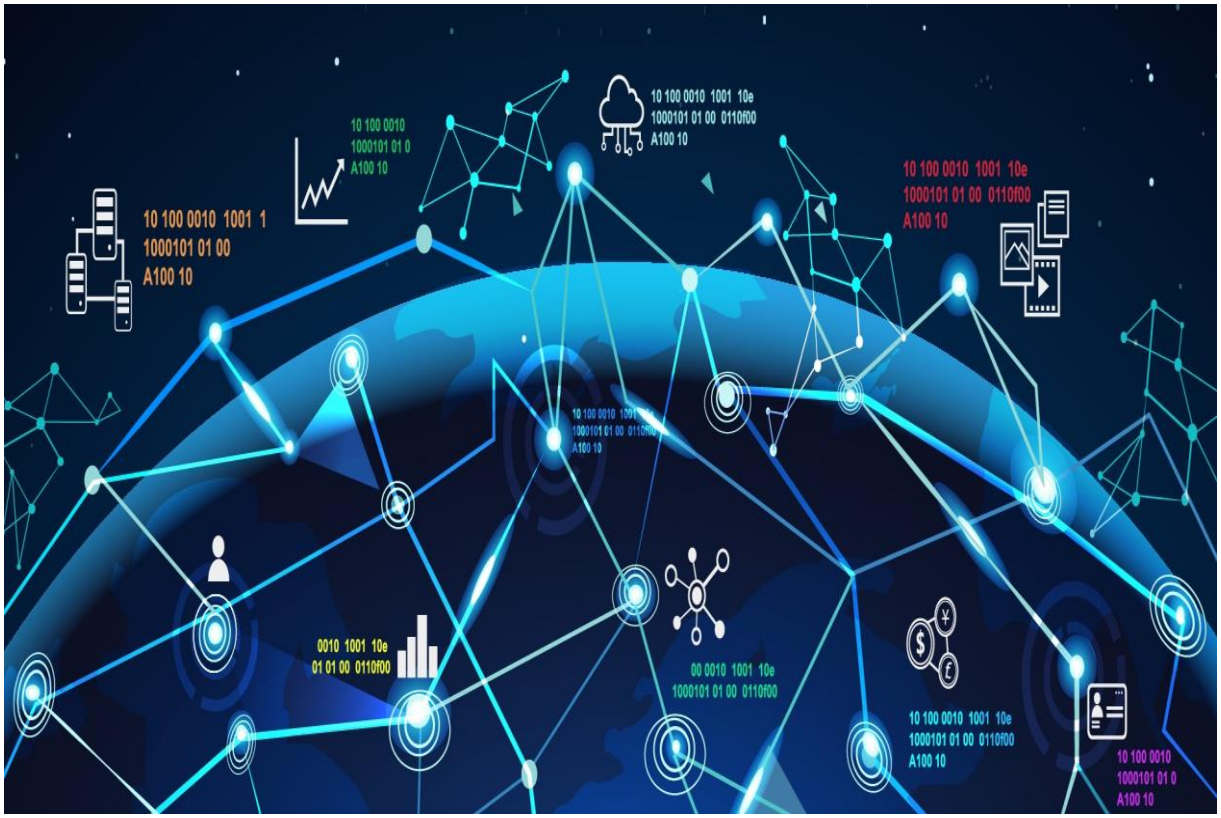
TechyEdz Solutions

Training | Consulting | Developement | Outsourcing



Data Analytics

9032803832

9032803832

contact@techyedz.com

www.techyedz.com

Data Analytics

Course Overview:

Data Analytics course curriculum provides extensive knowledge of Data Collection, Extraction, Cleansing, Exploration, and Transformation. Alongside the Data Mining, Data Integration is done with feature Engineering to build Prediction models for Data Visualization and deploying the solution. You name the skill set and our trainers are always there to handle the new generation tools with latest versions. As a part of the Data Analytics training, the range of skills and tools that are emphasized in the course include Statistical Analysis, Text Mining, Regression Modelling, Hypothesis Testing, Predictive Analytics, Machine Learning, Deep Learning, Neural Networks, Natural Language Processing, Predictive Modelling, R Studio, Tableau, Spark, Hadoop, programming languages like R and Python.

