Data Science Syllabus

Module-1 Python Syllabus

I. Python Introduction and Basics

- Variables
- Data types
- Global Variables
- Scope
- Operators
- Assignment Statement
- Indentation
- Interpreter
- Comments

II. Conditional Statements

- if
- if - elif
- if - elif - else
- Nested if

- Shorthand if
- Shorthand else
- match Statements

III. Iterative Statements

- for loop
- while loop
- Nested loop
- else   Class in loops

IV. Control Statements

- break Statement
- continue Statement
- pass Statement

V. Advanced Data Structures

- Lists
- Tuples & Sequences
- Sets
- Dictionaries

VI. Python Functions

- Default Arguments
- Keyword Arguments
- Positional or Keyword Arguments
- Positional - Only - Arguments
- Keyword - Only - Arguments
- Arbitrary Arguments
- Anonymous/Lambda
- Recursion

VII. Python Modules and Packages

- Standard Modules
- Executing modules as scripts
- Module Search Path
- Importing * From a Package
- Intra -package References
- Packages in Multiple Directories

VIII. File Input/Output

- Output Formatting
- Reading Files
- Writing Files

IX. Errors & Exceptions

- Handling Exceptions
- Raging Exceptions
- Exception Chaining
- Try Except

X. Object Oriented Programming

- Class
- Object
- Constructor
- Inheritance
- Polymorphism
- Abstraction
- Encapsulation

XI. Advanced Topics

- Iterator
- Generator
- Decorator
- Closure
- Regex

XII. Misc

- enum
- Arrays
- Multidimensional Arrays

Module-2 Python Libraries

I.Pandas

- Pandas Series
- Pandas DataFrames
- Pandas Read CSV
- Pandas Read JSON
- Pandas Analyzing data
- Cleaning Data

    - Cleaning Data
    - Cleaning Empty Cells
    - Cleaning Wrong Format
    - Cleaning Wrong Data
    - Removing Duplicates
- Correlations
- Plotting

II.Numpy

- NumPy - Ndarray Object
- NumPy - Array Attributes
- NumPy - Array Creation Routines
- NumPy - Array from Numerical Ranges
- NumPy - Advanced Indexing
- NumPy - Broadcasting
- NumPy - Iterating over Array
- NumPy - Array Manipulation
- NumPy - Binary Operators
- NumPy - String Functions
- NumPy - Mathematical Functions

- NumPy - Arithmetic Operations
- NumPy - Statistical Functions
- NumPy - Sort, Search & Counting Functions
- NumPy - Copies & Views
- NumPy - Matrix Library
- NumPy - Linear Algebra

III.Matplotlib

- Matplotlib Pyplot
- Matplotlib Plotting
- Matplotlib Markers
- Matplotlib Line
- Matplotlib Labels
- Matplotlib Grid
- Matplotlib Subplot
- Matplotlib Scatter
- Matplotlib Bars
- Matplotlib Histograms
- Matplotlib Pie Charts

IV.Scikit-learn

- Use of Scikit-learn in Data Science Life Cycle
    - Introduction to Data Science Life Cycle
    - Scikit-learn for Data Preprocessing
    - Treating missing values
    - Treating Outliers
    - Feature Engineering
    - Dimensionality Reduction
- Use of Scikit-Learn in Model Building
    - Introduction to Model Building and Evaluation
    - Regression
    - Classification
    - Clustering
- Machine Learning pipeline using scikit-learn!
    - Introduction
    - Understanding Problem Statement
    - Building a prototype model
    - Data Exploration and Preprocessing

- Encode the categorical variables
- Scale the data
- Model Building
- Feature Importance
- Identifying features to build the ML pipeline
- Pipeline Design
- Building Pipeline
- Predict the Target

