

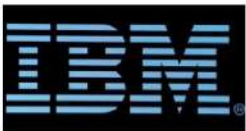


Rocket Data Analytics operates as an affiliate of Data Minds Analytics pvt ltd

DATA SCIENCE

Professional Course

In Collaboration with

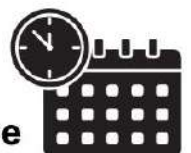


LIVE



CLASSROOM
ONLINE
HYBRID MODEL

3 MONTHS- Certificate Course



MODULE 01

PYTHON PROGRAMMING



FUNDAMENTALS

Python - Basics

- Historical Context
- Python Installation
- Overview of Integrated
- Development Environments (IDEs)
- Basics: Identifiers, Statements,
- Comments, Variables
- Memory Management
- Types of Data Types
- Integers, Float, Complex,
- Boolean, String
- Operators :
- Arithmetic, Relational,
- Logical, Assignment, Bitwise

Input, Output, and Import

- Python Input and Output
- Importing Modules

Namespaces and Scope

- Python Namespace and Scope
- Global, Local, and
- Non-local Variables

Control Flow

- Python Flow Control Statements
- if Statement
- if-else Statement
- if-elif-else Statement
- Nested if Statement
- for Loop
- while Loop
- break, continue &
- pass Statements

Date and Time

- Working with Dates and Times
- Formatting Dates and Times
- Current Date and Time
- Timestamps to Datetime Conversion
- Timing Functions with the
- time Module
- Introducing the sleep Function

DATA STRUCTURES

Lists-Introduction

- Creating and accessing elements
- List Operations
- Modifying, appending
- deleting elements
- List comprehensions

Tuples-Introduction

- Tuple Operations:
- Accessing elements
- Concatenation and repetition

Sets - Introduction

- Set Operations:
- Intersection, union, and difference
- Membership testing

Dictionaries -Introduction

- Key-value pairs and uniqueness.
- Creating and accessing elements.
- Dictionary Operations:
- Modifying, deleting, and
- iterating over elements.
- Dictionary comprehensions

String Basics

- String manipulation methods.
- String Operations:
- Concatenation, slicing, and formatting
- Common string functions

ADVANCE

Functions

- Python Functions
- Defining and Calling Functions
- Global Keyword
- Function Arguments
- Recursion
- Anonymous (Lambda) Functions
- Modules and Packages
- Working with random Module
- Working with math Module
- Documenting Functions
- (Docstrings)
- User-Defined Functions

File Handling

- File Input and Output (I/O)
- File Handling Techniques
- Working with Directories
- Handling Exceptions with Files
- User-Defined Exceptions

Object-Oriented Program

- OOP Concepts
- Classes and Objects
- Inheritance
- Operator Overloading

Advanced Topics

- Iterators
- Generators
- Closures
- Decorators
- *args and **kwargs
- Properties
- Regular Expressions (RegEx)

MODULE 02



NUMPY

Introduction to NumPy

- What is NumPy?
- History of NumPy

Understanding ndarrays

- What is an ndarray (N-dimensional array)?
- Creating NumPy arrays
- Array functions
- Numerical arrays
- Homogeneous arrays
- Diagonal arrays
- Random number generation with NumPy

Array Attributes

- Overview of array attributes
- Exploring and understanding
- array attributes

Multi-Dimensional Arrays

- Creating multi-dimensional arrays
- Extracting data from
- multi-dimensional arrays

Indexing and Slicing

- Basics of indexing
- Basics of slicing
- Boolean indexing
- Random indexing

Reshaping and Resizing

- Reshaping arrays
- Resizing arrays
- Transposing arrays

Vector Operations &

Array Functions

- Vector multiplication
- Array operations
- Broadcasting rules

PANDAS

Introduction to Pandas

- What is Pandas?
- History and evolution of Pandas

Series and DataFrames

- Introduction to Series and DataFrames
- Creating Series and DataFrames
- Essential operations on Series and DataFrames

Data Structures in Pandas

- Understanding the various data structures in Pandas
- Working with Series, DataFrame, and Index objects

Data Manipulation with Pandas

- Loading and saving data in
- different formats (CSV, Excel, SQL)
- Data cleaning and handling missing values
- Data transformation and manipulation techniques

