DATAhill
Learning Intelligence

DATA SCIENCE

with DATA ANALYTICS, MACHINE LEARNING, NLP, DL & AI using PYTHON & R

INTRODUCTION TO DATA SCIENCE
- What is a Data Science?
- Who is a Data Scientist?
- Who can become a Data Scientist?
- What is an Artificial Intelligence?
- What is a Machine Learning?
- What is a Deep Learning?
- Artificial Intelligence Vs Machine Learning Vs Deep Learning
- Real Time Process of Data Science
- Data Science Real Time Applications
- Technologies used in Data Science
- Prerequisites Knowledge to Learn Data Science

INTRODUCTION TO MACHINE LEARINING
- What is a Machine Learning?
- Machine Learning Vs Statistics
- Traditional Programming Vs Machine Learning
- How Machine Will Learn like Human Learning
- Machine Learning Engineer Responsibilities
- Types of Machine Learning
  - Supervised learning
  - Un-Supervised learning
  - Reinforcement Learning

CORE PYTHON PROGRAMMING
- PYTHON Programming Introduction
- History of Python
- Python is Derived from?
- Python Features
- Python Applications
- Why Python is Becoming Popular Now a Day?
- Existing Programming Vs Python Programming
- Writing Programs in Python

- ➢ Top Companies Using Python
- ➢ Python Programming Modes
  - o Interactive Mode Programming
  - o Scripting Mode Programming
- ➢ Flavors in Python, Python Versions
- ➢ Download & Install the Python in Windows & Linux
- ➢ How to set Python Environment in the System?
- ➢ Anaconda - Data Science Distributor
- ➢ Downloading and Installing Anaconda, Jupyter Notebook & Spyder
- ➢ Python IDE - Jupyter Notebook Environment
- ➢ Python IDE – Spyder Environment
- ➢ Python Identifiers(Literals), Reserved Keywords
- ➢ Variables, Comments
- ➢ Lines and Indentations, Quotations
- ➢ Assigning Values to Variables
- ➢ Data Types in Python
- ➢ Mutable Vs Immutable
- ➢ Fundamental Data Types: int, float, complex, bool, str
- ➢ Number Data Types: Decimal, Binary, Octal, Hexa Decimal & Number Conversions
- ➢ Inbuilt Functions in Python
- ➢ Data Type Conversions
- ➢ Priorities of Data Types in Python
- ➢ Python Operators
  - o Arithmetic Operators
  - o Comparison (Relational) Operators
  - o Assignment Operators
  - o Logical Operators
  - o Bitwise Operators
  - o Membership Operators
  - o Identity Operators
- ➢ Slicing & Indexing
  - o Forward Direction Slicing with +ve Step
  - o Backward Direction Slicing with -ve Step
- ➢ Decision Making Statements
  - o if Statement
  - o if-else Statement
  - o elif Statement
- ➢ Looping Statements
  - o Why we use Loops in python?
  - o Advantages of Loops
  - o for Loop

- o Nested for Loop
- o Using else Statement with for Loop
- o while Loop
- o Infinite while Loop
- o Using else with Python while Loop
- ➢ Conditional Statements
  - o break Statement
  - o continue Statement
  - o Pass Statement

ADVANCED PYTHON PROGRAMMING

- ➢ Advanced Data Types: List, Tuple, Set, Frozenset, Dictionary, Range, Bytes & Bytearray, None
- ➢ List Data Structure
  - o List indexing and splitting
  - o Updating List values
  - o List Operations
  - o Iterating a List
  - o Adding Elements to the List
  - o Removing Elements from the List
  - o List Built-in Functions
  - o List Built-in Methods
- ➢ Tuple Data Structure
  - o Tuple Indexing and Splitting
  - o Tuple Operations
  - o Tuple Inbuilt Functions
  - o Where use Tuple
  - o List Vs Tuple
  - o Nesting List and Tuple
- ➢ Set Data Structure
  - o Creating a Set
  - o Set Operations
  - o Adding Items to the Set
  - o Removing Items from the Set
  - o Difference Between discard() and remove()
  - o Union of Two Sets
  - o Intersection of Two Sets
  - o Difference of Two Sets
  - o Set Comparisons
- ➢ Frozenset Data Structure
- ➢ Dictionary Data Structure
  - o Creating the Dictionary
  - o Accessing the Dictionary Values

