Data Science Training

Course Duration: 1.5 Months • Class Time: 1.5 Hrs

1. Introduction

- Prelude
- The problem landscape
- Defining Data Science
- Demystifying Data Science, Decision Science, AI, ML and DL
- Overview of Data Scientist's Toolbox

2. Data Science Tool Box

- Python - Quick recap ? Python 2.7.x or 3.x?
- Installation and setup
- Data types, functions and important packages
- Data manipulation & Data Engineering
- Data Visualization

3. Probability and Statistics

- Theoretical foundations of statistics, with practical applications
- Describing data, populations, and sampling
- Understanding variables and their measurement scales
- Analyzing data distribution, central tendencies (mean, median, mode), and dispersion (variance, standard deviation)
- Exploring probability distributions like Gaussian, Bernoulli, Binomial, and Poisson
- Conducting statistical tests (z-test, t-test, chi-square test) and understanding errors (Type 1 and Type 2)
- Analyzing correlations, including Pearson and Spearman's rank
- Key rules and concepts in probability, such as addition, multiplication, permutation, and combination

4. Numpy

- Introduction to Numpy
- Random Data Generation
- Numpy Array, Indexing & Operations
- Array Data Structures in Numpy
- Array operations and methods
- Course Assignment

5. Pandas

- Importing Datasets
- Data Wrangling
- Exploratory Data Analysis and Model Development

6. Basic SQL for Data Science

- Introduction to SQL
- SQL Queries
- Joins and Subqueries
- Aggregation and Filtering
- Working with SQL Databases

7. Scipy and Seaborn

- Introduction to Scipy
- Numerical Computations
- Exploratory Data Analysis
- Model Generation

8. Plotting, Charting & Data Visualization

- Principles of Information Visualization
- Basic Charting
- Charting Fundamentals
- Applied Visualizations

9. Tableau Basics

- Introduction to Tableau
- Download and Install Tableau Public
- Load Data from Excel
- Creating Charts and Graphs
- Basic Visual Analysis

10. Exploratory Data Analysis (EDA) and Hypothesis Testing

- Overview of the Machine Learning methodology
- Exploratory Data Analysis (EDA)
- Introduction to Feature Engineering
- Statistical Inference, Probability Distributions
- Hypothesis Testing

11. Machine Learning Introduction

- Machine Learning Introduction
- ML core concepts
- Unsupervised and Supervised Learning
- Clustering, Classification, and Regression
- Supervised Vs Unsupervised

Supervised Learning

a. Linear Regression

- Introduction and Concept
- Best Fit Line and its Equation
- Model Training and Evaluation

b. Logistic Regression

- Introduction to Classification
- Sigmoid Curve and its Application
- Model Evaluation Techniques

c. Support Vector Machine (SVM)

- Introduction to SVM
- Kernel Trick and Hyperplanes
- Model Training and Evaluation

d. K-Nearest Neighbors (KNN)

- KNN Algorithm Overview
- Distance Metrics and Classification
- Model Evaluation and Tuning

AI Tool: AutoML for Model Building

Unsupervised Learning

a. Clustering Overview

- Understanding Unsupervised Learning
- Clustering vs Classification

b. K-Means Algorithm

- Introduction to K-Means
- K-Means Theory and its Working
- K-Means Algorithm Steps
- Model Training and Evaluation in Python

c. Principal Component Analysis (PCA)

- Introduction to PCA
- Dimensionality Reduction with PCA
- Understanding Eigenvectors and Eigenvalues
- PCA in Data Preprocessing
- Implementing PCA in Python

12. Text Mining In Python

- Basic Natural Language Processing
- Working with NLTK
- Text Preprocessing
- Text Cleaning and Regular Expression
- Regex Introduction
- Regex Codes
- Text Extraction with Python Regex
- Stop Word Removal
- Stemming
- Lemmatization
- POS Tagging
- Text Classification

13. Prompt Engineering for Data Science

- Introduction to Prompt Engineering
- Techniques for Crafting Effective Prompts
- Applications of LLM in Data Science
- Iterative Improvement
- Integration

AI Tool: ChatGpt, Gemini, DeepSeek

14. ML Web App Development with Streamlit

- Introduction to Streamlit
- Setting up Streamlit
- Building an ML Web App
- Interactive Visualizations
- Deploying Streamlit Apps

15. FastAPI and ML Deployment

- Introduction to FastAPI
- Building an API for ML Models
- Handling Requests & Responses
- Asynchronous Processing
- Deployment & Scaling

16. Projects

- Exploratory Data Analysis (EDA) and Hypothesis Testing
- Regression Analysis
- Sentiment Analysis
- Classification-based Projects
- Clustering based project
- Real-time ML Model Deployment