Techno India NJR Institute of Technology
Data ScienceTraining Module

Total Time: 1 Months

Day 1: (2 Hours)

Introduction to Python Programming:

- Introduction to Python Features.
- Python Installation and Package Installation.
- Python Indentation and Comments.
- Python Variables and Keywords.
- Python Casting and Strings.
- Python Operators.
- Python Data Structures: Boolean Variables.
- Python Lists.
- Python Tuples.
- Python Dictionary.
- Python Sets.

Day 2: (2 Hours)

Introduction to Python Programming:

- Python Functions.
- Python Conditionals.
- Python Loops.
- Python Classes, Objects and Inheritance.
- Constructors, Destructors and More Examples.
- Python Assert and Exception.
- Python Eval Function, Zip and Date.
- 5 Python Use Cases.

Day 3:NumPy, Pandas and Matplotlib:

NumPy (0.5 Hours)

- Creating arrays and Data Types.
- Indexing and Slicing.
- Boolean and Fancy Indexing.
- Conditional Logic and Statistical Methods.
- File Input and Output with Arrays.
- Linear Algebra and Random Number Generation.

Pandas (1.5 Hours)

- Series and DataFrames.
- Indexing, Selecting, and Filtering.
- Sorting and Ranking.
- Descriptive Statistics.
- Hierarchical Indexing.
- Missing Values and Duplicate Value Detection.
- File Formats.
- Merging Datasets.
- Reshaping and Transformation.
- Group Operations.

Day 4: Matplotlib (1 Hour)

- Types of Plots.
- Legends and Titles.
- Scatterplots.
- Bar Charts and Histograms.
- Stack Plots.
- Pie Charts.
- Loading, Getting, and Converting Data.
- Customization of Colours and Styles.
- Subplot, Basemap, and 3D Bar Charts.

Introduction to Statistics, Probability (1 Hour)

- **Descriptive Statistics:**

  - Variables.
  - Qualitative and Quantitative Variables.
  - Univariate/Bivariate/Multivariate Data.
  - Measures of Central Tendency.
  - Measures of Variability.
  - Measures of Position.
  - Data Patterns of Spread.
  - Skewness and Kurtosis.
  - Outliers Detection in Data.

- **Estimation:**

  - Point Estimate and Interval Estimate.
  - Confidence Intervals.
  - Margin of Error.
  - Critical Value Estimation.

Day 5: Inferential Statistics (1 Hour)

- Hypothesis Testing.
- Decision Errors and Rules.
- Proportions.
- Difference Between Proportions.
- Hypothesis Test for Mean.
- Difference Between Means.
- Difference Between Paired Means.
- Regression Slope.
- Test of Variance/ANOVA.

Probability (1 Hour)

- Introduction to Probability.
- Probability Basic Rules.
- Rule of Multiplication and Addition.
- Random Variables.
- Discrete Probability Distribution.
- Continuous Probability Distribution.
- Independence of Random Variables.
- Linear Transformation of Random Variables.
- Simulation of Random Events.
- Central Limit Theorem.
- Sampling Distribution of Mean and Proportion.

Day 6: Probability (2 Hours)

- Binomial Distribution.
- Hyper-Geometric Distribution.
- Negative Binomial.
- Geometric Distribution.
- Poisson Distribution.
- Normal Distribution.
- T-Distribution.
- Chi-Square Distribution.

Day 7: Linear Algebra (2 Hours)

- Vectors and Linear Combinations.
- Vector Dot and Cross Products.
- Elimination Method for Using Matrices.
- Transformations and Matrix Multiplication.
- Finding Inverses and Determinants.
- Transpose of Matrix.
- Orthogonal Components.
- Orthonormal Bases and Gram-Schmidt Process.
  - 5 Industry Use-Cases.

Day 8: Data Mining(2 Hours)

- Data Mining Introduction.
- Sources of Data Mining.
- Data Mining Implementation Process.
- Knowledge Discovery in Databases.
- Data Mining Tools.
- Applications of Data Mining.
- 5 Industry Use-Cases.

Day 9: (2 Hours)

- Exploratory Data Analysis.
- Feature Engineering Process.
- 5 Industry Use-Cases.

Day 10: (2 Hours)

- Data Visualization Using Excel.
- Data Visualization Using Tableau.
- 5 Industry Use-Cases.

Day 11: (2 Hours)

Introduction to Machine Learning

- What is Machine Learning?
- Need for Machine Learning.
- Prerequisites.
- Types of Machine Learning.
- Supervised Learning and Unsupervised Learning.
- Reinforcement Learning.
- Life Cycle of Machine Learning.
- Data Processing Steps.
- 5 Industry Use-Cases.

Day 12: (2 Hours)

Supervised Learning: Regression

- Linear Regression.
- Logistic Regression.
- Polynomial Regression.
- Ridge Regression.

Day 13: (2 Hours)

- Lasso Regression.
- Support Vector Regression.
- Decision Tree Regression.
- Random Forest Regression.

Day 14: (2 Hours)

Supervised Learning: Classification

- Binary Classifier.
- Multiclass Classifier.
- Linear Models
    - Logistic Regression.
    - Support Vector Machines.

- 5 Industry Use-Cases.

Day 15: (2 Hours)

Supervised Learning: Classification

- Non-linear Models
    - K-Nearest Neighbours.
    - Kernel SVM.
    - Naïve Bayes.

Day 16: (2 Hours)

- Decision Tree Classification.
- Random Forest Classification.
- Gradient Descent.
- 5 Industry Use-Cases.

Day 17: (2 Hours)

Unsupervised Learning

- Clustering.
- Association Rules.
- Dimensionality Reduction.

Day 18: (2 Hours)

- Natural Language Processing.
- Survival Analysis.
- Time-Series Forecasting.

- 5 Industry Use-Cases.

Day-19: (2 Hours)

- Introduction to Deep Learning.
- Introduction to Neural Networks.

Day-20: (2 Hours)

- Gradient Descent and Derivatives.
- Vectorization Techniques.

Day-21: (2 Hours)

- Activation Functions.
- Backpropagation.

Day-22: (2 Hours)

- Deep Layer Neural Networks.
- CNN

Day-23: (2 Hours)

- Bias/Variance and Regularization.
- RNN

Day-24: (2 Hours)

- LSTM
- Gradient Checking Implementations.

Day-25: (2 Hours)

- Introduction to Reinforcement Learning.
- Deployment of ML Models.

Day-26: (2 Hours)

- Deployment of ML Models.

Day-27: (2 Hours)

- 2 Latest Industry Use - Cases.

Day-28: (2 Hours)

- 2 Latest Industry Use - Cases.

Day-29: (2 Hours)

- 2 Latest Industry Use - Cases.

Day-30: (2 Hours)

- 2 Latest Industry Use - Cases.

Day 31 to 37: Industry Use-Cases. (To be discussed after sign-up).

Day 38: Interview Preparation and Career Counselling.

Day 39: Interview Preparation and Career Counselling.

Day 40: Assessments.