

Data Science Course Content

CHAPTER 1: INTRODUCTION TO DATA SCIENCE

- Market trend of Data Science
- Opportunities for Data Science
- What is the need for Data Scientists
- What is Data Science
- Data Science Venn Diagram
- Data Science Use cases
- Knowing the roles of a Data Science practitioner
- Data Science – Skills set
- Understanding the concepts & definitions of:
 - Artificial Intelligence
 - Machine Learning – Deep Learning
 - NLP
 - Computer Vision

CHAPTER 2: DATA AND TOOLS

- What is Business Intelligence?
- What is ETL?
- Layers of a Data Warehouse
- OLAP VS OLTP
- Facts and Dimensions
- Big Data tools and it's uses
- Big Data Stack
- Understanding Structured text Data
- Understanding Unstructured text Data

CHAPTER 3: DATA SCIENCE DEEP DIVE

- Understanding Descriptive vs Predictive vs Prescriptive Analytics
- Difference between Analytics vs. Analysis
- Data Science Project Lifecycle
- Technology Stack Involved in the Lifecycle
 - Machine Learning tools
 - Development tools
 - Languages
 - Data Platforms
- CRISP – Cross-industry standard process for Data Mining
- 5WIH – The questions that kick start a ML project

- 80-20 Rule of Data Analytics
- Supervised Vs Unsupervised Learning
- Data Science – Use case bubble
- Data Mining

CHAPTER 4: DATA

- Data Wrangling or Data Munging
- Data Categorization basics
- Different Types of Data
- Types of Data Collection
- Data Sources
- Data Collection plan
- Data Quality Issues
- Types of Data Error
- Ration Scale Vs Interval Scale
- Predictors/Features vs Predictions/Labels
- Understanding Imbalance in Data

CHAPTER 5: STATISTICS & PROBABILITY

- What is Statistics
- Sample Vs Population
- Measure of central vs Dispersion
- Frequency Distribution
- Cumulative Frequency Distribution
- Mean, Median, Mode
- Quartiles/Percentile
- Range, Variance, Standard Deviation, Co-efficient of Variation
- 68-95-99 Rule of SD
- Z Score (Standard Score)
- P-Value
- Maximum Likelihood Estimation
- Probability vs Likelihood
- PDF vs PMF
- Normal Distribution of Data
- Skewness & it's types
- Kurtosis & it's types
- Kth Central Moments
- Co-Variance/Joint Probability Distribution
- Correlation
- Entropy
- ANOVA
- Chi-Square
- F tests
- Types of Data Distribution



Real-time Practicals:

1. Hands on- Lab using pen and paper Only

CHAPTER 6: SETUP

- Anaconda & Python
- Understanding Jupiter Notebooks
- Python Package Installation
- Tableau Installation
- Oracle Database & Server

CHAPTER 7: DATA SOURCING, EXPLORATORY DATA ANALYSIS & READINESS

- Concept of List, Data frame, Dictionary
- Connecting to Databases using Python
- Importing data from csv, text, Excel
- Converting JSON, XML, to Data frame
- Understanding EDA
- Frequency Distribution
- Analyzing NA, blanks
- Using SQL concepts inside Python



Real-time Practicals:

1. Hands on- Lab using python.

CHAPTER 8: DATA TRANSFORMATION/WRANGLING

- Handling missing Values
- Handling Outliers
- Normalization techniques
- Standardization techniques
- Regularization techniques
- Feature Extraction
- Train Test data selection



Real-time Practicals:

1. Hands on- Lab using python.

CHAPTER 9: DATA SCIENCE CONCEPTS

- No Free Lunch
- Hypothesis vs Null Hypothesis
- BIAS VS Variance tradeoff
- Local Vs Global Minima/Maxima
- Bias – Loss/ Loss-Cost Function

CHAPTER 10: LINEAR REGRESSION

- Understanding Regression math
- Linear Algebra concepts
- Least Mean Square
- Analyzing Co-relation
- Heat Maps, Pair Plots, Distribution Graphs
- Simple Vs Multiple Linear regression
- Train Test data selection



Real-time Practicals:

1. Hands on- Lab & Model Implementation using python.

CHAPTER 11: POLYNOMIAL REGRESSION

- Understanding the math
- Polynomial Algebra concepts
- Degree of Polynomial



Real-time Practicals:

1. Hands on- Lab & Model Implementation using python.

CHAPTER 12: CLASSIFICATION

- Overfitting/ Under fitting/ Optimal Fits
- Handling Categorical Data inside
- Confusion Matrix
- Type I & Type II errors
- Precision Vs Accuracy
- AUC/ROC curve

CHAPTER 13: LOGISTIC REGRESSION

- Understanding the statistics behind Logistic Sigmoid
- Logistic regression math



Real-time Practicals:

1. Hands on- Lab & Model Implementation using python.

CHAPTER 14: RANDOM FOREST

- Understanding the Decision Tree & Bagging
- Math behind Classification and Regression in tree
- Decision Tree concepts
- Using Random Forest for Regression
- K fold Cross Validation
- Model Optimizers
- Hyper parameter Tuning
- Building a Decision Trees Model in R

CHAPTER 15: NAÏVE BAYES THEOREM

- Understanding the Naïve Bayes theorem
- Bayesian Vs Gaussian theorems
- Using naïve Bayes for Regression
- Model Optimizers
- Hyper parameter Tuning



Real-time Practicals:

1. Hands on- Lab & Model Implementation using python.

CHAPTER 16: NLP FOR MACHINE LEARNING FEATURING

- Label Encoding
- One hot encoding
- Synonym treatment
- Stemming
- Lemmatization
- Stop words
- Parts Of Speech Tagging
- TF-IDF and its math Behind



Real-time Practicals:

1. Hands on- Lab & Model Implementation using python.

CHAPTER 17: SUPPORT VECTOR MACHINE

- Understanding the SVM Concept
- Hyper plane and Kernel
- Using SVM for Regression
- Grid Search
- Model Optimizers
- Hyper parameter Tuning



Real-time Practicals:

1. Hands on- Lab & Model Implementation using python.

CHAPTER 18: GRADIENT BOOSTING MACHINE & XGBOOST

- Understanding the Boosting Concept
- Hyper plane and Kernel
- Learning Rate
- Model Optimizers
- Hyper parameter Tuning



Real-time Practicals:

1. Hands on- Lab & Model Implementation using python.

CHAPTER 19: K MEANS CLUSTERING ALGORITHM

- Understanding Nearest Neighbors concept
- Statistics behind K Means Clustering Algorithm



Real-time Practicals:

1. Hands on- Lab & Model Implementation using python.

CHAPTER 20: KERAS TENSOR FLOW – MLP DEEP LEARNING (NEURAL NETWORKS)

- Understanding Deep learning
- MLP Vs other Deep Learning
- How Neural Network works & Architecture
- Activation functions.
- Model Optimizers
- Hyper parameter Tuning
- Best Practice and when to use DL



Real-time Practicals:

1. Hands on- Lab & Model Implementation using python.

CHAPTER 21: H2O.AI

- Introduction to H2O.ai
- Pros and Cons
- Available models in H2O.ai



Real-time Practicals:

1. Hands on- Lab & Model Implementation using python.

CHAPTER 22: SAMPLING & DIMENSION REDUCTION (DR)

- Introduction to Sampling
- Over sampling and Under sampling
- SMOTE/SMOTENC & Near Miss
- Pros and Cons of sampling
- Introduction to DR
- PCA & it's code

CHAPTER 23: DEPLOYMENT OF MODEL TO PRODUCTION

- Introduction to Pyinstaller
- Pickle and Joblib



Real-time Practicals:

2. Hands on- Lab & Model deployment using python.

CHAPTER 24: TABLEAU BASICS

- Introduction to Tableau
- Data sources
- Exploratory Data Analysis
- Clustering Analysis and Inferences using Tableau
- Creating visualizations



Real-time Practicals:

3. Hands on- Lab using Tableau.

Contact Info:



+919884412301 | +919884312236



[Know more about Data Science](#)



info@credosystemz.com



New # 30, Old # 16A, Third Main Road,
Rajalakshmi Nagar, Velachery, Chennai (Opp. to
MuruganKalyanaMandapam)

[BOOK A FREE DEMO](#)