- o Updating Dictionary Values
- o Deleting Elements Using del Keyword
- o Iterating Dictionary
- o Properties of Dictionary Keys
- o Built-in Dictionary Functions
- o Built-in Dictionary Methods
- ➢ List Vs Tuple Vs Set Vs Frozenset Vs Dict
- ➢ Range, Bytes, Bytearray & None
- ➢ Python Functions
  - o Advantage of Functions in Python
  - o Creating a Function
  - o Function Calling
  - o Parameters in Function
  - o Call by Reference in Python
  - o Types of Arguments
    - ▪ Required Arguments
    - ▪ Keyword Arguments
    - ▪ Default Arguments
    - ▪ Variable-Length Arguments
- ➢ Scope of Variables
- ➢ Python Built-in Functions
- ➢ Python Lambda Functions
- ➢ String with Functions
  - o Strings Indexing and Splitting
  - o String Operators
  - o Python Formatting Operator
  - o Built-in String Functions
- ➢ Python File Handling
  - o Opening a File
  - o Reading the File
  - o Read Lines of the File
  - o Looping through the File
  - o Writing the File
  - o Creating a New File
  - o Using with Statement with Files
  - o File Pointer Positions
  - o Modifying File Pointer Position
  - o Renaming the File & Removing the File
  - o Writing Python Output to the Files
  - o File Related Methods
- ➢ Python Exceptions
  - o Common Exceptions
  - o Problem without Handling Exceptions

- o except Statement with no Exception
- o Declaring Multiple Exceptions
- o Finally Block
- o Raising Exceptions
- o Custom Exception
- ➢ Python Packages
  - o Python Libraries
  - o Python Modules
    - ▪ Collection Module
    - ▪ Math Module
    - ▪ OS Module
    - ▪ Random Module
    - ▪ Statistics Module
    - ▪ Sys Module
    - ▪ Date & Time Module
  - o Loading the Module in our Python Code
    - ▪ import Statement
    - ▪ from-import Statement
  - o Renaming a Module
- ➢ Regular Expressions
- ➢ Command Line Arguments
- ➢ Object Oriented Programming (OOPs)
  - o Object-oriented vs Procedure-oriented Programming languages
  - o Object
  - o Class
  - o Method
  - o Inheritance
  - o Polymorphism
  - o Data Abstraction
  - o Encapsulation
- ➢ Python Class and Objects
  - o Creating Classes in Python
  - o Creating an Instance of the Class
- ➢ Python Constructor
  - o Creating the Constructor in Python
  - o Parameterized Constructor
  - o Non-Parameterized Constructor
  - o In-built Class Functions
  - o In-built Class Attributes
- ➢ Python Inheritance
  - o Python Multi-Level Inheritance
  - o Python Multiple Inheritance

- o Method Overriding
- o Data Abstraction in Python
- ➢ Graphical User Interface (GUI) Programming
- ➢ Python TKinter
  - o Tkinter Geometry
    - ▪ pack() Method
    - ▪ grid() Method
    - ▪ place() Method
  - o Tkinter Widgets

DATA ANALYSIS WITH PYTHON NUMPY
- ➢ NumPy Introduction
  - o What is NumPy
  - o The Need of NumPy
- ➢ NumPy Environment Setup
- ➢ N-Dimensional Array (Ndarray)
  - o Creating a Ndarray Object
  - o Finding the Dimensions of the Array
  - o Finding the Size of Each Array Element
  - o Finding the Data Type of Each Array Item
  - o Finding the Shape and Size of the Array
  - o Reshaping the Array Objects
  - o Slicing in the Array
  - o Finding the Maximum, Minimum, and Sum of the Array Elements
  - o NumPy Array Axis
  - o Finding Square Root and Standard Deviation
  - o Arithmetic Operations on the Array
  - o Array Concatenation
- ➢ NumPy Datatypes
  - o NumPy dtype
  - o Creating a Structured Data Type
- ➢ Numpy Array Creation
  - o Numpy.empty
  - o Numpy.Zeros
  - o NumPy.ones
- ➢ Numpy Array from Existing Data
  - o Numpy.asarray
- ➢ Numpy Arrays within the Numerical Range
  - o Numpy.arrange
  - o NumPy.linspace
  - o Numpy.logspace
- ➢ NumPy Broadcasting

- Broadcasting Rules
- NumPy Array Iteration
  - Order of Iteration
    - F-Style Order
    - C-Style Order
  - Array Values Modification
- NumPy String Functions
- NumPy Mathematical Functions
  - Trigonometric Functions
  - Rounding Functions
- NumPy Statistical functions
  - Finding the Min and Max Elements from the Array
  - Calculating Median, Mean, and Average of Array Items
- NumPy Sorting and Searching
- NumPy Copies and Views
- NumPy Matrix Library
- NumPy Linear Algebra
- NumPy Matrix Multiplication in Python

DATA ANALYSIS WITH PYTHON PANDAS

- Pandas Introduction & Pandas Environment Setup
  - Key Features of Pandas
  - Benefits of Pandas
  - Python Pandas Data Structure
    - Series
    - DataFrame
    - Panel
- Pandas Series
  - Creating a Series
    - Create an Empty Series
    - Create a Series using Inputs
  - Accessing Data from Series with Position
  - Series Object Attributes
  - Retrieving Index Array and Data Array of a Series Object
  - Retrieving Types (dtype) and Size of Type (itemsize)
  - Retrieving Shape
  - Retrieving Dimension, Size and Number of Bytes
  - Checking Emptiness and Presence of NaNs
  - Series Functions
- Pandas DataFrame
  - Create a DataFrame
    - Create an Empty DataFrame
    - Create a DataFrame using Inputs

