ETHANS TECH

#PUNE'S NO. 1 TRAINING

DATA SCIENCE CURRICULUM

We Are Open For Career Counselling At Our Branches

For any query, contact us at-

Pimple-8698585003 | Baner-7620182669

Kharadi-8551013133 | Noida-9319222386

DATA SCIENCE TRAINING SYLLABUS

Module-1: Python Basics and Foundation

- What is Python and how it is different from other programming?
- Why companies are preferring Python as generic programming Language?
- Where to use Python in Day to Day IT life?
- Difference between about Python2 and Python3
- Download and Installing Python from python.org
- Set up Python on Windows, Mac and Linux Machine.
- Understanding different Python IDE's like IDLE, Pycharm etc.
- What are the unique features of Python as a language?
- Use Python on Interactive shell and Programming Env.
- Write your first Python ScriptUnderstand Variables and Predefined - Keywords in Python

Python Builtin Classes and Objects

- What is Python Built-in Classes and Objects?
- Int, Float and Complex Class and Object in Python.
- What are operators and how to use the in number objects.
- Str class and Objects in Python.
- How to use operators in string objects.List class and Objects in Python.
- How to use operators in List objectsTuple class and Object
- How to use operators in Tuple objectsDictionary Class and Object
- How to use operators in dict objectsSet and frozenset class and object
- Bool class and ObjectsNone Class and Objects
- Different data Structures, data processing Techniques to learn.

Programming Techniques in Python

- Learn conditional statements and Loops
- What are conditional statement?
- What is if, else, elif blocks and how to use indentations in blocks.
- Different syntax available in if-elif-else blocks.
- How to define loops in Python?
- What is for and while loops?
- How to define condition on while loop.
- What is iterator object and how to use them loops.
- Control the loops using break and continue
- How to iterate through the various objectUnderstand sequence and iterable objects

User Defined functions and Built-in Class function in Python

- What are functions?
- Difference between Builtin and User defined functions in Python.
- How to create user defined functions using def.
- What is return statement and use of return in functions.
- Defined user defined functions using parameters - positional and keyword parameter.
- Parameterize User defined function, through named and unnamed parameters
- Introducing Lambda functions.
- Str class and Objects and its functionsList class and Objects and its functions
- Tuple class and and its functions
- Dictionary Class and its functions
- Set, frozenset class and object and its functions

Understand Text File Handling with Python

- Handle different format types of files in Python.
- How to read, write and append data using Python
- What is file class and what are different attributes and functions available?
- read, readlines, readline, write, writelines, flush, close functions in Python.
- What is context manager and how to use them.
- Defined context manager for file handling.
- Process file data.CSV, DAT, TXT, file handling.
- File pointer and seek the pointerIntroduction to Python - Builtin Modules
- os module in Python and functions available in os and os.path

DATA SCIENCE TRAINING SYLLABUS, CONT1.

Module-2 – Python Advance

Python Libraries/Modules and Packages

- Discussion on Python Modules.User Defined Modules, Python inbuilt Modules/libraries and 3rd Party
- Modules
- Python inbuilt libraries - os, sys, logging, datetime, time, zip, json, csv etc
- Create your own Python Modules
- Configuration of Python UDM and Define PYTHONPATH
- Create your own Python Packages
- understand __init__.py File for package initialization
- Subpackage creation and modules usecases on sub packages.

Python Exceptional Handling Features

- What is exception class and different builtin exceptions.
- How to handle Python run time exceptions
- ZeroDivisionException, NameError, TypeError and Generic Exception Class
- Handling various exceptions using try....except...elseTry-finally clause
- Argument of an Exception and create self exception class
- Python Standard Exceptions
- Raising an exceptions
- User-Defined Exceptions

Objected Oriented - Python Programming

- Object oriented features in Python
- How to create classes, object and methods
- Understand real world examples on OOP
- Encapsulation, Inheritance, Polymorphism features in Oops
- Implement Object oriented with Python
- Creating Classes and Objects, Destroying Objects
- Accessing attributes, Built-In Class Attributes
- Overriding Methods, Data Hiding
- Overloading Operator

Python Debugging Techniques

- Pycharm debugger
- Understanding Breakpoint, continue, step in, step out and step over
- Assert statement for debugging
- Debug Python programs using pdb debugger
- Step by Step Executions of program.

