****



**MANIPAL INSTITUTE OF TECHNOLOGY**

Manipal University, Manipal

Karnataka -576104

**Course Plan**

**Department : CSE**

**Course Name & code : Data Warehousing and Data Mining CSE433**

**Semester & branch : Seventh, CSE**

**Name of the faculty : Dr. Geetha M**

**Mr. Prakash Kalingrao Aithal**

**No of contact hours/week : 4**

**ASSESSMENT PLAN:**

1. **In Semester Assessments - 50 %**

* Written tests : Two written tests for 20 marks each.

Each is of 1 hour duration

Surprise quizzes : Five assignments of 2 marks each

1. **End Semester Examination - 50 %**

* Written examination of 3 hours duration (Max. Marks: 50 )

|  |  |
| --- | --- |
| **Portions for Assignment** | |
| **Assignment no.** | **Topics** |
| **1** | L1 – L10 |
| **2** | L11 – L20 |
| **3** | L21 – L30 |
| **4** | L31-L40 |
| **5** | L41-L48 |
| **Portions for Sessional Test** | |
| **Test no.** | **Topics** |
| **1** | L1 – L20 |
| **2** | L21 – L40 |

**Course Plan**

|  |  |
| --- | --- |
| **L. No.** | **Topics** |
| **L0** | Introduction to the Course |
| **L1** | Introduction to Data Mining |
| **L2** | Data Preprocessing: Descriptive Data Summarization |
| **L3** | Measuring data similarity and dissimilarity |
| **L4** | Data cleaning and Transformation |
| **L5** | Data Integration |
| **L6** | Data reduction |
| **L7** | Association Rules: Introduction |
| **L8** | Apriori Algorithm for finding frequent item sets |
| **L9** | Improving the Efficiency of Apriori |
| **L 10** | Partition algorithm |
| **L11** | Pincer-search algorithm |
| **L12** | Mining Closed and Max Patterns |
| **L 13** | Pattern Evaluation |
| **L 14** | Classification: Introduction |
| **L 15** | Classification by Decision Tree Induction |
| **L16** | Pruning Technique |
| **L 17** | Classification: Model Evaluation and Selection |
| **L18** | Overview of cluster analysis |
| **L19** | Partitioning methods: k-Means |
| **L 20** | Evaluation of Clustering |
| **L 21** | Introduction to Data Warehousing |
| **L 22** | Multidimensional Data Model |
| **L23** | OLAP operations |
| **L 24** | Data Warehouse Design and Usage |
| **L25** | Data Warehouse implementation |
| **L 26** | Data Generalization by Attribute-Oriented induction |
| **L27** | Other Methods in Pattern Mining: Mining Frequent Itemsets using the Vertical Data Format |
| **L 28** | FP – Growth without generating candidate generation |
| **L29** | Incremental algorithm |
| **L 30** | Border algorithm |
| **L31** | Other Methods in Classification: Bayesian classification |
| **L32** | Rule-Based classification |
| **L33** | Back Propagation |
| **L 34** | Classification by Back Propagation(cont.) |
| **L 35** | Support Vector Machines |
| **L 36** | Lazy learners |
| **L 37** | Prediction |
| **L38** | Techniques to improve Classification Accuracy |
| **L 39** | Other Methods in Clustering: k-Medoids |
| **L40** | Hierarchical methods |
| **L41** | Density-Based methods |
| **L42** | Outlier Analysis |
| **L43** | Introduction to web mining, web structure mining |
| **L44** | Web content mining, web usage mining |
| **L45** | text mining |
| **L46** | Data Cube Computation: Concepts and Methods |
| **L47** | Processing advanced kinds of queries by exploring cube technology |
| **L48** | Multidimensional data analysis in Cube Space |

**References:**

**1.** **Jiawei Han and Micheline Kamber, “Data Mining- Concepts and Techniques”,3rd Edition, Morgan Kaufmann Publishers, 2012.**

2. **Arun K. Pujari, ”Data Mining Techniques”, University press, 2006**

**Submitted by: Dr. Geetha Maiya**

**(Signature of the faculty)**

**Date: 31-07-2015**

**Approved by:**

**(Signature of HOD)**

**Date: 31-07-2015**

**\*\*\*\*\*\*\*\*\***