|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Course Code** | 21CSS203T | **Course Name** | FOUNDATIONS OF DATA SCIENCE | **Course Category** | *S* | *Engineering sciences* | L | T | P | C |
| 2 | 0 | 0 | 2 |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Pre-requisite Courses** | | *Nil* | | | **Co-requisite Courses** |  | | | | | | | | | | | | | | | | | | | |
| **Course Offering Department** | | | | *Data Science and Business Systems* | | | | | | | | | | | | | | | | | | | |
|  | | | |  | | | | | | | | | | | | | | | | | | | |
| |  |  |  | | --- | --- | --- | | **Course Learning Rationale (CLR):** | | *The purpose of learning this course is to:* | | **CLR-1 :** | Learn the data science process and the different stages involved | | | **CLR-2 :** | Know the basic concepts of Machine Learning. | | | **CLR-3 :** | Learn about handling data. | | | **CLR-4 :** | Knowthe basic concepts of Distributed data storage | | | **CLR-5 :** | Know the basic concepts of Text mining, data visualization and Creating an interactive dashboard | | | | | | | | | |  | **Program Learning Outcomes (PLO)** | | | | | | | | | | | | | | | | |
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
| Engineering Knowledge | Problem Analysis | Design & Development | Analysis, Design, Research | Modern Tool Usage | Society & Culture | Environment & Sustainability | Ethics | Individual & Team Work | Communication | Project Mgt. & Finance | Life Long Learning | PSO - 1 | PSO - 2 | PSO – 3 |
| **Course Learning Outcomes (CLO):** | | | | *At the end of this course, learners will be able to:* | | | |
| **CO-1 :** | | Learn the basic concepts of data science and data science process. | | | | | | *1* | *1* | *-* | *-* | *-* | *-* | *-* | *-* | *-* | *-* | *-* | *-* | *1* | *-* | *-* |
| **CO-2 :** | | Know the basic concepts of machine learning models and Machine learning types. | | | | | | *1* | *1* | *-* | *-* | *1* | *-* | *-* | *-* | *-* | *-* | *-* | *-* | *1* | *-* | *-* |
| **CO-3 :** | | Learn about handling the data. | | | | | | *2* | *1* | *-* | *-* | *-* | *-* | *-* | *-* | *-* | *-* | *-* | *-* | *1* | *-* | *-* |
| **CO-4 :** | | *Know the basic concepts of Distributed data storage* | | | | | | *2* | *1* | *-* | *-* | *-* | *-* | *-* | *-* | *-* | *-* | *-* | *-* | *1* | *-* | *-* |
| **CO-5 :** | | *Know the basic concepts of Text mining, data visualization and Creating an interactive dashboard* | | | | | | *2* | *1* | *-* | *1* | *-* | *-* | *-* | *-* | *-* | *-* | *-* | *-* | *1* | *-* | *-* |

|  |
| --- |
| Unit-1 Introduction to Data science: Benefits and uses of Data science, Facets of data, The data science process- The big data ecosystem and data science- An introduction to distributed file system. The Data science and process- Defining research goals and creating-Cleansing, integrating, and transforming data- Exploratory data analysis- Build the models- Presenting the findings and building applications, |
| Unit-2 Modeling: Machine Learning- Applications of machine learning in data science - machine learning in the data science process - Tools used in machine learning:Python- The modeling process- Types of machine learning- Supervised learning, unsupervised learning, Semi-supervised learning. |
| Unit-3 Handling data : Problem faced when handling large data-General techniques for handling large volume of data- General programming tips for dealing large data sets- Discussion: Case study 1: Predicting malicious URLs – Case 2:Building a recommender system inside a database |
| Unit-4 : Data storage: Distributing data storage and processing with frameworks- Hadoop-spark- case study: Assessing risk when loaning money. Join the NoSQL movement: Introduction to NoSQL- Case study : Prediction of disease. Rise of graph databases: Introducing connected data and graph database- Introducing Neo4j: a graph database-Connected data example: a recipe recommendation engine. |
| Unit-5 Text mining and analytics: Text mining in the real world- Text mining techniques: Bag of words- Stemming and lemmatization- - Case study: Classifying Reddit posts. Data Visualization to the end user: Data visualization options-Crossfilter, the JavaScript MapReduce library- Creating an interactive dashboard with dc.js – Dashboard development tools.  . |

|  |  |  |
| --- | --- | --- |
| **Learning Resources** | 1. Davy Cielen, Arno Meysman, Mohamed Ali – Introducing Data Science: Big Data, Machine Learning, and, more, using Python tools, ManningPublications,2016 2. Nina Zumel, John Mount,―Practical Data Science with R, ManningPublications,2014 | 1. Jure Leskovec, Anand Rajaraman, Jeffrey D.Ullman, ― Mining of Massive Datasets, Cambridge University Press,2014 2. MarkGardener,―Beginning R TheStatistical Programming Language, JohnWiley & Sons, Inc, 2012 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  |  |  |  |
|  | **Bloom’s**  **Level of Thinking** | **CLA – 1**  **(50%)** | **CLA – 2 (10%)** | **Final exam**  **(40% Weightage)** |
| **Theory** | **Theory** | **Theory** |
| **Level 1** | **Remember** | *50%* | *50%* | *50%* |
| **Level 2** | **Understand** | *50%* | *50%* | *50%* |
| **Level 3** | **Apply** | **-** | **-** | **-** |
| **Level 4** | **Analyze** | ***-*** | ***-*** | ***-*** |
| **Level 5** | **Evaluate** | **-** | **-** | **-** |
| **Level 6** | **Create** | ***-*** | ***-*** | ***-*** |
|  | Total | 100 % | 100 % | 100 % |

|  |  |  |
| --- | --- | --- |
| Course Designers |  |  |
| Experts from Industry | Experts from Higher Technical Institutions | Internal Experts |
| *1. Rijo Jackson Tom, Senior Engineer : Innovation, Augusta Hitech, rijo.tomjackson@augustahitech.com* |  | 1. Dr. A. Shanthini, SRMIST |
|  |  | 2. Dr. G. Vadivu, SRMIST |