Introduction of Regular Expression

- What are regular expressions?
- match, compile, search and findall Function
- Matching vs searching operations
- Search and Replace feature using RE
- Extended Regular Expressions
- Wildcard characters and work with them

Python - Database Interaction

- Introduction of Python Library sqlite3
- Creating Databases and Tables
- CRUD Operations
- Creating a Database Object.
- Cursor Object and query execution.
- DML and DDL Operations with Databases
- Performing TransactionsHandling Database Errors
- Disconnecting Database

Introduction to pip and package installation

- Introduction to pip and package installation.
- What is pip?
- Install, download, uninstall, upgrade and search packages using pip.
- Install package using pip and Pycharm
- What is pip, easy_install
- Set up the environment to install packages?
- Install packages for XLS interface and XLS parsing with Python
- Create XLS reports with Python

DATA SCIENCE TRAINING SYLLABUS, CONT 2

Module-3 – Python for Analytics

Introduction to Anaconda Distribution

- What is Anaconda Distribution?
- Jupyter Notebook
- How it is different from Python Distribution?
- How to install Anaconda?
- conda repository pip and conda to get new package
- pip and conda commands
- set Virtual environment using conda
- Integrating Anaconda with Pycharm.

Introduction to numpy and statistical analysis

- numpy performance test with Python
- Introduction to numpy arrays
- Introduction to numpy functions
- Dealing with Flat files using numpy
- Mathematical functions
- Statistical function
- Operations with arrays

Introduction to pandas framework and statistical analysis

- What is Python - Pandas framework.
- Creating Series and data filter/transformation
- Creating Data Frames on pandas
- Grouping and Sorting of dataset
- read data from multiple sources. csv, xls, json etc
- Data analysis with data set
- Practical use cases using data analysis
- mini projects

Introduction to Python Matplotlib Library

- What is matplotlib and its gallery.
- Create different graph from dataset
- pie, bar, line, horizontal bar and different for of graphs.
- read data from multiple sources. csv, xls, json and plot graphs
- Data analysis with data set and plot graphs
- Practical use cases using data analysis
- Mini Projects

DATA SCIENCE TRAINING SYLLABUS, CONT 3

Module-4 - Python for Machine Learning

Introduction to Anaconda Distribution

- What is Anaconda Distribution?
- How it is different from Python Distribution?
- How to install Anaconda?
- conda repositoryAnaconda Navigator
- pip and conda to get new package
- pip and conda commands
- set Virtual
- Integrating Anaconda with Pycharm

Using Git and GitHub

- Setting up Your GitHub Account
- Configuring Your First Git Repository
- Making Your First Git Commit
- Pushing Your First Commit to GitHub
- Git and GitHub Workflow Step-by-Step
-

Introduction to SQL and DataBases

- SQL/RDBMS database management
- SQL Queries
- CRUD Operations
-

Introduction to numpy and statistical Analysis

- What is numpy?
- Numpy performance test
- Introduction to numpy arrays and Matrices
- Indexing and Selection
- Introduction to numpy function
- Numpy Operations Array with Array, Array with Scalars, Universal Array Functions
- Dealing with Flat files using numpy
- Mathematical functions
- Statistical functions

Introduction to Pandas and Data Analysis

- Integrating Anaconda with Pycharm
- What is Pandas
- Creating Series
- Creating Data Frames
- Grouping, Sorting
- Group by Operations
- Merging, Joining and Concatenating DataFrame
- Pandas Operations
- Data Input and Output from a variety of data formats like csv, excel, db, json and html
- Missing Data (Imputation)
- Data analysis with data set
- Practical use cases using data analysis

Statistics and Probability

- Type of Dataset: Numerical,Categorical and Ordinal
- Mean, Median and Mode
- Variance and Standard Deviation
- Probability Density Function(PDF) and Probability Mass Function (PMF)
- Percentiles and Moments
- Covariance and Correlation
- Conditional Probability
- Bayes' Theorem Module

Data Visualization using Matplotlib

- Plotting for exploratory data analysis (EDA)
- Line Graph on time Series
- Pie Chart, Bar and Horizontal Bar Graph
- Introduction to IRIS dataset
- 2D scatter plot
- Pair plots
- Histogram and Introduction to PDF(Probability Density Function)
- CDF(Cumulative Distribution Function)

DATA SCIENCE TRAINING SYLLABUS, CONT 4

Linear Algebra and Calculus

- Introduction to Vectors(2-D, 3-D, n-D) , Row Vector and Column Vector
- Dot Product and Angle between 2 Vectors
- Projection and Unit Vector
- Equation of a line (2-D), Plane(3-D) and Hyperplane (n-D), Plane Passing through origin, Normal to a Plane
- Distance of a point from a Plane/Hyperplane, Half-Spaces
- Equation of a Circle (2-D), Sphere (3-D) and Hypersphere (n-D)
- Equation of an Ellipse (2-D), Ellipsoid (3-D) and Hyperellipsoid (n-D)
- Square ,Rectangle

Machine Learning Introduction

- What is Machine Learning?
- Machine Learning Process
- Different Categories of Machine Learning: Supervised,Unsupervised and Reinforcement
- Scikit-Learn Overview
- Scikit-Learn cheat-sheet

Classification - k-Nearest Neighbor(KNN)

- Classification and Regression
- Application, Advantages and Disadvantages
- Distance Metric - Euclidean, Manhattan, Chebyshev, Minkowski
- Measuring accuracy using Cross-Validation, Stratified k-fold, Confusion Matrix, Precision, Recall, F1-score.

