes, Classifier Evaluation, Rule Induction – Sequential Covering, Rule Learning, Classification Based on Associations, Naïve Bayesian Classification , Naïve Bayesian Text Classification – Probabilistic Framework, Naïve Bayesian Model . Unsupervised Learning – Basic Concepts, K-means Clustering – K-means Algorithm, Representation of Clusters, Hierarchical Clustering – Single link method, Complete link Method, Average link method, Strength and Weakness.

UNIT III	INFORMATION RETRIEVAL AND WEB SEARCH	9 Hours

Basic Concepts of Information Retrieval, Information Retrieval Methods – Boolean Model, Vector Space Model and Statistical Language Model, Relevance Feedback, Evaluation Measures, Text and Web Page Preprocessing – Stopword Removal, Stemming, Web Page Preprocessing, Duplicate Detection, Inverted Index and Its Compression – Inverted Index, Search using Inverted Index, Index Construction, Index Compression, Latent Semantic Indexing – Singular Value Decomposition, Query and Retrieval, Web Search, Meta Search, Web Spamming.

UNIT IV	LINK ANALYSIS AND WEB CRAWLING	9 Hours

Link Analysis – Social Network Analysis, Co-Citation and Bibliographic Coupling, Page Rank Algorithm, HITS Algorithm, Community Discovery-Problem Definition, Bipartite Core Communities, Maximum Flow Communities, Email Communities. Web Crawling – A Basic

Crawler Algorithm- Breadth First Crawlers, Preferential Crawlers, Implementation Issues – Fetching, Parsing, Stopword Removal, Link Extraction, Spider Traps, Page Repository, Universal Crawlers, Focused Crawlers, Topical Crawlers, Crawler Ethics and Conflicts.

UNIT V	OPINION MINING AND WEB USAGE MINING	9 Hours

Opinion Mining – Sentiment Classification – Classification based on Sentiment Phrases, Classification Using Text Classification Methods, Feature based Opinion Mining and Summarization – Problem Definition, Object feature extraction, Feature Extraction from Pros and Cons of Format1, Feature Extraction from Reviews of Format 2 and 3, Comparative Sentence and Relation Mining, Opinion Search and Opinion Spam. Web Usage Mining – Data Collection and Preprocessing- Sources and Types of Data, Key Elements of Web usage Data Preprocessing, Data Modeling for Web Usage Mining, Discovery and Analysis of Web usage Patterns -Session and Visitor Analysis, Cluster Analysis and Visitor Segmentation, Association and Correlation Analysis, Analysis of Sequential and Navigation Patterns.

UNIT VI	RECENT TRENDS	

Recent Trends in Web Mining

TOTAL PERIODS: 45

Course Outcomes:

At the end of the course, Students can able to

- Design different application areas for Web mining.
- Understand the need for web mining
- Design and implement the different methods of web-based data
- Understand the methods of Web usage mining

Textbooks:

1. Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data by Bing Liu (Springer Publications)

Reference Books:

1. Data Mining: Concepts and Techniques, Second Edition Jiawei Han, Micheline Kamber (Elsevier Publications)
2. Web Mining:: Applications and Techniques by Anthony Scime
3. Mining the Web: Discovering Knowledge from Hypertext Data by Soumen Chakrabarti