Data Science Course Content

Module 1: Introduction to Data Science

- What is Data Science? Overview and Applications
- Data Science vs. Data Analytics vs. Machine Learning vs. AI
- The Data Science Lifecycle: Data Collection, Cleaning, Analysis, and Visualization
- Roles and Responsibilities of a Data Scientist
- Tools and Technologies in Data Science (Python, R, Jupyter, SQL, etc.)

Module 2: Data Collection and Preprocessing

- Data Sources and Types (Structured vs. Unstructured)
- Data Collection Techniques (APIs, Web Scraping, Databases)
- **Data Cleaning:**
  - Handling Missing Values
  - Removing Duplicates
  - Outlier Detection and Removal
  - Data Transformation (Normalization, Standardization)
- Feature Engineering (Creating new features, Encoding Categorical Data)

Module 3: Exploratory Data Analysis (EDA)

- Importance of EDA in Data Science
- Descriptive Statistics: Mean, Median, Mode, Standard Deviation
- Data Visualization Techniques: Histograms, Box Plots, Scatter Plots
- Correlation Analysis and Heatmaps
- Outlier Detection
- Handling Imbalanced Data (Resampling, SMOTE)

Module 4: Statistical Foundations for Data Science

- Probability Theory Basics
- Hypothesis Testing (p-values, t-tests, chi-square tests)
- Confidence Intervals and Significance Levels
- Bayes' Theorem and Conditional Probability
- Introduction to Regression Analysis (Linear and Logistic)

Module 5: Introduction to Machine Learning

- Overview of Machine Learning: Supervised vs. Unsupervised Learning
- Key Algorithms for Supervised Learning:
  - Linear Regression, Logistic Regression, Decision Trees, Random Forest
  - Support Vector Machines (SVM), K-Nearest Neighbors (KNN)
- Key Algorithms for Unsupervised Learning:
  - K-Means Clustering, Hierarchical Clustering
  - Principal Component Analysis (PCA)

Module 6: Basic Artificial Intelligence (AI) Concepts

- **Introduction to AI:** What is AI, History, and Types (Narrow AI vs. General AI)
- **AI Techniques:**
  - Search Algorithms (A*, Minimax, BFS, DFS)
  - Rule-based Systems (Expert Systems)
  - Heuristic Search and Optimization
- **Applications of AI:** Robotics, Natural Language Processing, Computer Vision, Autonomous Vehicles

Module 7: Advanced Machine Learning

- Ensemble Learning: Random Forest, Gradient Boosting (XGBoost, LightGBM)
- Model Tuning and Hyperparameter Optimization (Grid Search, Random Search)
- Overfitting and Underfitting
- Cross-validation and Model Evaluation Techniques (Accuracy, Precision, Recall, F1 Score, ROC-AUC)
- Feature Selection and Dimensionality Reduction (L1/L2 regularization)

Module 8: Deep Learning and Neural Networks

- Introduction to Deep Learning
- Overview of Neural Networks: Architecture, Layers, Activation Functions
- Training Neural Networks with Backpropagation
- Convolutional Neural Networks (CNNs) for Image Data
- Recurrent Neural Networks (RNNs) for Sequential Data (Time Series, Text)
- Introduction to TensorFlow and Keras for Deep Learning

Module 9: Natural Language Processing (NLP)

- Text Preprocessing: Tokenization, Lemmatization, Stop-word Removal
- Word Embeddings: Word2Vec, GloVe
- Sentiment Analysis
- Text Classification and Clustering
- Named Entity Recognition (NER)
- Introduction to Transformers and BERT

Module 10: Big Data and Cloud Computing

- Introduction to Big Data Concepts
- Working with Big Data Tools: Hadoop, Spark
- Cloud Computing for Data Science: AWS, Google Cloud, Azure
- Storing and Processing Big Data (HDFS, S3, Databricks)
- Introduction to Distributed Computing

Module 11: Data Visualization and Communication

- Importance of Data Visualization in Data Science
- Tools for Data Visualization: Matplotlib, Seaborn, Plotly, Tableau
- Interactive Dashboards in Dash and Streamlit
- Communicating Insights to Non-Technical Stakeholders
- Storytelling with Data

Module 12: Model Deployment and Production

- Model Deployment Techniques: Flask, Django for REST APIs
- Using Docker for Containerization
- Cloud Deployment: AWS SageMaker, Google AI Platform
- Model Monitoring and Maintenance
- Model Drift and Retraining

Module 13: Project Work

- **Mini Projects:**
  - Predictive Analytics (e.g., House Price Prediction, Customer Churn)
  - Text Classification (e.g., Spam Email Detection)
  - Image Classification (e.g., Cats vs. Dogs)

- **Capstone Project:**
  - Real-world Data Science Project (e.g., Sales Forecasting, Recommendation Systems)
  - Full Project Workflow: Data Collection, Cleaning, EDA, Model Building, Deployment
  - Final Report and Presentation

Module 14: Career Preparation for Data Science

- Resume Building for Data Scientists
- Portfolio Creation (GitHub, Kaggle Competitions)
- Interview Preparation (DSA, Machine Learning, Case Studies)
- Mock Interviews and Coding Challenges