

MACHINE LEARNING

Python Programming:

- Resources
- Installing Anaconda and Jupyter Notebook
- Opening Jupyter Notebook
- Basic Python Programming
- Exercise: Variables and Data Types
- Strings
- Exercise: Python Strings
- Tuples
- Lists
- Exercise: List and Tuples
- Sets
- Dictionary
- Exercise: Sets
- Operators in Python
- Input function
- Conditional statements [if, if-else, and elif Statements]
- Python Loops
- Comprehensions
- Python Functions
- Doc string, Local and global variables
- Functional Programming Concepts
- Exercise: 'if' 'for' loops

Data Manipulation with NumPy:

- Introduction to Numpy
- Creating a Numpy Array
- Operations on 1D array
- Multidimensional Array
- Array Manipulation
- Exercise: Numpy Functions
- Trigonometric and Linear algebra functions

Data Manipulation with Pandas:

- Introduction to Pandas
- Creating Data Frame
- Describing DataFrame
- Combining DataFrame
- Grouping a DataFrame
- Exercise: Data Manipulation using Pandas

Data Visualisation:

- Matplotlib
- Seaborn

Exploratory Data Analysis:

- EDA Introduction
- Graphical and Non Graphical
- Data Gathering
- Data Cleaning
- Data Cleaning part-2
- Exercise: Data Cleaning
- Univariate Analysis
- Bivariate Analysis
- Multivariate Analysis
- Exercise: Exploratory Data Analysis

Feature Engineering and Encoding:

- Feature Encoding
- Label Encoding
- One Hot Encoding
- One Hot Encoding using pandas
- One Hot Encoding using scikit-learn
- Feature Encoding (Repeated content)
- Exercise: Feature Encoding
- Outlier Documentation

Introduction to Machine Learning:

- Understanding Machine Learning
- Machine Learning and its types

Supervised Learning Models - Linear Regression:

- Linear Regression
- Math Behind Linear Regression
- Mean Absolute Error
- Mean Square Error and Root Mean Squared Error
- Difference between MAE-MSE and RMSE
- Linear Regression in Python
- Linear Regression on Excel
- Assumptions of Linear Regression
- Multiple Linear Regression
- Multicollinearity and Influence Analysis
- Lasso and Ridge Regularization
- Lasso and Ridge Regression
- Exercise: Linear Regression

Supervised Learning Models - Decision Tree:

- Introduction to Decision Trees
- Pruning a Tree
- Pruning a Tree in Python
- Credit Risk using Decision Tree
- Exercise: Decision Tree

Supervised Learning Models - Logistic Regression:

- Introduction to Logistic Regression
- Sigmoid Function
- Logit Function
- Introduction to Multivariate Analysis using Logistic Regression
- Cleaning Dataset for Logistic Regression: Part-1
- Cleaning Dataset for Logistic Regression: Part-2

- Data Preparation
- Model Building
- Confusion Matrix
- Manual Elimination
- Metrics
- ROC
- Logistic Regression Code Implementation
- Exercise: Logistic Regression

Supervised Learning Models - Classification (KNN):

- K-Nearest Neighbors Classifier
- KNN in Python Part 1
- KNN in Python Part 2
- Exercise: KNN Classifier

Supervised Learning Models - Model Evaluation and Metrics:

- Data Preparation
- Model Building
- Confusion Matrix
- Manual Elimination Features
- Metrics Beyond Accuracy
- ROC AUC
- Bias Variance Trade-off
- Exercise: Model Evaluation

Ensemble Technique:

- Introduction to Random Forest
- Random Forest
- Implementation of Random Forest
- Bootstrap Aggregation (Bagging)
- Ensemble Technique - Boosting
- Ensemble Technique - AdaBoost
- Ensemble Technique - XGBoost
- Exercise: Random Forest

Advance Machine Learning:

- Introduction to Cross Validation
- Introduction to K Fold CV
- Introduction to Stratified K Fold CV
- Implementation of Stratified K Fold CV
- Using Grid Search in Python
- Exercise: Random Forest Regression using Grid Search CV

AI - NLP, Neural Networks & Deep Learning:

- NLP Introduction
- Text Mining vs Text Analytics
- Introduction to NLTK
- Text Segmentation
- Text Word Tokenization
- Lemmatization
- Stop Words
- Dependency Parsing
- Bag of Words Model
- Restaurant Review with NLP
- Exercise: Spam Classification

Add-on Projects:

