## **Academy of Information Technology**

Enhance your Data Science Carrer

Tools Covered















### Objective:

- The objective of Data Science is to extract actionable insights and knowledge from large and complex datasets.
- It employs statistical analysis, machine learning algorithms, and data visualization techniques.
- Data science aims to uncover patterns, trends, and relationships in the data. The goal is to drive informed decision-making and solve real-world problems.
- The objective is to transform raw data into valuable information. This information can be used to optimize processes, improve efficiency, and gain a competitive advantage.
- Data Science bridges the gap between data and knowledge.
- It enables organizations to make data-driven decisions and predictions
- This can lead to better outcomes and innovation.

# • Key Features in the Training Duration:

- 6 Months Class Duration: 1.30-Hrs (Monday to Saturday)
- Help on Doubt Clearance, Monitoring Session, Career Guidance, Interviewpreparation & Mock interviewsProjects materials are provided with Lab Exercises, Data sets, Codes, Quizzes, Case studies on real data.

# Modules

- 1) Database (SQL,Excel)
- 2) Data Analiysis
- 3) Python
- 4) Tableau
- 5) Power BI
- 6) Machine Learning



# Python Introduction and Setting Up the Environment

- ✓ Introduction to programming
- R or Python?
- Why Python for Data Science?
- Different job roles with Python
- ✓ Different Python IDEs
- Downloading and setting up the Python environment

Hands-On - Installing Python - IDLE



#### **Python Basic Syntax and Data Types**

- ✓ Python input and output operations.
- ✓ Comments
- ✓ Variables, rules for naming variables
- ☑ Basic Data Types in Python
- ✓ Typecasting in python

Hands-On - Using comments, variables, data types, and typecasting in python program



#### **Operators in Python**

- ✓ Arithmetic operators
- Assignment operators
- Comparison operators
- ✓ Logical operators
- Identity operators
- Membership Operators
- ✓ Bitwise Operators

Hands-On - Working with different data types in a program



#### **Strings in Python**

- Creating strings
- String formatting
- Indexing
- ✓ Slicing
- ✓ String methods

Hands-On - Performing string operations



#### Lists

- ✓ Creating lists
- ✓ Properties of lists
- ✓ List indexing
- ✓ List slicing
- ✓ List of lists
- ✓ List Methods
- ✓ Adding, Updating & removing elements from lists

Hands-On - Slicing, Indexing, and using methods on lists



- ✓ Syntax to create tuples
- ✓ Tuple properties
- ✓ Indexing on tuples
- ✓ Slicing on tuples
- ✓ Tuple methods

#### **Hands-On - Working with Tuples**



#### Sets

- ✓ The syntax for creating sets
- Updating sets
- Set operations and methods
- ☑ Difference between sets, lists, and tuples

#### Hands-On - Performing set operations in a program



#### **Dictionaries**

- ✓ The syntax for creating Dictionaries
- Storing data in dictionaries
- Dictionaries keys and values
- Accessing the elements of dictionaries
- Dictionary methods

#### Hands-On - Creating dictionaries and using dictionaries methods



#### **Python Conditional Statements**

- Setting logic with conditional statements
- ✓ If statements
- If -else statements
- ✓ If-elif-else statements

Hands-On - Setting logic in programs using conditional statements

#### **Loops in Python**

- Iterating with python loops
- while loop
- ✓ for loop
- range
- ✓ break
- ✓ continue
- pass
- ✓ enumerate
- ✓ zip
- assert

Hands-On - Iterating with loops in python

#### Getting Started with HackerRank use cases and working on them

- ✓ Solving Level by Level Challenges
- ✓ Assignments to acquire Bronze and Silver Level badges

#### **List and Dictionaries comprehension**

- Why List comprehension
- ✓ The syntax for list comprehension
- ✓ The syntax for dict comprehension

Hands-On - Using List and Dictionary comprehension



### **Functions**

- ✓ What are Functions
- Creating functions
- Passing Arguments
- ✓ Keyword Arguments
- ✓ Variable Keyword length arguments (\*\*kargs) ✓ Return keyword in python
- ✓ Passing function as an argument
- ✓ Global and local variables

- ✓ Modularity and code reusability
- Calling functions
- ✓ Positional Arguments
- √ Variable-length arguments (\*args)
- Passing function in return
- **√** Recursion