Indexing and Selection

- Basics of indexing and selecting data in Pandas
- Advanced indexing techniques
- Boolean indexing

Grouping and Aggregation

- Grouping data with Pandas
- Aggregating data using different functions
- Transformations and filtering within groups

Merging and Joining DataFrames

- Combining DataFrames using merge and join operations
- Concatenating DataFrames

Time Series Analysis with Pandas

- Handling time and date data
- Resampling and frequency conversion
- Time shifting and lagging

Data Visualization with Pandas:

- Plotting with Pandas
- Exploratory data analysis using visualizations

Advanced Topics in Pandas:

- Handling categorical data

MODULE 03



MATPLOTLIB

Overview of Matplotlib

Matplotlib Basics

- Installing Matplotlib
- Basic plotting with Matplotlib
- Line plots, scatter plots, and bar plots

Customizing Matplotlib Plots

- Adding titles and labels
- Setting colors and styles
- Adding legends and annotations

Multiple Plots and Subplots

- Creating multiple plots in a single figure
- Working with subplots

Advanced Matplotlib Plots:

- Histograms
- Box plots
- Pie charts
- 3D plots
- Violin plots

SEABORN

Overview of Matplotlib

Seaborn Introduction

- Installing Seaborn
- Overview of Seaborn's capabilities

Seaborn Plots for Univariate Data

- Distplot for distribution visualization
- Countplot for categorical data
- Boxplot and violin plot for summarizing distributions

Seaborn Plots for Bivariate Data:

- Scatter plots and regression plots
- Pair plots for pairwise relationships
- Heatmaps for correlation visualization

Styling and Customization in Seaborn

- Themes and color palettes
- Customizing Seaborn plots

Additional Seaborn Features

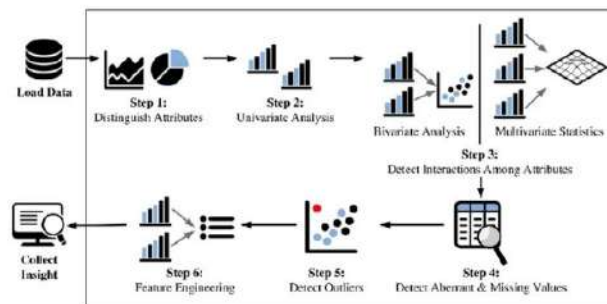
- FacetGrid for multi-plot grids
- Categorical plots for complex categorical relationships

Real-world Data Visualizations

MODULE 04



DATA SCIENCE Exploratory Data Analysis



Exploratory Data Analysis

- Understanding the impact of EDA on decision-making
- Identifying patterns and trends in the data
- Assessing normality and skewness

Type castings

- Converting data types for compatibility and efficiency
- Addressing issues with incorrect data types

Outlier Analysis and Treatment

- Identifying outliers through statistical methods & visualization
- Understanding the impact of outliers on analysis
- Outlier Treatment
- Techniques for handling outliers (e.g., trimming)
- Deciding when to remove or transform outliers

Discretization / Binning / Grouping

Missing Values

- Identifying Missing Values
- Techniques for recognizing and quantifying missing data
- Understanding the reasons behind missing values
- Imputation Techniques
- Strategies for filling or imputing missing values
- appropriate imputation method based on the data

Introduction to Data Preprocessing

- Overview of Data Preprocessing

Handling Duplicates

- Identifying and removing duplicate records
- Strategies for handling duplicate values

Data Exploration Techniques

- Univariate, bivariate,
- Histograms, box plots, scatter plots
- Multivariate analysis
- Correlation and covariance analysis
- Heatmaps and pair plots

Feature Engineering

- Creating new features for better model performance
- Techniques such as encoding, scaling & transformations

Encoding: Dummy Variable Creation

- Creating dummy variables for categorical data
- Handling multicollinearity issues

Scaling: Standardization Normalization

- Scaling features to a standard normal distribution
- Considering the suitability of normalization for different algorithms

Communication and Reporting

- Communicating EDA Findings
- Creating clear and compelling data narratives
- Designing visualizations for effective communication
- Documenting and sharing EDA insights
- Preparing reports for stakeholders and team members



MODULE 06

MACHINE LEARNING



Foundations of Machine Learning

Introduction to Machine Learning

- Overview of machine learning
- Setting up development environments (Python, Jupyter, sklearn libraries)

Supervised Learning

- Definition of supervised learning
- Explanation of the difference between supervised and unsupervised learning.

