



PGP- DATA SCIENCE Course Module

Foundations

Introduction to programming using Python (3 weeks)

Introduction

- Introduction to Python
- Basic Programming syntax
- Variables
- Basic Arithmetic & logical operators (int, float)
- Data Types

Conditional Statement in Python

- If
- If-else
- elif

Advance Python

- Functions
- Methods
- Map Function
- Reduce
- Filter
- Lambda

Data Structures

- List
- Tuple
- Dictionary
- Array
- List Comprehension

Iteration (loops)

- While Loop
- For Loop

Python as OOP Language

- OOPS Concept- Class, objects, Detailed Introduction
- Inheritance-Multi level Inheritance, Single level Inheritance
- Encapsulation
- Polymorphism
- Generators
- Iterators

Exploring Data Analysis (1 Week)

- Reading the Data
- Cleaning the Data
- Data Visualization in Python
- Summary statistics (mean, median, mode, variance, standard deviation)
- Seaborn
- Matplotlib
- Population VS sample
- Univariate and Multivariate statistics
- Types of variables – Categorical and Continuous
- Coefficient of correlations, Skewness and kurtosis



Statistical Methods for Decision Making (1 week)

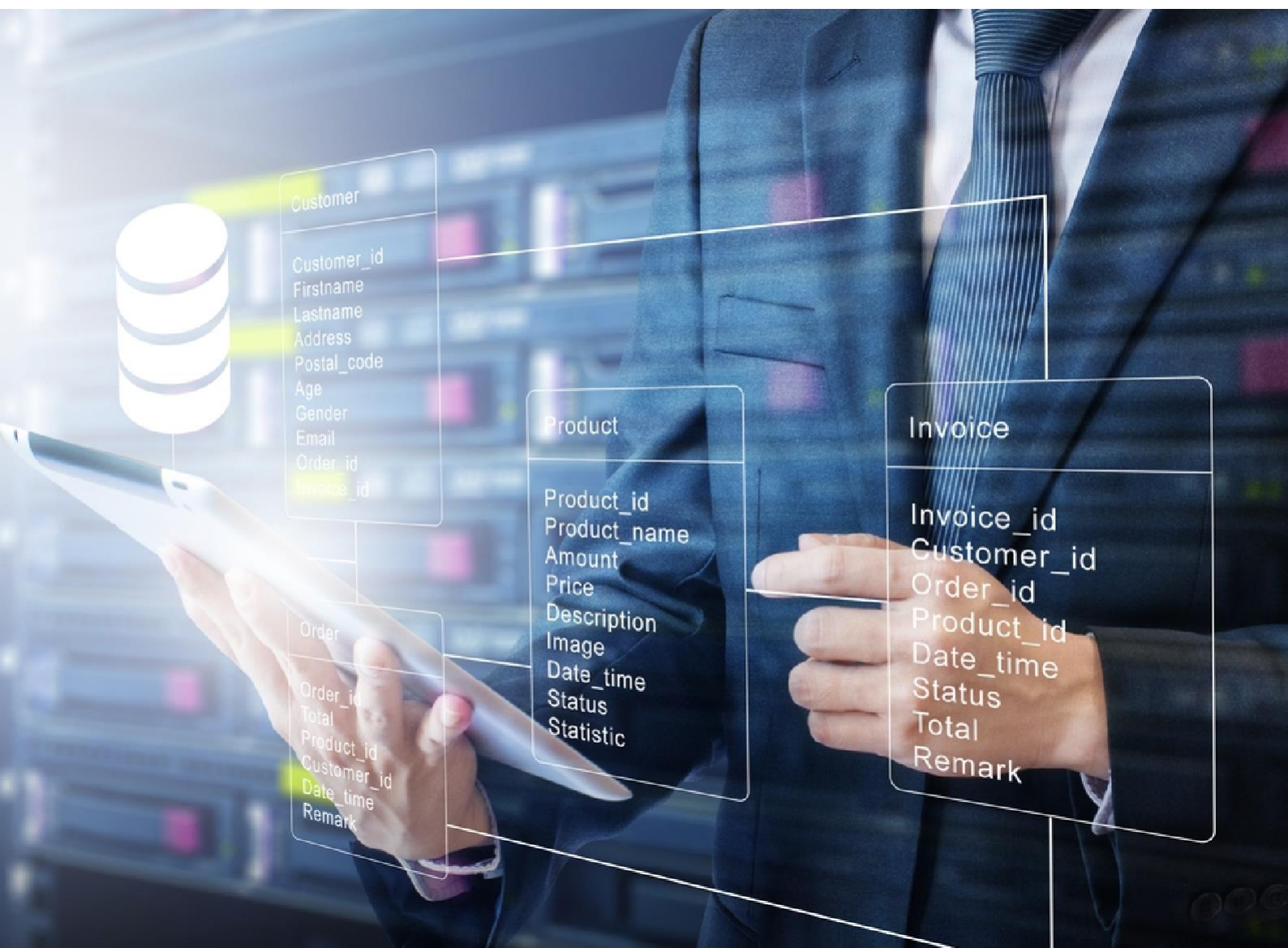
- Brief Introduction To Statistics
- Probability distribution
- Normal distribution
- Poisson's distribution
- Bayes' theorem
- Central limit theorem
- Hypothesis testing
- One Sample T-Test
- Two Sample T-Test
- Anova and Chi-Square
- Pearson Co-relation
- Co-Variance
- Chebyshiv-Inequality Formula

