Comprehensive Data Science

📌 Module 1: Introduction to Data Science

- What is Data Science?

- History and Evolution of Data Science

- Key Components: Data, Algorithms, Models, and Insights

- Overview of Data Science Tools and Technologies

- Understanding the Data Science Life Cycle

- Roles and Responsibilities of a Data Scientist

📌 Module 2: Mathematics and Statistics for Data Science

- **Mathematics for Data Science:**

  - Linear Algebra (Matrices, Vectors, Eigenvalues, Eigenvectors)

  - Calculus (Derivatives, Partial Derivatives, Gradients)

  - Probability Theory (Bayes Theorem, Probability Distributions)

- **Statistics for Data Science:**

  - Descriptive Statistics (Mean, Median, Mode, Variance, Standard Deviation)

  - Inferential Statistics (Hypothesis Testing, Confidence Intervals)

  - Correlation and Causality

  - Sampling Techniques (Random Sampling, Stratified Sampling)

  - Statistical Tests (t-tests, chi-square tests)

📌 Module 3: Python Programming for Data Science

- Introduction to Python Programming

- Python Libraries for Data Science: NumPy, Pandas, Matplotlib, Seaborn

- Data Structures in Python: Lists, Tuples, Dictionaries, Sets

- File Handling in Python: Reading and Writing Data Files (CSV, Excel, JSON)

- Control Flow and Loops

- Functions and Lambda Expressions

- Error Handling and Debugging in Python

📌 Module 4: Data Preprocessing and Cleaning

- Importance of Data Cleaning in Data Science

- Handling Missing Values (Mean, Median Imputation, Dropping)

- Data Transformation (Scaling, Normalization, Standardization)

- Handling Categorical Data (Label Encoding, One-Hot Encoding)

- Detecting and Handling Outliers

- Feature Engineering and Feature Selection

- Data Aggregation and Grouping

📌 Module 5: Data Visualization

- Introduction to Data Visualization

- Basic Visualizations: Bar Charts, Line Graphs, Pie Charts, Histograms

- Advanced Visualizations: Heatmaps, Boxplots, Violin Plots

- Using Matplotlib for Static Visualizations

- Using Seaborn for Statistical Plots

- Interactive Visualizations with Plotly

- Visualizing Large Datasets with Bokeh and D3.js

- Best Practices for Data Visualization: Choosing the Right Plot for the Right Data

---

📌 Module 6: Exploratory Data Analysis (EDA)

- What is Exploratory Data Analysis?

- Steps in EDA: Data Inspection, Data Cleaning, Statistical Summaries

- Univariate and Multivariate Analysis

- Analyzing Distribution of Data

- Correlation Analysis: Pearson, Spearman, and Kendall Correlations

- Outlier Detection and Analysis

- Feature Importance and Dimensionality Reduction Techniques (PCA, LDA)

📌 **Module 7: Machine Learning Fundamentals**

- What is Machine Learning?

- Types of Machine Learning:

  o Supervised Learning

  o Unsupervised Learning

  o Reinforcement Learning

- Overview of ML Algorithms:

  o Linear Regression

  o Logistic Regression

  o Decision Trees

  o k-Nearest Neighbors (k-NN)

  o Naive Bayes

  o Support Vector Machines (SVM)

- Evaluating Model Performance:

  o Accuracy, Precision, Recall, F1-Score

  o Confusion Matrix, ROC Curve, AUC

- Overfitting and Underfitting

📌 **Module 8: Supervised Learning Algorithms**

- **Linear Regression:**

  o Simple Linear Regression

  o Multiple Linear Regression

  o Evaluation Metrics (MSE, RMSE, MAE)

- **Logistic Regression:**
    - Binary and Multinomial Logistic Regression
    - Cost Function and Gradient Descent
    - Evaluating Classification Models

- **Decision Trees:**
    - Building Decision Trees
    - Gini Index and Entropy
    - Pruning and Overfitting

- **Random Forest:**
    - Random Forest Algorithm
    - Feature Importance in Random Forest
    - Tuning Random Forest Models

