| Course Code: | UDSCO0511 |
|---|---|
| Course Name: | Data Science for Everyone |

| L | T | P | Credit |
|---|---|---|---|
| 3 | | | 3 |

**Course Prerequsites:**

Basic Knowledge of Computer Science

**Course Description:**

This course deals with the principles of data science

**Course Outcomes:** **After the completion of the course the student will be able to -**

| | |
|---|---|
| **CO1** | Demonstrate the core concept of data science along with mathematical foundations needed to handle the data. |
| **CO2** | Explain Python-based toolkits and machine learning essentials |
| **CO3** | Collect, explore, clean and manipulate data |

**CO-PO Mapping:**

| | PO1 | PO2 | PO3 | PO4 | PO5 | PO6 | PO7 | PO8 | PO9 | PO10 | PO11 | PO12 | PSO1 | PSO2 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **CO1** | 3 | 2 | 2 | 2 | 1 | | | | | | | 1 | 3 | 2 |
| **CO2** | 3 | 2 | 2 | 2 | 3 | | | | | | | 1 | 3 | 2 |
| **CO3** | 3 | 2 | 2 | 2 | 3 | | | | | | 2 | 1 | 3 | 1 |

**Assessment Scheme:**

| SN | Assessment | Weightage | Remark |
|---|---|---|---|
| 1 | In Semester Evaluation 1 (ISE1) | 10% | Assignment, Test, Quiz, Seminar, Presentation, etc. |
| 2 | Mid Semester Examination (MSE) | 30% | 50% of course contents |
| 3 | In Semester Evaluation 2 (ISE2) | 10% | Assignment, Test, Quiz, Seminar, Presentation, etc. |
| 4 | End Semester Examination (ESE) | 50% | 100% course contents |

**Course Contents:**

| UNIT 1 | INTRODUCTION TO DATA SCIENCE: | 6 Hours |
|---|---|---|

The data science Venn diagram, Terminology, Data science case studies, Summary, Types of Data, Flavors of Data: Structured versus unstructured data, Quantitative and qualitative data, The four levels of data: Nominal level, Ordinal level, Interval level, and Ratio level

| UNIT 2 | THE 5 STEPS OF DATA SCIENCE: | 6 Hours |
|---|---|---|

Introduction to data science, Overview of the five steps, Explore the data, obtain the data, model the data, communicate and visualize the results.

| | | |
|---|---|---|
| **UNIT 3** | **CONCEPTS OF DATA SCIENCE:** | **6 Hours** |
| Traits of Big data, Web Scraping, Analysis vs Reporting, Introduction to Programming, Tools for Data Science, Toolkits using Python: Matplotlib, NumPy, Scikit-learn, NLTK | | |
| **UNIT 4** | **MACHINE LEARNING:** | **6 Hours** |
| Overview of Machine learning concepts – Over fitting and train/test splits, Types of Machine learning – Supervised, Unsupervised, Reinforcement learning variables | | |
| **UNIT 5** | **VISUALIZING DATA: BAR CHARTS, LINE CHARTS, SCATTERPLOTS:** | **6 Hours** |
| Working with data: Reading Files, Scraping the Web, Using APIs (Example: Using the Twitter APIs), Cleaning and Munging, Manipulating Data, Rescaling | | |
| **UNIT 6** | **DATA SCIENCE FOR EVERYONE:** | **6 Hours** |
| **CASE STUDIES:**<br>DS in Healthcare, Finance, Bio-Informetics, Robotics, IoT analytics, Weather forecasting, Stockmarket prediction, Object Recognition, Real Time Sentiment Analysis etc. | | |

**Text Books:**


1. Joel Grus, "Data Science from Scratch: First Principles with Python", O'Reilly Media
2. Aurélien Géron, "Hands-On Machine Learning with Scikit-Learn and Tensor Flow: Concepts, Tools, and Techniques to Build Intelligent Systems", 1st Edition, O'Reilly Media
3. Jain V.K., “Data Sciences”, Khanna Publishing House, Delhi


**Reference Books:**


1. Jain V.K., “Big Data and Hadoop”, Khanna Publishing House, Delhi.
2. Jeeva Jose, “Machine Learning”, Khanna Publishing House, Delhi.
3. Chopra Rajiv, “Machine Learning”, Khanna Publishing House, Delhi.
4. Ian Goodfellow, Yoshua Bengio and Aaron Courville, "Deep Learning", MIT Press
5. Jiawei Han and Jian Pei, "Data Mining Concepts and Techniques", Third Edition, Morgan Kaufmann Publishers