

INDUSTRY ALIGNED CURRICULUM (IAC)

Module 1

Data Science Fundamentals

Data Science Overview

Define Data Science | Data Science Life Cycle | Data Science Importance | Python and Its importance | Roles and responsibilities of Data Scientist | Various Applications of Data Science | Linux Essentials | Git, Version | Control Essentials | Case Study – Spotify | Case Study – LinkedIn | Case Study – Uber | Data Science in Education | Case Study-Customer | Personality Analysis

Job Description-Roles and Responsibilities of leading corporates

Google | Infosys | Virtusa | Accenture | IBM

Business Use Cases

Credit Card Fraud Detection | Customer Segmentation | Pattern recognition

Data Analytics Overview

Data Analytics Process and Its steps | Skills and Tools Required for Data Analysis | Challenges in Data Analytic Processes | Data Visualization Technique | Exploratory Data Analysis Technique | Hypothesis Testing

Module 2

Statistical Analysis and Business Applications

Statistical vs Non-Statistical Analysis | Major categories of Statistics | Population-Sample | Statistical Analysis Process | Distribution | Bell Curve | Descriptive Statistics | Hypothesis Testing | Data Types and Variable types | Types of Frequencies | Chi- Square test | Correlation matrix | Inferential Statistics

Module 3

Numeric Computing with Python

Numpy Fundamentals | Numpy DataTypes | Working with Arrays using Numpy | Boolean and fancy Indexing using Numpy | Generating Data With Numpy

Module 4

Data Manipulation, Collecting & Feature Engineering

Data Manipulation

Pandas and its Features | Data Structures | Series, Panel | Dataframes | Data Operations | Data Standardization | Pandas SQL Operations

Collecting Data

Understanding the domain and the dataset | Importing and Exporting Data | Basic Insights from dataset

Data Cleaning

Feature Engineering | Normalization, Indicator Variables and Binning

Module 5

Scientific Computing with Python

SciPy and its characteristics | SciPy Sub-Package | Sub-package Linear Algebra | Sub-package Optimization | Sub-package Statistics | Sub-package Weave and IO | Sample Programs

Module 6

Encapsulating the data

Descriptive Statistics | Grouping | Variance & Covariance | ANOVA | ANCOVA | Correlation

Module 7

Data Visualization & Types of Plots

Data Visualization in Python using Matplotlib

Data Visualization | Considerations of Data Visualization | Factors of Data Visualization | Python libraries for Data Visualization | Steps to create a plot | Line Properties | Alpha and Annotation | Multiple Plots

Types of Plots

Seaborn module | Pandas built in functions for data visualization | Basic Visualization Tools | Specialized Visualization Tools | Advanced Visualization Tools | Plotly and Cufflinks | Maps and Geospatial Data

Module 8

Tools and Techniques Tableau

Tableau Introduction | Working with Tableau | Deep diving with data and connection
| Creating charts | Mapping data in Tableau | Dashboards and stories

Module 9

Development of Model

What is Model | How to Build a Model | Types of Learning | Supervised Learning |
Unsupervised Learning | Reinforcement Learning

Module 10

Supervised & Unsupervised Learning

Mathematical Modeling and performance of Regression

Simple Linear Regression | Multiple Linear Regression | R-squared and MSE for
sample Evaluation | Logistic Regression | Ridge Regression | Lasso Regression |
Polynomial Regression | Bayesian Linear Regression | Prediction and Decision
Making

Supervised Learning-Mathematical Modeling and performance of Classification Algorithms

Nearest Neighbor | Gaussian Naive Bayes | Decision Trees | Support Vector Machine
(SVM) | Random Forest | Fine-tuning the model | Preprocessing and Pipelines

Unsupervised Learning

Clustering data | Visualization | Decorrelating and Dimension Reduction | Interpre-
tation - NMF | K-means and Hierarchical clustering | Agglomerative and Divisive
Hierarchical Clustering | Density Based Clustering | Singular Value Decomposition |
Independent Component Analysis

Association Rule Learning

Apriori | Eclat | F-P Growth Algorithm

Module 11

Dimensionality Reduction

Principal Component Analysis

Introduction | Eigen values and Vectors | Applications | Implementation | t-Distrib-
uted Stochastic Neighbor Embedding (t-SNE) | Linear Discriminant Analysis

Module 12

Evaluation of Model

Model Evaluation Metrics for Supervised Learning | Model Evaluation Metrics for
UnSupervised Learning | Over-fitting and Under-fitting | Model Selection | Grid
Search | Fine Tuning the model

Module 13

Web Scraping & Natural Language Processing

Natural Language Processing with Scikit Learn

NLP Introduction & NLP Techniques | Tokenization, Stemming, Lemmatization | NLP -
Benefits, Applications | Modules to Load content and category | Implementation of
Extraction | Feature Extraction Techniques | Twitter Sentiment Analysis

Web Scraping with Beautiful Soup

Web Scraping and its advantages | Web Scraping process | Beautiful Soup and its
Features | BS4, Requests Library

Module 14

Recommender Systems

Overview | Amazon App Case Study | Health and Fitness Case Study | Location
based Recommendation Case Study | Simple Recommender | Content Based
Recommender

Module 15

Neural Nets and Deep Learning Overview

ANN | Neural Networks | TensorFlow | Convolution Neural Network | Recurrent
Neural Network | Keras