- **Support Vector Machines (SVM):**
    - Linear vs Non-Linear SVM
    - Kernel Trick in SVM
    - Tuning Hyperparameters in SVM

📌 **Module 9: Unsupervised Learning Algorithms**

- **K-Means Clustering:**
    - Working with Clusters
    - Elbow Method for Optimal Clusters
    - K-Means Algorithm
    - K-Means++ for Initial Centroids

- **Hierarchical Clustering:**
    - Agglomerative vs Divisive Clustering
    - Dendrograms and Cut-off Thresholds

- **Principal Component Analysis (PCA):**
    - Dimensionality Reduction with PCA

- o   Eigenvectors and Eigenvalues

- o   Applications of PCA in Feature Extraction

- **DBSCAN (Density-Based Spatial Clustering):**

  - o   Understanding DBSCAN Clustering Algorithm

  - o   Handling Noise in Data

📌 **Module 10: Neural Networks and Deep Learning**

- Introduction to Neural Networks

- Perceptron and Multi-Layer Perceptron

- Activation Functions: ReLU, Sigmoid, Tanh

- Backpropagation Algorithm

- Introduction to Deep Learning: CNNs, RNNs, GANs

- Working with Keras & TensorFlow

- Image Classification with Convolutional Neural Networks (CNN)

- Time Series Prediction with Recurrent Neural Networks (RNN)

📌 **Module 11: Natural Language Processing (NLP)**

- Introduction to Natural Language Processing

- Text Preprocessing: Tokenization, Lemmatization, and Stopwords Removal

- Bag-of-Words and TF-IDF for Feature Extraction

- Sentiment Analysis with NLP

- Text Classification using Naive Bayes and SVM

- Word Embeddings: Word2Vec, GloVe

- Named Entity Recognition (NER)

- Building Chatbots with NLP

- Topic Modeling with Latent Dirichlet Allocation (LDA)

📌 **Module 12: Model Deployment & Productionizing**

- Introduction to Model Deployment

- Preparing Models for Deployment

- Deployment using Flask or FastAPI

- Dockerizing Machine Learning Models

- Cloud Deployment with AWS, Azure, or Google Cloud

- Real-time Model Prediction & Monitoring

- Version Control and Model Retraining

📌 Module 13: Big Data Technologies for Data Science

- Introduction to Big Data

- Hadoop Ecosystem: HDFS, MapReduce

- Apache Spark for Big Data Processing

- Working with PySpark for Data Analysis

- NoSQL Databases (MongoDB, Cassandra)

- Distributed Computing for Machine Learning

📌 Module 14: Time Series Analysis

- Introduction to Time Series Data

- Time Series Decomposition (Trend, Seasonality, Residuals)

- ARIMA (AutoRegressive Integrated Moving Average) Model

- Exponential Smoothing Methods

- Forecasting with Prophet (from Facebook)

- Handling Missing Data in Time Series

📌 Module 15: Advanced Topics in Data Science

- Reinforcement Learning Overview

- Recommendation Systems (Collaborative Filtering, Content-Based)

- Anomaly Detection in Data

- Graph Theory & Network Analysis

- Advanced Deep Learning Models: GANs, Autoencoders

- Generative Models

📌 Module 16: Data Science Projects and Case Studies

- **Project 1:** Sales Forecasting with Time Series Analysis

- **Project 2:** Customer Segmentation using Clustering Algorithms

- **Project 3:** Sentiment Analysis of Social Media Data

- **Project 4:** Image Classification with Deep Learning

- **Project 5:** Predictive Maintenance in Manufacturing

- **Project 6:** Building a Recommendation System

📌 Module 17: Data Science Career Path and Certification

- Preparing for Data Science Interviews

- Building a Data Science Portfolio

- Certifications in Data Science (Coursera, edX, DataCamp, etc.)

- Job Roles in Data Science: Data Scientist, Data Analyst, Machine Learning Engineer, AI Engineer

- Building a Resume & Job Search Strategies