Regression

- Correlation
- Scatter Diagram
- Correlation coefficient
- Correlation analysis
- Correlation coefficient
- Regression

Linear Regression

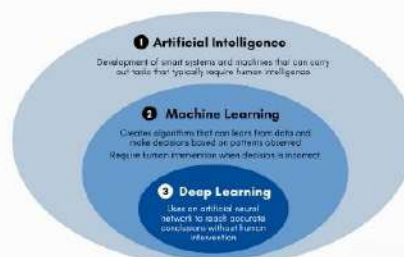
- Simple Regression
- Linear Equation - coefficients, intercept
- Residuals, Least Squares Method:
- Assumptions of Linear Regression
- Homoscedasticity, Heteroscedasticity
- Multicollinearity
- Polynomial Regression

Naive Bayes

- Probability
- Bayes Rule
- Naïve Bayes Classifier



ARTIFICIAL INTELLIGENCE VS MACHINE LEARNING VS DEEP LEARNING



Math Fundamentals

- Python programming basics
- Linear algebra and calculus essentials

Model Selection

LIBRARIES

Classification

- Definition of classification.
- Understanding the concept of class labels.
- Binary and multiclass classification.

Logistic regression

- Types of Logistic regression
- Logit and Log-Likelihood
- Sigmoid function
- Analysis of logistic regression results
- Multiple Logistic regression
- Evaluation metrics
- Confusion matrix
- AUC / ROC for binary classifier

k-Nearest Neighbors (k-NN)

- Deciding the K value
- Thumb rule in choosing the K value.
- Building a KNN model by
- splitting the data

MODEL EVALUATION METRICS

- Mean Squared Error
- Root Mean Squared Error
- RSquare
- Adjusted RSquare
- Mean Absolute Error
- Mean Absolute Percentage Error
- Confusion matrix
- Accuracy, Precision,
- Recall, F1 Score
- Sensitivity, Specificity
- False Positive, False Negative
- True Positive, True Negative
- AUC - Area Under Curve
- ROC - (Receiver operating characteristics)

MODULE 06



MACHINE LEARNING



Decision Tree

- Elements of classification tree
- Root node, Child Node, Leaf Node, etc.
- Greedy algorithm
- Measure of Entropy
- Information gain
- Gini Index
- Decision Tree C5.0 and understanding
- Various arguments
- Checking for Underfitting and
- Overfitting in Decision Tree
- Pruning – Pre and Post Prune techniques
- Generalization and Regulation Techniques to avoid overfitting in Decision Tree

Kernel Method - SVM

- Hyperplanes
- Best Fit "boundary"
- Linear Support Vector Machine using
- Maximum Margin
- SVM for Noisy Data
- Non- Linear Space Classification
- Non-Linear Kernel Tricks
- Linear Kernel
- Polynomial, Sigmoid, Gaussian RBF
- SVM for Multi-Class Classification
- One vs. All, One vs. One
- Directed Acyclic Graph (DAG) SV

Advanced Topics in Machine Learning

Ensemble Techniques

- Ensemble Learning Introduction:
- Types of Ensemble Methods
- Bagging (Bootstrap Aggregating)
- Boosting
- Stacking

Cross-validation Techniques

- Need for Cross-Validation
- K-Fold Cross-Validation
- Leave One Out Cross Validation
- Stratified cross validation method
- Cross-Validation for Hyperparameter Tuning
- Grid search and randomized search
- For optimal hyperparameters.
- Model Selection and Comparison
- Extension to Monte Carlo Cross-Validation

Bagging - Random Forest

- Introduction to Random Forest
- Random Forest Works - Majority Voting
- Advantages of Random Forest - oob
- Random Forest Variations
- Comparison with Other Algorithms
- Evaluation and Interpretability

Boosting- AdaBoost

- Motivation for Boosting
- Feature Importance in Boosting
- Weak Learners and Strong Learners
- Decision Stump
- Assigning weights to misclassified instances
- Over fitting & Under fitting Challenges

