AI & Data Science Pro: Python, R, SQL & Generative AI

Duration: 15 days (120 hours)

Course Overview

This all-in-one program offers a deep dive into the world of Artificial Intelligence, Data Science, and Generative AI, blending foundational theory with hands-on practice across Python, R, SQL, and cutting-edge AI frameworks. Starting with AI fundamentals and Python programming, learners progress through data analysis, machine learning, deep learning, and advanced neural architectures including CNNs, RNNs, and Transformers. The R programming track complements this journey with robust data manipulation, visualization, and machine learning techniques. The Generative AI specialization introduces Large Language Models (LLMs), prompt engineering, Retrieval-Augmented Generation (RAG), and fine-tuning with open-source tools like LangChain and MLflow. Finally, the SQL module equips learners with essential skills for data querying, transformation, and exploratory analysis. Whether you're aspiring to become a data scientist, AI engineer, or generative AI specialist, this course provides the tools and knowledge to thrive in the AI-driven future.

Course Contents

Part 1: Complete Artificial Intelligence

Module 1: Introduction to AI and Python Basics

- What is AI?
- History and evolution of AI
- AI applications in various fields
- Introduction to machine learning and deep learning
- Python installation and setup
- Basic syntax and data types
- Control structures: if statements, loops
- Functions and modules

Module 2: Python for Data Analysis & Pre-processing

- Introduction to NumPy, Pandas, Matplotlib, Seaborn
- Basic operations with NumPy
- Data manipulation with Pandas
- Data cleaning techniques
- Handling missing and categorical data
- Exploratory Data Analysis (EDA) with visualizations

Module 3: Data Preprocessing and Transformation

- Importance of data preprocessing
- Handling missing values
- Label encoding and one-hot encoding
- Data normalization techniques: Standard Scaler, MinMax Scaler
- Splitting data: Train, test, and validation sets

Module 4: Introduction to Machine Learning

- Types of machine learning
- Lifecycle of a data science project
- Overview of supervised and unsupervised learning \
- Simple Linear Regression
- Multiple Linear Regression
- Evaluation metrics: R-squared, RMSE

Module 5: Supervised Learning - Classification

- Logistic Regression: Concept and intuition
- Confusion matrix and performance metrics
- Support Vector Machine (SVM)
- Decision Trees
- Random Forests

Module 6: Feature Selection and Dimensionality Reduction

- Importance of feature selection
- Univariate feature selection techniques
- Recursive Feature Elimination (RFE)
- Principal Component Analysis (PCA)
- Hands-on PCA implementation

Module 7: Unsupervised Learning and NLP Basics

- Introduction to clustering
- K-means clustering
- Elbow method
- Introduction to NLP
- Tokenization, stop words, stemming, lemmatization

Module 8: Introduction to Deep Learning

- Need and applications of deep learning
- Working of Artificial Neural Networks (ANN)
- Introduction to TensorFlow and Keras
- Keras model building: Construct, compile, evaluate
- Activation functions
- Loss functions and optimization techniques

Module 9: Neural Network Basics and Shallow Networks

- Machine learning problem setup (Neural Networks)
- Neural network mindset
- Vectorization for efficient computation
- Building a neural network with one hidden layer
- Forward propagation and backpropagation

Module 10: Deep Neural Networks

- Building and training deep neural networks
- Computation in deep learning
- Initialization methods
- Regularization techniques to prevent overfitting

Module 11: Optimization and Hyperparameter Tuning (Deep Learning)

- Advanced optimization techniques
- Random minibatching and learning rate decay
- Batch normalization
- Grid search and randomized search (Hyperparameter tuning)

Module 12: Convolutional Neural Networks (CNNs)

- Understanding pooling and convolutional layers
- Building deep CNNs for image classification
- Advanced tricks and methods in deep CNNs
- Transfer learning with pretrained models

Module 13: Computer Vision Applications with CNNs

- Using CNNs for object detection tasks
- Applying CNNs for face recognition
- Implementing neural style transfer for art generation

Module 14: Recurrent Neural Networks (RNNs) and NLP

- Introduction to RNNs
- Variants of RNNs for sequential data modeling
- NLP applications with deep learning models
- Word embeddings for text analysis

Module 15: Advanced Sequence Models and Transformers

- Enhancing sequence models with attention mechanisms
- Speech recognition and audio data processing
- Understanding transformer networks
- Applications of transformers in NLP

Part 2: R Programming for Data Science & Machine Learning

Module 1: Introduction To R Basics:

- Introduction to R Programming Basics
- Arithmetic, Variables, and Basic Data Types in R
- Vector Basics, Operations, Comparison Operators, Indexing, and Slicing in R

Module 2: R Matrices

- Introduction to R Matrices
- Creating, Arithmetic, Operations, Selection, and Indexing in R Matrices
- Factor and Categorical Matrices in R

Module 3: R Data Frames & Lists

- Introduction to R Data Frames
- Data Frame Basics, Indexing, Selection, and Operations in R
- List and its Operations in R

Module 4: Data Processing With R

- Importing data from CSV Files, Excel Files
- Connecting to SQL Databases
- Web Scraping with R

Module 5: R Programming Concepts

- Introduction to R Programming Concepts
- Logical Operators, Conditional Statements, Loops, and Functions in R
- Built-in R Features, Apply Functions, Math Functions
- Regular Expressions
- Dates and Timestamps in R

