DATA SCIENCE

Course Code: BEECE1C202 / BEECA1C202
Course Title: DATA SCIENCE
Semester: II
Credits: 04

Rationale

As organizations generate vast amounts of data, there's a growing need for professionals who can analyze, interpret, and extract actionable insights from this data. Data Science skills are in high demand across various industries, including healthcare, finance, retail, and technology. Data Science course provide a broad set of skills that are applicable in multiple domains. They cover areas such as data cleaning, statistical analysis, machine learning, and data visualization. These skills are transferable across industries and job roles, enhancing career flexibility.

Course Outline

Contents	No. of Lectures
UNIT-1 **Introduction** Definition of Data Science, Evolution of Data Science, Big Data and Data Science hype, Datafication, Data Science Roles, Applications of Data Science in various fields, Data Collection Strategies Populations and samples, Data Preprocessing: Need for Preprocessing the Data, Data Cleaning, Data Integration and Transformation, Data Reduction.	10
UNIT-II **Exploratory Data Analytics** Data Objects, Types of Data, Attributes and measurements (Nominal, Ordinal, Binary, Numeric: Discrete vs. Continuous), Symmetric vs. Asymmetric attributes, Basic Statistical Descriptions of Data: Central Tendency: Mean, Median, Mode, Dispersion: Range, Quartiles, Variance, Standard Deviation, Interquartile Range, Probability Distributions, Statistical Relationships and Tests: Correlation and covariance, Chi-square test.	10
UNIT-III **Introduction to Python for Data Analysis** Python: Introduction to Python, Integrated development environment (IDE), How to Install, Introduction to Jupyter Notebook, Python scripting basics: variables, conditional statements, loops, functions, Fundamental libraries for data Scientists, Numpy and Pandas, Reading and Handling with Different Formats, Read and write csv, Descriptive statistics using pandas, working with text data, Treating Missing Values, Indexing, Selecting, and Filtering, Data Normalization.	10

UNIT-IV	10
Data Visualization Tools in Python Introduction to Matplotlib, Basic plots using matplotlib, Introduction to Seaborn: Seaborn functionalities and usage, Spatial Visualizations in Python with Folium, Visualizing data: Histograms, Boxplots, Scatterplots.	
UNIT-V	10
Introduction to Machine Learning Machine Learning, Types of Machine Learning: Supervised, Unsupervised, Semi-supervised and Reinforcement learning, Steps for model building, Loading the Dataset, Test-Train Spilt, Building the Prediction Model, Evaluation Metrics.	

Laboratory work:
Implementation of various data analysis techniques.

Course outcomes
On completion of this course, the students will be able to:

CO1: Understand the fundamental concepts of big data and data science.
CO2: Understand the fundamental concepts of statistics and analyze different types of data used in the data science.
CO3: Evaluate the data analysis techniques for applications handling large data and demonstrate the data science process.
CO4: Visualize and present the inference using various tools.
CO5: Understand concept of machine learning used in the data science process.

Text books:

1. Davy Cielen, Arno D.B.Meysman, Mohmaed Ali, Introducing Data Science: Big Data, Machine Learning, and More, Using Python Tools , Manning Publications, 2016.
2. Cathy O'Neil and Rachel Schutt, Doing Data Science Straight Talk from The Frontline. O'Reilly, 2014.
3. Joel Grus, Data Science from Scratch: First Principles with Python, O'Reilly, 1st edition, 2015.

Reference books:

1. Peter Bruce, Andrew Bruce and Peter Gedeck, Practical Statistics for Data Scientists, O'Reilly Media, 2nd Edition, May 2020.
2. Pratap Dangeti , Statistics for Machine Learning, Packet, July 2017.