Module-3 Implementation of ML Algorithms Using Python

- Installation IDE (integrated development environment)
- Implementation All Machine Learning Algorithms using python libraries like Pandas, Numpy, Matplotlib and Scikit
- Implementation Classification Algorithms using python
- Implementation Regression Algorithms using python
- Implementation Clustering Algorithms using python
- Implementation Association Algorithms using python
- Implementation Reinforcement Algorithms using python

Module-4 Statistics Syllabus

- What is Statistics?
- Uses of Statistics
- Need of Statistics
- Mean, Median, Mode
- Variance, Standard deviation
- Sensitivity
- Information Gain
- Data Distributions
- Central Limit Theorem
- Entropy
- Statistical parameters to represent data
- Uses of probability
- Need of probability
- Data Distributions
- Central Limit Theorem
- Bayesian Inference
- Density Concepts
- Normal Distribution Curve

- Association and Dependence
- Causation and Correlation
- Covariance
- what is pValue
- Chi-Square test
- What is ANOVA
- Correlation vs Regression
- Uses of Correlation and Regression

Module-5 Calculus Syllabus

- Proof by Induction
- Limits at Infinity
- Limits at a Specific Point
- Continuity of a Function
- Using the Derivative to Graph Functions
- Finding Concavity with the Second Derivative
- Comparing the Graphs of f(x) and f’(x)
- Maxima and Minima of Functions
- Applications of Maxima and Minima

- Graphing Using Maxima and Minima
- Rate of Change
- Calculus Notation
- Computation Map
- Linear Functions
- Derivative
- Product & Quotient Rule
- Exponential & Logarithm Rule
- Sine and Cosine Functions
- Sigmoid Function
- Hyperbolic Function
- Chain Rule
- Differentiation
- Multivariable Functions
- Partial Differentiation
- Sigma Notation
- Taylor/Mclaurin Series Expansion
- Integration
- Area Between Curves

- Distance Measurement
- Argmax – Argmin

Module-6 Linear Algebra Syllabus

- Matrices
- solving linear equations
- vector spaces
- linear independence
- basis and rank
- linear
- mappings
- affine spaces
- norms
- inner products
- orthogonality
- orthonormal basis
- inner product of functions
- orthogonal projections

- Vectors
- Matrices
- Transpose of a matrix
- Inverse of a matrix
- Determinant of a matrix
- Trace of a matrix
- Dot product
- Eigenvalues
- Eigenvectors

Machine Learning Syllabus

Module-7 Supervised Learning

I. Machine Learning

- What is Machine Learning
- Use of Machine Learning
- Machine Learning Applications
- Types in Machine Learning

- What is Supervised Learning
- What is Unsupervised Learning
- Difference Between Supervised and Unsupervised Learning

II. Classifications

- What is Classification
- Types of Classification
- K- Nearest Neighbors
- Logistic Regression
- Naive Bayes
- Decision Tree Classification
- Random Forest Classification
- Support Vector Machines
- Implementation using Python

III. Regression

- What is Regression
- Types of Regression
- Linear Regression
- Multinomial Regression

- Polynomial Regression
- Lasso Regression
- Ridge Regression
- Implementation using Python

Module-8 Unsupervised Learning

IV. Clustering

- What is Clustering
- Methods of Clustering
- Partitioning Methods
    - K-Means Clustering
    - K-Medoid Clustering
- Hierarchical Methods
    - Agglomerative Clustering
    - Divisive Clustering
    - BIRCH
    - Chameleon
    - Probabilistic Hierarchical Clustering
- Density Based Methods

    - DBSCAN
    - OPTICS
    - DENCLUE

- Grid Based Methods
    - STING
    - CLIQUE
- Implementation using Python

V. Association

- What is Association
- Types of Association
- Applications of Association
- Apriori Algorithm
- Eclat Algorithm
- Q-Learning
- F-P Growth Algorithm
- Implementation using Python

Module-9 Reinforcement Learning

VI. Reinforcement Learning

- What is Reinforcement Learning
- Types of Association
- Applications of Reinforcement Learning
- Upper Confidence Bound
- Thompson Sampling
- Implementation using Python