**Data Warehousing & Mining**

This course is aimed at training candidates for the course on *Data Warehousing & Mining* and aims at building the following key competencies amongst the candidates:

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Program Name** | **B.Tech. in Computer Science and Information Technology** | | | | | | | | | | |
| **Qualification Pack** |  | | | | | | | | | | |
| **Course Name** | **Data Warehousing & Mining** | | | | **Course Code** | | | | **BTCS04DEA3** | | |
| **Version No** | **1.0** | | | |  | | | |  | | |
| **Pre-requisite** | * **Adequate knowledge of statistics and mathematics.** * **Working knowledge on software.** * **Knowledge of writing code in computer programming language.** | | | | | | | | | | |
| **Course Outcome** | * To understand the role of data mining (DM) and data warehousing (DW). * To understand the nature of data and transform as per the need of the analysis. * To compute and analyze data using data mining techniques. * To understand and evaluate various data mining tools. * To understand process of data warehousing and creating data cubes. * To create solutions for the real world problems. | | | | | | | | | | |
| **Total Credits / L:T:P:S** | 4/ 1:1:1:1 | | | | | | | | | | |
| **Teaching &**  **Examination Scheme** | **Teaching Scheme** | | | | | **Examination Scheme** | | | | | |
|  | **L 15** | **T 15** | **P 15** | **S 15** | | CAT 50 | CAP 40 | TEE 50 | | TEP 40 | SA 20 |

## The Course Encompasses

|  |  |  |  |
| --- | --- | --- | --- |
| **Sr. No** | **Module/Units** | **Key Learning Outcomes** | **Instructional Objectives** |
| **1** | **Introduction to Data Mining (DM)**  Theory Duration (hh.mm): 2.00 | The **candidates** should be able to:   * LO1: Understand scope of DM. | The candidates become familiar with the concepts DM |

|  |  |  |  |
| --- | --- | --- | --- |
|  |  | * LO2: Understand use of DM in various data set. * LO3: Understand various techniques of DM |  |
| **2** | **Knowing Data and Data Preprocessing**  Theory Duration (hh.mm): 3.00 | The **Candidates** should be able to:   * LO1: Understand of data and its importance in data set. * LO2: Examine data and checking properties on the available dataset. Verifying the important predictors | The candidates could be able to understand nature of data. Candidates will also be able to use basic statistics to understand characteristics of the data sets. |
|  |  | * LO3: Understand Basis concepts of Skewness, Kurtosis and distribution of data. |  |
|  |  | * LO4: Understand methods for data preprocessing and dimension reduction. |  |
|  |  | * LO 5: Students will be sensitized to the concept of data visualization. |  |
| **3** | **Data Warehousing and Data Cube Technology** | The **candidates** should be able to:   * LO1: Understand scope of data warehousing. * LO2: Understand concept of data cube. | The candidates could be able to understand data warehousing and data  cube. |
|  | Theory Duration (hh.mm): 5.00 |  |
| **4** | **Mining Frequent Patterns, Associations and Correlations** | The **candidates** should be able to: | The candidates can get conversant with mining frequent patterns, |

|  |  |  |  |
| --- | --- | --- | --- |
|  | Theory Duration (hh.mm): 2.00 | * LO1: Understand concept of mining frequent data patterns. * LO2: Understand need of association rule. | association rule and correlations. |
| **5** | **Classification Basics, Cluster Analysis and Neural Networks**  Theory Duration (hh.mm): 3.00 | The **candidates** should be able to:   * LO1: Understand concept of classification through rule based classification and decision tree. | The candidates will be sensitized on the concept of classification cluster analysis. They will be conversant with use of neural networks for data  mining |
|  |  | * LO2: Understand cluster analysis. |  |
|  |  | * LO3: Understand use of neural networks for data mining. |  |

**Module/Unit wise Syllabus Details**

|  |  |  |  |
| --- | --- | --- | --- |
| **Sr. No** | **Module/Units** | **Detailed Topic wise Syllabus** | **References** |
| **1** | **Introduction to Data Mining**  Theory Duration (hh.mm): 02.00 | Introduction of DM, what kind of data to be mined? What kinds of patters can be mined? Various techniques of data mining Overview of statistical perspective of BI, DM and BA. Overview of Data Modeling, Applications of BA and BI – concepts of learning knowledge, Knowledge discovery and Analytical Intelligence. | Chapter 1 |
| **2** | **Knowing Data and Data Preprocessing** | Data objects & attributes types, testing  of normality, Outlier detection and | Chapter 2, 3 and 12 |

|  |  |  |  |
| --- | --- | --- | --- |
|  | Theory Duration (hh.mm): 03.00 | handling, Handling Missing values, Review of Skewness, Kurtosis, data exploration and preliminary steps for data set. Data cleaning, data integration, Data reduction and data discretization. |  |
| **3** | **Data Warehousing and Data Cube Technology**  Theory Duration (hh.mm): 05.00 | Data warehouse: basic concepts, Data warehouse Modeling : Data Cube and OLAP, Data warehouse design and usage, implementation. Data cube technology: Data cube computation- methods. | Chapter 4 and 5 |
| **4** | **Mining Frequent Patterns, Associations and Correlations** | Basic concepts, frequent item set mining methods, Introduction to association rule, Discovering association rules in traditional Databases. | Chapter 6 and 7 |
|  | Theory Duration (hh.mm): 02.00 |  |  |
| **5** | **Classification Basics, Cluster Analysis and Neural Networks**  Theory Duration (hh.mm): 3.00 | Basics of classification methods, classification methods – decision tree, rule based and Bayes classification. Cluster analysis: basic concepts and methods, advanced methods like neural networks and Support vector Machine. | Chapter 8, 9 and 10 |