Introduction

- Introduction to DBMS
- An Introduction to Relational Database
- Concepts and SQL Accessing

Working on My SQL work bench

- Data Servers MYSQL/RDBMS Concepts
- Extraction, Transformation and Loading (“ETL”) Processes
- Retrieve data from Single Tables-(use of SELECT Statement) and the power of WHERE and ORDER by Clause. Retrieve and Transform data from multiple Tables using JOINS and Unions
- Introduction to Views Working with Aggregate functions, grouping and summarizing Records Writing Sub queries



Machine Learning Techniques

(Supervised Learning+ Unsupervised Learning)- (6 Weeks)

Regression

Introduction

Introduction to Regression

Looking at regression through the perspective of machine learning

Brief Introduction to Regression Techniques

Brief Introduction to Best Fit line in Regression

Linear Regression

Introduction to Linear Regression

Accuracy scores as a metric of model performance

Measuring the importance of individual variables in a regression model

Review - testing for individual significance vs joint significance

Using the adjusted R^2 to compare model with different number of independent variables

Approaches to feature selection

Forward and backward selection

Parameter tuning and Model evaluation

Extending linear regression

Data transformations and normalization

L1 & L2(LASSO AND RIDGE)

Logistic

Regression

- Introduction to Logistic Regression
- Log transformation of dependent and independent variables
- Dealing with categorical independent variables
- One hot encoding vs dummy variable
- Modelling probabilistic dependent variables
- The sigmoid function and odds ratio
- The concept of logit
- The failure of OLS in estimating parameters for a logistic regression
- Introduction to the concept of Maximum likelihood estimation
- Advantages of the maximum likelihood approach sigmoid function
- Modelling a logistic regression problem with a case study
- Making predictions and evaluating parameters

Case Study

Case study on Linear Regression

Case study on Logistic Regression

Featurization

Featurization, Model Selection & Tuning

- Feature engineering
- Model selection and tuning
- Model performance measures
- Regularising Linear models
- ML pipeline
- Bootstrap sampling
- Grid search CV
- Randomized search CV
- K fold cross-validation



Classification

Introduction

Introduction to Classification

Looking at Classification through the perspective of machine learning

Brief Introduction to Classification Techniques

Balancing Data set

Binary classification vs Multi class classification

Classification Techniques

CART - Extending decision trees to regressing problems.

