Advance AI & ML for IA

Duration: 5 Days

Day 1: Python, NumPy and Pandas for Data Science

Module 1: Python Fundamentals

- Python syntax and data types
- Lists, tuples and dictionaries
- Loops and conditionals
- Functions and modules
- File handling
- Error handling

Lab Exercises:

1. Write Python scripts covering data types, loops, functions and file operations.
2. Build a small utility script that reads a CSV file and prints summary information.

Module 2: NumPy and Pandas for Data Analysis

- NumPy arrays and operations
- Array slicing and indexing
- Mathematical operations
- Pandas Series and DataFrame
- Data loading and inspection
- Exploratory data analysis

Lab Exercises:

3. Perform array operations and mathematical computations using NumPy.
4. Load a real dataset with Pandas, inspect it and generate an exploratory data analysis report.

Day 2: AI/ML Foundations and Data Preparation

Module 3: AI/ML Foundations

- AI, ML Foundation
- Generative AI overview
- Real-world use cases
- Types of ML
- ML project lifecycle
- Model evaluation basics

Lab Exercises:

5. Map real-world problems to the correct ML type and document the reasoning.

6. Set up an ML project structure and configure the development environment.

Module 4: Data Preparation and Feature Engineering

- Missing values handling
- Duplicates and outliers
- Categorical encoding
- Normalization and scaling
- Train-validation-test split
- Avoiding data leakage

Lab Exercises:

7. Clean a real dataset by fixing missing values, duplicates and outliers.
8. Apply encoding and scaling then build a reusable preprocessing pipeline.

Day 3: Supervised ML and Advanced Model Optimization

Module 5: Supervised Machine Learning

- Linear regression
- Logistic regression
- Decision trees
- K-nearest neighbours
- Support vector machines
- Overfitting and underfitting

Lab Exercises:

9. Train regression and classification models on a provided dataset and compare performance.
10. Plot learning curves to diagnose overfitting and underfitting across models.

Module 6: Advanced ML and Model Optimization

- Random forest
- Gradient boosting
- Cross-validation
- Hyperparameter tuning
- Class imbalance handling
- Explainable AI basics

Lab Exercises:

11. Tune multiple models using cross-validation and grid search on an imbalanced dataset.
12. Explain model predictions using feature importance and explainability techniques.

Day 4: Unsupervised Learning, Deep Learning and NLP

Module 7: Unsupervised Learning and Deep Learning

- Unsupervised Learning and Deep Learning Foundation
- K-means clustering
- PCA and dimensionality reduction
- Anomaly detection
- Neural networks basics
- Activation and loss functions
- Dropout and regularization

Lab Exercises:

13. Apply K-means clustering and reduce dimensions with PCA on a provided dataset.
14. Build and train a neural network model and evaluate training vs validation loss.

Module 8: NLP, Embeddings and Transformer Models

- Text preprocessing
- Tokenization
- TF-IDF and embeddings
- Attention mechanism
- Transformer architecture
- LLM fundamentals

Lab Exercises:

15. Generate sentence embeddings and calculate document similarity scores.
16. Build a basic semantic search application on a provided text corpus.

Day 5: LLM Development, RAG Architecture and Deployment

Module 9: LLM Application Development and RAG

- Prompt design
- Zero-shot and few-shot
- Structured output
- RAG architecture
- Vector databases
- Chunking and retrieval

Lab Exercises:

17. Design prompts and create an LLM test suite covering hallucination and safety cases.
18. Build an end-to-end RAG pipeline that retrieves chunks and generates grounded answers.

Module 10: Advanced RAG, Evaluation and Deployment

- Hybrid retrieval
- Query rewriting
- Conversation memory
- RAG evaluation metrics
- Answer faithfulness
- API deployment and monitoring

Lab Exercises:

19. Improve the RAG pipeline with hybrid retrieval and evaluate answer faithfulness.
20. Expose the RAG system through a REST API and present the final capstone project.