- Breast Cancer Wisconsin (Diagnostic) Project using KNN.
- https://www.kaggle.com/uciml/breast-cancer-wisconsin-data

- Iris Species
- https://www.kaggle.com/uciml/iris

Classification - Naive Bayes

- Conditional probability
- Independent vs Mutually exclusive events
- Bayes Theorem with examples
- Exercise problems on Bayes Theorem
- Naive Bayes algorithm
- Toy example: Train and test stages
- Naive Bayes on Text data
- Laplace/Additive Smoothing
- Log-probabilities for numerical stability
- Bias and Variance tradeoff
- Feature importance and interpretability
- Code example

Logistic Regression

- Geometric intuition of Logistic Regression
- Sigmoid function: Squashing
- Mathematical formulation of Objective function
- Weight vector
- L2 Regularization: Overfitting and Underfitting
- L1 regularization and sparsity
- Probabilistic Interpretation: Gaussian Naive Bayes
- Loss minimization interpretation
- Hyperparameter Search: Grid search and random search
- Column Standardization
- Feature importance and Model interpretability
- Collinearity of features
- Train & Run time space & time complexity
- Non-linearly separable data & feature engineering
- Code sample: Logistic regression, GridSearchCV, RandomSearchCV
- Extensions to Logistic Regression: Generalized linear models(GLM)

Digit Recognizer
https://www.kaggle.com/c/digit-recognizer

Titanic: Machine Learning from Disaster
https://www.kaggle.com/c/titanic

DATA SCIENCE TRAINING SYLLABUS, CONT 5

Linear Regression

- What is Linear Regression
- Geometric intuition of Linear Regression
- Mathematical formulation
- Real world Cases
- Code sample for Linear Regression

Projects

Predicting Boston House Prices https://www.kaggle.com/schirmerchad/bostonhoustingmlnd

Insurance forecast
https://www.kaggle.com/mirichoi0218/insurance

Classification - SVM (Support Vector Machine)

- Classification and Regression
- Separating line, Margin and Support Vectors
- Linear SVC Classification
- Polynomial Kernel – Kernel Trick
- Gaussian Radial Basis Function (rbf)
- Grid Search to tune hyper-parameters.
- Support Vector Regression
-

1. Breast Cancer Wisconsin (Diagnostic) Project using KNN
https://www.kaggle.com/uciml/breast-cancer-wisconsin-data

2. Iris Species
https://www.kaggle.com/uciml/iris

Classification - Decision Tree

- CART (Classification and Regression Tree)
- Advantages and Disadvantages and its applications
- Decision Tree Learning algorithms – ID3, C4.5, C5.0 and CART
- Gini Impurity, Entropy and Information Gain
- Decision Tree Regression
- Visualizing a Decision Tree using graphviz module.
- Regularization using tuning hyper-parameters using GridSearch CV.

Projects

IBM HR Analytics Employee Attrition and Performance
https://www.kaggle.com/pavansubhasht/ibm-hr-analytics-attrition-dataset

Zomato Restaurants Data
https://www.kaggle.com/shrutimehta/zomato-restaurants-data

Unsupervised Learning

- Unsupervised learning
- Metrics for Clustering
- K-Means: Geometric intuition, Centroids
- K-Means: Mathematical formulation: Objective function
- K-Means Algorithm.
- How to initialize: K-Means++
- Failure cases/Limitations
- K-Medoids
- Determining the right K

Projects

1. Analyze Lending Club's issued loans https://www.kaggle.com/wendykan/lending-club-loan-data

2. Credit Card Dataset for Clustering
https://www.kaggle.com/arjunbhasin2013/ccdata

DATA SCIENCE TRAINING SYLLABUS, CONT 6

Density based clustering

- DBSCAN (Density based clustering) Technique
- Density based clustering
- Eps: Density
- Core, Border and Noise points
- Density edge and Density connected points.
- DBSCAN Algorithm
- Hyper Parameters: MinPts and EpsA
- Advantages and Limitations of DBSCAN

Project

Compares socio-economic info with suicide rates by year and country
https://www.kaggle.com/russellyates88/suicide-rates-overview-1985-to-2016