- ✓ Lambda
- ✓ Lambda with filter
- ✓ Lambda with map
- ✓ Lambda with reduce

Hands-On - Working with lambda, filter, map, and reduce in python



✓ Creating and using generators

Hands-On - Creating and using generators



- Creating modules
- ✓ Importing functions from a different module
- ✓ Importing Variables from different modules
- Python built-in modules

**Hands-On - Creating and importing Modules** 



#### **Exceptions and Error handling**

- Syntax errors
- ✓ Logical errors
- $\checkmark$  Handling errors using try, except and finally

Hands-On - Handling Errors with try and except



#### **Classes and Objects (OOPS)**

- ✓ Creating classes & Objects
- ✓ Understanding \_\_init\_\_ constructor method
- ✓ Different types of methods
- ✓ Class methods
- ✓ Inheritance
- Overriding parent methods
- ✓ Understanding Types of inheritance
- ✓ Multiple Inheritance
- ✓ Polymorphism

- Attributes and methods
- ✓ Class and instance attributes
- ✓ Instance methods
- ✓ Static methods
- Creating child and parent class
- ✓ The super() function
- ✓ Single inheritance
- ✓ Multilevel Inheritance
- Operator overloading

Hands-On - Creating classes, objects. Creating methods and attributes. Working with different methods. Using inheritance and polymorphism.



#### **Date and Time**

- ✓ date module
- ✓ datetime module
- formatting date and time
- ✓ strftime()

Hands-On - working with date and time



#### Regex

- Understanding the use of regex
- ✓ re.search()
- ✓ re.compile()
- ✓ re.find()
- ✓ re.split()
- re.sub()
- ✓ Meta characters and their use

Hands-On - Using regular expression to search patterns



#### ✓ Opening file

- ✓ Opening different file types
- Read, write, close files
- ✓ Opening files in different modes

Hands-On - Reading, Writing, Appending, opening, and closing files.



#### APIs the Unsung Hero of the Connected World

- ✓ Introduction to APIs
- Accessing Public APIs

Hands-On - Accessing Public Weather APIs and People in Space API



#### Python for Web Development - Flask

- ✓ Introduction to Python Web Framework Flask
- ✓ Installing Flask
- ☑ Working on GET, POST, PUT, METHODS using Python Flask Framework
- ☑ Working on Templates, render\_template function



#### **Hands-On Projects**

- Web Scraping- Dynamic Website with multiple pages along with Data Analysis
- ✓ Sending Automated Emails
- ☑ Building a Virtual Assistant with Frontend Interface



## Module 2 : Data Analysis



#### **Packages**

- Creating packages
- ✓ Importing modules from the package
- Different ways of importing modules and packages
- Working on Numpy, Pandas, and Matplotlib

Hands-On - Creating and importing



#### Web Scraping

- Introduction to web scraping: Tools, libraries, and ethical considerations
- Scraping data from websites using libraries like BeautifulSoup and requests: HTML parsing, locating elements, and extracting data
- Handling different types of data on websites: Tables, forms, etc.
- ✓ Storing scraped data in appropriate formats: CSV, JSON, or databases

Hands-On - Working on Scraping Data from Static Dynamic Websites



# Exploratory data analysis (EDA) using Pandas and NumPy

- ✓ Introduction to Pandas, a Python library for data manipulation and analysis.
- ✓ Overview of NumPy, a fundamental package for scientific computing with Python.
- ☑ Explanation of key data structures in Pandas: Series and DataFrame.
- ☑ Hands-on exploration of data using Pandas to summarize, filter, and transform data.
- ☑ Data cleaning techniques, handling missing values, and dealing with outliers.
- ✓ Statistical analysis of data using NumPy functions



# Data Visualization using Matplotlib, Seaborn, and Plotly

- Introduction to data visualization and its importance in data analysis.
- Overview of Matplotlib, a popular plotting library in Python.
- Exploring different types of plots: line plots, scatter plots, bar plots, histogram, etc.
- ✓ Customizing plots with labels, titles, colors, and styles.
- ✓ Introduction to Seaborn, a Python data visualization library based on Matplotlib.
- Advanced plotting techniques with Seaborn: heatmaps, pair plots, and categorical plots.
- Introduction to Plotly, an interactive plotting library for creating web-based visualizations.
- $\checkmark$  Creating interactive and dynamic visualizations with Plotly.

