Data Science and Artificial Intelligence Course

Course Overview

This comprehensive course covers fundamental to advanced concepts in Data Science and Artificial Intelligence (AI). It equips students with the skills to analyze data, build predictive models, and create AI-driven solutions using Python, machine learning techniques, and deep learning methods.

Course Modules

Part 1: Foundations of Data Science

Module 1: Introduction to Data Science and AI

- **Overview of Data Science and AI**
  - What is Data Science?
  - Definition and Scope of AI
  - Applications in Various Industries
  - The Role of Data Science and AI in Modern Business
- **Data Science Workflow**
  - Steps in the Data Science Process: Problem Definition, Data Collection, Data Cleaning, Analysis, Modeling, Evaluation, and Deployment

Module 2: Python Programming for Data Science

- **Setting Up Python Environment**
  - Installing Python, IDEs (Jupyter Notebook, VS Code)
  - Managing Packages with pip and conda
- **Basic Syntax and Data Types**
  - Python Syntax, Variables, and Basic Data Types (Integers, Strings, Floats, Booleans)
- **Control Structures: Loops and Conditionals**
  - If Statements, For Loops, While Loops, List Comprehensions
- **Functions and Modules**
  - Defining Functions, Scope, Arguments, Return Values
  - Importing and Using Modules
- **Data Structures: Lists, Tuples, Dictionaries, and Sets**
  - Lists and Tuples: Creation, Indexing, and Methods
  - Dictionaries: Key-Value Pairs, Accessing, Adding, Updating
  - Sets: Unique Elements, Operations

Module 3: Essential Mathematics and Statistics

- **Basic Algebra and Calculus**
  - Algebraic Expressions, Solving Equations
  - Differential and Integral Calculus Concepts for Machine Learning
- **Probability Theory**
  - Basic Probability Concepts, Conditional Probability, Bayes' Theorem
- **Descriptive Statistics**
  - Mean, Median, Mode, Variance, Standard Deviation
- **Inferential Statistics**
  - Hypothesis Testing, Confidence Intervals
  - Regression Analysis: Linear Regression, Logistic Regression

Part 2: Data Manipulation and Visualization

Module 4: Data Manipulation with Pandas

- **Introduction to Pandas**
  - What is Pandas?
  - Data Structures: Series and DataFrames
- **Data Cleaning and Preparation**
  - Handling Missing Data, Data Transformation
  - Data Aggregation and Grouping
- **Data Transformation and Aggregation**
  - Merging, Joining DataFrames
  - Data Aggregation Methods: GroupBy, Pivot Tables

Module 5: Data Visualization

- **Introduction to Matplotlib**
  - Creating Basic Plots: Line, Bar, Scatter, Histograms
- **Customizing Plots**
  - Adding Labels, Legends, Annotations, and Colors
- **Advanced Visualization with Seaborn**
  - Creating Statistical Plots: Box Plots, Violin Plots, Pair Plots
- **Interactive Visualization with Plotly**
  - Creating Interactive Graphs and Dashboards

Part 3: Machine Learning

Module 6: Introduction to Machine Learning

- **Overview of Machine Learning**

- ○ What is Machine Learning? Types of Learning: Supervised, Unsupervised
- **Setting Up Scikit-learn**
  - ○ Installing and Configuring Scikit-learn
  - ○ Understanding Scikit-learn API

Module 7: Supervised Learning

- **Linear Regression**
  - ○ Simple and Multiple Linear Regression
- **Logistic Regression**
  - ○ Classification Problems, Implementing Logistic Regression
- **Decision Trees and Random Forests**
  - ○ Building and Evaluating Decision Trees
  - ○ Ensemble Methods: Random Forests
- **Support Vector Machines (SVM)**
  - ○ Concepts of SVM, Kernel Tricks, Hyperparameter Tuning
- **Model Evaluation and Validation**
  - ○ Metrics: Accuracy, Precision, Recall, F1 Score, ROC-AUC