- Column Selection, Addition & Deletion
- Row Selection, Addition & Deletion
- DataFrame Functions
- Merging, Joining & Combining DataFrames
- Pandas Concatenation
- Pandas Time Series
  - Datetime
  - Time Offset
  - Time Periods
  - Convert String to Date
- Viewing/Inspecting Data (loc & iloc)
- Data Cleaning
- Filter, Sort, and Groupby
- Statistics on DataFrame
- Pandas Vs NumPy
- DataFrame Plotting
  - Line: Line Plot (Default)
  - Bar: Vertical Bar Plot
  - Barh: Horizontal Bar Plot
  - Hist: Histogram Plot
  - Box: Box Plot
  - Pie: Pie Chart
  - Scatter: Scatter Plot

DBMS - Structured Query Language
- Introduction & Models of DBMS
- SQL & Sub Language of SQL
- Data Definition Language (DDL)
- Data Manipulation Language (DML)
- Data Query/Retrieval Language (DQL/DRL)
- Transaction Control Language (TCL)
- Data Control Language (DCL)
- Installation of MySQL & Database Normalization
- Sub Queries & Key Constraints
- Aggregative Functions, Clauses & Views

Importing & Exporting Data
- Data Extraction from CSV (pd.read_csv)
- Data Extraction from TEXT File (pd.read_table)
- Data Extraction from CLIPBOARD (pd.read_clipboard)
- Data Extraction from EXCEL (pd.read_excel)
- Data Extraction from URL (pd.read_html)
- Writing into CSV (df.to_csv)

- Writing into EXCEL (df.to_excel)
- Data Extraction from DATABASES
    - Python MySQL Database Connection
        - Import mysql.connector Module
        - Create the Connection Object
        - Create the Cursor Object
        - Execute the Query

DATA VISUALIZATION WITH PYTHON MATPLOTLIB

- Data Visualization Introduction
- Tasks of Data Visualization
- Benefit of Data Visualization
- Plots for Data Visualization
- Matplotlib Architecture
- General Concept of Matplotlib
- MatPlotLib Environment Setup
- Verify the MatPlotLib Installation
- Working with PyPlot
- Formatting the Style of the Plot
- Plotting with Categorical Variables
- Multi-Plots with Subplot Function
- Line Graph
- Bar Graph
- Histogram
- Scatter Plot
- Pie Plot
- 3Dimensional - 3D Graph Plot
- mpl_toolkits
- Functions of MatPlotLib
- Contour Plot, Quiver Plot, Violin Plot
- 3D Contour Plot
- 3D Wireframe Plot
- 3D Surface Plot
- Box Plot
    - What is a Boxplot?
    - Mean, Median, Quartiles, Outliers
    - Inter Quartile Range (IQR), Whiskers
    - Data Distribution Analysis
    - Boxplot on a Normal Distribution
    - Probability Density Function
    - 68–95–99.7 Rule (Empirical rule)

Data Analysis Project using Python Programming

MACHINE LEARNING

- ➢ What is Machine Learning
- ➢ Importance of Machine Learning
- ➢ Need for Machine Learning
- ➢ Statistics Vs Machine Learning
- ➢ Traditional Programming Vs Machine Learning
- ➢ How Machine Learning like Human Learning
- ➢ How does Machine Learning Work?
- ➢ Machine Learning Engineer Responsibilities
- ➢ Life Cycle of Machine Learning
  - o Gathering Data
  - o Data preparation
  - o Data Wrangling
  - o Analyze Data
  - o Train the model
  - o Test the model
  - o Deployment
- ➢ Features of Machine Learning
- ➢ History of Machine Learning
- ➢ Applications of Machine Learning
- ➢ Types of Machine Learning
  - o Supervised Machine Learning
  - o Unsupervised Machine Learning
  - o Reinforcement Learning

Supervised Machine Learning
- ➢ How Supervised Learning Works?
- ➢ Steps Involved in Supervised Learning
- ➢ Types of supervised Machine Learning Algorithms
  - o Classification
  - o Regression
- ➢ Advantages of Supervised Learning
- ➢ Disadvantages of Supervised Learning

Unsupervised Machine Learning
- ➢ How Unsupervised Learning Works?
- ➢ Why use Unsupervised Learning?
- ➢ Types of Unsupervised Learning Algorithm
  - o Clustering
  - o Association
- ➢ Advantages of Unsupervised Learning
- ➢ Disadvantages of Unsupervised Learning
- ➢ Supervised Vs Unsupervised Learning

- ➢ Reinforcement Machine Learning
- ➢ How to get Datasets for Machine Learning?
  - o What is a Dataset?
  - o Types of Data in Datasets
  - o Popular Sources for Machine Learning Datasets

