Full Stack
DATA SCIENCE & AI

PYTHON

Introduction to Data Science

- Introduction to Data Science
- Discussion on Course Curriculum
- Introduction to Programming

Python – Basics

- Introduction to Python: Installation and Running
  (Jupyter Notebook, .py file from terminal, Google Colab)
- Data types and type conversion
- Variables
- Operators
- Flow Control : If, Elif, Else
- Loops
- Python Identifier
- Building Funtions (print, type, id, sys, len)

Python - Data Types & Utilities

- List, List of Lists and List Comprehension
- List creation
- Create a list with variable
- List mutable concept
- len() || append() || pop()
- insert() || remove() || sort() || reverse()
- Forward indexing
- Backward Indexing
- Forward slicing
- Backward slicing
- Step slicing

Set

- SET creation with variable
- len() || add() || remove() || pop()
- union() | intersection() || difference()

Tuple

- TUPLE Creation
- Create Tuple with variable
- Tuple Immutable concept
- len() || count() || index()
- Forward indexing
- Backward Indexing

Dictionary and Dictionary comprehension

- create a dictionary using variable
- keys:values concept
- len() || keys() || values() || items()
- get() || pop() || update()
- comparision of datastructure
- Introduce to range()
- pass range() in the list
- range() arguments
- For loop introduction using range()

Functions

- Inbuilt vs User Defined
- User Defined Function
- Function Argument
- Types of Function Arguments
- Actual Argument
- Global variable vs Local variable
- Anonymous Function | LAMBDA

Packages

Map Reduce

OOPs

Class & Object:

- what is mean by inbuild class
- how to creat user class
- crate a class & object
- __init__ method
- python constructor
- constructor, self & comparing objects
- instane variable & class variable

Methods:

- what is instance method
- what is class method
- what is static method
- Accessor & Mutator

Python DECORATOR:

- how to use decorator
- inner class, outerclass
- Inheritence

Polymorphism:

- duck typing
- operator overloading
- method overloading
- method overridding
- Magic method
- Abstract class & Abstract method
- Iterator
- Generators in python

Python - Production Level

- Error / Exception Handling
- File Handling
- Docstrings
- Modularization

Pickling & Unpickling

Pandas

- Introduction, Fundamentals, Importing Pandas, Aliasing, DataFrame
- Series – Intro, Creating Series Object, Empty Series Object, Create series from List/Array/Column from DataFrame, Index in Series, Accessing values in Series
- NaN Value
- Series – Attributes (Values, index, dtypes, size)
- Series – Methods – head(), tail(), sum(), count(), nunique() etc.,
- Date Frame
- Loading Different Files
- Data Frame Attributes
- Data Frame Methods
- Rename Column & Index
- Inplace Parameter
- Handling missing or NaN values
- iLoc and Loc
- Data Frame – Filtering
- Data Frame – Sorting
- Data Frame – GroupBy
- Merging or Joining
- Data Frame – Concat
- DataFrame - Adding, dropping columns & rows
- DataFrame - Date and time
- DataFrame - Concatenate Multiple csv files

Numpy

- Introduction, Installation, pip command, import numpy package, ModuleNotFoundError, Famous Alias name to Numpy
- Fundamentals – Create Numpy Array, Array Manipulation, Mathematical Operations, Indexing & Slicing
- Numpy Attributes
- Important Methods- min(),max(), sum(), reshape(), count_nonzero(), sort(), flatten() etc.,
- adding value to array of values
- Diagonal of a Matrix
- Trace of a Matrix
- Parsing, Adding and Subtracting Matrices
- "Statistical Functions: numpy.mean()
- numpy.median()
- numpy.std()
- numpy.sum()
- numpy.min()"
- Filter in Numpy