Module 8: Unsupervised Learning

- **K-Means Clustering**
  - ○ Introduction to Clustering Algorithms, Implementing K-Means
- **Hierarchical Clustering**
  - ○ Agglomerative and Divisive Clustering Methods
- **Principal Component Analysis (PCA)**
  - ○ Dimensionality Reduction Techniques, Eigenvalues, Eigenvectors
- **Anomaly Detection**
  - ○ Techniques for Identifying Outliers in Data

Part 4: Deep Learning

Module 9: Introduction to Deep Learning

- **Overview of Neural Networks**
  - ○ Architecture of Neural Networks: Neurons, Layers, Activation Functions
- **Setting Up TensorFlow and Keras**
  - ○ Installing TensorFlow and Keras
  - ○ Building and Training Neural Networks
- **Evaluating Neural Network Models**
  - ○ Model Performance Metrics: Loss Functions, Optimizers

Module 10: Advanced Deep Learning Techniques

- **Convolutional Neural Networks (CNNs)**
  - Image Classification, Feature Extraction Techniques
- **Recurrent Neural Networks (RNNs)**
  - Sequence Modeling, Applications in Text and Time Series
- **Long Short-Term Memory (LSTM) Networks**
  - Advanced RNN Architecture for Long-Term Dependencies
- **Autoencoders**
  - Encoder-Decoder Architecture, Applications in Data Compression

Part 5: Natural Language Processing (NLP)

Module 11: Introduction to NLP

- **Overview of NLP**
  - What is NLP? Applications in Real-World Scenarios
- **Text Preprocessing Techniques**
  - Tokenization, Stop Words Removal, Lemmatization, Stemming
- **Sentiment Analysis**
  - Techniques for Analyzing Sentiment in Text Data
- **Text Classification**
  - Categorizing Text Data into Different Classes

Module 12: Advanced NLP Techniques

- **Word Embeddings: Word2Vec, GloVe**
  - Techniques for Representing Words in Vector Space
- **Transformers and BERT**
  - Introduction to Transformers, BERT Architecture, Fine-Tuning Models
- **Sequence-to-Sequence Models**
  - Building Models for Translation and Text Generation
- **Applications in Language Translation and Chatbots**
  - Implementing Translation Systems, Building Conversational Agents

Part 6: Tools and Technologies

Module 13: Big Data Technologies

- **Introduction to Big Data**
  - What is Big Data? Characteristics and Technologies
- **Hadoop and Spark**
  - Overview of Hadoop Ecosystem, Spark for Big Data Processing
- **Data Processing with PySpark**
  - Using PySpark for Large-Scale Data Processing

- **Integrating Big Data with Machine Learning**
  - Combining Big Data Technologies with ML Algorithms

Module 14: Model Deployment and Monitoring

- **Introduction to Model Deployment**
  - Deploying Machine Learning Models for Production Environments
- **Deploying Models with Flask and Django**
  - Building APIs for Model Deployment
- **Model Monitoring and Management**
  - Techniques for Monitoring Model Performance, Updating Models
- **Using Docker for Deployment**
  - Containerizing Applications with Docker for Consistent Environments

Capstone Project

Capstone Project: Real-World Data Science and AI Project

- **Project Overview**
  - End-to-End Data Science and AI Project
  - From Data Collection to Model Deployment
- **Project Phases**
  - **Data Collection and Cleaning:** Gather Data, Perform Initial Exploration
  - **Model Building and Evaluation:** Develop Models, Evaluate Performance
  - **Deployment and Monitoring:** Deploy Models, Implement Monitoring Solutions
  - **Presentation and Interpretation of Results:** Present Findings, Provide Recommendations

Assessment and Certification

- **Quizzes and Exams**
  - Regular Assessments to Test Knowledge and Understanding
- **Practical Lab Assessments**
  - Hands-On Exercises and Mini-Projects
- **Final Project Evaluation**
  - Assessment of Capstone Project Based on Criteria
- **Certification of Completion**
  - Awarded Upon Successful Completion of the Course