|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Walchand College of Engineering, Sangli**  *(Government Aided Autonomous Institute)* | | | | | |
| **AY 2023-24** | | | | | |
| **Course Information** | | | | | |
| **Programme** | | | B.Tech. (Computer Science and Engineering) | | |
| **Class, Semester** | | | Third Year B. Tech., Sem V | | |
| **Course Code** | | |  | | |
| **Course Name** | | | Data Science | | |
| **Desired Requisites:** | | | Probability and Statistics | | |
|  | | | | | |
| |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | | **Teaching Scheme** | | **Examination Scheme (Marks)** | | | | | **Lecture** | 3 Hrs/week | **ISE** | **MSE** | **ESE** | **Total** | | **Tutorial** | - | 20 | 30 | 50 | 100 | | **Practical** | - |  | | | | | **Interaction** | - | **Credits: 3** | | | | | | | | | |
|  | | | | | |
| **Course Objectives** | | | | | |
| **1** | To provide the knowledge and expertise to become a proficient data scientist. | | | | |
| **2** | To critically evaluate data visualizations based on their design and use for communicating. | | | | |
| **3** |  | | | | |
| **Course Outcomes (CO) with Bloom’s Taxonomy Level** | | | | | |
| **CO1** | Acquaint core concepts and technologies in Data Science. | | | Understanding | |
| **CO2** | Demonstrate data collection and management using different technologies. | | | Applying | |
| **CO3** | Analyse and interpret large data sets in the context of real-world problems. | | | Analyzing | |
|  | | | | | |
| **Module** | | **Module Contents** | | | **Hours** |
| I | | **Module 1: Introduction to core concepts and technologies**  Introduction, Terminology, data science process, data science toolkit, Types of data, Example applications | | | 4 |
| II | | **Module 2 Data Collection and Management**  Introduction, Sources of data, Data collection, Exploring and fixing data, Data storage and management, Using multiple data sources. | | | 7 |
| III | | **Module 3 Data Preprocessing**  Data Cleaning, Data Integration, Data Reduction, Data Transformation and Data Discretization. | | | 8 |
| IV | | **Module 4 Data Visualization**  Introduction, Types of data visualization, Data for visualization: Data types, Data encodings, Retinal variables, Mapping variables to encodings, visual encodings. | | | 6 |
| V | | **Module 5 Data Analysis**  Introduction, Terminology and concepts, Introduction to statistics, Central tendencies and distributions, Variance, Distribution properties and arithmetic, Samples/CLT, Correlation, Linear Regression, Least Squares, Residuals, Regression Inference, classification, classifiers. | | | 8 |
| VI | | **Module 6 Recent trends**  Recent trends in various data collection and analysis techniques, various visualization techniques, Case Study, application development methods used in data science. | | | 6 |
|  | | | | | |
| **Text Books** | | | | | |
| 1 | | Adhikari Ani and DeNero John. Computational and Inferential Thinking, The Foundations of Data Science, UC Berkeley. | | | |
| 2 | | Jiawei Han, Micheline Kamber and Jian Pei. Data Mining Concepts and Techniques. Morgan Kaufmann, Third Edition. | | | |
|  | | | | | |
| **References** | | | | | |
| 1 | | O’Neil Cathy and Schutt Rachel. Doing Data Science, Straight Talk From The Frontline. O’Reilly. | | | |
| 2 | | Leskovek Jure, Rajaraman Anand and Ullman Jeffrey. Mining of Massive Datasets. v2.1, Cambridge University Press. | | | |
| 3 | |  | | | |
|  | | | | | |
| **Useful Links** | | | | | |
| 1 | |  | | | |
| 2 | |  | | | |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **CO-PO Mapping** | | | | | | | | | | | | | | | |
|  | **Programme Outcomes (PO)** | | | | | | | | | | | | **PSO** | | |
|  | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 1 | 2 | 3 |
| **CO1** | 3 |  |  |  |  |  |  |  |  |  |  |  | 1 |  |  |
| **CO2** | 1 | 2 |  |  |  |  |  |  |  | 1 |  |  | 1 |  |  |
| **CO3** | 1 | 2 |  |  |  |  |  |  |  | 1 |  |  | 1 |  |  |
| The strength of mapping is to be written as 1,2,3; Where, 1:Low, 2:Medium, 3:High  Each CO of the course must map to at least one PO. | | | | | | | | | | | | | | | |

|  |
| --- |
| **Assessment** |
| Two components of In Semester Evaluation (ISE), One Mid Semester Examination (MSE) and one End Semester Examination (ESE) having 20%, 30% and 50% weights respectively.   |  |  | | --- | --- | | Assessment | Marks | | ISE1 | 10 | | MSE | 30 | | ISE2 | 10 | | ESE | 50 |   ISE 1 and ISE 2 are based on assignment/declared test/quiz/seminar etc.  MSE: Assessment is based on 50% of course content (Normally first three modules)  ESE: Assessment is based on 100% course content with 70-80% weightage for course content (normally last three modules) covered after MSE. |