Skill Ascent
by D BASE SOLUTIONS PVT. LTD.,

DATA SCIENCE

DURATION
150 Days

1. INTRODUCTION TO DATA SCIENCE

- Introduction to Data Science
- What is Data Science?
- Why Data Science?
- Impact of Data Science
- Future of Data Science
- Data Science Life Cycle

2. PYTHON & DEVELOPMENT ENVIRONMENT

PRE-CORE PYTHON

- Introduction to Pre-Core Python

NOTEBOOKS & IDES

- Introduction to Pre-Core Python
- Introduction to Jupyter Notebook
- Overview of Data Science Real-Time IDEs
- Introduction to Google Colab Notebook

OPERATING SYSTEM BASICS

- Introduction to UNIX Operating System

3. PYTHON LIBRARIES FOR DATA SCIENCE

- NumPy for Data Science
- Pandas for Data Science
- Matplotlib for Data Science
- Seaborn for Data Science

LIVE APPLICATION

- Live Application Implementation

4. DATA VISUALIZATION

CORE VISUALIZATION

- Basic Plotting for Data Visualization
- Data Manipulation for Visualization
- 1D Data Analysis
  - Histograms
  - Boxplots
  - Violin Plots

5. POWER BI

- Introduction to Power BI
- Data Extraction Process
- Data Transformations
- Data Modelling and DAX
- Data Visualization with Analytics
- Power BI Q&A and Data Insights

LIVE APPLICATIONS

- Visualization of World GDP and Carbon Dioxide Emissions
- Geographic Overlays using Folium Library

6. DATA ANALYSIS USING EXCEL & SQL

EXCEL FOR DATA ANALYSIS

- Introduction to Excel
- Functions, Formulas, and Charts
- Pivot Tables and Lookups
- Ranges and Tables
- Data Cleaning
  - Text Functions
  - Dates and Times
- Conditional Formatting
- Sorting and Filtering
- Subtotals with Ranges
- Data Visualization in Excel
- Advanced Excel with AI Features

SQL FOR DATA ANALYSIS

- SQL Overview and Process
- RDBMS Concepts
- SQL Databases
- What is Database?
- What is DBMS and RDBMS?
- Sub Languages in SQL
- SQL Syntax
  - Data Types
  - Operators
- SQL Commands
  - Create
  - Select
  - Insert
  - Delete
  - Drop

- Where Clause
- AND / OR Operators
- Like, Top, Limit, ROWNUM
- Order By, Group By, Distinct
- Constraints
- Joins
- Indexes
- Alter and Truncate
- Properties of Transactions
- Select with Where Clause
- SQL Connectivity with Python

7. STATISTICS

- Basics of Statistics
- Types of Statistics
- Population and Sample
- Measures of Central Tendency
- Percentiles and Dispersion

STATISTICS WITH PYTHON

- Range
- Sample Variance
- Standard Deviation
- Correlation and Causation
- Hypothesis Testing
- Parametric Tests
- Non-Parametric Tests

8. PROBABILITY

- What is Probability?
- Importance of Probability in Machine Learning
- Basics of Probability
- Random Variables
- Probability Distributions
- Maximum Likelihood Estimation

- Bayes Theorem
- Information Theory
- Cross Entropy
- Information Gain

9. LINEAR ALGEBRA

- Scalar and Vector
- Vector Addition
- Vector Subtraction
- Scalar Multiplication
- Dot Product
- Cross Product
- Matrices
- Types of Matrices
- Transpose of Matrix
- Matrix Addition and Subtraction
- Eigen Values and Eigen Vectors

10. CALCULUS (MATHS FOR DATA SCIENCE)

- What is Calculus?
- Limits and Differential Calculus
- Limits and Continuity
- Evaluating Limits
- Functions and Derivatives
- Continuous Functions
- Derivatives of Powers and Polynomials
- Introduction to Multivariate Calculus

11. DATA STRUCTURES & ALGORITHMS (PYTHON)

- What are Data Structures?
- Big O Notation
- Arrays in Python
- Static vs Dynamic Arrays
- Linked Lists
- Doubly Linked Lists
- Hash Tables
- Hash Maps
- Collision Handling
- Chaining Technique
- Stack Implementation
- Queue Implementation
- Tree Data Structure
- Binary Tree
- Binary Search Tree (BST)
- Graph Introduction
- Linear Search
- Binary Search
- Sorting Algorithms
  - Bubble Sort
  - Quick Sort
  - Insertion Sort
  - Merge Sort
  - Shell Sort
- Recursion in Python
- Additional DSA Exercises

12. MACHINE LEARNING

INTRODUCTION

- What is Machine Learning?
- Types of Machine Learning
  - Supervised Learning
  - Unsupervised Learning

- Applications of Machine Learning
- Types of Data
  - Continuous
  - Categorical