Data Preprocessing in Machine Learning
- ➢ Why do we need Data Preprocessing?
  - o Getting the Dataset
  - o Importing Libraries
  - o Importing Datasets
  - o Finding Missing Data
    - ▪ By Deleting the Particular Row
    - ▪ By Calculating the Mean
  - o Encoding Categorical Data
    - ▪ LableEncoder
    - ▪ OneHotEncoder
  - o Splitting Dataset into Training and Test Set
  - o Feature Scaling
    - ▪ Standardization
    - ▪ Normalization

Classification Algorithms in Machine Learning
- ➢ What is the Classification Algorithm?
- ➢ Types of Classifications
  - o Binary Classifier
  - o Multi-class Classifier
- ➢ Learners in Classification Problems
  - o Lazy Learners
  - o Eager Learners
- ➢ Types of ML Classification Algorithms
  - o Linear Models
    - ▪ Logistic Regression
    - ▪ Support Vector Machines
  - o Non-linear Models
    - ▪ K-Nearest Neighbors
    - ▪ Naïve Bayes
    - ▪ Decision Tree Classification
    - ▪ Random Forest Classification
    - ▪ Kernel SVM
- ➢ Evaluating a Classification Model
  - o Confusion Matrix
    - ▪ What is a Confusion Matrix?

- True Positive
- True Negative
- False Positive – Type 1 Error
- False Negative – Type 2 Error
  - Why need a Confusion matrix?
  - Precision
  - Recall
  - Precision vs Recall
  - F1-score
  - Confusion Matrix in Scikit-Learn
  - Confusion Matrix for Multi-Class Classification
- Log Loss or Cross-Entropy Loss
- AUC-ROC curve
- Use cases of Classification Algorithms

K-Nearest Neighbor(KNN) Algorithm in Machine Learning

- Why do we Need a K-NN Algorithm?
- How does K-NN work?
  - What is Euclidean Distance
  - How it Calculates the Distance
- How to Select the Value of K in the K-NN Algorithm?
- Advantages of KNN Algorithm
- Disadvantages of KNN Algorithm
- Python Implementation of the KNN Algorithm
- Analysis on Social Network Ads Dataset
- Steps to Implement the K-NN Algorithm
  - Data Pre-processing Step
  - Fitting the K-NN algorithm to the Training Set
  - Predicting the Test Result
  - Test Accuracy of the Result (Creation of Confusion Matrix)
  - Visualizing the Test Set Result.
  - Improve the Performance of the K-NN Model

Naïve Bayes Classifier Algorithm in Machine Learning

- Why is it Called Naïve Bayes?
  - Naïve Means?
  - Bayes Means?
- Bayes' Theorem
  - Posterior Probability
  - Likelihood Probability
  - Prior Probability
  - Marginal Probability
- Working of Naïve Bayes' Classifier

- ➤ Advantages of Naïve Bayes Classifier
- ➤ Disadvantages of Naïve Bayes Classifier
- ➤ Applications of Naïve Bayes Classifier
- ➤ Types of Naïve Bayes Model
  - o Gaussian Naïve Bayes Classifier
  - o Multinomial Naïve Bayes Classifier
  - o Bernoulli Naïve Bayes Classifier
- ➤ Python Implementation of the Naïve Bayes Algorithm
- ➤ Steps to Implement the Naïve Bayes Algorithm
  - o Data Pre-processing Step
  - o Fitting Naive Bayes to the Training set
  - o Predicting the Test Result
  - o Test Accuracy of the Result (Creation of Confusion matrix)
  - o Visualizing the Test Set Result
  - o Improve the Performance of the Naïve Bayes Model

Decision Tree Classification Algorithm in Machine Learning

- ➤ Why use Decision Trees?
- ➤ Types of Decision Trees
  - o Categorical Variable Decision Tree
  - o Continuous Variable Decision Tree
- ➤ Decision Tree Terminologies
- ➤ How does the Decision Tree Algorithm Work?
- ➤ Attribute Selection Measures
  - o Entropy
  - o Information Gain
  - o Gini index
  - o Gain Ratio
- ➤ Algorithms used in Decision Trees
  - o ID3 Algorithm → (Extension of D3)
  - o C4.5 Algorithm→ (Successor of ID3)
  - o CART Algorithm → (Classification & Regression Tree)
- ➤ How to Avoid/Counter Overfitting in Decision Trees?
  - o Pruning Decision Trees
  - o Random Forest
- ➤ Pruning: Getting an Optimal Decision tree
- ➤ Advantages of the Decision Tree
- ➤ Disadvantages of the Decision Tree
- ➤ Python Implementation of Decision Tree
- ➤ Steps to Implement the Decision Tree Algorithm
  - o Data Pre-processing Step
  - o Fitting a Decision-Tree Algorithm to the Training Set
  - o Predicting the Test Result