Matplotlib

- Introduction
- Pyplot
- Figure Class

- Axes Class
- Setting Limits and Tick Labels
- Multiple Plots
- Legend
- Different Types of Plots:
- Line Graph
- Bar Chart
- Histograms,
- Scatter Plot
- Pie Chart
- 3D Plots
- Working with Images
- Customizing Plots

Seaborn

- catplot() function
- stripplot() function
- boxplot() function
- violinplot() function
- pointplot() function
- barplot() function
- Visualizing statistical relationship with Seaborn relplot() function
- scatterplot() function
- regplot() function
- Implot() function
- Seaborn Facetgrid() function
- Multi-plot grids
- Statistical Plots:
- Color Palettes:
- Faceting:
- Regression Plots:
- Distribution Plots
- Categorical Plots:
- Pair Plots

Scipy

- Signal and Image Processing (scipy.signal, scipy.ndimage):
- Linear Algebra (scipy.linalg):
- Integration (scipy.integrate)
- Statistics (scipy.stats):
- Spatial Distance and Clustering (scipy.spatial):

Statsmodels

- Linear Regression (statsmodels.regression):
- Time Series Analysis (statsmodels.tsa):
- Statistical Tests (statsmodels.stats)
- Anova (statsmodels.stats.anova):
- Datasets (statsmodels.datasets):

Mathematics

Set Theory

- Data Representation & Database Operations

Combinatorics

- Feature Selection
- Permutations and Combinations for Sampling
- Hyperparameter Tuning
- Experiment Design
- Data Partitioning and Cross-Validation

Probability

- Basics
- Theoretical Probability
- Empirical Probability
- Addition Rule
- Multiplication Rule
- Conditional Probability
- Total Probability
- Probability Decision Tree
- Bayes Theorem
- Sensitivity & Specificity in Probability
- Bernouli Naïve Bayes, Gausian Naïve Bayes, Multinomial Naïve Bayes

Distributions

- Binomial, Poisson, Normal Distribution, Standard Normal Distribution
- Guassian Distribution, Uniform Distribution
- Z Score
- Skewness
- Kurtosis
- Geometric Distribution
- Hyper Geometric Distribution
- Markov Chain

Linear Algebra

- Linear Equations
- Matrices(Matrix Algebra: Vector Matrix Vector matrix multiplication Matrix matrix multiplication)
- Determinant
- Eigen Value and Eigen Vector

Euclidean Distance & Manhattan Distance

Calculus

- Differentiation
- Partial Differentiation
- Max & Min

Indices & Logarithms

STATISTICS

Introduction

- Population & Sample
- Reference & Sampling technique

Types of Data

- Qualitative or Categorical – Nominal & Ordinal
- Quantitative or Numerical – Discrete & Continuous
- Cross Sectional Data & Time Series Data

Measures of Central Tendency

- Mean, Mode & Median – Their frequency distribution

Descriptive statistic Measures of symmetry

- skewness (positive skew, negative skew, zero skew)
- kurtosis (Leptokurtic, Mesokurtic, Platrykurtic)

Measurement of Spread

- Range, Variance, Standard Deviation

Measures of variability

- Interquartile Range (IQR):
- Mean Absolute Deviation (MAD)
- Coefficient of variation
- Covariance

Levels of Data Measurement

- Nominal, Ordinal, Interval, Ratio

Variable

- Types of Variables.
- Categorical Variables - Nomial variable & ordinal variables
- Numerical Variables: discreate & continuous
- Dependent Variable
- Independent Variable
- Control Moderating & Mediating

Frequency Distribution Table

- Nominal, Ordinal, Interval, Ratio

Types of Variables.