DATA PREPARATION

- Data Exploration and Visualization
- Descriptive Statistics
- Inferential Statistics
- Data Distributions
- Correlation and Covariance
- Handling Missing Values
- Scatter Plots and Heatmaps
- Data Normalization Techniques
- Data Imputation Techniques

13. REGRESSION

- Introduction to Regression
- Simple Linear Regression
- Multiple Linear Regression
- Assumptions of Linear Regression
- Regularization Techniques
  - Lasso Regression
  - Ridge Regression
- Polynomial Regression
- Stepwise Regression
- ElasticNet Regression
- R-Squared and Adjusted R-Squared

14. CLASSIFICATION

- Introduction to Classification
- Types of Classifiers

LINEAR CLASSIFIERS

- Logistic Regression
- Multinomial Logistic Regression

NON-LINEAR CLASSIFIERS

- Decision Trees
  - CART Algorithm
  - ID3 Algorithm
- Random Forest
- Support Vector Machines
  - Kernel Trick
  - Soft Margin SVM
  - Multi-Class SVM
- Naive Bayes
- K-Nearest Neighbors (KNN)

NEURAL NETWORKS FOR CLASSIFICATION

- Perceptron Algorithm
- Multilayer Perceptron (MLP)
- Backpropagation Algorithm
- Activation Functions

15. MODEL EVALUATION

- Confusion Matrix
- Accuracy
- Precision
- Recall
- F1-Score
- ROC Curve
- AUC

16. FEATURE ENGINEERING & MODEL SELECTION

- Feature Selection Techniques
- Hyperparameter Tuning
- Model Selection
- Bias–Variance Tradeoff
- Cross Validation
- Leave-One-Out Cross Validation

17. ENSEMBLE LEARNING

- Ensemble Methods
- Bagging Algorithms
- Boosting Algorithms
  - XGBoost
  - Gradient Boosting
  - LightGBM
  - CatBoost
  - AdaBoost
- Stacking Technique
- Blending Technique

18. CLUSTERING

- Introduction to Clustering
- K-Means Clustering
- Hierarchical Clustering

19. DIMENSIONALITY REDUCTION

- Introduction to Dimensionality Reduction
- Principal Component Analysis (PCA)
- Singular Value Decomposition (SVD)
- t-SNE
- Linear Discriminant Analysis (LDA)
- Truncated SVD

20. DEEP LEARNING

- What is Deep Learning?
- Difference between Machine Learning and Deep Learning
- Biological Neural Networks
- Artificial Neural Networks (ANN)
- Convolutional Neural Networks (CNN)
- Recurrent Neural Networks (RNN)

CNN & COMPUTER VISION

- Image Basics
- Image Preprocessing using OpenCV
- CNN Architecture
- Image Classification Case Study
- Transfer Learning Case Study

21. NATURAL LANGUAGE PROCESSING (NLP)

- Introduction to Text Processing
- Text Preprocessing using NLTK and SpaCy
- Vectorization Techniques

PROJECTS

- Text Classification Project
- Sequence Tagging Project

ADVANCED NLP

- RNNs
- LSTMs
- Autoencoders

22. GENERATIVE AI & LLMS

- Understanding Generative AI
- Applications of Generative AI
- Types of Generative Models
  - GANs
  - VAEs
  - Autoregressive Models
- Introduction to Large Language Models
- Transformer Architecture
- Open-Source LLMs
- HuggingFace and Pre-trained Models
- Limitations of LLMs
- Advantages and Disadvantages of LLM Architectures
- Using Open-Source LLMs for Projects
- Responsible AI Development

22. GENERATIVE AI & LLMS

- Understanding Generative AI
- Applications of Generative AI
- Types of Generative Models
  - GANs
  - VAEs
  - Autoregressive Models
- Introduction to Large Language Models
- Transformer Architecture
- Open-Source LLMs
- HuggingFace and Pre-trained Models
- Limitations of LLMs
- Advantages and Disadvantages of LLM Architectures
- Using Open-Source LLMs for Projects
- Responsible AI Development

23. PROMPT EGINEERING

- What is Prompt Engineering?
- Importance of Prompt Engineering
- Prompt Engineering Principles
- Prompt Design Strategies
- Types of Prompting
- Best Practices for Prompt Creation
- Parameter Tuning
- Prompting Techniques for Generative Models

24. WORD EMBEDDINGS & VECTOR DATABASES

- Word Embedding Introduction
- Word Embedding Techniques
- Capturing Word Relationships
- Sentence Embeddings

VECTOR DATABASES

- Introduction to Vector Databases
- Types of Vector Databases

25. CHUNKING MECHANISMS

- Chunking Concept
- Document Chunking
- Traditional Chunking
- Advanced Chunking
- Character Splitting
- Recursive Character Splitting
- Document-Based Chunking
- Semantic Chunking
- Agentic Chunking