Hands-on: Instagram Reach Analysis



#### **Database Access**

- ✓ Introduction to databases.
- ✓ Why SQL?
- ✓ Execution of an SQL statement.
- ✓ Installing MySQL
- ✓ Load data.
- ✓ Use, Describe, Show table.
- ✓ Select.
- ✓ Limit, Offset.
- ✓ Order By.
- ✓ Distinct.
- ✓ Where, Comparison Operators, NULL.
- ✓ Logic Operators.
- ☑ Aggregate Functions: COUNT, MIN, MAX, AVG, SUM.
- ✓ Group By.
- ✓ Having.
- Order of Keywords.
- ✓ Join and Natural Join.
- ✓ Inner, Left, Right, and Outer Joins.
- ✓ Sub Queries/Nested Queries/Inner Queries.
- ✓ DML: INSERT
- ✓ DML: UPDATE, DELETE
- ✓ DML: CREATE,TABLE
- ✓ DDL: ALTER, ADD, MODIFY, DROP
- ✓ DDL: DROP TABLE, TRUNCATE, DELETE
- ✓ Data Control Language: GRANT, REVOKE

Hands-on - Working on SQL Queries



- Excel Introduction
- ✓ Workbook Window
- ✓ Create & Open Workbooks
- ✓ MS Excel Online
- ✓ Excel vs Google Sheets
- ✓ Office Button
- ✓ Ribbon and Tabs
- ✓ Features of Tabs
- Quick Access Toolbar
- ✓ Mini Toolbar
- ✓ Title, Help, Zoom, View



#### **Excel Worksheet**

- Worksheet, Row, Column
- ✓ Moving on Worksheet
- ✓ Enter Data
- Select Data
- ✓ Delete Data
- ✓ Move Data
- ✓ Copy Paste Data
- ✓ Spell Check
- ✓ Insert Symbols



#### **Excel Calculation**

- ✓ Addition
- Sigma Addition
- ✓ Subtraction
- ✓ Calculate Average
- ✓ Sigma Average



#### **Excel Fill Handle**

- ✓ Fill Handle
- **▼** Fill Handle with Text
- Text with Numbers
- ✓ Fill Handle with Dates



#### **Excel Formula**

- Fill Handle in Formula

- ✓ Instruction for Typing

# 0

#### **Quick Excel Functions**

- Excel IF
- ✓ If Function
- ✓ If with Calculations
- Excel COUNTIF
- Advanced If
- ✓ WHAT IF Analysis



#### **Excel Charts and visualizations**

- ✓ Introduction to Excel Charts
- ✓ Dynamic Advanced Charts
- ✓ Pivot Table with Dashboard
- Advanced Pivot Table Tips & Tricks



#### **Excel Advanced**

- Excel Macros
- Excel sumif
- Excel vlookup
- Excel ISNA
- ✓ Find & Remove Duplicates
- Create drop-down List
- ✓ Merge cells in Excel

# Tableau

- ☑ Building bar charts and line charts
- Creating pie charts and scatter plots
- ✓ Designing basic maps and geographic visualizations
- ✓ Using filters to subset data
- ✓ Sorting data by different criteria
- Applying quick filters for interactive exploration
- ✓ Adding labels, tooltips, and colors to visualizations
- Formatting axes and gridlines
- ✓ Customizing visual elements for better presentation
- ✓ Combining multiple visualizations into a dashboard

- ✓ Publishing dashboards to Tableau Public or Tableau Server
- ✓ Embedding dashboards in websites or presentations
- ✓ Presenting and sharing dashboards effectively



#### Power BI

- ✓ Overview of Power BI and its features
- ✓ Understanding the Power BI interface
- Connecting to data sources
- Importing and transforming data
- Creating bar charts and line charts
- Designing pie charts and scatter plots
- Building basic tables and matrices
- Using filters and slicers to subset data
- Adding interactivity to visualizations
- ✓ Sorting and formatting data
- Building interactive dashboards with multiple visualizations
- Adding filters and slicers for user interactivity
- ✓ Formatting and organizing dashboard elements
- ✓ Publishing reports to the Power BI Service
- Sharing reports and dashboards with others
- Configuring security and access controls