- Categorical Variables - Nomial variable & ordinal variables
- Numerical Variables: discreate & continuous
- Dependent Variable
- Independent Variable
- Control Moderating & Mediating

Frequency Distribution Table

- Relative Frequency, Cumulative Frequency
- Histogram
- Scatter Plots
- Range
- Calculate Class Width:
- Create Intervals
- Count Frequencies
- Construct the Table

Correlation, Regression & Collinearity

- Pearson & Spearman Correlation Methods
- Regression Error Metrics

Others

- Percentiles, Quartiles, Inner Quartile Range
- Different types of Plots for Continuous, Categorical variable
- Box Plot, Outliers
- Confidence Intervals
- Central Limit Theorem
- Degree of freedom

Bias and Variance in ML

Entropy in ML

Information Gain

Surprise in ML

Loss Function & Cost Function

- Mean Squared Error, Mean Absolute Error – Loss Function
- Huber Loss Function
- Cross Entropy Loss Function
-

Inferential Statistics

- Hypothesis Testing: One tail, two tail and p-value
- Formulation of Null & Alternate Hypothesis
- Type-I error & Type-II error
- Statistical Tests:
- Sample Test
- ANOVA Test
- Chi-square Test
- Z-Test & T-Test

SQL

Introduction

- DBMS vs RDBMS
- Intro to SQL
- SQL vs NoSQL
- MySQL Installation

Keys

- Primary Key
- Foreign Key

Constraints

- Unique
- Not NULL
- Check
- Default
- Auto Increment

CRUD Operations

- Create
- Retrieve
- Update
- Delete

SQL Languages

- Data Definition Language (DDL)
- Data Query Language
- Data Manipulation Language (DML)
- Data Control Language
- Transaction Control Language

SQL Commands

- Create
- Insert
- Alter, Modify, Rename, Update
- Delete, Truncate, Drop
- Grant, Revoke
- Commit, Rollback
- Select

SQL Clause

- Where
- Distinct
- OrderBy
- GroupBy
- Having
- Limit

Operators

- Comparison Operators
- Logical Operators
- Membership Operators
- Identity Operators

Wild Cards

Aggregate Functions

SQL Joins

- Inner Join & Outer Join
- Left Join & Right Join
- Self & Cross Join
- Natural Join

EDA & ML

EDA

- Univariate Analysis
- Bivariate Analysis
- Multivariate Analysis

Data Visualisation

- Various Plots on different datatypes
- Plots for Continuous Variables
- Plots for Discrete Variables
- Plots for Time Series Variables

ML Introduction

- What is Machine Learning?
- Types of Machine Learning Methods
    - Supervised Learning
    - Unsupervised Learning
    - Reinforcement Learning)
- Classification problem in general
- Validation Techniques: CV,OOB
- Different types of metrics for Classification
- Curse of dimensionality
- Feature Transformations
- Feature Selection
- Imabalanced Dataset and its effect on Classification
- Bias Variance Tradeoff

Important Element of Machine Learning

Multiclass Classification

- One-vs-All
- Overfitting and Underfitting
- Error Measures
- PCA learning
- Statistical learning approaches
- Introduce to SKLEARN FRAMEWORK

Data Processing

- Creating training and test sets, Data scaling and Normalisation
- Feature Engineering – Adding new features as per requirement, Modifying the data
- Data Cleaning – Treating the missing values, Outliers
- Data Wrangling – Encoding, Feature Transformations, Feature Scaling
- Feature Selection – Filter Methods, Wrapper Methods, Embedded Methods
- Dimension Reduction – Principal Component Analysis (Sparse PCA & Kernel PCA), Singular Value Decomposition
- Non Negative Matrix Factorization

Regression

- Introduction to Regression
- Mathematics involved in Regression
- Regression Algorithms:
- Simple Linear Regression
- Multiple Linear Regression
- Polynomial Regression
- Lasso Regression
- Ridge Regression
- Elastic Net Regression

Evaluation Metrics for Regression:

- Mean Absolute Error (MAE)
- Mean Squared Error (MSE)
- Root Mean Squared Error (RMSE)
- R^2
- Adjusted R^2

Classification

- Introduction
- K-Nearest Neighbors
- Logistic Regression:
    - Implementation and Optimizations
    - Stochastic gradient descent algorithms
    - Finding the optimal HyperParameters through Grid Search
- Support Vector Machines (Linear SVM):

