BRIGHT INFOTECH

Data Science Course Details

1. Introduction to Data Science

- **What is Data Science?**
  - Definition and Importance
  - Data Science vs. Data Analytics vs. Machine Learning
- **Applications of Data Science**
  - Real-world Examples

2. Python for Data Science

- **Python Basics**
  - Variables, Data Types, and Operators
  - Control Structures: Conditional Statements and Loops
  - Functions and Modules
- **Python Libraries for Data Science**
  - NumPy for Numerical Computing
  - Pandas for Data Manipulation
  - Matplotlib and Seaborn for Data Visualization

3. Data Collection and Cleaning

- **Data Collection**
  - Importing Data from CSV, Excel, and Databases
  - Web Scraping with BeautifulSoup and Scrapy
  - APIs for Data Collection
- **Data Cleaning**
  - Handling Missing Values
  - Data Transformation and Normalization
  - Dealing with Outliers
  - Data Encoding and Feature Engineering

4. Exploratory Data Analysis (EDA)

- **Descriptive Statistics**
  - Measures of Central Tendency: Mean, Median, Mode
  - Measures of Dispersion: Range, Variance, Standard Deviation
- **Data Visualization**
  - Univariate and Bivariate Analysis

    - Visualization Techniques: Histograms, Bar Charts, Box Plots, Scatter Plots
    - Correlation Analysis

5. Introduction to Machine Learning

- **Machine Learning Basics**
    - Supervised vs. Unsupervised Learning
    - Steps in a Machine Learning Project
- **Supervised Learning**
    - Linear Regression
    - Logistic Regression
    - Decision Trees
    - Support Vector Machines
    - Model Evaluation Metrics: Accuracy, Precision, Recall, F1 Score
- **Unsupervised Learning**
    - K-Means Clustering
    - Hierarchical Clustering
    - Principal Component Analysis (PCA)

6. Advanced Machine Learning

- **Ensemble Methods**
    - Random Forest
    - Gradient Boosting Machines (GBM)
    - XGBoost
- **Deep Learning**
    - Introduction to Neural Networks
    - Convolutional Neural Networks (CNNs) for Image Data
    - Recurrent Neural Networks (RNNs) for Sequential Data
    - Frameworks: TensorFlow and Keras

7. Natural Language Processing (NLP)

- **Introduction to NLP**
    - Text Preprocessing: Tokenization, Lemmatization, Stopwords
    - Bag of Words and TF-IDF
- **NLP Techniques**
    - Sentiment Analysis
    - Topic Modeling
    - Named Entity Recognition (NER)
    - Word Embeddings: Word2Vec, GloVe

8. Big Data Technologies

- **Introduction to Big Data**
    - Characteristics of Big Data
    - Hadoop Ecosystem
- **Big Data Processing with Spark**

- Introduction to Apache Spark
- Spark DataFrames and SQL
- Machine Learning with Spark MLlib

9. Data Visualization and Reporting

- **Advanced Data Visualization**
  - Interactive Visualizations with Plotly and Bokeh
  - Dashboards with Dash
- **Data Storytelling**
  - Effective Communication of Insights
  - Creating Reports and Presentations

10. Capstone Project

- **Project Planning and Design**
  - Identifying a Data Science Problem
  - Data Collection and Cleaning
- **Model Building and Evaluation**
  - Feature Selection and Engineering
  - Model Training and Tuning
- **Deployment**
  - Deploying Models with Flask/Django
  - Creating APIs for Model Inference

11. Ethics and Legal Considerations

- **Data Privacy and Security**
  - Understanding Data Privacy Laws (e.g., GDPR)
  - Ethical Implications of Data Science