

Gautam Buddha University School of Management
Greater Noida
Programme: MBA
Course: Data Mining

Instructor: Sonica Upadhyay
Email: upadhyaysonica@gmail.com
Department: Business Management
Sessions: 30

Year/Sem: 2019-20/ II Sem
Phone: 8860124930
Credit: 2
Each Session: 60 minutes

Introduction: Data Mining in Big Data is to process, structure and derive values from massive blocks of data. As an example: the analysis of structured and unstructured data helps companies to create models that can analyze customer opinions, buying patterns, or company's competitors. Data Mining is the mining, or discovery, of new information in terms of patterns or rules from vast amounts of data. To be useful, data mining must be carried out efficiently on large files and databases.

Aims and Objectives: This course is designed to help the students get the feel to discover structure inside unstructured data, extract meaning from noisy data, discover patterns, and use all this information to better understand trends, pattern, classification and ultimately predict customer behavior, market and competition, so that company uses its own data more meaningfully to better position itself on the new waves.

Pedagogy: Class time will be used to expose various concepts and illustrate their applications to Data Mining. The course is well supported with various examples and cases. The emphasis in the class will be on learning by doing. Success (or failure) in this course will, therefore, to a large extent be determined by the amount of effort that you put in outside class.

Homework may be assigned and is due at the beginning of class on the designated submission date.

Learning Outcomes: By the end of the course, it is expected that the students will be able to

- Understand basic Data Mining concepts.
- Develop skills in structuring and analyzing problems
- Develop skills on Data Mining Tool:- Rapid-Miner

Evaluation Scheme:

Assignments/Cases/ Presentations:	15%
Mid-Term Exam:	15%
End-Sem Exam:	70%

Text Book:

Pang-Ning Tan, Vipin Kumar, Michael Steinbach

Reference Book:

Han and Kamber, Ian H.Witten & Eibe Frank, Gordon S.Linoff & J.A. Berry

Session Plan: (Each session of 60 minutes)

Session No.	Topics to be covered	Readings
1 - 3	Module I : Introduction to Data Mining Introduction, Scope of Data Mining, What is Data Mining, How does Data Mining works, Architecture of Data Mining, Application:- Data Mining tools.	Pang-Ning Tan(TB)
4-7	Module II : Data Preprocessing Overview of Data Preprocessing, Data Cleaning, Data Integration and Transformation, Data Reduction , Discretization and concept	Pang-Ning Tan(TB) & Han and Kamber(RB)
8-11	Module III : Data Mining Techniques Data Mining vs Database management system, Association rule, Classification , Clustering	Pang-Ning Tan(TB) & Gordon S.Linoff & J.A. Berry(RB)
12-15	Module IV: Classification Pattern mining, Pattern Based Classification, Pattern Mining Application:- Mining Quality phrases from Text Data	Pang-Ning Tan(TB) & Han and Kamber(RB)
16-20	Module V : Classification Overview , Cluster Analysis, Introduction to similarity measures for cluster analysis	Pang-Ning Tan(TB) & Han and Kamber(RB)
21-24	Types of Clustering Partitioning based Clustering method , Hierarchical Clustering Method, K-Means(Density Based Clustering)	Vipin Kumar(TB) & Han and Kamber(RB)
25-27	Application of Data Mining Business Application using Data Mining, Risk Management and Targeted marketing, Customer Profile and feature construction, Medical Application(Diabetic screening), Scientific Application, other application	Han and Kamber(RB)
27-30	Module VI : Data Mining Lab Tool:- Rapid Miner	

* TB- Text Book, RB- Reference Book