

OBJECTIVES

- Acquire advanced Data Analysis skills.
- Stay Industry relevant and grow in your career.
- Create AI/ML solutions for various business problems.
- Build and deploy production grade AI/ML applications.
- Apply AI/ML methods, techniques and tools immediately

AI is a multidisciplinary field that requires a range of skills in statistics, mathematics, predictive modeling and business analysis. An AI professional should feel at ease to build the algorithms necessary, work with various data sources (often in disparate forms) and an innate ability to ask the right questions and find the right answer. This module helps lay out the canvas on which the rest of the modules are built.

Unit 1: Introduction to Data Science and AI & ML

- Data Science, AI & ML
- Use Cases in Business and Scope
- Scientific Method
- Modeling Concepts
- CRISP-DM Method

Unit 2: R Essentials (Tutorial)

Programming

- Commands and Syntax
- Packages and Libraries
- Introduction to Data Types
- Data Structures in R - Vectors, Matrices, Arrays, Lists, Factors, Data Frames
- Importing and Exporting Data. Control structures and Functions

Descriptive Statistics

- Data exploration (histograms, bar chart, box plot, line graph, scatter plot)
- Qualitative and Quantitative Data
- Measure of Central Tendency (Mean, Median and Mode),
- Measure of Positions (Quartiles, Deciles, Percentiles and Quantiles),
- Measure of Dispersion (Range, Median, Absolute deviation about median, Variance and standard deviation), Anscombe's quartet
- Other Measures: Quartile and Percentile, Interquartile Range

Unit 3: Statistical Analysis

Initial Data Analysis

- Relationship between attributes: Covariance, Correlation Coefficient, Chi Square
- Measure of Distribution (Skewness and Kurtosis), Box and Whisker Plot (Box Plot and its parts, Using Box Plots to compare distribution) and other statistical graphs

Norway, Adolph Tidemands Gate 36 ,N-2000 Lillestrøm,

Unit 4: Data Pre-processing and Preparation

- Data Munging, Wrangling
- Plyr packages
- Cast/Melt

Unit 5: Data Quality and Transformation

- Data imputation
- Data Transformation (minmax, log transform, z-score transform etc.,).
- Binning, Classing and Standardization.
- Outlier/Noise& Anomalies

Unit 6: Handling Text Data

- Bag-of-words
- Regular Expressions
- Sentence Splitting and Tokenization
- Punctuations and Stop words, Incorrect spellings
- Properties of words and Word cloud
- Lemmatization and Term-Document TxD computation
- Sentiment Analysis (Case Study)

Unit 7: Principles of Big Data

- Introduction to Big Data
- Challenges of processing Big Data (Volume, Velocity and Variety perspective)
- Use Cases

Unit 8: Big Data Frameworks – Hadoop, Spark and NoSQL

- Processing, Storage and Programming Framework
- Hadoop eco-system Components and their functions
- Essential Algorithms (Word count, Page Rank, IT-IDF)
- Spark: RDDs, Streaming and Spark ML
- NoSQL concepts (CAP, ACID, NoSQL types)