Hands-on: Instagram Reach Analysis



### Module 3: Statistics



#### **Descriptive Statistics**

- Data-types of data
- A measure of central tendency Mean-Median-Mode
- A measure of shape Variance- Standard deviation, Range, IQR
- The measure of shape Skewness, and kurtosis
- Covariance
- ✓ Correlation Pearson correlation & Spearman's rank correlation
- Probability Events, Sample Space, Mutually exclusive events, Mutually exclusive events
- ✓ Classical and Conditional Probability
- ✓ Probability distribution Discrete and Continuous
- ✓ Uniform Distribution
- 🗹 Expected values, Variance, and means
- ✓ Gaussian/Normal Distribution
- Properties, mean, variance, empirical rule of normal distribution
- Standard normal distribution and Z-score



#### **Inferential Statistics**

- ✓ Central Limit Theorem
- ☑ Hypothesis testing Null and Alternate hypothesis
- ▼ Type I and Type II error
- ✓ Critical value, significance level, p-value
- ✓ One-tailed and two-tailed test
- ✓ T-test one sample, two-sample, and paired t-test
- √ f-test
- ✓ One way and two way ANOVA
- ✓ Chi-Square test



## Module 4: Machine Learning



#### Introduction to Machine Learning

- Introduction to Machine Learning and its types (supervised, unsupervised, reinforcement learning)
- Setting up the development environment (Python, Jupyter Notebook, libraries: NumPy, Pandas, Scikit-learn)
- Overview of the Machine Learning workflow and common data preprocessing techniques



#### Introduction to data science and its applications

- Definition of data science and its role in various industries.
- Explanation of the data science lifecycle and its key stages.
- Overview of the different types of data: structured, unstructured, and semi-structured.
- Discussion of the importance of data collection, data quality, and data preprocessing..



#### **Data Engineering and Preprocessing**

- ☑ Introduction to Data Engineering: Data cleaning, transformation, and integration
- ☑ Data cleaning and Handling missing values: Imputation, deletion, and outlier treatment
- Feature Engineering techniques: Creating new features, handling date and time variables, and encoding categorical variables
- ✓ Data Scaling and Normalization: Standardization, min-max scaling, etc.
- ✓ Dealing with categorical variables: One-hot encoding, label encoding, etc.



#### Model Evaluation and Hyperparameter Tuning

- Cross-validation and model evaluation techniques
- ☑ Hyperparameter tuning using GridSearchCV and RandomizedSearchCV
- ✓ Model selection and comparison



#### **Supervised Learning - Regression**

- ✓ Introduction to Regression: Definition, types, and use cases
- Linear Regression: Theory, cost function, gradient descent, residual analysis, Q-Q Plot, Interaction Terms, and assumptions
- Polynomial Regression: Adding polynomial terms, degree selection, and overfitting
- Lasso and Ridge Regression: Regularization techniques for controlling model complexity
- Evaluation metrics for regression models: Mean Squared Error (MSE), R-squared, and Mean Absolute Error (MAE)

Hands-On - House Price Prediction



#### **Supervised Learning - Classification**

- ✓ Introduction to Classification: Definition, types, and use cases
- ✓ Logistic Regression: Theory, logistic function, binary and multiclass classification
- ☑ Decision Trees: Construction, splitting criteria, pruning, and visualization
- ☑ Random Forests: Ensemble learning, bagging, and feature importance
- ☑ Evaluation metrics for classification models: Accuracy, Precision, Recall, F1-score, and ROC curves
- ☑ Implementation of classification models using scikit-learn library

Hands-On - Heart Disease Detection & Food Order Prediction



#### **SVM, KNN & Naive Bayes**

- Support Vector Machines (SVM): Study SVM theory, different kernel functions (linear, polynomial, radial basis function), and the margin concept. Implement SVM classification and regression, and evaluate the models.
- K-Nearest Neighbors (KNN): Understand the KNN algorithm, distance metrics, and the concept of K in KNN. Implement KNN classification and regression, and evaluate the models.
- Naive Bayes: Learn about the Naive Bayes algorithm, conditional probability, and Bayes' theorem. Implement Naive Bayes classification, and evaluate the model's performance

