



PGP-

DATA SCIENCE

Course Module

Foundations

Introduction to Programming using Python

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

Introduction to Programming using R

- Introduction to R Language
- How to install R
- Documentation in R
- Hello world
- Package in R
- Data Types in R
- Data structures
- Conditional statement in R
- Loops in R
- Subsetting
- Reading Data from csv,excel files
- Creating a vector and vector operation
- Initializing data frame
- Control structure
- Data Visualization in R
- Creating Bar Chart
- Creating Histogram and box plot
- Plotting with base graphics
- Plotting and coloring in R
- Machine Learning Algorithms Using R

Database Management System using My SQL Workbench

Introduction

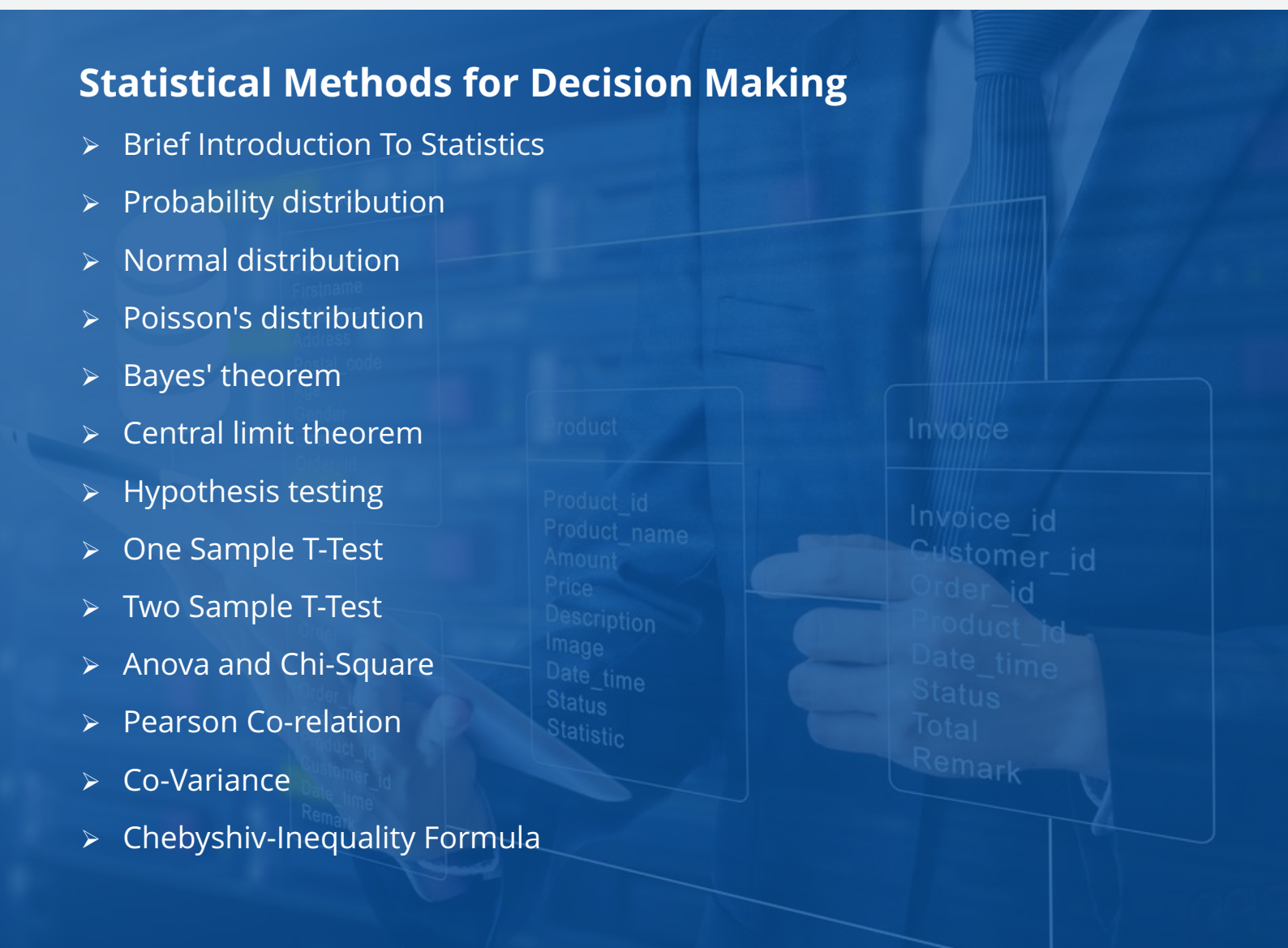
- Introduction to DBMS
- An Introduction to Relational Database
- Concepts and SQL Accessing

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

Statistical Methods for Decision Making

- 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



The background of the lower half of the page features a person in a dark suit and tie, holding a tablet. The tablet displays a list of database tables. The 'Product' table is highlighted with a white border. The 'Invoice' table is also visible, partially obscured by the 'Product' table's highlight.

