Integrated Programme in Data Science – Machine Learning & Big Data Analytics

Training Mode : Instructor – Led Classroom Training
Number of Hours : 180 to 200 hrs (Hands-On Training)
Machine Learning in : R / Python / Spark ML,Tableau
Big Data Engineering : Hadoop Components

About Data Science / Big Data Analytics Instructor Led Classroom Training:

The Data Science – Machine Learning and Big Data Analytics integrated program is an intensive course designed that makes aspiring Data Scientists, an expert at understanding the real business problem, designing the analysis and applying statistical modelling techniques to derive business insights from data. Our training course content and materials is developed by analytics experts and taught by qualified faculty with extensive experience in the industry and research in their respective domains ho help ensure that students derive the maximum value from the Data Science training chosen.

Analytics Path Data Science training program has been designed to help meet the growing needs for these *"Data Scientists"* with the program comprising the focus area of both Big Data and Basic to Advanced Machine Learning Techniques.

Module1: Business Statistics & Application

- Descriptive Statistics: Measures of Central Tendency, Measures of Spread, Probability Distributions (Discrete & Continuous), Distributions – Normal, Binomial and Poisson, Probability Density Functions, Sampling Distributions, Central Limit Theorem, t-distribution
- Inferential Statistics: Confidence Intervals, Measures of Relationships – Correlation, Covariance, Associations & Odds Ratio
- Probability Refresher

Module2: Data Exploration / Visualization & Data Pre-Processing

- Charts & Graphs – Histogram, Bar chart, Pie chart, Box Plots, Scatter Plots, Line Graphs
- Data Pre-Processing – Data Types & Conversions, Binning & Normalization, Min-Max Scaling, Imputation, Dimensionality Reduction, Outlier Detection and Management, Handling missing values

Module3: MACHINE LEARNING Techniques & Algorithms – *Supervised and Unsupervised*

- Essentials to Machine Learning: Regression & Classification, Training, Validation & Testing, Measures of Performance
- Linear Regression, Logistic Regression(Classification), Decision Trees, Bagging and Random Forest, Boosting, Cross Validation
- Clustering(Segmentation), Dimensionality Reduction Techniques – PCA & SVD, Factor Analysis, Recommender Systems, Association Rules, Forecasting – Time Series
- Deep Learning
- Neural Networks, Support Vector Machines
- Text Mining, Social Network Analysis and Natural Language Processing (NLP), Naïve Bayes classifier for Text Classification

Module4: Big Data Analytics

- Introduction to Big Data, Hadoop – Distributed File System, Map Reduce Applications, Hadoop Ecosystem Components – Hive, Sqoop, Hbase
- Apache Spark MLlib, PySpark & SparkR

Plot No. 28, 4th Floor, Suraj Trade Center, **Opp. Cyber Towers,** Hitech City, Hyderabad - 500081, Telangana.
Tel: 040 - 66828899, Mob:+91 7842828899 Email: info@analyticspath.com