- o Test Accuracy of the Result (Creation of Confusion matrix)
- o Visualizing the Test Set Result
- o Improve the Performance of the Decision Tree Model

Random Forest Classifier Algorithm in Machine Learning
- ➢ Working of the Random Forest Algorithm
- ➢ Assumptions for Random Forest
- ➢ Why use Random Forest?
- ➢ How does Random Forest Algorithm Work?
  - o Ensemble Techniques
  - o Bagging (Bootstrap Aggregation)
- ➢ Applications of Random Forest
- ➢ Disadvantages of Random Forest
- ➢ Python Implementation of Random Forest Algorithm
- ➢ Steps to Implement the Random Forest Algorithm:
  - o Data Pre-processing Step
  - o Fitting the Random Forest Algorithm to the Training Set
  - o Predicting the Test Result
  - o Test Accuracy of the Result (Creation of Confusion Matrix)
  - o Visualizing the Test Set Result
  - o Improving the Performance of the Random Forest Model

Logistic Regression Algorithm in Machine Learning
- ➢ Logistic Function (Sigmoid Function)
- ➢ Assumptions for Logistic Regression
- ➢ Logistic Regression Equation
- ➢ Type of Logistic Regression
  - o Binomial Logistic Regression
  - o Multinomial Logistic Regression
  - o Ordinal Logistic Regression
- ➢ Python Implementation of Logistic Regression (Binomial)
- ➢ Steps to Implement the Logistic Regression:
  - o Data Pre-processing Step
  - o Fitting Logistic Regression to the Training Set
  - o Predicting the Test Result
  - o Test Accuracy of the Result (Creation of Confusion Matrix)
  - o Visualizing the Test Set Result
  - o Improve the Performance of the Logistic Regression Model

Support Vector Machine Algorithm
- ➢ Types of Support Vector Machines
  - o Linear Support Vector Machine
  - o Non-Linear Support Vector Machine

- ➤ Hyperplane in the SVM Algorithm
- ➤ Support Vectors in the SVM Algorithm
- ➤ How does SVM Works?
  - o How does Linear SVM Works?
  - o How does Non-Linear SVM Works?
- ➤ Python Implementation of Support Vector Machine
- ➤ Steps to Implement the Support Vector Machine:
  - o Data Pre-processing Step
  - o Fitting Support Vector Machine to the Training Set
  - o Predicting the Test Result
  - o Test Accuracy of the Result (Creation of Confusion Matrix)
  - o Visualizing the Test Set Result
  - o Improve the Performance of the Support Vector Machine Model

Regression Algorithms in Machine Learning
- ➤ Terminologies Related to the Regression Analysis
  - o Dependent Variable
  - o Independent Variable
  - o Outliers
  - o Multi-collinearity
  - o Under fitting and Overfitting
- ➤ Why do we use Regression Analysis?
- ➤ Types of Regression
  - o Linear Regression
  - o Logistic Regression
  - o Polynomial Regression
  - o Support Vector Regression
  - o Decision Tree Regression
  - o Random Forest Regression
  - o Ridge Regression
  - o Lasso Regression

Linear Regression in Machine Learning
- ➤ Types of Linear Regression
  - o Simple Linear Regression
  - o Multiple Linear Regression
- ➤ Linear Regression Line
  - o Positive Linear Relationship
  - o Negative Linear Relationship
- ➤ Finding the Best Fit Line
  - o Cost Function
  - o Gradient Descent

- o Model Performance
- o R-Squared Method
- ➢ Assumptions of Linear Regression

Simple Linear Regression in Machine Learning
- ➢ SLR Model
- ➢ Implementation of Simple Linear Regression Algorithm using Python
  - o Data Pre-processing Step
  - o Fitting Simple Linear Regression to the Training Set
  - o Predicting the Test Result
  - o Test Accuracy of the
  - o Visualizing the Test Set Result.
  - o Try to Improve the Performance of the Model

Multiple Linear Regression in Machine Learning
- ➢ MLR Equation
- ➢ Assumptions for Multiple Linear Regression
- ➢ Implementation of Multiple Linear Regression model using Python
  - o Data Pre-processing Step
  - o Fitting Multiple Linear Regression to the Training Set
  - o Predicting the Test Result
  - o Test Accuracy of the
  - o Visualizing the Test Set Result.
  - o Try to Improve the Performance of the Model

Backward Elimination
- ➢ What is Backward Elimination?
- ➢ Steps of Backward Elimination
- ➢ Need for Backward Elimination: An optimal Multiple Linear Regression model
- ➢ Implement the Steps for Backward Elimination method

Polynomial Regression in Machine Learning
- ➢ Need for Polynomial Regression
- ➢ Equation of the Polynomial Regression Model
- ➢ Implementation of Polynomial Regression using Python
- ➢ Steps for Polynomial Regression:
  - o Data Pre-processing
  - o Build a Linear Regression Model
  - o Build a Polynomial Regression Model
  - o Visualize the Result for Linear Regression Model
  - o Visualize the Result for Polynomial Regression Model

