

III Year Sem IIB. Tech CSE	ELECTIVE	L	T	P	C
Code: CS3605	Data Mining	2	2	0	4

UNIT I

Types of data mining problems. The process of data mining Statistical evaluation of big data: Statistical prediction, performance measures, pitfalls in data mining evaluation.

UNIT II

Data preparation: Data models, data transformations, handling of missing data, time dependent data, and textual data. Data reduction: feature selection, principal components, smoothing data, and sampling.

UNIT III

Classification: Issues Regarding Classification, Classification by Decision Tree Induction, Bayesian Classification, and Classification by Back propagation, SVM etc.,

UNIT IV

Prediction: Regression Analysis, Fuzzy logic. Classifier Accuracy. Solution analyses: graphical trend analyses, comparison and evaluation of classifiers. Case studies Future trends: Text mining, Web Mining.

UNIT V

Association Rule Mining, Mining Single & Multi-Dimensional Association Rules from Transactional Databases, Candidate and Frequent Item set generation & pruning, Rule Generation and Pruning.

UNIT VI

Cluster Analysis Introduction: Types of Data in Cluster Analysis, A Categorization of Major Clustering, Methods, Partitioning Methods, K Mean clustering, Density Based Methods etc. Practical sessions using open source software.

Text Books:

- 1) Jiawei Han, MichelineKamber,“Data Mining: Concepts and Techniques”, Morgan Kaufmann Publishers editor, 2006.
- 2) David J. Hand, HeikkiMannila, PadhraicSmyth , “Principles of Data mining” MIT Press

Reference Books

- 1) “Advances in Knowledge Discovery and Data Mining”. AAAI/MIT Press, 1996. Selected papers from conferences and journals, conference tutorials.

Video Reference:

Title	Expert Name	Affiliation	Weblink
Data Mining & Knowledge Discovery	Dr. S. Srinath	IIT Madras	http://nptel.ac.in/courses/106106093/35