

Pre-Requisite(s): 19CSE305 Machine Learning, 19CSE304 Foundations of data Science, 19CSE352 Business Analytics

Course Objectives

- This course serves as an excellent introduction to web mining using different data mining/ML techniques
- The course enables the students to apply different tools and techniques to discover patterns hidden from innumerable webpages in terms of its structure, usage and content so as to design and build more intelligent applications for site management, automatic personalization, recommendation, and user profiling.

Course Outcomes

CO1: Understand the scope of Web mining, identifying the opportunities and the challenges

CO2: Learn to apply different data mining/ML techniques in web mining

CO3: Learn graph based representation of WWW

CO4: Understand techniques for web crawling for web contents to build useful statistics like page ranking

CO-PO Mapping

PO/ PSO	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2
CO														
CO1	3	2										2	3	2
CO2	2	2	3	1				2					3	2
CO3	2	2	2				2					2	3	2
CO4	3	2	2	1	3			2				2	3	2
CO5	3	2	2	1	1	2							3	2

Syllabus

Unit 1

World Wide Web – Data Mining Vs Web Mining – Data Mining Foundations: Association rules and Sequential Patterns – Machine Learning in Data Mining, Web Mining: Web Structure Mining, Web Content Mining, and Web Usage Mining. Web Structure Mining: Web Graph - Extracting pattern from hyperlinks – Mining Document Structure – PageRank.

Unit 2

Web Content Mining: Text and Web Page Pre-processing – Inverted Indices – Latent Semantic Indexing – Web Spamming – Social Network Analysis – Web Crawlers – Structured Data Extraction – Opinion mining and Sentiment Analysis.

Unit 3

Web usage Mining: Data collection and Pre-processing – Data Modelling – Discovery and Analysis of Web Usage – Recommender System and Collaborative Filtering – Query log mining.

Text Book(s)

Liu B. Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data. Springer Science & Business Media; 2007.

Reference(s)

Markov Z, Larose DT. Data Mining the Web: Uncovering Patterns in Web Content, Structure, and Usage. John Wiley & Sons; 2007.

Evaluation Pattern

Assessment	Internal	External
Periodical 1 (P1)	15	
Periodical 2 (P2)	15	
*Continuous Assessment (CA)	20	
End Semester		50

*CA – Can be Quizzes, Assignment, Projects, and Reports