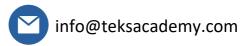


Data Science







Module 01 - Data Science Project Lifecycle

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

Module 02 - 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 notebook
- Set up GitHub
- Descriptive Stats in Python
- Pandas and Matplotlib / Seaborn





Module 03 - Probability And Hypothesis Testing

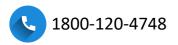
Topics

- Random Variable
- Probability
- Probability Distribution
- Normal Distribution
- SND
- Expected Value
- Sampling Funnel
- Sampling Variation
- CLT

- Introduction to Hypothesis Testing
 Hypothesis Testing with and
 2 -
- 2 proportion test
- 2 sample t-test
- Anova and Chisquare case studies

Module 04 - Exploratory Data Analysis -1

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







Module 05 - Linear Regression

Topics

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

Module 06 - Logistic Regression

Topics

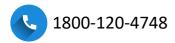
- Multiple Logistic Regression
- Confusion matrix
 - 1. False Positive, False Negative
 - 2. True Positive, True Negative
 - 3. Sensitivity, Recall, Specificity, F1 score
- Receiver operating characteristics curve (ROC curve)

Module 07 - Deployment

- R shiny
- Streamlit

Module 08 - Data Mining Unsupervised Clustering

- Supervised vs Unsupervised learning
- Data Mining Process







- Hierarchical Clustering / Agglomerative Clustering
 - Measure of distance
 - 1. Numeric Euclidean, Manhattan, Mahalanobis
 - 2. Categorical Binary Euclidean, Simple
 Matching Coefficient, jacquard's Coefficient
 - 3. Mixed Gower's General Dissimilarity Coefficient
 - Types of Linkages
 - 1. Single Linkage / Nearest Neighbour
 - 2. Complete Linkage / Farthest Neighbour
 - 3. Average Linkage
 - 4. 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 general intuition for DBSCAN
- Different parameters in DBSCAN
- Metrics used to evaluate the performance of a model
- Pro's and Con's of DBSCAN





Module 09 - Dimension Reduction Techniques

Topics

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

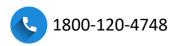
Module 10 - Association Rules

Topics

- What is Market Basket / Affinity Analysis
- Measure of association
- Support
- Confidence
- Lift Ratio
- Apriori Algorithm

Module 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







Module 12 - Introduction To Supervised Machine Learning

- · Workflow from data to deployment
- Data nuances
- · Mindsets of modeling

Module 13 - Decision Tree

Topics

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

Module 14 - Exploratory Data Analysis – 2

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





Module 15 - Feature Engineering

Topics

- Recursive Feature Elimination
- PCA

Module 16 - Model Validation Methods

Topics

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

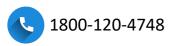
Module 17 - Ensembled Techniques

Topics

- Bagging
- Boosting
- Random Forest
- XGBM
- LGBM

Module 18 - Feature Engineering

- 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







Module 19 - Regularization Techniques

Topics

- Lasso Regression
- Ridge Regression

Module 20 - Neural Networks

Topics

- 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 Propagation
- Introduction to concepts of CNN

Module 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

Description: Learn how to extract data from Social Media, and download user reviews from E-commerce and Travel websites. Generate various visualizations using the downloaded data.

Topics

- Extract Tweets from Twitter
- Extract user reviews of the products from Amazon, Snapdeal, and TripAdvisor

Description: Learn how to perform text analytics using Python and work with various libraries that aid in data extraction, text mining, sentiment analysis and

Topics

- Install Libraries from Shell
- · Extraction and text analytics in Python

Module 22 - Natural Language Processing

- Sentiment Extraction
- · Lexicons and Emotion Mining







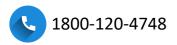
Module 23 - Naive Bayes

Topics

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

Module 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

nowledge To S **Module 25 – Survival Analysis**

· Concept with a business case

Module 25 – Survival Analysis

 End to End project Description with deployment using R and Python







Why Teks Academy?



Training from Industry Experts



3 Capestone Projects



Internship Certificate



Resume Development



Interview Skills
Training



Communication Skills Training

12000+

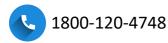
80%

ISO

Student Trained

Placement Record

Certified







Download Syllabus