Advantages of using CART.

The Bayes theorem. Prior probability.

KNN CLASSIFIER

The Gaussian NAÏVE'S BAYES Classifier.

Assumptions of the Naive Bayes Classifier.

Functioning of the Naïve's Bayes algorithm.

Evaluating the model - Precision, Recall, Accuracy metrics and k-fold cross validation

Random Forest

Voting Classifier

ROC Curve and AUC for binary classification for Naive Bayes.

Extending Bayesian Classification for Multiclass Classification

Support Vector Machine

KNN

Decision Trees

- Entropy and Ginny
- Information Gain
- Decision trees – Simple decision trees. Visualizing decision trees and nodes and splits.
- Working of the Decision tree algorithm.
- Importance and usage of Entropy and Gini index.
- Manually calculating entropy using gini formula and working out how to split decision nodes
- Evaluating decision tree models.
- Accuracy metrics – precision, recall and confusion matrix
- Interpretation for accuracy metric.
- Building a a robust decision tree model. k-fold cross validation - Advantages against simple train test split.

Case Study

Case study on Classification Data Set

Ensemble Techniques (1 week)

- Bagging
- Boosting(GRADIENT BOOSTING,ADA BOOSTING,XG BOOSTING)

Unsupervised Learning

Introduction

What is Unsupervised Learning?

The two major Unsupervised Learning problems - Dimensionality reduction and clustering.

Clustering Algorithms

The different approaches to clustering – Heirarchical and K means clustering.

Heirarchical clustering - The concept of agglomerative and divisive clustering.

Agglomerative Clustering – Working of the basic algorithms.

Distance matrix - Interpreting dendograms.

Choosing the threshold to determine the optimum number of clusters.

Case Study

The relationship between unsupervised and supervised learning.

Case study on Dimensionality reduction followed by a supervised learning model.

Case study on Clustering followed by classification model.

K-Means Algorithm

The K-means algorithm.

Measures of distance – Euclidean, Manhattan and Minowski distance.

The concept of within cluster sums of squares.

Using the elbow plot to select optimum number of cluster's.

Case study on k-means clustering.

Comparison of k means and agglomerative approaches to clustering.

PCA (Principal Component Analysis)

- Noise in the data and dimensional reduction.
- Capturing Variance - The concept of a principal components.
- Assumptions in using PCA.
- The working of the PCA algorithm.
- Eigen vectors and orthogonality of principal components.
- What is complexity curve?
- Advantages of using PCA.
- Bulid a model using Principal components and comparing with normal model. What is the difference?
- Putting it all together.

Data Visualization Using Tableau (2 week)