- Linear support vector machines
- Scikit-learn implementation

- Linear Classification
- Kernel-based classification
  - Radial Basis Function
  - Polynomial Kernel
  - Sigmoid Kernel
  - Custom Kernels
- Non-linear examples
- 2 features forms straight line & 3 features forms plane
- Hyperplane and Support vectors
- Controlled support vector machines
- Support vector Regression
- Kernel SVM (Non-Linear SVM)
- Naives Bayes:
  - Bayes theorem
  - Naive Bayes Classifiers
  - Naive Bayes in scikit learn (Bernoulli Naive Bayes, Mulitnomial Naive Bayes, Guassian Naive Bayes)"
- Decision Trees:
  - Binary Decision Trees
  - Binary decisions
  - CART Algorithm
  - Impurity measures (Gini impurity index, Cross-entropy impurity index, Misclassification impurity index)
  - Feature importance
  - Decision tree classification with scikit-learn
- Random Forest / Bagging:
  - Random Forests and Features importance in Random Forest
  - AdaBoost
  - Gradient tree boosting
  - Voting classifier
  - Ensemble:Bagging
  - Ensemble:Boosting"
- Ada Boost
- Gradient Boost
- XG Boost
- Evaluation Metrics for Classification:
  - Confusion Matrix
  - Accuracy & F1 Score
  - Precision & Recall
  - Sensitivity & Specificity
  - True Positive Rate, False Positive Rate
  - ROC & ROC_AUC

Clustering

Introduction

K-Means Clustering:

- Finding the optimal number of clusters
- Optimizing the inertia
- Cluster instability
- Elbow method

Hierarchical Clustering

Agglomerative clustering

DBSCAN Clustering

Association Rules

- Market Basket Analysis
- Apriori Algorithm

Recommendation Engines

- Collaborative Filtering:
- User based collaborative filtering
- Item based collaborative filtering
- Recommendation Engines

Time Series & Forecasting

- What is Time series data
- Different components of time series data
- Stationary of time series data
- ACF, PACF
- Time Series Models:
- AR
- ARMA
- ARIMA
- SARIMAX

Model Selection & Evaluation

Over Fitting & Under Fitting

- Biance-Variance Tradeoff
  - Cross Validation:
  - Stratified Cross validation
  - K-Fold Cross validation
- Hyper Parameter Tuning
- Joblib And Pickling

Others

- Dummy Variable, Onehotencoding
- gridsearchcv vs randomizedsearchcv

ML Pipeline

ML Model Deployment in Flask

PowerBI

Introduction

- Power BI for Data scientist
- Types of reports
- Data source types
- Installation

Basic Report Design

- Data sources and Visual types
- Canvas and fields
- Table and Tree map
- Format button and Data Labels
- Legend,Category and Grid
- CSV and PDF Exports

Visual Sync, Grouping

- Slicer visual
- Orientation,selection process
- Slicer:Number,Text,slicer list
- Bin count,Binning

Hierarchies, Filters

- Creating Hierarchies
- Drill Down options
- Expand and show
- Visual filter,Page filter,Report filter
- Drill Thru Reports

Power Query

- Power Query transformation
- Table and Column Transformations
- Text and time transformations
- Power query functions
- Merge and append transformations

DAX Functions

- DAX Architecture,Entity Sets
- DAX Data types,Syntax Rules
- DAX measures and calculations
- Creating measures
- Creating Columns

Deep Learning

Deep learning at Glance

- Introduction to Neural Network
- Biological and Artificial Neuron
- Introduction to perceptron
- Perceptron and its learning rule and drawbacks
- Multilayer Perceptron, loss function
- Neural Network Activation function

