**Undergraduate Course Description Document**

**Semester: ODD Semester Year: 2023**

**Course Coordinator: Pooja Gupta, VE-418,** [**pooja.gupta@dituniversity.edu.in**](mailto:pooja.gupta@dituniversity.edu.in)

**Course Instructor: Pooja Gupta**

|  |  |
| --- | --- |
| 1. **Department offering the course** | **Computer Science and Engineering** |
| 1. **Course Code** | **CSF346** |
| 1. **Course Title** | **Data Mining and Data Warehousing** |
| 1. **Credits (L:T:P:C)** | **2:0:1:3** |
| 1. **Contact Hours (L:T:P)** | **2:0:2** |
| 1. **Prerequisites (if any)** | **None** |
| 1. **Course Basket** | **Departmental Elective** |

1. **Course Summary**

This course aims to provide a basic understanding of Data ware housing and mining concepts,

implementation of Data Mining algorithms. This course introduces students to enterprise data and the process and technologies to integrate data from a variety of sources.

1. **Course Objectives**

This course will covers the basic concepts of Data Warehouse and Data Mining techniques, Examine the types of the data to be mined and apply pre-processing methods on raw data. It also discover interesting patterns, analyze supervised and unsupervised models and estimate the accuracy of the algorithms

1. **Course Outcomes**

**On successful completion of the course, students will be able to achieve the following:**

* Explain and evaluate the various data mining algorithms
* Discover and measure interesting patterns from different kinds of databases.
* Apply the techniques of clustering, classification, association finding, feature selection and visualization to real world data.

1. **Curriculum Content**

|  |
| --- |
| **Unit 1: OVERVIEW (5L)**  Motivation (for Data Mining), Data Mining-Definition & Functionalities. Data  Warehousing: Overview, Definition, Delivery Process, Difference between Database System and Data Warehouse, Multi-Dimensional Data Model, Data Cubes, Stars, Snow Flakes, Fact Constellations, Concept hierarchy, Process Architecture, 3 Tier Architecture, Data Marting. ROLAP, MOLAP, HOLAP |
| **Unit 2:** **DATA PRE-PROCESSING (5L)**  Data Cleaning: Missing Values, Noisy Data, Binning, Clustering, Regression, Inconsistent Data, Data Integration and Transformation. Data Reduction: Data Cube Aggregation, Dimensionality reduction, Data Compression, Numerosity Reduction, Clustering, Discretization and Concept hierarchy generation |
| **Unit 3:** **CONCEPT DESCRIPTION (5L)**  Definition, Data Generalization, Analytical Characterization, Analysis of attribute relevance, Mining Class comparisons, Statistical measures in large Databases. Measuring Central Tendency, Measuring Dispersion of Data, Apriori Algorithm, Mining Multilevel Association rules from Transaction Databases |
| **Unit- 4: CLASSIFICATION (5L)**  What is Classification, Issues regarding Classification, Decision tree, Bayesian Classification, Classification by Back propagation |
| **Unit- 5:** CLUSTER ANALYSIS **(5L)**  Data types in cluster analysis, Partitioning methods. Hierarchical Clustering- CURE and Chameleon, Density Based Methods-DBSCAN, OPTICS, Grid Based Methods-STING, CLIQUE, Outlier Analysis |

**Text Books:** 1. “Data Mining Concepts and Techniques”, Jiawei Han and Micheline Kamber, Elsevier, Third Edition, 2012.

**Reference Books:** 1. “Data-Mining. Introductory & Advanced Topics”, Margaret H.Dunham, Pearson Education, India, 3rd edition, 2012.

**Teaching and Learning Strategy**

All materials (ppts, assignments, labs, etc.) will be uploaded in MS Teams.

**Evaluation Scheme**

|  |  |
| --- | --- |
| **Evaluation Instrument** | **Weightage** |
| Quizzes | 15 % |
| Graded Lab | 25% |
| Project | 10% |
| Mid Term | 20% |
| End Term | 30% |

**List of Experiments**

|  |  |
| --- | --- |
| **S. No.** | **EXPERIMENT NAME** |
| **1** | Build Data Warehouse and Explore WEKA. List all the categorical (or nominal) attributes and the real-valued attributes separately. What attributes do you think might be crucial in making the credit assessment? |
| **2** | Perform data preprocessing tasks and demonstrate performing association rule mining on data sets. |
| **3** | Demonstrate performing classification on data sets. |
| **4** | Demonstrate performing clustering on data sets. |
| **5** | Demonstrate performing Regression on data sets |
| **6** | Task 1: Credit Risk Assessment. Sample Programs using German Credit Data |
| **7** | Task 2: Sample Programs using Hospital Management System -Simple Project on Data Preprocessing. |

Tools/Software for experiments: Python / R Programming Language

|  |  |  |  |
| --- | --- | --- | --- |
| **FFCBCS Scheme** | | **Assessment** | |
| **Category** | **Marks** | **Category** | **Marks** |
| Continuous Lab Assessment | 25 | Graded Lab1 | 5 |
| Graded Lab 2 | 5 |
| Graded Lab 3 | 5 |
| Graded Lab 4 | 5 |
| Practical File (Graded Lab-5) | 5 |
| Project | 10 | **Corresponding Coursera** Course | 10 |

**Coursera Course 1: Data Analysis Using Python**

<https://www.coursera.org/learn/data-analysis-python/home/week/1>