

Data Science (AI & ML) Course Content

MODULE 1: Introduction to Python - Data Science

- Installation of Anaconda setup (Data Science Development Environment)
- Installation of Pycharm
- Working with Python List, List operation, Functions
- Python Tuple, working and functions
- Sets and Dictionary -operations and Working with them
- Python More on Strings
- Python Dates and Times
- More on functions
- Advanced Python Lambda
- List Comprehensions

MODULE 2: Data Analysis

1. Data Wandering

- All about files
- importing and exporting data with CSV files
- XLRD module - working with xls .xlsx formats
- Json data
- XML data
- Relational data Bases
- Sql in python
- Data quality Analysis

2. DATA MANIPULATION - Cleaning - Munging - Cleansing Data with Python

- Data Manipulation steps (Sorting, filtering, duplicates, merging, appending, subsetting, derived variables, sampling, Data type conversions, renaming, formatting etc.)
- Data manipulation tools (Operators, Functions, Packages, control structures, Loops, arrays etc.)
- Python Built-in Functions (Text, numeric, date, utility functions)
- Python User Defined Functions
- Stripping out extraneous information
- Normalizing data
- Formatting data
- Important Python modules for data manipulation (Pandas, Numpy, re, math, string, date time etc.)

3. DATA VISUALIZATION

- Introduction exploratory data analysis
- Descriptive statistics, Frequency Tables and summarization
- Univariate Analysis (Distribution of data & Graphical Analysis)
- Bivariate Analysis (Cross Tabs, Distributions & Relationships, Graphical Analysis)
- Creating Graphs- Bar/pie/line chart/histogram/ boxplot/ scatter/ density etc)
- Important Packages for Exploratory Analysis (Numpy Arrays, Matplotlib, seaborn, Pandas and scipy.stats etc)

4. DATA ANALYSIS WITH PANDAS

- The Series Data Structure
- Querying a Series
- The Data-Frame Data Structure
- Data-Frame Indexing and Loading
- Querying a Data-Frame
- Indexing Data-frame
- Understanding business problem
- Selecting columns from Pandas Data Structures
- Treating with missing values, outliers, NaN values
- Creating new columns
- Aggregate data (use: groupby, merge, pivot, lambda)
- Identifying unique values in data
- Filter Data
- Using basic functionality of Pandas API

MODULE 3: Mathematics

1. STATISTICS

- Basic Statistics - Measures of Central Tendencies and Variance
- Building blocks - Probability Distributions - Normal distribution - Central Limit Theorem
- Inferential Statistics -Sampling - Concept of Hypothesis Testing
- Statistical Methods - Z/t-tests(One sample, independent, paired), Anova, Correlations and Chi-square
- Important modules for statistical methods: Numpy, Scipy, Pandas

2. PROBABILITY

- Probability , Conditional Probability
- Basic of Probability, Independent and Dependent events
- Conditional Probability and Bayes Theorem
- Continuous Probability Distributions
- Mean, Median, Mode, Range
- Determination of statistical techniques
- Standard Deviation, Variance, Covariance, Correlation
- outliers
- Distribution of Data – Normal, Binomial, Gaussian
- Different types of Data
- Continuous , Categorical, Range
- Testing of Hypothesis – which covers
- Level of Significance (LOS), Level of Confidence, P-Value, T test, Z-test, ANOVA Test, CHI - Square Test

MODULE 4: Machine Learning

1. SUPERVISED LEARNING AND MODEL BUILDING

- Process of Machine Learning
- Model Building based on Data sets
- Splitting Data: Training and Test sets
- Regression Analysis (Linear, Multiple, Logistics Regression)
- Classification concepts and Distance Functions
- K-nn Algorithm concept and demonstration with data sets
- Bayes Classification concept and demonstration with data sets
- Decision Tree Algorithm concept and demonstration with data sets
- Random Forests - Ensembling Techniques and Algorithms

2. UNSUPERVISED LEARNING AND MODEL BUILDING

- Unsupervised Learning and Clustering Techniques
- Centroid-based Clustering: K- Mean Algorithm concept and demonstration
- Hierarchical Clustering concepts and Applications
- Density-based Clustering: DBSCAN Algorithm concept and demonstration

3. DIMENSION REDUCTION TECHNIQUES

- Dimension Reduction Introduction
- Why Dimension Reduction Required
- LDA (Linear Discriminant Analysis) concept and applications
- PCA (Principle Component Analysis) concept and applications

4. TIME SERIES FORECASTING: SOLVING FORECASTING PROBLEMS

- Introduction - Applications
- Time Series Components(Trend, Seasonality, Cyclicity and Level) and Decomposition
- Classification of Techniques(Pattern based - Pattern less)
- vBasic Techniques - Averages, Smoothing
- Advanced Techniques - AR Models, ARIMA

5. DATA SCIENCE PROJECTS WITH DATA SETS

- Applying different algorithms to solve the business problems and bench mark the results.