Product
Product_id
Product_name
Amount
Price
Description
Image
Date_time
Status
Statistic

Invoice
Invoice_id
Customer_id
Order_id
Product_id
Date_time
Status
Total
Remark



Exploring Data Analysis

- 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

Machine Learning Techniques

Supervised Learning - Regression

- Looking at regression through the perspective of machine learning
- 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
- Log transformation of dependent and independent variables
- Case study: -
- Dealing with categorical independent variables
- One hot encoding vs dummy variable regression
- Case study on linear regression
- 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
- Modelling a logistic regression problem with a case study
- Making predictions and evaluating parameters

Ensemble Techniques

- Bagging
- Boosting
- Bagging & Boosting Examples

Machine Learning Model Deployment using Flask

- Introduction to Model Deployment
- Introduction to Flask in Python
- How to deploy Applications in Flask?
- Types of Model deployment

Machine Learning Techniques

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

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

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)

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

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

- Introduction to Visualization, Rules of Visualization
- 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

- Building Dash Boards and story boards
- Building Dash Boards and Story Boards in Data Studio

Data Visualization Using Power Bi

- Introduction to Microsoft Power BI
- The key features of Power BI workflow
- Desktop application
- BI service
- File data sources
- Sourcing data from the web (OData and Azure)
- Building a dashboard
- Data visualization
- Publishing to the cloud
- DAX data computation
- Row context
- Filter context
- Analytics pane
- Creating columns and measures
- Data drill down and drill up
- Creating tables
- Binned tables
- Data modeling and relationships
- Power BI components such as Power View, Map, Query, and Pivot

Data Visualization Using Google Data Studio

- Introduction to Visualization
- Introduction to Google Data Studio
- How Data Studio Works?
- Data Types, Sources, Connections, Loading, Reshaping
- Data Aggregation
- Working with Continuous and Discrete Data
- Report Edit Mode in Data Studio.
- Using Filters in Data Studio
- Using Calculated Fields and parameters
- Creating Tables and Charts



Internship Module

Module 1: Natural Language Processing and Speech Recognition

Lesson 1 - Introduction to Natural Language Processing

Lesson 2 - Feature Engineering on Text Data Lesson

Lesson 3 - Natural Language Understanding Techniques

Lesson 4 - Natural Language Generation

Lesson 5 - Natural Language Processing Libraries

Lesson 6 - Natural Language Processing with Machine Learning and Deep Learning

Lesson 7 - Introduction of Speech Recognition

Lesson 8 - Signal Processing and Speech Recognition Models

Lesson 9 - Speech to Text

Lesson 10 - Text to Speech

Lesson 11 - 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 - Text 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

Lesson 1 - Introduction To AI And Deep Learning

Lesson 2 - Artificial Neural Network Lesson

Lesson 3 - Deep Neural Network and Tools Lesson

Lesson 4 - Deep Neural Net Optimization, Tuning, and Interpretability

Lesson 5 - Convolutional Neural Net(CNN)