- o Predicting the Final Result with the Linear Regression Model
- o Predicting the Final Result with the Polynomial Regression Model
- ➢ Support Vector Regression (SVR)
- ➢ Decision Tree Regression
- ➢ Random Forest Regression
- ➢ Ridge Regression
- ➢ Lasso Regression
- ➢ Linear Regression Vs Logistic Regression
- ➢ Classification vs Regression

Clustering Algorithms in Machine Learning

- ➢ Types of Clustering Methods
  - o Partitioning Clustering
  - o Density-Based Clustering
  - o Distribution Model-Based Clustering
  - o Hierarchical Clustering
  - o Fuzzy Clustering
- ➢ Clustering Algorithms
  - o K-Means Algorithm
  - o Mean-shift Algorithm
  - o DBSCAN Algorithm
  - o Expectation-Maximization Clustering using GMM
  - o Agglomerative Hierarchical Algorithm
  - o Affinity Propagation
- ➢ Applications of Clustering

Hierarchical Clustering Algorithm in Machine Learning

- ➢ Hierarchical Clustering Technique Approaches
- ➢ Why Hierarchical Clustering?
- ➢ Agglomerative Hierarchical Clustering
- ➢ How the Agglomerative Hierarchical Clustering Work?
- ➢ Measure for the Distance between two Clusters
  - o Single Linkage
  - o Complete Linkage
  - o Average Linkage
  - o Centroid Linkage
- ➢ Working of Dendrogram in Hierarchical Clustering
- ➢ Hierarchical Clustering Example with Scratch Data
- ➢ Python Implementation of Agglomerative Hierarchical Clustering
- ➢ Steps for Implementation of Agglomerative Hierarchical Clustering using Python
  - o Data Pre-processing

- o Finding the Optimal Number of Clusters using the Dendrogram
- o Training the Hierarchical Clustering Model
- o Visualizing the Clusters

K-Means Clustering Algorithm in Machine Learning
- ➢ What is K-Means Algorithm?
- ➢ How does the K-Means Algorithm Work?
- ➢ How to Choose the Value of "K Number of Clusters" in K-Means Clustering?
  - o Elbow Method
  - o Within Cluster Sum of Squares (WCSS)
- ➢ K-Means Clustering Example with Scratch Data
- ➢ Python Implementation of K-means Clustering Algorithm
- ➢ Steps to Implement of K-means Clustering Algorithm
  - o Data Pre-processing
  - o Finding the Optimal Number of Clusters using the Elbow Method
  - o Training the K-means Algorithm on the Training Dataset
  - o Visualizing the Clusters

Association Rules in Machine Learning
- ➢ Association Rules
- ➢ Pattern Detection
- ➢ Market Basket Analysis
- ➢ Support, Confidence, Expected Confidence, Lift
- ➢ Finding Item Sets with High Support
- ➢ Finding Item Rules with High Confidence or Lift

Apriori Algorithm in Machine Learning
- ➢ Apriori Algorithm
- ➢ How does Apriori Algorithm Works?
- ➢ Apriori Algorithm Example
- ➢ Implementation of Apriori Algorithm using Python
- ➢ Limitations of Apriori Algorithm

Dimensionality Reduction & Model Selection Boosting
- ➢ Dimensionality Reduction
  - o Principal Component Analysis (PCA)
  - o Linear Discriminant Analysis (LDA)
  - o Kernel PCA
- ➢ Model Selection Boosting
  - o Model Selection

- Grid Search
- K-Fold Cross Validation
  - o XGBoost

STATISTICS

- ➤ Mean, Median and Mode
- ➤ Data Variability, Range, Quartiles
- ➤ IQR, Calculating Percentiles
- ➤ Variance, Standard Deviation, Statistical Summaries
- ➤ Types of Distributions – Normal, Binomial, Poisson
- ➤ Probability Distributions & Skewness
- ➤ Data Distribution, 68–95–99.7 rule (Empirical rule)
- ➤ Descriptive Statistics and Inferential Statistics
- ➤ Statistics Terms and Definitions, Types of Data
- ➤ Data Measurement Scales, Normalization, Standardization
- ➤ Measure of Distance, Euclidean Distance
- ➤ Probability Calculation – Independent & Dependent
- ➤ Entropy, Information Gain
- ➤ Regression

NATURAL LANGUAGE PROCESSING

- ➤ Natural Language Processing Introduction
  - o What is NLP?
  - o History of NLP
  - o Advantages of NLP
  - o Disadvantages of NLP
- ➤ Components of NLP
  - o Natural Language Understanding (NLU)
  - o Natural Language Generation (NLG)
  - o Difference between NLU and NLG
- ➤ Applications of NLP
- ➤ How to build an NLP Pipeline?
- ➤ Phases of NLP
  - o Lexical Analysis and Morphological
  - o Syntactic Analysis (Parsing)
  - o Semantic Analysis
  - o Discourse Integration
  - o Pragmatic Analysis
- ➤ Why NLP is Difficult?
- ➤ NLP APIs
- ➤ NLP Libraries
- ➤ Natural Language Vs Computer Language

