

5 DAYS TRAINING CONTENT

STATISTICAL DATA MINING AND PROCESSING

1.0 Introduction to Statistical Data Mining and Processing

- Overview of data mining and its significance in modern industries.
- Fundamental concepts and terminologies.
- Differences between data mining and data processing.

2.0 Statistics in Business Intelligence

- Data cleaning and pre-processing techniques using Power BI Software.
- Handling missing data, outliers and noise using Power BI software.
- Data transformation and normalization using Power BI software.

3.0 Exploratory Data Analysis (EDA) for Big Data using Power BI Software

- Descriptive statistics (mean, median, mode, variance and standard deviation)
- Data visualization techniques (histograms, box plots, scatter plots and heatmaps).
- Identifying patterns and relationships in the data.

4.0 Data Mining Techniques

- Classification (decision trees, random forests, support vector machines and k-nearest neighbors)
- Clustering (k-means, hierarchical clustering and DBSCAN).
- Association rule mining (Apriori and FP-Growth algorithms).
- Anomaly detection techniques.

5.0 Tools and Software for Data Mining

- Introduction to popular data mining tools (R, Python (with libraries like pandas, NumPy, SciPy, scikit-learn) and SAS).
- Hands-on sessions and practical examples using these tools.

6.0 Case Studies and Real-World Applications

- Analyzing case studies from various industries (e.g., finance, Maritime Transport, marketing).
- Practical applications of statistical data mining and processing.
- Lessons learned and best practices.