- Book Recommendation Feedback System
- IPL Probability Prediction
- Heart Disease Predictor

Mini Project: Develop a prediction model using Machine Learning

SQL

Installation and Demonstration:

- Installation
- Case Study Part 1
- Case Study Part 2

Theoretical concepts of SQL:

- Tables and DBMS
- Types of SQL commands
- PostgreSQL

Basic SQL commands:

- CREATE
- INSERT
- Import data from File
- SELECT statement
- SELECT DISTINCT
- WHERE
- Logical Operators
- UPDATE
- DELETE
- ALTER Part 1
- ALTER Part 2

Restoring DB:

- Restore and Back-up
- Debugging restoration
- Creating DB using CSV files

Selection commands: Filtering:

- Module Introduction
- IN
- BETWEEN
- LIKE

Selection commands: Ordering:

- Side Lecture: Commenting in SQL
- ORDER BY
- LIMIT
- AS

Aggregate Commands:

- COUNT
- SUM
- AVERAGE
- MIN MAX

Group By Commands:

- Group By
- HAVING
- CASE WHEN

JOINS:

- Module Introduction
- Joining concepts
- Preparing data for Joins
- Inner Join
- Left Join
- Right Join
- Full Outer Join
- Self Join
- Cross Join
- Intersect, Intersect all, Union all
- Except
- Union

Subqueries:

- Part 1 - Subquery in WHERE clause
- Part 2 - Subquery in FROM clause
- Part 3 - Subquery in SELECT clause

Views and Indexes:

- Views
- Index

String Functions:

- Module Introduction
- LENGTH
- UPPER LOWER
- REPLACE
- TRIM LTRIM RTRIM
- CONCATENATION
- SUBSTRING
- LIST AGGREGATION

Mathematical Functions:

- CEIL FLOOR
- RANDOM
- SETSEED
- ROUND
- POWER

Date-Time Functions:

- CURRENT DATE TIME
- AGE
- EXTRACT

Pattern(String) Matching:

- Pattern Matching Basics
- Regular Expressions Part 1
- Regular Expressions Part 2

Data Type conversion functions:

- Converting Numbers or Dates to String
- Converting String to Numbers or Dates

User Access Control Functions:

- User Access Control - Part 1
- User Access Control - Part 2

Windows Function:

- Introduction to Window functions
- Introduction to Row number
- Implementing Row number in SQL
- RANK and DENSERANK
- NTILE function
- AVERAGE function
- COUNT
- SUM TOTAL
- RUNNING TOTAL
- LAG and LEAD

Interview related theoretical concepts:

- Module Introduction
- Tablespace
- Primary key and Foreign Key
- ACID compliance
- Truncate
- Normalization
- First Normal Form
- Second Normal Form
- Third Normal Form and BCNF
- Fourth and Fifth Normal Form
- TIP 1 (EXPLAIN)
- TIP 2
- TIP 3
- TIP 4 (VACCUUM)
- TIP 5
- TIP 6 (STRING FUNCTIONS)
- TIP 7 (JOINS)
- TIP 8 (SCHEMAS)

Mini Project: Develop a project using SQL

TABLEAU

Introduction:

- Tableau Resources
- What is Tableau and why is it used
- Tableau products
- Installing Tableau Desktop and Tableau Public

Connecting with Data:

- About the data
- Connecting to data
- Live vs Extract
- Combining data from multiple tables introduction
- Relationships in tableau
- Joins in tableau
- Types of joins
- Union in tableau
- Physical logical layer and Datamodels

Making Visualizations in Tableau:

- The visualization screen - Sheet
- Types of data in tableau Part 1 (Dimensions and Measures)
- Types of data in tableau Part 2 (Discrete Vs Continuous)
- Changing data type in tableau
- Bar chart
- Line chart
- Scatter plot

Customizing Visualizations using Marks card:

- Marks Card
- Dropping dimensions and measures on the marks card
- Dropping dimensions on Line Chart
- Adding marks in scatterplot

Other important viz in Tableau:

- Text tables, heat map and highlight tables
- Pie charts
- Area charts
- Creating Custom hierarchies
- Tree map
- Dual Combination Charts

Binning and Grouping data:

- Creating Bins
- Histogram
- Grouping data

Filtering and Sorting data:

- Filtering data
- Dimension filters
- Measure filters
- Date-Time filters
- Filter Options
- Types of filters and order of operation
- Customizing Visual Filters
- Sorting Options

Geo Map Charts:

