

## 20CS009 ADVANCED DATA MINING

Hours Per Week :

L	T	P	C
3	-	-	3

Total Hours :

L	T	P	WA/RA	SSH/HSB	CS	SA	S	BS
45	-	-	15	30	-	5	5	-

### Course Description and Objectives:

This course enables the students with the advanced data mining techniques such as semi-supervised learning, web mining, social network mining and time series analysis. This course includes both theoretical and skill components. The theoretical part shares the methodological and algorithmic knowledge. Whereas the skill part encompasses with data analytics aids to solve real world problems.

### Course Outcomes:

Upon completion of this course, student should able to :

- ✓ Comprehend advanced Data Mining paradigms in semi-supervised learning.
- ✓ Analyze the algorithmic constructs in web and sentiment mining.
- ✓ Apply the advanced data mining techniques for crucial decision making.
- ✓ Investigate the extracted patterns for web usage mining.
- ✓ Evaluate the performance of data mining models on various kinds of data.

### SKILLS:

- ✓ *Pre process the given data.*
- ✓ *Find the correlation among the attributes.*
- ✓ *Apply classification, association rule mining and clustering algorithms on data.*
- ✓ Evaluate the classification and clustering methods performance.

## UNIT - I

**SEMI-SUPERVISED LEARNING:** Learning from labeled and unlabeled Examples, EM algorithm with naïve bayesian classification; Co-training; Self-training; Transductive support Vector Machines; Graph-based methods; Learning from positive and unlabeled examples, Applications of PU learning, Theoretical foundation; Building classifiers-two-step approach; Building classifiers; Biased-SVM, Building classifiers: Probability estimation.

## UNIT - II

**WEB MINING:** Text pre-processing, 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, Web Spamming.

## UNIT- III

**SOCIAL NETWORK ANALYSIS AND OPINION, SENTIMENT MINING:** Social Network Analysis, Co-citation and bibliographic coupling, PageRank, HITS, Community discovery, Opinion mining and sentiment analysis, The problem of opinion mining, Document sentiment classification, Sentence subjectivity and sentiment classification, Aspect-based opinion mining.

## UNIT - IV

**WEB USAGE MINING:** Web usage mining - data collection and pre-processing; Data modeling for Web usage Client - Server mining; Discovery and analysis of web usage patterns; Recommender systems and collaborative filtering; Query log mining.

## UNIT -V

**TIME SERIES ANALYSIS:** Time series regression and exploratory data analysis, ARIMA models, Spectral analysis and filtering.

### TEXTBOOKS:

1. Bing Liu, "Web Data Mining", 2<sup>nd</sup> edition, Springer, 2011.
2. Robert H. Shumway and David S. Stoffer, "Time series analysis and its applications with R examples", 4<sup>th</sup> edition, Springer, 2011.

### REFERENCEBOOKS:

1. Jiawei Han, Micheline Kamber and Jian Pei, "Data Mining: Concepts and Techniques", 3<sup>rd</sup> edition, Morgan Kaufmann Publishers, 2011.
2. Bing Liu, "Sentiment analysis and opinion mining", 2<sup>nd</sup> edition, Morgan & Claypool Publishers, 2012,
3. Jure Leskovec, Anand Raja Raman and Jeffrey D Ullman, "Mining of Massive Datasets", 5<sup>th</sup> edition, Stanford University, 2014.