Course Outline:

1. Data Science Project Lifecycle

- Recap of Demo
- Introduction to Types of Analytics
- Project life cycle
- An introduction to our E learning platform

2. Introduction To Basic Statistics Using R And Python

- Data Types
- Measure Of central tendency
- Measures of Dispersion
- Graphical Techniques
- Skewness & Kurtosis
- Box Plot
- R
- R Studio
- Descriptive Stats in R
- Python (Installation and basic commands) and Libraries
- Jupyter note book
- Set up Github
- Descriptive Stats in Python
- Pandas and Matplotlib / Seaborn

3. Probability And Hypothesis Testing

- Random Variable
- Probability
- Probility Distribution
- Normal Distribution
- SND
- Expected Value
- Sampling Funnel
- Sampling Variation
- CLT
- Confidence interval
- Assignments Session-1 (1 hr)
- Introduction to Hypothesis Testing
- Hypothesis Testing with examples
    - 2 proportion test
    - 2 sample t test
- Anova and Chisquare case studies

4. Exploratory Data Analysis -1

- Visualization
- Data Cleaning
- Imputation Techniques
- Scatter Plot
- Correlation analysis
- Transformations
- Normalization and Standardization

5. Linear Regression

- Principles of Regression
- Introduction to Simple Linear Regression
- Multiple Linear Regression

6. Logistic Regression

- Multiple Logistic Regression
- Confusion matrix
    - False Positive, False Negative
    - True Positive, True Negative
    - Sensitivity, Recall, Specificity, F1 score
- Receiver operating characteristics curve (ROC curve)

7. Deployment

- R shiny
- Streamlit

8. Data Mining Unsupervised Clustering

- Supervised vs Unsupervised learning
- Data Mining Process
- Hierarchical Clustering / Agglomerative Clustering
    - Measure of distance
        - Numeric - Euclidean, Manhattan, Mahalanobis
        - Categorical - Binary Euclidean, Simple Matching Coefficient, Jaquard's Coefficient
        - Mixed - Gower's General Dissimilarity Coefficient
    - Types of Linkages
        - Single Linkage / Nearest Neighbour
        - Complete Linkage / Farthest Neighbour
        - Average Linkage
        - Centroid Linkage
- Visualization of clustering algorithm using Dendrogram

K-Means

- Non-Hierarchial
- Measurement metrics of clustering - Within Sum of Squares, Between Sum of Squares, Total Sum of Squares
- Choosing the ideal K value using Scree plot / Elbow Curve

DBSCAN

- A geneal intuition for DBSCAN
- Different parameters in DBSCAN
- Metrics used to evaluate the performance of model
- Pro's and Con's of DBSCAN

9. Dimension Reduction Techniques

- PCA and tSNE
- Why dimension reduction
- Advantages of PCA
- Calculation of PCA weights
- 2D Visualization using Principal components
- Basics of Matrix algebra

10. Association Rules

- What is Market Basket / Affinity Analysis
- Measure of association
- Support
- Confidence

- Lift Ratio
- Apriori Algorithm

11. Recommender System

- User-based collaborative filtering
- Measure of distance / similarity between users
- Driver for recommendation
- Computation reduction techniques
- Search based methods / Item to item collaborative filtering
- Vulnerability of recommender systems

12. Introduction To Supervised Machine Learning

- Workflow from data to deployment
- Data nuances
- Mindsets of modelling

13. Decision Tree

- Elements of Classification Tree - Root node, Child Node, Leaf Node, etc.
- Greedy algorithm
- Measure of Entropy
- Attribute selection using Information Gain
- Implementation of Decision tree using C5.0 and Sklearn libraries

14. Exploratory Data Analysis - 2

- Encoding Methods
  - OHE
  - Label Encoders
  - Outlier detection-Isolation Fores
- Predictive power Score

15. Feature Engineering

- Recurcive Feature Elimination
- PCA

16. Model Validation Methods

- Splitting data into train and test
- Methods of cross validation
- Accuracy methods

17. Ensembled Techniques

- Bagging
- Boosting
- Random Forest
- XGBM
- LGBM

18. KNN And Support Vector Machines

- Deciding the K value
- Building a KNN model by splitting the data
- Understanding the various generalization and regulation techniques to avoid overfitting and underfitting
- Kernel tricks

19. Regularization Techniques

- Lasso Regression
- Ridge Regression

20. Neural Networks

- Artificial Neural Network
- Biological Neuron vs Artificial Neuron
- ANN structure
- Activation function
- Network Topology
- Classification Hyperplanes
- Best fit “boundary”
- Gradient Descent
- Stochastic Gradient Descent Intro
- Back Propogation
- Intoduction to concepts of CNN

21. Text Mining

- Sources of data
- Bag of words
- Pre-processing, corpus Document-Term Matrix (DTM) and TDM
- Word Clouds
- Corpus level word clouds
    - Sentiment Analysis
    - Positive Word clouds
    - Negative word clouds
    - Unigram, Bigram, Trigram
- Vector space Modelling
- Word embedding

- Document Similarity using Cosine similarity
- Extract Tweets from Twitter
- Extract user reviews of the products from Amazon, Snapdeal and TripAdvisor
- Install Libraries from Shell
- Extraction and text analytics in Python