- How to make a Map chart
- Considerations before making a Map chart
- Marks Card for Customizing Maps
- Layers in a Map
- Customizing Maps using Map Menu
- Visual Toolbar on a Map
- Custom Background Images
- Territories in Maps
- Data blending for Missing Geocoding

Calculations and Analytics:

- Calculated Fields in Tableau
- Functions in Tableau
- Table Calculation Theory
- Table Calculations in Tableau
- Level of Detail Expressions -1
- Level of Detail Expressions -2
- LOD Expressions Example
- Analytics Pane - 1
- Analytics Pane - 2

Parameters and Sets:

- Parameters
- Understanding Sets in Tableau
- Creating Sets in Tableau

Dashboards:

- Creating a Dashboard
- Dashboard Actions for Interactivity
- Story

Appendix:

- Adding SQL Data Source
- Adding a Cloud-Based Data Source

Mini Project: Develop a project using Tableau

EXCEL AND VISUALIZATION

Learning Basics:

- Excel Intro
- Worksheet Basics
- Entering Values Formulas
- Data Format
- Data handling basics cut copy paste
- Saving Printing basics
- Basic Formula Operations

Important Functions and Formulas:

- Mathematical Formula I
- Mathematical Formula II
- Textual Function I
- Textual Function II
- Logical-Functions
- Date Functions
- Lookup and Reference Functions I [VLOOKUP, MATCH, INDEX]
- Lookup and Reference Functions II [VLOOKUP, MATCH, INDEX]

Data Tools:

- Data Tools I
- Data Tools II
- Data Format
- Pivot Table

Introduction To Charts:

- Importance of Data Visualization
- Categories of messages to be conveyed
- Elements of Charts
- Easiest way to create charts

Important Chart Types:

- Bar Column Chart
- Chart Format I
- Chart Format II
- Line Chart
- Area Chart
- Pie Doughnut Chart
- Reasons to avoid pie chart
- Scatter Chart I
- Scatter Chart II
- Frequency Distribution Histogram
- WaterFall Chart
- Sparklines
- Heat Map
- Charts
- Pivot Charts

Advance Topics:

- Named Ranges
- Indirect Function
- Macros
- Shortcuts
- Advance Excel
- Data Table
- Goal Seek

New Additions in Excel:

- Importing Data From Website
- Maps

Use Cases:

- Infographics I
- Infographics II
- Transportation
- Tax

Introduction to Statistics:

- Introduction to Statistics
- Data Statistics
- Population VS Sample Statistics
- Sampling Techniques
- Descriptive & Inferential Statistics
- Measures of Central Tendency
- Mode and Skewness
- Variance
- Standard Deviation
- Coefficient of Variation
- Range and IQR
- Bi-Variate & Multi-Variate Analysis

Key Statistical Methods to Know:

- Probability statistics
- Permutation & Combination statistics
- Event Statistics
- Probability Mass function and Cumulative Distribution Function
- Binomial Distribution
- Poisson Distribution
- Continuous Probability Distribution
- Uniform Distribution
- Normal & Standard Distribution
- Central Limit Theorem
- Degrees of Freedom
- Hypothesis Testing
- Confidence Level & Confidence Interval
- Critical Values
- Examples on Hypothesis
- Type I & Type II Errors
- Significance Level and P-Value
- Two Sample Independent t-Test
- One Sampled t-Test
- ANOVA Statistics
- Chi-Square Statistics
- AB Testing

Mini Project: Analyze data using Excel

MODULE PROJECTS

Introduction: In this module, we will have four hands-on projects — Excel Module, SQL Module, Tableau Module, and ML Module.

These projects are designed to evaluate your practical understanding of key concepts from each module. Successfully clearing these will bring you one step closer to entering the placement process.

Excel Module Project:

- Covers basic to intermediate Excel functionalities.
- Tests include data formatting, formula creation, data handling, and chart preparation.
- Tasks simulating real-world business scenarios.

SQL Module Project:

- Evaluates your ability to write and optimize SQL queries.
- Topics include SELECT, JOINS, GROUP BY, and subqueries etc.
- Case study-based question to mimic real-time data challenges.

Tableau Module Project:

- Assess your dashboard building and data visualization skills.
- Includes tasks like creating bar charts, filters, maps, and interactive dashboards etc.
- Scenario-based tests to reflect real-world analytics tasks.

ML Module Project:

- Tests core Python and Machine Learning understanding and implementation.
- Include questions on data preprocessing, EDA and model building.
- Build prediction models from real world datasets.