**Module wise List of Activities/ Experiments/Practical/Tutorials**

|  |  |  |  |
| --- | --- | --- | --- |
| **Sr. No** | **Module/Units** | **Description** | **Equipment required Code** |
| **1** | **Introduction to Data Mining** | * Performing basic statistics operations on dataset. | Experiments will be done on WEKA , XLMiner and/or any |

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | * Creating graphs. | various | charts | and | other Mining. | tool | for | Data |
| **2** | **Knowing Data and Data** | * Finding out extreme values | | | |  | | | |
|  | **Preprocessing** | (outliers).   * Handling outlier values. * Missing values handling. * Reducing dimensions. * Finding predictors for the study. * Preprocessing data sets | | | |
| **3** | **Data Warehousing and Data Cube Technology** | * Creating data cubes | | | |  | | | |
| **4** | **Mining Frequent Patterns, Associations and Correlations** | * Performing association rule * Mining frequent patters | | | |  | | | |
| **5** | **Classification Basics, Cluster Analysis and Neural Networks** | * Decision tree classification. * Cluster analysis of the created data set * Using neural networks and support vector machine for data mining | | | |  | | | |

**Text Books/Reference Books\***

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Sr. No** | **Title of the Book** | **Author** | **Edition / volume** | **Text (T) Reference ®** |
| **1.** | Data Mining – Concepts and Techniques | Jiawei Han,  Micheline Kamber and Jian Pei | Elsevier, Third Edition | **T1** |
| **2** | “Data Mining for Business Intelligence – Concepts, Techniques  and Applications” | Galit Shmueli, Nitin  R. Patel and Peter C.  Bruce | Wiley India, 2009 (Reprint  – 2016) | **R1** |

**List of Unique Equipment required**

|  |  |  |  |
| --- | --- | --- | --- |
| **Sr. No** | **Module/Units** | **Description of Equipment** | **Equipment**  **Code** |
| **1** | **Introduction to Data**  **Mining** | Experiments will be done on WEKA , XLMiner and/or any other tool for Data Mining |  |
| **2** | **Knowing Data and Data**  **Preprocessing** |
| **3** | **Data Warehousing and**  **Data Cube Technology** |
| **4** | **Mining Frequent Patterns,**  **Associations and Correlations** |
| **5** | **Classification Basics, Cluster Analysis and Neural**  **Networks** |  |

**Assessment Matrix**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Sr**  **. N**  **o** | **Module/Unit** | **Learning Outcome ID** | **Le ar nin g Ou tco**  **me** | **Written Test** | **Practical Experiment** | **Lab Experiment** | **Tutorial** | **Project** | **Seminar** | **Presentation** | **Research assignments** | **Case Study analysis** | **Group Discussions** | **Role Play** | **Prototype making** | **Other (Pl explain)** |
| 1 | **Introduction to Data**  **Mining** | LO1 |  | √ |  | √ |  |  |  | √ |  | √ |  |  |  |  |
| LO2 |  | √ |  | √ |  |  |  |  |  |  |  |  |  |  |
| LO3 |  | √ |  | √ |  |  |  |  |  |  | √ |  |  |  |
| 2 | **Knowing Data and Data Preprocessin g** | LO1 |  |  | √ | √ |  |  |  |  |  |  |  |  |  |  |
| LO2 |  |  | √ | √ |  |  |  |  |  |  |  |  |  |  |
| LO3 |  |  | √ | √ |  |  |  |  | √ |  | √ |  |  |  |
| LO4 |  |  | √ | √ |  |  |  |  | √ |  |  |  |  |  |
| LO5 |  |  |  | √ |  |  |  |  |  | √ |  |  |  |  |
| 3 | **Data**  **Warehousing and Data** | LO1 |  | √ | √ | √ |  | √ |  |  | √ |  |  |  |  |  |
| LO2 |  | √ | √ | √ |  | √ |  |  |  |  | √ |  |  |  |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Cube Technology** |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 4 | **Mining Frequent Patterns, Associations and**  **Correlations** | LO1 |  | √ | √ | √ |  |  |  |  | √ | √ | √ |  |  |  |
| LO2 |  | √ | √ | √ |  |  |  |  | √ | √ |  |  |  |  |
| 5 | **Classification Basics, Cluster Analysis and Neural Networks** | LO1 |  | √ | √ | √ |  |  |  |  |  |  |  |  |  |  |
| LO2 |  | √ | √ | √ |  |  |  |  |  |  |  |  |  |  |
| LO3 |  | √ | √ | √ |  |  |  |  |  |  |  |  |  |  |

Internal Theory Assessment –

* Unit Tests – 15 Marks each, 3 number, Best 2
* Assignments – 10 Marks each, 2 number Internal Practical Assessment –
  + Journal Completion – 10 Marks
  + Completion of Experiment / Activities – 10 Marks each, 5 number, Best three Term end Practical –
* Viva Voce on internal practical submission – 10 Marks
* Performance in practical experiment / Activity – 20 Marks
* Presentation in viva/experiment – 10 Marks Skill Assessment –
* Completion of Skill Journal – 5 Marks
* Completion of Activities / Projects during Skill Sessions – 10 Marks
* Viva-voce – 5 Marks

## Weightage of Units for Examination

|  |  |
| --- | --- |
| Unit | % weightage |
| 1 | 20 |
| 2 | 20 |
| 3 | 20 |
| 4 | 20 |
| 5 | 20 |