22. Natural Language Processing

- Sentiment Extraction
- Lexicons and Emotion Mining

23. Naive Bayes

- Probability – Recap
- Bayes Rule
- Naive Bayes Classifier
- Text Classification using Naive Bayes

24. Forecasting

- Introduction to time series data
- Steps of forecasting
- Components of time series data
- Scatter plot and Time Plot
- Lag Plot
- ACF - Auto-Correlation Function / Correlogram
- Visualization principles
- Naive forecast methods
- Errors in forecast and its metrics
- Model Based approaches
    - Linear Model
    - Exponential Model
    - Quadratic Model
    - Additive Seasonality
    - Multiplicative Seasonality
- Model-Based approaches
- AR (Auto-Regressive) model for errors
- Random walk
- ARMA (Auto-Regressive Moving Average), Order p and q
- ARIMA (Auto-Regressive Integrated Moving Average), Order p, d and q
- Data-driven approach to forecasting
- Smoothing techniques
    - Moving Average
    - Simple Exponential Smoothing
    - Holts / Double Exponential Smoothing
    - Winters / HoltWinters

- De-seasoning and de-trending
- Forecasting using Python and R

25. Survival Analysis

- Concept with a business case

26. End To End Project Description With Deployment

- End to End project Description with deployment using R and Python

27. Assignments

- Basic Statistics
  - Data types Identification and probability
  - Expected values, Measures of central tendencies
  - Skewness and Kurtosis & Boxplot
  - Practice Mean, Median, Varience, Standard Deviation and Graphical representations in R
  - Creating Python Objects
  - Practice Mean, Median, Varience, Standard Deviation and Graphical representations in Python
  - Confidence intervals and distributions
- Hypothesis Testing
  - Buyer ratio
  - Customer Order Form
  - Cutlets
  - Pantaloons
  - Lab TAT
- Linear regression
  - Prediction of weight based on Calories consumed
  - Delivery Time period Vs Sorting time
  - Employee Churn rate Vs Salary
  - Salary Prediction
- R shiny and Flask
  - Practice R shiny and Python Flask for Linear Regression assignments
- Multiple Linear Regression
  - 50 startups case study
  - Computer data Case study
  - Toyota Corolla
- Logistic Regression
  - Term deposit case study
  - Elections results Case study
- Multinomial Regression
  - Student Program Case study
- Hierarchical Clustering

- Crime data
- Eastwest Airlines

- K means Clustering
  - Insurance policy
  - Crime data
- PCA
  - Dimension Reduction for Wine data
- Network Analytics
  - Node Properties practice in R
- Association Rules
  - Association Rules for Book store
  - Association Rules for Mobile store
  - Association Rules for Retail Transactions
- Recommendation Engine
  - Recommend Jokes for subscribers
- Text mining, Web Extraction
  - Extraction of tweets from twitter
  - Reviews from ecommerce websites
- Text mining
  - Sentiment Analysis on extracted data
- NLP
  - Emotion mining by extracting a speech or novel from web
- Naive Bayes
  - Spam and Ham classifications
- KNN Classifier
  - Types of Glass
  - Classification of Animals
- Decision Tree and Random Forest
  - Fraud Check
  - Sales prediction of an Organization
- XGB and GLM
  - Social Networks Ads
- Lasso and Ridge Regression
  - Practice Lasso and Ridge with multiple Linear Assignments
- ANN
  - Forest Fires case study
- SVM
  - Classification of Alphabets
- Survival analysis
  - Prediction of Patient survival probability
- Forecasting model based
  - Airlines Forecasting
  - Forecasting of sales for a soft drinks case study
- Forecasting
  - Forecasting of Bike shares

- Forecasting of Solar power consumption

Prerequisites:

- Some programming experience (preferred)
- Structured thinking approach
- Passion for solving problems
- Willingness to learn statistical concepts

Who Can attend:

- Professionals or fresher's who are really serious about making a career in Data Analytics can do this course
- Individuals from any domain who possess logical thinking about mathematical and analytical skills.
- People who are working on business intelligence tools, data warehousing and reporting tools.
- Statisticians, Economists, Mathematicians
- Software programmers
- Business Analysts
- Six Sigma Consultants
- Digital Marketing professionals
- Freshers from any stream with good analytical and logical skills.

Number of Hours: 50hrs

Certification:

1. Associate Certified Analytics Professional (aCAP)

2. Certified Analytics Professional (CAP)

Key Features:

- One to One Training
- Online Training
- Fastrack & Normal Track
- Resume Modification

- Mock Interviews
- Video Tutorials
- Materials
- Real Time Projects
- Virtual Live Experience
- Preparing for Certification