

CERTIFICATION IN INTRODUCTION TO DATA SCIENCE



Course Code : OCIT0010

Data science is an interdisciplinary field that uses scientific methods, processes, algorithms and systems to extract knowledge and insights from data in various forms, both structured and unstructured, similar to data mining. Data Scientists are better statistician than computer programmer and better computer programmer than the Statistician. Students who come out of school education can shape their career by intersecting in this specialization to get the huge job opportunities in the data science field. After globalization, career evangelist comes up with a new era called "Data Era".

Curriculum

Module 1: Introduction to Data Science Methodologies

Data types, Introduction to Data Science tools, Statistics, Approach to business problems, Numerical Categorical, R, Python, WEKA, RapidMiner, Hypothesis testing: Z, T, F test Anova, ChiSq

Module 2: Correlation/ Association Regression Categorical Variables

Introduction to Correlation Spearman Rank correlation, OLS Regression - Simple and Multiple Dummy Variables, Multiple Regression, Assumptions violation - MLE estimates, Using UCI ML repository dataset or built in R dataset

Module 3: Data Preparation

Data preparation and Variable identification, Advanced regression, Parameter Estimation/Interpretation, Robust Regression, Accuracy in Parameter Estimation, Using UCL, ML repository dataset or built in R dataset

Module 4: Logistic Regression

Introduction to Logistic Regression, Logit Function, Training-validation approach, Lift charts, Decline Analysis, Using UCL, ML repository dataset or built in R dataset

Module 5: Cluster Analysis Classification Models

Cluster Analysis Classification Models - Introduction to cluster Techniques, Distance Methodologies, Hierarchical and non- hierarchical Procedure, K-Means clustering, Introduction to decision trees/Segmentation with case study, Using UCL, ML repository dataset or built in R dataset.

Module 6: Introduction and to forecasting techniques

Introduction to Introduction to Time Series, Data and Analysis, Decomposition of Time Series, Trend and Seasonality detection and forecasting, Exponential Smoothing, Building R Dataset, Sales forecasting Case Study

Learning Outcomes

- After completing this course, students will be able to appreciate the need of Data Science in day to day life.
- They will be able to understand the process and components of Data Science project.
- Student will learn the importance of probability and statistics in Data Science