BIGDATAPEDIA
Learning will lead the future...

Data Science with Python

Duration:

3-4 Months

Contact us

Call or WhatsApp @ +91 9715-010-010

Course Overview

Introduction

This course provides techniques for data cleaning, visualizing the data, predictive modeling, machine learning and deep learning by using python, TensorFlow and keras deployed using docker.

Expectations and Goals

Understanding is matter, so please do not skip any classes and complete the assignments on given time.

Course Materials

- Recorded Videos
- Class Notes
- Practice Material
- Assignments
- Private Group Discussion
- Online eLibraries
- Interview References and Tips
- Sample Resumes

Fee Details

- Please refer the below nomination form details
  https://forms.gle/Gx42goMcmDVjVVJN8
- Payment can fold by Two Installments (While joining and After One Month).

Note*

If you are not interested and informed within a month, we will refund your payment with 100% without any queries.

Course Content Details

Month	Topic	Content

Month 1	Python Programming	Python Basics
		• Installation & Environment setup
		• Basic commands in Python
		• Data Types & Operators
		• Data Structures in python
		  ○ List, Tuples, Dictionary, and sets
		• Conditional statements
		  ○ IF, IF-Else, nested IF, nested IF-Else conditional
		• Loops
		  ○ for loop and do while
		• User Defined Functions
		• Lambda, map, filter and reduce
		Python Libraries
		• Pandas
		• Numpy
		• MatplotLib
		• Sklearn
		• Seaborn

	Statistics using Python	Basic Statistics
		• Descriptive statistics
		• Inferential statistics
		• Measure of central tendency
		  ○ Mean, Median and Mode
		• Measure of Dispersion
		  ○ Range, Variance, standard deviation, and coefficient of variation
		• Frequency distribution
		• Introduction to Probability
		• Practice Session & Assignments
		Probability Distribution
		• Discrete distribution:
		• Bi-nominal distribution
		• Poisson distribution
		• Multinomial distribution Continuous distribution:
		• Normal distribution
		• Uniform normal distribution
		• T-student distribution
		• Exponential distribution
		• Chi- square distribution
		• F- distribution
		Sampling Techniques
		• Random sampling:

    - Sample with replacement
    - Sample without replacement
    - Training, testing and hold out dataset
- Stratified sampling
- Sequential or systematical sampling
- Clustering sampling techniques

Data Manipulation

- Data importing
- Working with datasets
- Manipulating the data sets
- Subset the data
- Sort the data
- Creating new variables
- Bin's creation
- Identifying & removing duplicates
- Exporting the datasets into external files
- Data Merging
- Pivot table analysis
- Data visualization through matplotlib, seaborn
- Histogram
- Bar Plot
- Pie Chart
- Scatter Matrix Pandas
- Scatter matrix Violin
- Plots
- Line Graphs
- Graphs

Hypothesis Testing

- Why hypothesis testing needed...?
- Null hypothesis testing
- Alternative hypothesis testing

Variable Reduction Techniques

- Correlation
- Pearson correlation
- Rank Correlation
- VIF/Multi collinearity
- PCA
- Chi-Square Technique
- Information value
- Cluster based method
- Tree based method
- Lasso regression method
- Stepwise regression method

Month	Topic	Content

Month 2		Data Pre-Processing (EDA)
		• Data sanity checks
		• Anomalies detection
		• Missing Value detections & treatments
		• Outliers' detection and outlier's treatment
		  ○ Boxplot
		  ○ QQ-plot
		  ○ IQR method
		• Variable transformation techniques
		• Exploratory Data Analysis
		• Uni-variate analysis
		• Bi-variate analysis
		• Multi-variate analysis
		• EDA Analysis

	PySpark	PySpark DataFrame Introduction DataFrame Transformation DataFrame Actions Spark-SQL Spark Mllib Implementation

	Machine Learning	Supervised Learning
		• Regression
		  ○ Linear Regression
		  ○ Multiple linear Regression
		  ○ Rigid Regression
		  ○ Lasso Regression
		  ○ Elastic Net Regression
		  ○ Polynomial Regression • Time series Analysis:
		  ○ Need of time series
		  ○ Moving average method
		  ○ Holt-winter method
		  ○ ARIMA method
		  ○ Model Evolution metrics
		  ○ Use case with Regression models-Project and Assignments
		• Classification
		  ○ Logistic Regression
		  ○ Decision Tree
		  ○ Decision Tree Regressors
		  ○ Decision Tree Classifier
		  ○ Naive Bayes
		  ○ KNN
		  ○ KNN-Regressors
		  ○ KNN-Classifiers- Binary labels and multi labels
		  ○ Support Vector Machines

Month	Topic	Content
		○ Support vectors-Regressors ○ Support vectors-Classifiers ○ Ensemble learning ○ Bagging ○ Boosting ○ Random Forest ○ Random Forest -Regressor ○ Random Forest-Classifier ○ Extra Tree Network ○ Model Elevation metrics Un Supervised Learning • Clustering Algorithm ○ Hierarchical Clustering ○ Agglomerative Clustering ○ Non-Hierarchical Clustering K-Means
Month 3		Model Selection and Cross Validation • How to validate a model? • What is a best model? • Types of data • Types of errors • The problem of over fitting • The problem of under fitting • Bias Variance Tradeoff • Cross Validation • Boot Strapping
	Deep Learning	TensorFlow • Introduction to Tensor flow • Constant • Place holders • Variables Multi layers Neural Networks • Neurons • Weights • Activations • Networks of Neurons • Training Networks • Back propagation • Gradient Descent Artificial Neural Network (ANN) • Neural Networks Introduction • Neural Network Intuition • Neural Network and vocabulary • Neural Network algorithm • Math behind Neural Network algorithm ○ Building the Neural Networks

Month	Topic	Content
		○ Validating the Neural network model ○ Neural Network applications Convolutional Neural Network (CNN) • Feature learning ○ Convolution ○ Pooling • Classification learning • Flatten • Fully Connected • SoftMax
Month 4	Docker	Docker Architecture Docker Containerization Play with Docker Images Docker Volume Management Docker Network management Create Customize ML Models Images
	Final Projects	

Please feel free to reach us if you have any queries

Call or WhatsApp@ +91 9715-010-010