**Hands-On - Contact Tracing & Sarcasm Detection** 



#### **Ensemble Methods and Boosting**

- AdaBoost: Boosting technique, weak learners, and iterative weight adjustment
- Gradient Boosting (XGBoost): Gradient boosting algorithm, Regularization, and hyperparameter tuning
- ☑ Evaluation and fine-tuning of ensemble models: Cross-validation, grid search, and model selection
- ✓ Handling imbalanced datasets: Techniques for dealing with class imbalance, such as oversampling and undersampling

Hands-On - Medical Insurance Price Prediction



#### **Unsupervised Learning - Clustering**

- ✓ Introduction to Clustering: Definition, types, and use cases
- K-means Clustering: Algorithm steps, initialization methods, and elbow method for determining the number of clusters
- DBSCAN (Density-Based Spatial Clustering of Applications with Noise): Core points, density reachability, and epsilon-neighborhoods
- Evaluation of clustering algorithms: Silhouette score, cohesion, and separation metrics

Hands-On - Credit Card Clustering



#### **Unsupervised Learning - Dimensionality Reduction**

- Introduction to Dimensionality Reduction: Curse of dimensionality, feature extraction, and feature selection
- ✓ Principal Component Analysis (PCA): Eigenvectors, eigenvalues, variance explained, and dimensionality reduction
- ✓ Implementation of PCA using scikit-learn library

#### Hands-On - MNIST Data



#### **Recommendation Systems**

- ✓ Introduction to Recommendation Systems: Understand the concept of recommen-
- dation systems, different types (collaborative filtering, content-based, hybrid), and evaluation metrics.
- ✓ Collaborative Filtering: Explore collaborative filtering techniques, including user-based and item-based approaches, and implement a collaborative filtering model.
- ✓ Content-Based Filtering: Study content-based filtering methods, such as TF-IDF and cosine similarity, and build a content-based recommendation system.
- Deployment and Future Directions: Discuss the deployment of recommendation systems and explore advanced topics in NLP and recommendation systems.

#### **Hands-On - News Recommendation System**



#### **Reinforcement Learning**

- Introduction to Reinforcement Learning: Agent, environment, state, action, and reward
- Markov Decision Processes (MDP): Markov property, transition probabilities, and value functions
- ✓ Q-Learning algorithm: Exploration vs. exploitation, Q-table, and learning rate
- ✓ Hands-on reinforcement learning projects and exercises

#### Hands-On - Working with OpenAI Gym



#### Developing API using Flask / Webapp with Streamlit

- ✓ Introduction to Flask / Streamlit web framework
- ✓ Creating a Flask / Streamlit application for ML model deployment
- ✓ Integrating data preprocessing and ML model
- ✓ Designing a user-friendly web interface



#### **Deployment of ML Models**

- ☑ Building a web application for Machine Learning models: Creating forms, handling user input, and displaying results
- ☑ Deployment using AWS (Amazon Web Services): Setting up an AWS instance, configuring security groups, and deploying the application
- ✓ Deployment using PythonAnywhere: Uploading Flask application files, configuring WSGI, and launching the application



#### **Project Work and Consolidation**

- Work on a real-world Machine Learning project: Identify a problem, gather data, and define project scope
- Apply the learned concepts and algorithms: Data collection, preprocessing, model building, and evaluation
- Deployment of the project on AWS or PythonAnywhere: Showcase the developed application and share the project with others
- Presentation and discussion of the project: Demonstrate the project, explain design decisions, and receive feedback





#### **Natural Language Processing (NLP)**

- Introduction to NLP: Understand the basics of NLP, its applications, and challenges.
- Named Entity Recognition (NER): Understand the various approaches and tools used for NER, such as rule-based systems, statistical models, and deep learning.
- Text Preprocessing: Learn about tokenization, stemming, lemmatization, stop word removal, and other techniques for text preprocessing.
- Text Representation: Explore techniques such as Bag-of-Words (BoW), TF-IDF, and word embeddings (e.g., Word2Vec, GloVe) for representing text data.
- Sequential Models: Introduction to RNN, LSTM, Hands on Keras LSTM
- Sentiment Analysis: Study sentiment analysis techniques, build a sentiment analysis model using supervised learning, and evaluate its performance.

