

M.Sc. Computer Science Syllabus First Year (2018-23)

Data Mining and Data Warehousing

Semester II

Subject Code: MS21803

Lectures: 60

Objectives:

The syllabus aims in equipping students with,

- To understand principles behind data analytics.
- To explore and adapt data mining techniques.
- To provide hands on various analytical tools.

Unit 1: Introduction to Data Mining and Data Warehousing

8

Ch 1. Introduction to Data Mining

4

- Basic Data Mining Tasks
- DM versus Knowledge Discovery in Databases
- Data Mining Issues
- Data Mining Metrics
- Social Implications of Data Mining
- Overview of Applications of Data Mining

Ch 2. Introduction to Data Warehousing

4

- Architecture of DW
- OLAP and Data Cubes
- Dimensional Data Modeling-star, snowflake schemas
- Data Preprocessing – Need, Data Cleaning, Data Integration & Transformation, Data Reduction
- Machine Learning
- Pattern Matching

BOS Members :

Dr. Reena Bharati (Subject Expert)

Dr. Manisha Bharambe (Subject Expert)

Dr. Jyoti Yadav (Subject Expert)

Mr. Vishal Salke (Industry Expert)

Ms. Amruta Nambiar (Alumni)

Prof. Ashwini Kulkarni (Chairman)

Prof. Shubhangi Jagtap (Internal Faculty)

Prof. Alka Kalhapure (Internal Faculty)

(Handwritten signatures in blue ink: MB, R, Yadav, Salke, Nambiar, Ashwini, Shubhangi, Alka)



Unit 2: Data Mining Techniques	23
Ch 3. Association Rule Mining	4
<ul style="list-style-type: none"> Frequent item-sets and Association rule mining: Apriori algorithm, Use of sampling for frequent item-set, FP tree algorithm Graph Mining: Frequent sub-graph mining, Tree mining, Sequence Mining 	
Ch 4. Classification & Prediction	14
<ul style="list-style-type: none"> Linear regression Non-linear regression Decision tree learning: <ul style="list-style-type: none"> Construction, performance, attribute selection Issues: Over-fitting, tree pruning methods, missing values, continuous classes Classification and Regression Trees (CART) Bayesian Classification: <ul style="list-style-type: none"> Bayes Theorem, Naïve Bayes classifier 	
Ch 5. Clustering	5
<ul style="list-style-type: none"> k-means Expectation Maximization (EM) algorithm Hierarchical clustering, Correlation clustering 	

BOS Members :

Dr. Reena Bharati (Subject Expert)

Dr. Manisha Bharambe (Subject Expert)

Dr. Jyoti Yadav (Subject Expert)

Mr. Vishal Salke (Industry Expert)

Ms. Amruta Nambiar (Alumni)

Prof. Ashwini Kulkarni (Chairman)

Prof. Shubhangi Jagtap (Internal Faculty)

Prof. Alka Kalhapure (Internal Faculty)

A collection of handwritten signatures in blue ink, corresponding to the BOS members listed on the left. The signatures are: a stylized 'R' for Reena Bharati, 'M.B.' for Manisha Bharambe, 'Jyoti' for Jyoti Yadav, 'Vishal' for Vishal Salke, 'Amruta' for Amruta Nambiar, 'Ashwini' for Ashwini Kulkarni, 'Shubhangi' for Shubhangi Jagtap, and 'Alka' for Alka Kalhapure.



Unit 3:Text Mining and Web Mining	17
Ch 6. Accuracy Measures	5
• Precision, recall, F-measure, confusion matrix, cross-validation, bootstrap	
Ch 7. Text mining	5
• Precision, recall, F-Score	
• Document selection problem, document ranking problem	
• Text Mining Approaches	
• Text Mining Applications	4
Ch 8. Web Mining	
• Web usage mining	
• Web content mining	3
• Page Rank Algorithm	
Ch 9. Brief overview of advanced techniques	
• Active learning	
• Reinforcement learning	

***Contact hours – 12 hours**

Reference Books:

1. Han, Elsevier, *Data Mining: Concepts and Techniques*, ISBN:9789380931913/ 9788131205358
2. Margaret H. Dunham, S. Sridhar, *Data Mining – Introductory and Advanced Topics*, Pearson Education
3. G. K. Gupta, *Data Mining with Case Studies*, Publisher: Prentice-Hall of India Pvt.Ltd, 2nd edition, ISBN13 9788120343269.
4. Tom Mitchell, —*Machine Learning*, McGraw-Hill, 1997

BOS Members :

- Dr. Reena Bharati (Subject Expert)
 Dr. Manisha Bharambe (Subject Expert)
 Dr. Jyoti Yadav (Subject Expert)
 Mr. Vishal Salke (Industry Expert)
 Ms. Amruta Nambiar (Alumni)
 Prof. Ashwini Kulkarni (Chairman)
 Prof. Shubhangi Jagtap (Internal Faculty)
 Prof. Alka Kalhapure (Internal Faculty)