Exploring Features of NLTK

- o Open the Text File for Processing
- o Import Required Libraries
- o Sentence Tokenizing
- o Word Tokenizing
- o Find the Frequency Distribution
- o Plot the Frequency Graph
- o Remove Punctuation Marks
- o Plotting Graph without Punctuation Marks
- o List of Stopwords
- o Removing Stopwords
- o Final Frequency Distribution
- ➢ Word Cloud
  - o Word Cloud Properties
  - o Python Code Implementation of the Word Cloud
  - o Word Cloud with the Circle Shape
  - o Word Cloud Advantages
  - o Word Cloud Disadvantages
- ➢ Stemming
  - o Stemmer Examples
  - o Stemming Algorithms
    - ▪ Porter's Stemmer
    - ▪ Lovin's Stemmer
    - ▪ Dawson's Stemmer
    - ▪ Krovetz Stemmer
    - ▪ Xerox Stemmer
    - ▪ Snowball Stemmer
- ➢ Lemmatization
  - o Difference between Stemmer and Lemmatizer
  - o Demonstrating how a lemmatizer works
  - o Lemmatizer with default PoS value
  - o Demonstrating the power of lemmatizer
  - o Lemmatizer with different POS values
- ➢ Part-of-Speech (PoS) Tagging
  - o Why do we need Part of Speech (POS)?
  - o Part of Speech (PoS) Tags
- ➢ Chunking
  - o Categories of Phrases
  - o Phrase Structure Rules
- ➢ Chinking
- ➢ Named Entity Recognition (NER)
  - o Use-Cases
  - o Commonly used Types of Named Entity

- WordNet
- Bag of Words
  - What is the Bag-of-Words method?
  - Creating a basic Structure on Sentences
  - Words with Frequencies
  - Combining all the Words
  - Final Model of our Bag of Words
  - Applications & Limitations
- TF-IDF
  - Term Frequency
  - Inverse Document Frequency
  - Term Frequency - Inverse Document Frequency

Deploying a Machine Learning Model on a Web using Flask
- What is Model Deployment?
- What is Flask?
- Installing Flask on your Machine
- Understanding the Problem Statement
- Build our Machine Learning Model
- Create the Webpage
- Connect the Webpage with the Model
- Working of the Deployed Model

DEEP LEARNING INTRODUCTION
- What is Deep Learning?
- Deep learning Process
- Types of Deep Learning Networks
  - Deep Neural Networks
  - Artificial Neural Networks
  - Convolutional Neural Networks
  - Recurrent Neural Networks
- TensorFlow
  - History of TensorFlow
  - Components of TensorFlow
  - Use Cases/Applications of TensorFlow
  - Features of TensorFlow
- Installation of TensorFlow through pip & conda
- Advantage and Disadvantage of TensorFlow
- TensorFlow Playground
- Introduction to Keras, OpenCV & Theano
- Implementation of Deep Learning

ARTIFICIAL INTELLIGENCE INTRODUCTION

- What is Artificial Intelligence?
  - Why Artificial Intelligence?
  - Goals of Artificial Intelligence
  - What Comprises to Artificial Intelligence?
  - Advantages of Artificial Intelligence
  - Disadvantages of Artificial Intelligence
- Applications of Artificial Intelligence
- History of Artificial Intelligence
- Types of Artificial Intelligence
- Types of AI Agents
  - Simple Reflex Agent
  - Model-Based Reflex Agent
  - Goal-Based Agents
  - Utility-Based Agent
  - Learning Agent
- Search Algorithms in Artificial Intelligence
  - Search Algorithm Terminologies
  - Properties of Search Algorithms
  - Types of Search Algorithms
- Subsets of Artificial Intelligence
- Implementation of Artificial Intelligence

R PROGRAMMING

- Why R Programming is Important?
- Why Learn R?
- History of Python
- Features of R
- Applications of R
- Comparison between R and Python
- Which is Better to Choose
- Pros and Cons of R
- Companies using R
- R Packages
- Downloading and Installing R
- What is CRAN?
- Setting R Environment:
  - Search Packages in R Environment
  - Search Packages in Machine with inbuilt function and manual searching
  - Attach Packages to R Environment
  - Install Add-on Packages from CRAN