Module 6: Data Manipulation & Data Visualization With R

- Data Manipulation Overview with Dplyr, Pipe Operator, and Tidyr in R
- Overview of ggplot2 for Static Visualizations (Histograms, Scatterplots, Bar plots, Box plots, 2 Variable Plotting)
- Customizing Visualizations: Coordinates, Faceting, Themes
- Overview of Plotly and Interactive Visualizations in R

Module 7: Supervised Machine Learning With R

- Introduction to Machine Learning Concepts
- Regression Algorithms: Simple Linear Regression, Multiple Linear Regression
- Assumptions of Multiple Linear Regression
- Classification Algorithms: K-Nearest Neighbours
- Classification Algorithms: Decision Trees and Random Forests
- Classification Algorithms: Support Vector Machines

Module 8: Unsupervised Machine Learning & Deep Learning With R Topics:

- Unsupervised Learning: K-Means Clustering
- Natural Language Processing Basics in R
- Deep Learning Concepts using Neural Networks in R

Part 3: Generative AI Speciality

Module 1: Introduction To Generative AI

- Overview of Generative AI
- Architecture of Generative AI
- Applications of Generative AI with the Transformer Library
- Introduction to Generative Adversarial Networks (GANs)

Module 2: Introduction To Large Language Models (LLMs)

- Architecture of Large Language Models
- Types of Large Language Models (LLMs)
- Task-based Text AI LLMs: Translation, Summarization, Sentence Similarity, Automatic Speech Recognition, Text-to-Speech, etc.
- Major Text AI LLMs: LLaMA, Qwen, Cohere, Falcon LLM
- Image AI Models & Services: Object Detection, Image Segmentation, Image Retrieval, Image Captioning, Visual Q&A, Zero-shot Image Classification, etc.

Module 3: Prompt Engineering Using Open-Source Models

- Introduction to Prompt Engineering
- Prompt Engineering Techniques
- Text Prompting with LLaMA (Meta)
- Image Prompting with LLaMA (Meta)
- Code Prompting with LLaMA (Meta)

Module 4: Basic LLM Systems (RAG) Using Open-Source Models

- Introduction to Retrieval-Augmented Generation (RAG)
- Overview of LangChain
- Concepts of Embeddings, Retrieval, Chains, and Agents with LangChain
- Building simple LLM applications with LangChain (Implied through concepts and practicals)
- Creating chatbots with LangChain (Implied through concepts and practicals)
- Developing vector stores and retrievers using LangChain (Implied through concepts and practicals)
- Building agents with LangChain (Implied through concepts and practicals)
- Creating RAG and conversational RAG applications with LangChain (Implied through concepts and practicals)

Module 5: Advanced LLM Systems (Q&A) Using Open-Source Models

- Difference Between RAG and Question-Answering Systems
- Building a Question-Answering System Over Tabular Data with LangChain
- Building a Question-Answering System Over SQL Data with LangChain

Module 6: Fine-Tuning Techniques Using Open-Source Models

- Introduction to Quantization
- Optimization of Model Weights (Data Types)
- Modes of Quantization
- Fine-Tuning LLMs (e.g., Meta's LLaMA, Alibaba's Qwen, Google's Gemma)

Module 7: Model Evaluation Using Mlflow

- Introduction to MLflow
- Building and Tracking Machine Learning Models with MLflow
- MLflow Deployment Servers
- LLM Evaluation with MLflow

PART 4: SQL for Data Science

Module 1: Introduction to SQL and Data Science

- What is SQL? Why SQL for Data Science?
- Overview of Relational Databases
- SQL vs NoSQL: When and Why
- Setting up SQL Environment (SQLite / PostgreSQL / MySQL)
- Lab: Install and connect to a sample database
- Lab: Explore tables using SELECT, LIMIT, and DESCRIBE

Module 2: Data Retrieval and Filtering

- Basic Queries: SELECT, WHERE, ORDER BY
- Filtering with AND, OR, IN, BETWEEN, LIKE
- Sorting and Aliasing
- Lab: Query a dataset (e.g., Titanic or Iris) to extract specific insights
- Lab: Practice filtering and sorting data

Module 3: Aggregation and Grouping

- Aggregate Functions: COUNT, SUM, AVG, MIN, MAX
- Grouping Data with GROUP BY
- Filtering Groups with HAVING
- Lab: Analyze sales or customer data using aggregation
- Lab: Compare grouped statistics

Module 4: Joins and Subqueries

- Types of Joins: INNER, LEFT, RIGHT, FULL OUTER
- Joining Multiple Tables
- Subqueries: Scalar, Correlated, and Nested
- Lab: Join customer and transaction tables to derive insights
- Lab: Use subqueries to filter complex conditions

Module 5: Data Cleaning and Transformation

- Handling NULLs and Missing Data
- Using CASE, COALESCE, and NULLIF

- String and Date Functions
- Lab: Clean and transform messy data for analysis
- Lab: Format dates and standardize text fields

Module 6: SQL for Exploratory Data Analysis (EDA)

- Descriptive Statistics in SQL
- Creating Derived Columns
- Using CTEs (Common Table Expressions)
- Lab: Perform EDA on a real-world dataset (e.g., COVID, e-commerce)
- Lab: Generate summary tables and insights

Module 7 (Optional/Advanced): Window Functions and Optimization

- Introduction to Window Functions: ROW_NUMBER, RANK, LEAD, LAG
- Performance Tips: Indexing, Query Optimization Basics
- Lab: Use window functions to analyze trends over time
- Lab: Optimize a slow-running query