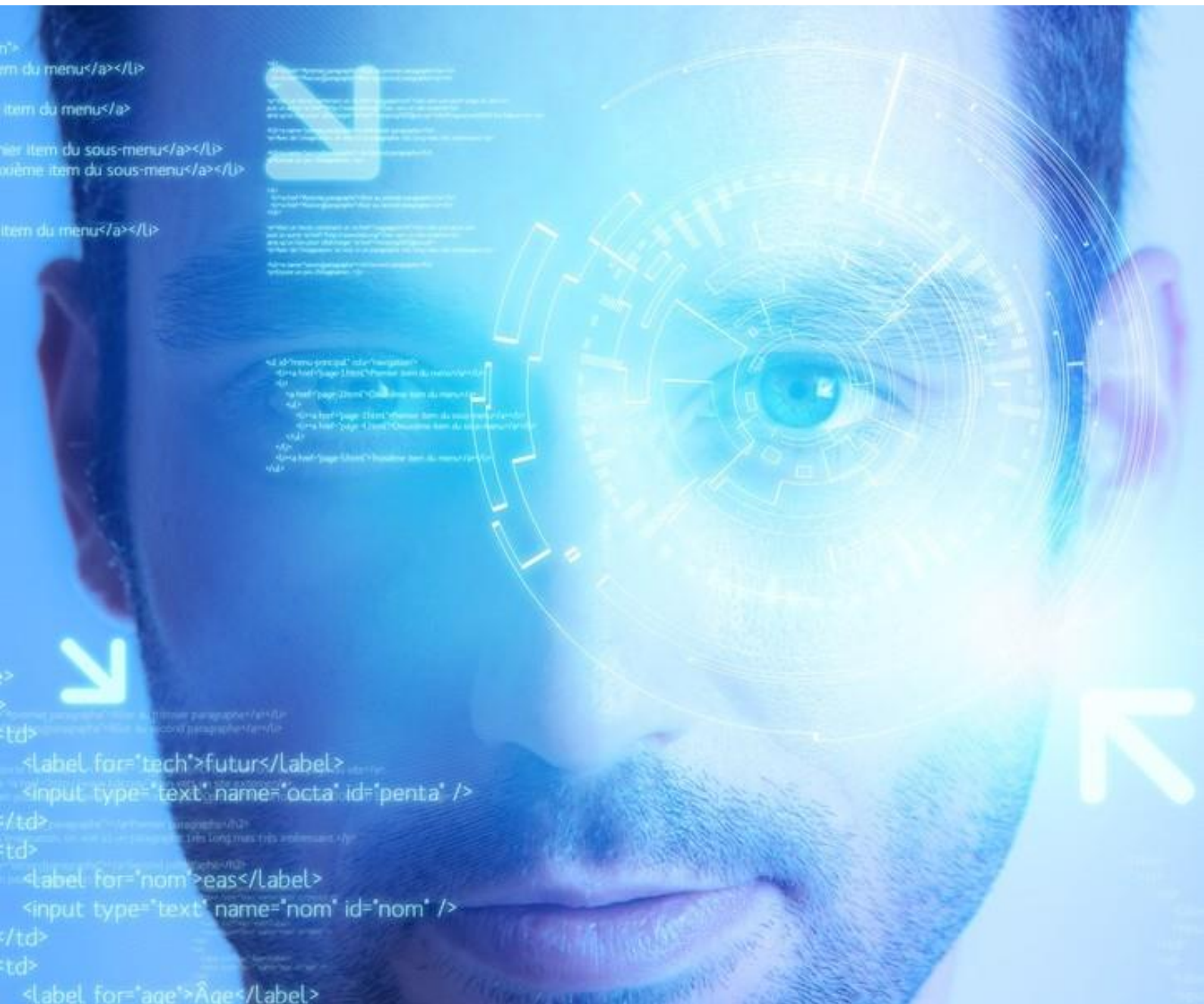
Introduction

- Introduction to Visualization, Rules of Visualization
- How to Import The Data In Tableau?
- Blending the Data

Visualization in Tableau

- Data Types, Sources, Connections, Loading, Reshaping
- Data Aggregation
- Working with Continuous and Discrete Data
- Using Filters
- Using Calculated Fields and parameters
- Creating Tables and Charts
- Building Dash Boards and story boards
- Sharing Your Work and Publishing for wider audience





Internship Module

Module 1: Natural Language Processing and Speech Recognition

Introduction to Natural Language Processing

Lesson 1 - Feature Engineering on Text Data Lesson

Lesson 2 - Natural Language Understanding Techniques

Lesson 3 - Natural Language Generation

Lesson 4 - Natural Language Processing Libraries

Lesson 5 - Natural Language Processing with Machine Learning and Deep Learning

Introduction of Speech Recognition

Lesson 1 - Signal Processing and Speech Recognition Models

Lesson 2 - Speech to Text

Lesson 3 - Text to Speech

Lesson 4 - Voice Assistant Devices



Module 2 : Text Mining And Sentimental Analysis

Lesson 1 - Text cleaning, regular expressions, Stemming, Lemmatization

Lesson 2 - Word cloud, Principal Component Analysis, Bigrams & Trigrams

Lesson 3 - Web scrapping, Text summarization, Lex Rank algorithm

Lesson 4 - Latent Dirichlet Allocation (LDA) Technique

Lesson 5 - Word2vec Architecture (Skip Grams vs CBOW)