- o Detach Packages from R Environment
- o Functions and Packages Help
- ➢ R Programming IDE
  - o RStudio
  - o Downloading and Installing RStudio
- ➢ Variable Assignment
  - o Displaying Variables
  - o Deleting Variables
- ➢ Comments
  - o Single Line
  - o Multi Line Comments
- ➢ Data Types
  - o Logical
  - o Integer
  - o Double
  - o Complex
  - o Character
- ➢ Operators
  - o Arithmetic Operators
  - o Relational Operators
  - o Logical Operators
  - o Assignment Operators
  - o R as Calculator
  - o Performing different Calculations
- ➢ Functions
  - o Inbuilt Functions
  - o User Defined Functions
- ➢ STRUCTURES
  - o Vector
  - o List
  - o Matrix
  - o Data frame
  - o Array
  - o Factors
- ➢ Inbuilt Constants & Functions
- ➢ Vectors
  - o Vector Creation
  - o Single Element Vector
  - o Multiple Element Vector
  - o Vector Manipulation
  - o Sub setting & Accessing the Data in Vector
- ➢ Lists
  - o Creating a List

- - Naming List Elements
  - Accessing List Elements
  - Manipulating List Elements
  - Merging Lists
  - Converting List to Vector
- Matrix
  - Creating a Matrix
  - Accessing Elements of a Matrix
  - Matrix Manipulations
  - Dimensions of Matrix
  - Transpose of Matrix
- Data Frames
  - Create Data Frame
  - Vector to Data Frame
  - Character Data Converting into Factors: StringsAsFactors
  - Convert the columns of a data frame to characters
  - Extract Data from Data Frame
  - Expand Data Frame, Column Bind and Row Bind
- Merging / Joining Data Frames
  - Inner Join
  - Outer Join
  - Cross Join
- Arrays
  - Create Array with Multiple Dimensions
  - Naming Columns and Rows
  - Accessing Array Elements
  - Manipulating Array Elements
  - Calculations across Array Elements
- Factors
  - Factors in Data Frame
  - Changing the Order of Levels
  - Generating Factor Levels
  - Deleting Factor Levels

Loading and Reading Data in R

- Data Extraction from CSV
  - Getting and Setting the Working Directory
  - Input as CSV File, Reading a CSV File
  - Analyzing the CSV File, Writing into a CSV File
- Data Extraction from URL
- Data Extraction from CLIPBOARD
- Data Extraction from EXCEL

- o Install "xlsx" Package
- o Verify and Load the "xlsx" Package, Input as "xlsx" File
- o Reading the Excel File, Writing the Excel File
- ➢ Data Extraction from DATABASES
  - o RMySQL Package, Connecting to MySql
  - o Querying the Tables, Query with Filter Clause
  - o Updating Rows in the Tables, Inserting Data into the Tables
  - o Creating Tables in MySql, Dropping Tables in MySql
  - o Using dplyr and tidyr package

Machine Learning using R
- ➢ Data Pre-processing
- ➢ Classification Algorithms
  - o K Nearest Neighbors Classification
  - o Naive Bayes Classification
  - o Decision Tree Classification
  - o Random Forest Classification
  - o Support Vector Machine Classification
  - o Logistic Regression
  - o Kernel SVM
- ➢ Regression Algorithms
  - o Simple Linear Regression
  - o Multiple Linear Regression
  - o Polynomial Regression
  - o Support Vector Regression
  - o Decision Tree Regression
  - o Random Forest Regression
- ➢ Clustering Algorithms
  - o K-Means Clustering
  - o Hierarchical Clustering
- ➢ Association Rule Algorithms
  - o Apriori
  - o Eclat
- ➢ Dimensionality-Reduction
  - o Principal Component Analysis
  - o Linear Discriminant Analysis
  - o Kernal PCA
- ➢ Model Selection & Boosting
  - o Grid Search
  - o K Fold Cross Validation
  - o XGBoost
- ➢ Natural Language Processing
- ➢ Deep Learning - Artificial Neural Networks

DATA MINING WEKA

- ➢ Explore Weka Machine Learning Toolkit
  - ○ Installation of WEKA
  - ○ Features of WEKA Toolkit
  - ○ Explore & Load data sets in Weka
- ➢ Perform Data Preprocessing Tasks
  - ○ Apply Filters on Data Sets
- ➢ Performing Classification on Data Sets
  - ○ J48 Classification Algorithm
  - ○ Decision Trees Algorithm
  - ○ K-NN Classification Algorithm
  - ○ Naive-Bayes Classification Algorithm
  - ○ Comparing Classification Results
- ➢ Performing Regression on Data Sets
  - ○ Simple Linear Regression Model
  - ○ Multi Linear Regression Model
  - ○ Logistic Regression Model
  - ○ Cross-Validation and Percentage Split
- ➢ Performing Clustering on Data Sets
  - ○ Clustering Techniques in Weka
  - ○ Simple K-means Clustering Algorithm
  - ○ Association Rule Mining on Data Sets
  - ○ Apriori Association Rule Algorithm
  - ○ Discretization in the Rule Generation Process
- ➢ Graphical Visualization in Weka
  - ○ Visualization Features in Weka
  - ○ Visualize the data in various dimensions
  - ○ Plot Histogram, Derive Interesting Insights