Optimization Techniques

- Loss Functions
- Gradient Descent
- Learning Rate and Hyperparameter Tuning
- Regularization Techniques:
- L1 (Lasso) and L2 (Ridge) regularization
- Early Stopping
- Hyperparameter Search Strategies
- Bias-Variance Tradeoff

Gradient Boosting

- Gradient Boosting
- Components of Gradient Boosting
- loss function, weak learners
- Loss Functions and Error Minimization
- Strategies for dealing with noisy data

Applications and Use Cases Live Projects

XGBoost, LightGBM and CatBoost:

MODULE 06

MACHINE LEARNING



*"predicting the future is'nt magit,
its Artificial Intelligence ! "*

Introduction to Unsupervised Learning:

- Definition of unsupervised learning and its contrast with supervised learning.
- Explanation of learning from unlabeled data and discovering inherent patterns.

Clustering

- Distance Metrics
- k-Means clustering
- Hierarchical Clustering
- Non-Hierarchical Clustering DBSCAN
- Clustering Evaluation metrics

K-Means Clustering:

- In-depth coverage of the K-means algorithm, its initialization methods, and convergence
- Practical implementation and examples.

Association Rules

- Association rules mining
- Market Basket Analysis
- Apriori Algorithm, Fp Growth
- Metrics - Support, Confidence, Lift

Recommender Systems

- User Based Collaborative Filtering
- Similarity Metrics
- Item Based Collaborative Filtering
- Search Based Methods
- SVD Method

Dimensionality Reduction:

Principal Component Analysis (PCA):

- In-depth coverage of PCA, including eigenvalue decomposition and feature extraction.
- Applications in reducing dimensionality.

t-Distributed Stochastic Neighbor Embedding (t-SNE):

- Explanation of t-SNE and its use for visualizing high-dimensional data.
- Comparison with other dimensionality reduction techniques.

Anomaly Detection

- Discussion of anomaly detection as a task in unsupervised learning.
- Techniques for identifying unusual patterns.

Natural Language Processing (NLP)

Tokenization and text processing

Introduction to language models

- Text Mining and Natural Language Processing (NLP)
- Sources of data
- Bag of words
- Pre-processing, corpus Document
- Term Matrix (DTM) & TDM
- Word Clouds
- Corpus-level word clouds
- Sentiment Analysis
- Positive Word clouds
- Negative word clouds
- Unigram, Bigram, Trigram
- Semantic network
- Extract, user reviews of the product/services from Amazon and tweets from Twitter
- Install Libraries from Shell
- Extraction and text analytics in Python
- LDA / Latent Dirichlet Allocation
- Topic Modelling
- Sentiment Extraction
- Lexicons & Emotion Mining

Applications and Use Cases Live Projects



ABOUT US



Rocket Data Analytics operates as an affiliate of Data Minds Analytics pvt ltd

Our professional courses are instructed by industry experts actively engaged in real-time practices, utilizing the latest teaching tools and techniques. The combination of our Learning Management System (LMS) and dedicated support mentors constitutes key elements that facilitate easy and simplified learning

- 3000 + Successfully trained students
- 1056 + Facilitated career transitions
- 10+ Industry 4.0 diverse range of Digital Transformation courses
- Flexible training options, including Classroom, Online, E-Learning, and Corporate Trainings.

DATA MINDS ANALYTICS PVT LTD

PROFESSIONAL COURSES

- | | | |
|---------------------|-------------------|---------------------------|
| ✓ Data Science | ✓ Data Analytics | ✓ Data Engineering |
| ✓ Gen - AI | ✓ Block Chain | ✓ Artificial Intelligence |
| ✓ Python Full Stack | ✓ Java Full Stack | ✓ R Language |
| ✓ Power BI / Apps | ✓ Aws / Devops | ✓ Azure Cloud |
| ✓ Cyber Security | ✓ ERP - SAP | ✓ Digital Marketing |

Our Collaborations & Placements

We foster meaningful collaborations with industry partners to enhance opportunities for our students. Our commitment to facilitating placements ensures that our graduates seamlessly transition into promising career paths, establishing a strong foundation for their professional journey.

 **accenture**



 **amazon**

 **cognizant**

Deloitte.



 **Google**



Infosys