Lesson 6 - ext classification, Document vectors, Text classification using Doc2vec

Module 3: Reinforcement Learning

Lesson 1 - Introduction to Reinforcement Learning

Lesson 2 - Reinforcement Learning Framework and Elements

Lesson 3 - Multi-Arm Bandit

Lesson 4 - Markov Decision Process

Lesson 5 - Solution Methods

Lesson 6 - Q-value and Advantage Based Algorithms



Module 4: Time Series Forecasting

Lesson 1 - What is Time Series?

Lesson 2 - Regression vs Time Series

Lesson 3 - Examples of Time Series data

Lesson 4 - Trend, Seasonality, Noise and Stationarity

Lesson 5 - Time Series Operations

Lesson 6 - Detrending

Lesson 7 - Successive Differences

Lesson 8 - Moving Average and Smoothing

Lesson 9 - Exponentially weighted forecasting model

Lesson 10 - Lagging

Lesson 11 - Correlation and Auto-correlation

Lesson 12 - Holt Winters Methods

Lesson 13 - Single Exponential smoothing

Lesson 14 - Holt's linear trend method

Lesson 15 - Holt's Winter seasonal method

Lesson 16 - ARIMA and SARIMA



Module 5

Introduction to AI And Deep Learning

- What is Deep Learning ?
- Universal Function Approximator
- Neural Network
- Difference between ML and DL
- Major Types of NN-ANN,CNN and RNN
- Perceptron In Neural Network
- Solving the INPUT
- Bias
- Epoch
- Batch size
- Front Propagation
- Back Propagation
- Weight Updation in Back Propagation Formula
- **Introduction Activation Function**
 - Types of Activation Function
 - Vanishing Gradient Descent
 - Exploding Gradient Descent
 - Under Fitting and Over Fitting
- **Introduction of Regularization Technique**
 - Regularization- L1 and L2 (Lasso and Rigid)
 - Regularization - Drop Out
- **Introduction of Weight Initialization Technique.**
 - Different Weight Initialization Technique
- **Introduction of Gradient Descent**
 - TYPES OF GRADIENT DESCENT
 - Convex and Non Convex Function
 - Saddle Point

- **Artificial Neural Network Lesson**
- **Deep Neural Network and Tools Lesson**
- **Deep Neural Net Optimization, Tuning, and Interpretability**
- **Convolutional Neural Net(CNN)**
- **Recurrent Neural Networks**

Module 6: Advanced Deep Learning and Computer Vision

Introduction to Computer Vision

Lesson 1 - Prerequisites for the course

Lesson 2 - RBM and DBNs

Introduction to Image processing in CNN

- Introduction to Convolutional Neural Networks
- Famous CNN architectures
- Transfer Learning
- Classification model in image processing
- Semantic segmentation
- Instance Segmentation
- Other variants of convolution
- Metric Learning

Introduction to sequence Data Processing in RNN

- Introduction to RNN
- Advantages of RNN
- Disadvantages of RNN

Introduction of LSTM and GRU

- Long-short term memory (LSTM)
- Model of LSTM
- Advantages of LSTM
- LSTM v/s GRU (Gate Recurrent Unit)
- Input gate, Output gate, forget gate
- Bidirectional Recurrent Neural Networks (BRNN)



(USA)

2-Industrial Park Drive, E-Waldorf, MD, 20602,
United States



(INDIA)

B-44, Sector-59, Noida Uttar Pradesh 201301

(USA)

+1-844-889-4054



(INDIA)

+91-92-5000-4000



info@careerera.com



www.careerera.com

CONTACT US