**Detailed Syllabus**

**Lecture-wise Breakup**

|  |  |  |  |
| --- | --- | --- | --- |
| **Subject Code** | 16B1NCI635 | **Semester Even** | **Semester** Even **Session 2019**  **Month from January to June** |
| **Subject Name** | Data and Web Mining | | |
| **Credits** | 4 | **Contact Hours** | 4 |

|  |  |  |
| --- | --- | --- |
| **Faculty (Names)** | **Coordinator(s)** | Neetu Sardana |
| **Teacher(s) (Alphabetically)** |  |

|  |  |  |
| --- | --- | --- |
| S. No. | **Course Objective** | **Cognitive Level**  **(Bloom’s Taxonomy)** |
| CO1 | Apply the pre-processing techniques to nominal, binary, categorical and ordinal data. | Apply Level  (Level III) |
| CO2 | Design a Data warehouse using star, snowflake and galaxy schema and perform OLAP operations like roll-up, drill-down, slicing and dicing, etc | Apply Level  (Level III) |
| CO3 | Apply a wide range of classification techniques like Naïve-bayes, decision tree, and KNN for the numerous application including fraud detection, target marketing, medical diagnosis, etc. | Apply Level  (Level III) |
| CO4 | Cluster the similar/dissimilar objects using different methods like partitioning, hierarchical and density based clustering. | Create Level  (Level VI) |
| CO5 | Analyze the transactional data for finding frequent and interesting patterns using association rule mining techniques like Apriori and FP-Growth. | Analyse Level  (Level IV) |
| CO6 | Analyze the link structure of web using page rank and HITS algorithms. | Analyse Level  (Level IV) |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Module No.** | **Subtitle of the Module** | **Topics in the module** | **Co’s** | **No. of Lectures for the module** |
| **1.** | Introduction | What Motivated Data Mining? Why Is It Important? What Is Data Mining? Data Mining—On What Kind of Data? Data Mining Functionalities—What Kinds of Patterns Can Be Mined? Are All of the Patterns Interesting? Classification of Data Mining Systems, Data Mining Task Primitives, Integration of a Data Mining System with a Database or Data Warehouse System, Major Issues in Data Mining. | CO1 | 3 |
|  | **Data Warehouse** |  |  |  |
| **2.** | Data Warehouse Concepts | Data Warehouse, Multidimensional Data Model, Data Warehouse Architecture, Data Warehouse Implementation, From Data Warehousing to Data Mining | CO2 | 1 |
| **3.** | Data Pre-processing | Data extraction, Data Cleaning, Data Integration and Transformation, Data Reduction, Loading into Staging area, Post Load Processing | CO2 | 1 |
| **4.** | Dimensional modeling and OLAP Technology | Defining Dimensional model, Granularity of Facts, Additivity of facts, Helper tables, Implementing Many–to-Many Relationship between fact and dimension tables, Implementing changing dimensions, Efficient Methods for Data Cube Computation, Further Development of Data Cube and OLAP Technology, Attribute-Oriented Induction | CO2 | 2 |
|  | **Data Mining** |  |  |  |
| **5.** | Classification Algorithms | Issues Regarding Classification and Prediction, Naive Bayes Classification, Usability and Complexity Analysis of Bayesian algorithm, Nearest Neighbour algorithm, Decision Tree based algorithm , Rule based Algorithm , Performance evaluation of classifiers: Precision recall, F Measure, Sensitivity, Sensibility,; Ensemble based techniques, Chi-square | CO3 | 9 |
| **6.** | Clustering Algorithms | Types of Data in Cluster Analysis, A Categorization of Major Clustering Methods, Partitioning Methods, Hierarchical Methods, Usability and Complexity Analysis of Agglomerative Hierarchical Algorithm, k-means Partitioning Algorithm, Density based clustering,: DBSCAN, BIRCH | CO4 | 6 |
| **7.** | Association Algorithms | Apriori Algorithm, Sampling Algorithm, Partitioning, Using multiple minimum supports , Rough set approach | CO5 | 6 |
|  | **Web Mining** |  |  |  |
| **8** | Searching , crawling and indexing Algorithms | Link Based Search Algorithm, Web Crawling, Indexing, Searching, Zone Indexing, Term-Frequency, Link Analysis Algorithm. | CO6 | 4 |
| **9** | Ranking Algorithms | Page rank, Hits ranking algorithms | CO6 | 3 |
| **10** | Web caching Algorithm | LRV, FIFO, LRU, Random, OPT | CO6 | 3 |
| **11** | Recommendation Algorithms | Collaborative Filtering, Item-to-Item recommendation, Memory Based Recommendation, | CO6 | 3 |
| **Total number of Lectures** | | |  | 41 |

|  |  |
| --- | --- |
| **Recommended Reading material:** Author(s), Title, Edition, Publisher, Year of Publication etc. ( Text books, Reference Books, Journals, Reports, Websites etc. in the IEEE format) | |
|  | Jiawei Han, Micheline Kamber, Data Mining, Morgan Kaufmann Publishers,Elsevier,2005 |
|  | Kimball R. and Ross M ,The Data Warehouse Toolkit”, Wiley |
|  | Pujari, Arun K,Data mining and statistical analysis using SQL, Universities press |
|  | Pang-Ning Tan, Michael Steinbach, Vipin Kumar, Introduction to Data Mining |
|  | Soumen Chakrabarti, Mining the Web:Discovering knowledge from hypertext data”, Morgan Kaufmann, Elsevier |
|  | Alex, Berson,StephenJ.Smith, Data Warehousing, data mining and OLAP , McGraw-Hill,2004 |
|  | InmonW.H.,Building the Data Warehouse ,4th Edition, Wiley |
|  | Anahory S. and Murray D, Data Warehousing in the Real World, Addison-Wesley |
|  | Margaret H. Dunham, Data Mining: Introductory and Advanced Topics, Prentice Hall,2003 |
|  | Mattison R. ,Web Warehousing and Knowledge Management”, Tata McGraw-Hill. |
|  | David Hand, HeikkiMannila and Padhraic Smyth ,Principles of Data Mining,PHI |
|  | Transactions on Database Systems (ACM) |
|  | IEEE Transactions on Knowledge & Data Engineering |
|  | The VLDB Journal The International Journal on Very Large Data Bases |