Training MLP: Backpropagation

Cost Function

Gradient Descent Backpropagation - Vanishing and Exploding Gradient Problem

Introduce to Py-torch

Regularization

Optmizers

Hyperparameters and tuning of the same

TENSORFLOW FRAMEWORK

- Introduction to TensorFlow
- TensorFlow Basic Syntax
- TensorFlow Graphs
- Variables and Placeholders
- TensorFlow Playground

ANN (Artificial Neural Network)

- ANN Architecture
- Forward & Backward Propagation, Epoch
- Introduction to TensorFlow, Keras
- Vanishing Gradient Descend
- Fine-tuning neural network hyperparameter
- Number of hidden layers, Number of neurons per hidden layer
- Activation function
- INSTALLATION OF YOLO V8, KERAS, THEANO

PY-TORCH Library

RNN (Recurrent Neural Network)

- Introduction to RNN
- Back Propagation through time
- Input and output sequences
- RNN vs ANN
- LSTM (Long Short-Term Memory)
- Different types of RNN: LSTM, GRU
- Biirectional RNN
- Sequential-to-sequential architecture (Encoder Decoder)
- BERT Transformers
- Text generation and classification using Deep Learning
- Generative-AI (Chat-GPT)

Basics of Image Processing

- Histogram of images
- Basic filters applied on the images

Convolutional Neural Networks (CNN)

- ImageNet Dataset
- Project: Image Classification
- Different types of CNN architectures
- Recurrent Neural Network (RNN)
- Using pre-trained model: Transfer Learning

Natural Language Processing (NLP)

Natural Language Processing (NLP)

- Text Cleaning
- Texts, Tokens
- Basic text classification based on Bag of Words

Document Vectorization

- Bag of Words
- TF-IDF Vectorizer
- n-gram: Unigram, Bigram
- Word vectorizer basics, One Hot Encoding
- Count Vectorizer
- Word cloud and gensim
- Word2Vec and Glove
- Text classification using Word2Vec and Glove
- Parts of Speech Tagging (PoS Tagging or POST)
- Topic Modelling using LDA
- Sentiment Analysis

Twitter Sentiment Analysis Using Textblob

- TextBlob
- Installing textblob library
- Simple TextBlob Sentiment Analysis Example
- Using NLTK's Twitter Corpus

Spacy Library

- Introduction, What is a Token, Tokenization
- Stop words in spacy library
- Stemming
- Lemmatization,
- Lemmatization through NLTK
- Lemmatization using spacy
- Word Frequency Analysis
- Counter
- Part of Speech, Part of Speech Tagging
- Pos by using spacy and nltk
- Dependency Parsing
- Named Entity Recognition(NER)
- NER with NLTK
- NER with spacy

Computer Vision

Human vision vs Computer vision

- CNN Architecture
- CONVOLUTION – MAX POOLING – FLATTEN LAYER – FULLY CONNECTED LAYER
- CNN Architecture
- Striding and padding
- Max pooling
- Data Augmentation
- Introduction to OpenCV & YoloV3 Algorithm

Image Processing with OpenCV

- Image basics with OpenCV
- Opening Image Files with OpenCV
- Drawing on Images, Image files with OpenCV
- Face Detection with OpenCV

Video Processing with OpenCV

- Introduction to Video Basics, Object Detection
- Object Detection with OpenCV

Reinforcement Learning

- Introduction to Reinforcement Learning
- Architecture of Reinforcement Learning
- •Reinforcement Learning with Open AI
- Policy Gradient Theory

OPEN AI

- Introduction to Open AI
- Generative AI
- Chat Gpt (3.5)
- LLM (Large Language Model)
- Classification Tasks with Generative AI
- Content Generation and Summarization with Generative AI
- Information Retrieval and Synthesis workflow with Gen AI

Time Series and Forecasting

- Time Series Forecasting using Deep Learning
- Seasonal-Trend decomposition using LOESS (STL) models.
- Bayesian time series analysis

MakerSuite Google

- PaLM API
- MUM models

Azure ML