Course code	Course Name	L-T-P Credits	Year of Introduction
CS402	**DATA MINING AND WAREHOUSING**	3-0-0-3	2016

Course Objectives:

- To introduce the concepts of data Mining and its applications
- To understand investigation ofe data using practical data mining tools.
- To introduce Association Rules Mining
- To introduce advanced Data Mining techniques

Syllabus:

Data Mining, Applications, Data Mining Models, Data Warehousing and OLAP, Challenges, Tools, Data Mining Principles, Data Preprocessing: Data Preprocessing Concepts, Data Visualization, Data Sets and Their Significance, Classification Models, Multi Resolution Spatial Data Mining, Classifiers, Association Rules Mining, Cluster Analysis, Practical Data Mining Tools, Advanced Data Mining Techniques, Web Mining, Text Mining, CRM Applications and Data Mining, Data warehousing.

Expected Outcome:

The Student will be able to :

  i.   identify the key process of Data mining and Warehousing

 ii.   apply appropriate techniques to convert raw data into suitable format for practical data mining tasks

iii.   analyze and compare various classification algorithms and apply in appropriate domain

 iv.   evaluate the performance of various classification methods using performance metrics

  v.   make use of the concept of association rule mining in real world scenario

 vi.   select appropriate clustering and algorithms for various applications

vii.   extend data mining methods to the new domains of data

Text Books:

1. Dunham M H, "Data Mining: Introductory and Advanced Topics", Pearson Education, New Delhi, 2003.
2. Jaiwei Han and Micheline Kamber, "Data Mining Concepts and Techniques", Elsevier, 2006.

References:

1. M Sudeep Elayidom, "Data Mining and Warehousing", 1st Edition, 2015, Cengage Learning India Pvt. Ltd.
2. Mehmed Kantardzic, "Data Mining Concepts, Methods and Algorithms", John Wiley and Sons, USA, 2003.
3. Pang-Ning Tan and Michael Steinbach, "Introduction to Data Mining", Addison Wesley, 2006.

Module	Contents	Hours	End Sem Exam. Marks
	Course Plan		
I	Data Mining:- Concepts and Applications, Data Mining Stages, Data Mining Models, Data Warehousing (DWH) and On-Line Analytical Processing (OLAP), Need for Data Warehousing, Challenges, Application of Data Mining Principles, OLTP Vs DWH, Applications of DWH	6	15%
II	Data Preprocessing: Data Preprocessing Concepts, Data Cleaning, Data integration and transformation, Data Reduction, Discretization and concept hierarchy.	6	15%
FIRST INTERNAL EXAM			
III	Classification Models: Introduction to Classification and Prediction, Issues regarding classification and prediction, Decision Tree- ID3, C4.5, Naive Bayes Classifier.	6	15%
IV	Rule based classification- 1R. Neural Networks-Back propagation. Support Vector Machines, Lazy Learners-K Nearest Neighbor Classifier. Accuracy and error Measures-evaluation. Prediction:-Linear Regression and Non-Linear Regression.	6	15%
SECOND INTERNAL EXAM			
V	Association Rules Mining: Concepts, Apriori and FP-Growth Algorithm. Cluster Analysis: Introduction, Concepts, Types of data in cluster analysis, Categorization of clustering methods. Partitioning method: K-Means and K-Medoid Clustering.	8	20
VI	Hierarchical Clustering method: BIRCH. Density-Based Clustering –DBSCAN and OPTICS. Advanced Data Mining Techniques: Introduction, Web Mining- Web Content Mining, Web Structure Mining, Web Usage Mining. Text Mining. Graph mining:- Apriori based approach for mining frequent subgraphs. Social Network Analysis:- characteristics of social networks. Link mining:- Tasks and challenges.	8	20
END SEMESTER EXAMINATION			

Question Paper Pattern

1. There will be *FOUR* parts in the question paper – **A, B, C, D**
2. **Part A**
   a. **Total marks : 40**
   b. *TEN* questions, each have **4 marks**, covering **all the SIX modules** (*THREE* questions from **modules I & II**; *THREE* questions from **modules III & IV**; *FOUR* questions from **modules V & VI**).
   All the TEN questions have to be answered.
3. **Part B**
   a. **Total marks : 18**
   b. *THREE* questions, each having **9 marks**. One question is from **module I**; one question is from **module II**; one question *uniformly* covers **modules I & II**.
   c. *Any TWO* questions have to be answered.
   d. Each question can have *maximum THREE* subparts.
4. **Part C**
   a. **Total marks : 18**
   b. *THREE* questions, each having **9 marks**. One question is from **module III**; one question is from **module IV**; one question *uniformly* covers **modules III & IV**.
   c. *Any TWO* questions have to be answered.
   d. Each question can have *maximum THREE* subparts.
5. **Part D**
   a. **Total marks : 24**
   b. *THREE* questions, each having **12 marks**. One question is from **module V**; one question is from **module VI**; one question *uniformly* covers **modules V & VI**.
   c. *Any TWO* questions have to be answered.
   d. Each question can have *maximum THREE* subparts.
6. There will be *AT LEAST* **60%** analytical/numerical questions in all possible combinations of question choices.