

Enterprise Big Data Analyst (EBDA) Training

Course Outline

Module 1: Introduction to Big Data Analysis

- What is Enterprise Big Data Analysis?
- The Objective of Enterprise Big Data Analysis
- The Data Analyst versus the Data Scientist
- The Big Data Analysis Toolbox
- Models, Algorithms and Intellectual Property

Module 2: The Data Analysis Process

- The Business objective
- Introduction
- Types of Business Objectives

Module 3: Data Ingestion – Importing and Reading Data

- Introduction
- Raw versus Processed Data
- Reading Local Data Sets
- Reading Online Data Set
- Reading Data Sets from Databases

Module 4: Data Preparation – Cleaning and Wrangling Data

- Tidy Data
- Data Inspection – Review your Data
- Data Cleaning
- Data Wrangling
- Data and R Files for this Chapter

Module 5: Data Analysis – Model Building

- Introduction to Data Analysis

- Exploratory Data Analysis
- Statistical Inference
- Correlation
- Regression

Module 6: Classification Techniques

- K-Nearest Neighbour (K-NN algorithm)
- Dimensions in the k-NN classifiers
- Naïve Bayes
- Naïve Bayes Classifier with multiple variables
- Laplace Smoothing
- Logistic Regression
- Classification Trees
- Building a Classification tree
- Model Overfitting and Accuracy

Module 7: Clustering Techniques

- Hierarchical Clustering
- Variations in hierarchical clustering
- Jaccard index
- K-Means Clustering

Module 8: Outlier Detection

- Grubbs Outlier detection
- K-NN Outlier Detection

Module 9: Data Presentation

- Introduction to Reproducible Research
- Codebooks
- Data Visualisation