|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Subject Code**  **20CST371/20CSP376** | **Data Mining** | **L** | **T** | **P** | **S** | **C** |
| Total Contact Hours: 60 Hours | **0** | **0** | **4** | **0** | **2** |
| **Pre-requisites:** | 20CST316 | | | | | |
| **Co-requisites:** | 20CST311, 20CST313, 20CSP314, 20CST315, 20CST316, 20CST371, 20CST372, 20CST373, 20CST374, 20CSR318, 20CSP312 | | | | | |
| **Anti-Requisites:** | 20CST319 | | | | | |

**Course Objectives**

* To understand the data analysis techniques.
* To understand the concepts behind machine learning.

**Course Outcomes**

* Understand the various stages of data mining process and OLAP with its characteristics.
* Classify Supervised and Unsupervised Learning and understand Regression & Classification techniques.
* Analyse regression & ANOVA approaches.
* Analyse supervised and predictive like K -Nearest Neighbors, Regression and Classification Trees etc.
* Develop the concept of big data mining with its characteristics & challenges.

**List of Experiments**

|  |
| --- |
| **UNIT-I** |
| **Experiment-1**  Demonstration of preprocessing on .arff file using student data .arff. |
| **Experiment-2**  To perform the statistical analysis of data. |
| **Experiment-3**  Demonstration of association rule mining using Apriory algorithm on supermarket data. |
| **Experiment-4**  Demonstration of FP Growth algorithm on supermarket data. |

|  |
| --- |
| **UNIT-II** |
| **Experiment-5**  To perform the classification by decision tree induction using WEKA tools. |
| **Experiment-6**  To perform classification using Bayesian classification algorithm using R. |
| **Experiment-7**  To perform the cluster analysis by k-means method using R. |

|  |
| --- |
| **UNIT-III** |
| **Experiment-8**  To perform the hierarchical clustering using R programming. |
| **Experiment-9**  Study of Regression Analysis using R programming. |
| **Experiment-10**  Outlier detection using R programming. |

**Text Book/Reference Book**

**TEXT BOOKS**

1. EMC Education Services, Data Science and Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data. John Wiley & Sons, 2015.
2. Jaiwei Han, Micheline Kamber, “Data Mining Concepts and Techniques”, Elsevier, 2006.
3. Michael Berthold, David J. Hand, Intelligent Data Analysis, Springer, 2007.

**REFERENCE BOOKS**

1. Cathy O’Neil and Rachel Schutt. Doing Data Science, Straight Talk From The Frontline. O’Reilly. 2014.
2. Introduction to Data Mining by Pang-Ning Tan, Michael Steinbach, and Vipin Kumar, 2005.
3. Ian H. Witten and Eibe Frank, Data Mining: Practical Machine Learning Tools and Techniques (Second Edition), Morgan Kaufmann, 2005.