DSOE4001 DATA SCIENCE AND ANALYTICS (3-0-0)

Course Objectives :

1. Introduce the fundamental concepts and lifecycle of data science.
2. Develop understanding of data preprocessing, visualization, and exploratory data analysis techniques.
3. Familiarize students with basic statistical and machine learning methods used in analytics.
4. Enable students to interpret analytical results for data-driven decision-making.
5. Create awareness about ethical issues, data privacy, and responsible use of data.

Module I: Introduction to Data Science (9 Hours)

Meaning and Scope of Data Science, Data Science vs. Data Analytics vs. Business Intelligence.
Types of Data: Structured, Semi-structured, Unstructured, Data Science Lifecycle (Data Collection, Cleaning, Modeling, Evaluation, Deployment), Sources of Data: Databases, Web, Sensors, IoT, Social Media.
Introduction to Big Data (Concepts of Volume, Variety, Velocity), Tools and Platforms for Data Science (Python, R, Excel, Tableau), Applications of Data Science in Engineering, Healthcare, Finance, E-commerce, and Smart Cities.

Module II: Data Preprocessing and Exploratory Data Analysis (9 Hours)

Data Cleaning: Handling Missing Values, Noise, and Outliers, Data Transformation and Normalization, Data Integration and Reduction, Introduction to Exploratory Data Analysis (EDA).
Descriptive Statistics: Mean, Median, Mode, Variance, Standard Deviation.
Data Visualization Techniques: Bar Charts, Histograms, Box Plots, Scatter Plots, Correlation Analysis, Introduction to Data Wrangling using Python/R (Conceptual Overview).

Module III: Statistical and Machine Learning Techniques (9 Hours)

Statistical Foundations: Probability Basics, Random Variables and Distributions, Hypothesis Testing, Regression Analysis (Simple Linear Regression).
Machine Learning Basics: Introduction to Machine Learning
Supervised Learning: Classification (K-Nearest Neighbors, Decision Trees – overview), Regression.
Unsupervised Learning: Clustering (K-Means – overview), Model Evaluation Metrics (Accuracy, Precision, Recall, Confusion Matrix – basic understanding).

Module IV: Advanced Analytics, Applications, and Ethics (9 Hours)

Introduction to Predictive and Prescriptive Analytics, Basics of Time Series Analysis (conceptual), Introduction to Artificial Intelligence in Data Science, Big Data Analytics, Data Visualization Tools (Tableau/Power BI), Case Studies in Engineering and Industry, Ethical Issues in Data Science, Data Privacy, Cybersecurity, and Regulatory Framework, Responsible AI and Bias in Data

Course Outcomes (COs)

After successful completion of the course, students will be able to:

CO-1: Define key concepts, terminology, and processes in data science and analytics.
CO-2: Explain data preprocessing, visualization, and statistical techniques used in data analysis.
CO-3: Apply basic statistical and machine learning methods to analyze datasets and extract insights.
CO-4: Analyze data patterns and evaluate model performance using appropriate metrics.
CO-5: Evaluate ethical implications of data usage and propose data-driven solutions to real-world engineering problems.

Textbooks

1. Joel Grus, Data Science from Scratch, O'Reilly Media (Available in India).
2. Jake VanderPlas, Python Data Science Handbook, O'Reilly Media.
3. Cathy O'Neil & Rachel Schutt, Doing Data Science, O'Reilly Media.

Reference Books

1. Trevor Hastie, Robert Tibshirani & Jerome Friedman, The Elements of Statistical Learning, Springer.
2. Gareth James et al., An Introduction to Statistical Learning, Springer India.
3. Jiawei Han, Micheline Kamber & Jian Pei, Data Mining: Concepts and Techniques, Elsevier India.
4. Aurélien Géron, Hands-On Machine Learning with Scikit-Learn, Keras & TensorFlow, O'Reilly
5. Foster Provost & Tom Fawcett, Data Science for Business, O'Reilly Media.