Lesson 6 - Recurrent Neural Networks

Lesson 7 – Autoencoders



Module 6: Advanced Deep Learning and Computer Vision

Lesson 1 - Course Introduction

Lesson 2 - Prerequisites for the course

Lesson 3 - RBM and DBNs

Lesson 4 - Variational AutoEncoder

Lesson 5 - Working with Deep Generative Models

Lesson 6 - Applications: Neural Style Transfer and Object Detection

Lesson 7 - Distributed & Parallel Computing for Deep Learning Models

Lesson 8 - Reinforcement Learning

Lesson 9 - Deploying Deep Learning Models and Beyond

Lesson 10 - Introduction to Image data

Lesson 11 - Introduction to Convolutional Neural Networks

Lesson 12 - Famous CNN architectures

Lesson 13 - Transfer Learning

Lesson 14 - Object detection

Lesson 15 - Semantic segmentation

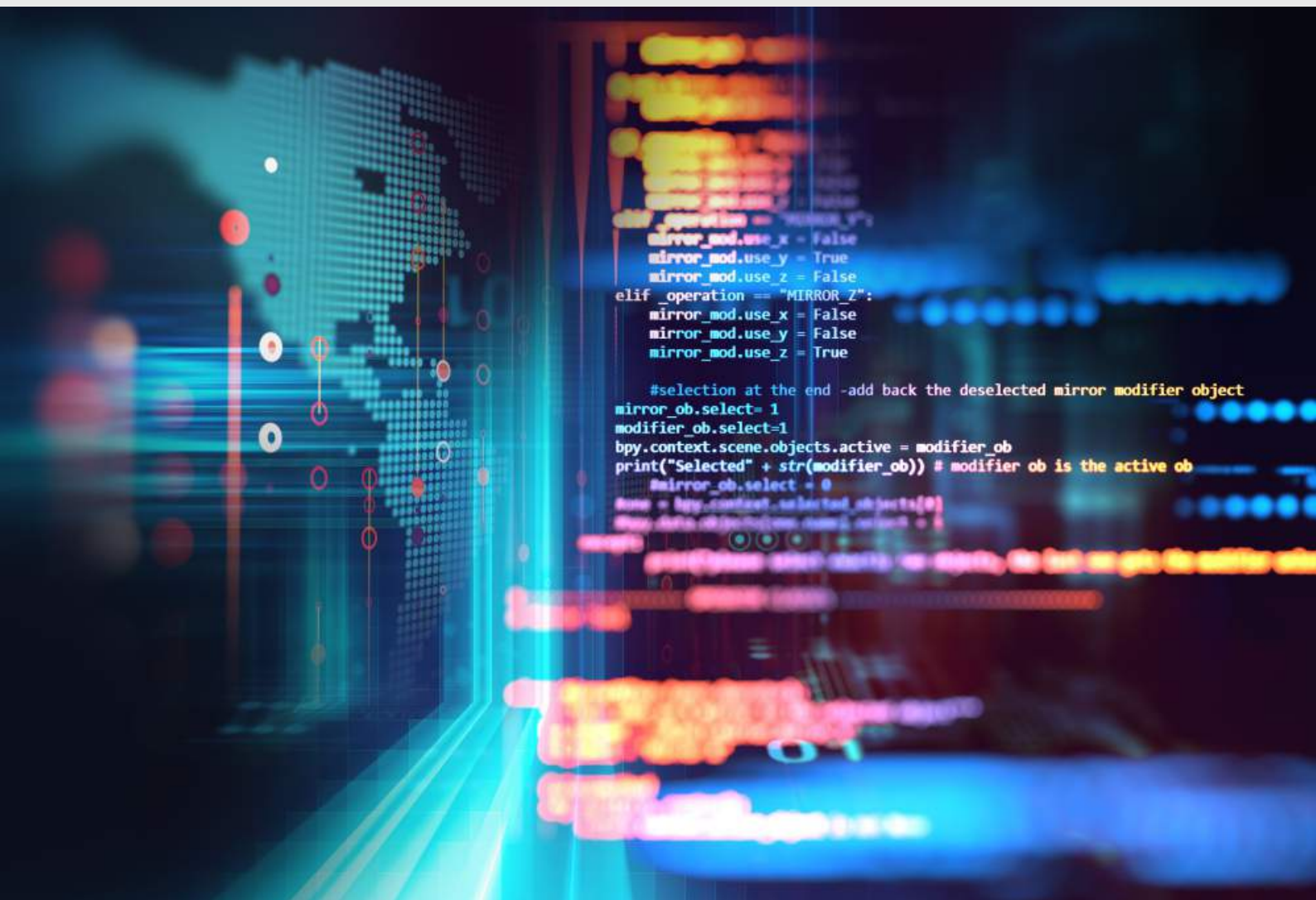
Lesson 16 - Instance Segmentation

Lesson 17 - Other variants of convolution

Lesson 18 - Metric Learning

Lesson 19 - Siamese Networks

Lesson 20 - Triplet Loss





(USA)

+1-844-889-4054



(Singapore)

+65-3159-1284



(USA)

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



(SINGAPORE)

Centennial Tower, Levels 21 & 34, 3 Temasek
Avenue, Singapore, 039190



info@careerera.com



www.careerera.com

CONTACT US