

Data Science Course Content

CHAPTER 1: INTRODUCTION TO DATA SCIENCE

- What is the need for Data Scientists
- Data Science Foundation
- Business Intelligence
- Data Analysis
- Data Mining
- Machine Learning
- Difference between Analytics vs. Data Science
- Analytics and its types
- Lifecycle Probability
- More about Analytics Project Lifecycle

CHAPTER 2: DATA

- Data Categorization basics
- Different Types of Data
- Types of Data Collection
- Forms of Data and Sources
- Data Quality and Changes
- Data Quality Issues
- Data Quality Story
- Data Architecture and its Components
- OLTP vs. OLAP
- How Data is being stored

CHAPTER 3: DATA SCIENCE DEEP DIVE

- What is Data Science
- Demand for Data Scientists
- What is Data Product
- Need for Data Science
- Large Scale Analysis Cost vs Storage
- Data Science – Skills
- Data Science – Use Cases

- Data Science Project Life Cycle & Stages
- What is Data Acquisition
- Sourcing the data
- Techniques
- Evaluating the input data
- Data formats
- Data Quantity
- Data Quality
- Resolution Techniques
- Data Transformation
- File format Conversions
- Anonymization

CHAPTER 4: INTRO TO R PROGRAMMING

- R Programming introduction
- What is Business Analytics
- Concept of Analytics
- Importance of R in analytics
- R Language community and eco-system
- R and its Usage
- How to install R and other packages
- How to Perform basic R Operations in command line
- Usage of IDE R Studio and various GUI

CHAPTER 5: R PROGRAMMING CONCEPTS

- R – Data types
- R – Built-in functions
- Sub setting methods
- Vectors
- Lists
- Matrices
- Dataframes
- Condition statements in R
- Function building in R

CHAPTER 6: MACHINE LEARNING

- Math behind machine learning
- Data Structures and it's classification
- Data Structures – Correlation and regression
- Data Structures – Prediction and Forecasting
- Case Study – Data Science trends

CHAPTER 7: PYTHON PROGRAMMING

- Introduction & Set up
- Data types and its uses
- Operators and statements
- Functions and methods
- Libraries and Packages – Data Science relevant

CHAPTER 8: DATA VISUALIZATION TECHNIQUES-R

- Building plots using ggplot2
- Types of plots



Real-time Practicals:
Replicate and build a plot from a news website

CHAPTER 9: DATA VISUALIZATION TECHNIQUES - PYTHON

- Building plots using seaborn library
- Types of plots in python
- Geo plots



Real-time Practicals
Build a geoplot to show data across the globe

CHAPTER 10: LINEAR REGRESSION

- Linear Regression Model
- Statistics behind Linear Regression
- Building a Linear Regression Model in Python
- Building a Linear Regression Model in R



Real-time Practicals:
Forecast the final grade results of students using the input variables in the academic data set provided.

CHAPTER 11: LOGISTIC REGRESSION

- Logistic Regression Model
- Statistics behind Logistic Regression
- Building a Logistic Regression Model in Python
- Building a Logistic Regression Model in R



Real-time Practicals:

Predict the survival of passengers travelling onboard using the Titanic dataset.

CHAPTER 12: K MEANS CLUSTERING ALGORITHM

- K Means Clustering Model
- Statistics behind K Means Clustering Algorithm
- Building a K Means Clustering Model in Python
- Building a K Means Clustering Model in R



Real-time Practicals 1:

Form clusters out of the wine chemical parameter data set given.

Real-time Practicals 2:

Cluster the Colleges dataset into two groups based on various data points provided about each college.

CHAPTER 13: K NEAREST NEIGHBORS ALGORITHM

- K Nearest Neighbors Model
- Statistics behind k Nearest Neighbors Algorithm
- Building a K Nearest Neighbors Model in Python
- Building a K Nearest Neighbors Model in R



Real-time Practicals:

Insurance domain – Based on the applicant data and their policy purchase status, build a KNN model to predict whether a new applicant will purchase the insurance policy or not.

CHAPTER 14: DECISION TREES ALGORITHM

- Decision Trees Model
- Statistics behind Decision Trees Algorithm
- Building a Decision Trees Model in Python

- Building a Decision Trees Model in R



Real-time Practicals:

Given Dataset contains various data points about the universities across the country and their category. Build a Decision Tree classifier using the dataset to classify a new university into the correct category.

CHAPTER 15: RANDOM FOREST ALGORITHM

- Random Forest Model
- Statistics behind Random Forest Algorithm
- Building a Random Forest Model in Python
- Building a Random Forest Model in R



Real-time Practicals:

A healthcare data set containing details about the occurrence of kyphosis in patients, the surgery age and whether kyphosis recurred for that patient – is provided. Build a Random Forest Model to predict whether other patients treated previously would be subjected to kyphosis again.

CHAPTER 16: NLP – TEXT ANALYTICS – OCR & TWITTER ANALYTICS

- What is Text Analytics & Uses
- Stemming & Lemmatizing
- Stop words & Parts Of Speech Tagging
- Social Media Analytics – Twitter Analytics
- What is OCR?
- Simple Image to Text Conversion
- Enhanced Image to Text Conversion
- Batch text extraction from PDF files using R



Real-time Practicals:

Pull Tweets from Twitter in R and build a wordcloud visualization.

CHAPTER 17: NLP – NAÏVE BAYES CLASSIFIER MODEL

- What is Naïve Bayes Classifier Model
- Statistics behind Naïve Bayes Classifier Model



Real-time Practicals 1:

Build a Spam detector Model using Naive Bayes Algorithm, which will detect the Spam messages from the organic ones.

Real-time Practicals 2:

Build a Naive Bayes Classification Model to assign star ratings to the text reviews submitted by users.

Contact Info:



+91 9884412301 | +91 9884312236



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

CREDO SYSTEMZ