#### **Hands-On - Real Time Sentiment Analysis**



## Module 6: Deep Learning



#### RISE OF THE DEEP LEARNING

- ✓ Introduction
- History of Deep Learning
- Perceptrons
- ✓ Multi-Level Perceptrons
- ✓ Representations
- ✓ Training Neural Networks
- Activation Functions



#### **Artificial Neural Networks**

- ✓ Introduction
- ✓ Deep Learning
- ✓ Understanding Human Brain
- ✓ In-Depth Perceptrons
- Example for perceptron
- ✓ Multi Classifier
- ✓ Neural Networks
- ✓ Input Layer
- Output Layer
- ✓ Sigmoid Function
- ✓ Introduction to Tensorflow and Keras
- CPU vs GPU

- ✓ Understanding Notations
- ✓ Activation Functions
- Feed-Forward Networks
- ✓ Online offline mode
- ✓ Bidirectional RNN
- ✓ Understanding Dimensions
- ✓ Back Propagation
- ✓ Loss function
- **✓** SGD
- ✓ Regularization
- ✓ Training for batches

**Hands-On - Facial Emotion Recognition** 

# 1

### **Convolution Neural Networks**

- ✓ Introduction to CNN
- ✓ Idea behind CNN
- ✓ Understanding Images
- ✓ Understanding Videos
- ✓ Convolutions
- ✓ Striding and Padding
- ✓ Max Pooling
- Edges, Gradients, and Textures
- ✓ Understanding Channels
- **√** Formulas
- ✓ Weight and Bias
- Feature Map
- ✓ Pooling
- ✓ Combining



#### **CNN - Transfer Learning**

- ✓ Introduction
- ✓ AlexNet
- ✓ GoogleNet
- ✓ ResNet
- ✓ Transfer learning using Keras

Hands-On - Face Mask Detection



#### **RNN - Recurrent Neural Networks**

- ✓ Introduction to RNNs
- Training RNNs
- ✓ RNN Formula
- ✓ Architecture
- **✓** Batch Data
- ✓ Simplified Notations
- ▼ Types of RNNs
- ✓ LSTM
- GRUs
- ▼ Training RNN
- One to many
- √ Vanishing Gradient problem

#### Hands-On - COVID-19 Cases Prediction



#### **Generative Models and GANs**

- ✓ Introduction to Generative Models:
- ✓ Understanding GANs (Generative Adversarial Networks)
- ✓ GAN Architecture
- GAN Training
- Evaluating GAN Performance
- GAN Variants and Applications



### **Module 7: Computer Vision**



#### **Computer Vision**

- ✓ Intro to OpenCV
- Reading and Writing Images
- ✓ Saving images
- ✓ Draw shapes using OpenCV
- √ Face detection and eye detection using OpenCV
- CNN with Keras
- **V** VGG





#### Real-Time Rain Prediction using ML

- ✓ Install necessary libraries
- ✓ Obtain an API key
- Fetch live weather data
- ✓ Preprocess the data
- Train a machine learning model
- ✓ Evaluate the model
- Integrate the model with Flask
- Display the results
- Test and debug
- Deploy the application
- Continuously update the weather data

# **⟨**/>

#### Real Time Drowsiness Detection Alert System

- ✓ Dataset collection
- Data preprocessing
- ✓ Feature extraction
- Labeling
- ✓ Model selection
- Model training
- ✓ Model evaluation
- ✓ Real-time implementation
- ✓ Alert mechanism
- Continuous improvement



#### **House Price Prediction using LSTM**

- Identify a reliable source for house price data
- ✓ Understand the website structure
- Perform web scraping
- Preprocess the scraped data
- Explore and preprocess additional data sources (if applicable)
- Define the problem
- ✓ Split the data
- ✓ Train the model
- ✓ Evaluate the model
- Fine-tune the model (optional)
- Deploy the model
- ✓ Continuously update the dataset and retrain the model



#### **Customizable Chabot using OpenAI API**

- Define chatbot goals and scope
- Gather training data
- Data preprocessing
- ✓ API integration
- ✓ Model customization
- User input handling
- Response generation
- ✓ Post-processing and filtering
- Error handling and fallback mechanisms
- Continuous improvement



#### Fire and Smoke Detection using CNN

- ✓ Data collection
- Data preprocessing
- ✓ Dataset augmentation
- ✓ Model architecture
- ✓ Model architecture
- ✓ Training
- ✓ Model evaluation
- √ Fine-tuning
- ✓ Real-time inference
- Thresholding and alerts
- ✓ Model optimization