



DATA SCIENCE

Complete Course Curriculum

Python • Statistics • Machine Learning
Deep Learning • NLP • SQL • Tableau

📍 Tilak Nagar Metro, New Delhi | 📞 9911144059, 9911144051 | 🌐 www.airoman.in



DATA SCIENCE

THE DISCIPLINE OF INTELLIGENT DISCOVERY

TOP DATA SCIENCE CAREERS

1

Data Scientist

2

Machine Learning
Engineer

3

Data Analyst

4

Data Engineer

5

AI Specialist

10

Marketing Analyst

9

NLP Engineer

8

Big Data Expert

7

Research Scientist

6

Business Analyst



Demand for Data Science Positions (Job Growth 2021–2031)

41.90% increase in jobs

DATA SCIENCE CURRICULUM

01

Python

02

Statistics

03

Machine Learning

04

Deep Learning +
Computer Vision

05

NLP

06

SQL

07

Tableau

COURSE OBJECTIVE & KEY FEATURES

Course Objective

- To understand the vital nature of data for organizations
- To learn the conceptual framework of machine learning
- To explore and analyze data using supervised and unsupervised learning techniques
- To develop and deploy knowledge learning models using Python
- To work on Unstructured Data using Nltk and building Modules
- Understanding Neural Networks and building deep networks using Tensorflow and Keras
- Working with image processing & computer vision

Key Features in the Training

- Duration: 5-6 Months | Class Duration: 2 Hrs (Mon-Fri)
- Online help on Doubt Clearance, Monitoring Session, Career Guidance, Interview Preparation & Mock Interviews
- Use Cases: Python & Statistics: 4, Machine Learning: 10, NLP: 2, DL: 3
- One Big Hackathon Challenge on Machine Learning
- Projects: Data Analysis, Regression, Classification, Sentiment Analysis / Chatbot, Face Recognition
- Assignments & Quizzes for each Module
- IBM Credentials and Certification after completion
- Guaranteed In-house Internship
- Nearly working on 20 use cases during your course
- Training materials with Lab Exercises, Datasets, Codes, Quizzes, Case Studies on real data
- Every online session Recorded video & live running notes
- Real time Training with live Scenarios and Applications
- Job Assistance after completion of the course

MODULE 1: PYTHON CORE & ADVANCED

Introduction & Walk Through the Course

- Introduction to Data Science
- Life cycle of data science
- Skills required for data science
- Applications of data science in different industries

Introduction to Python

- What is Python? Why does Data Science require Python?
- Installation of Anaconda
- Understanding Jupyter Notebook
- Basic commands in Jupyter Notebook
- Understanding Python Syntax

Data Types and Data Structures

- Variables and Strings
- Lists, Sets, Tuples and Dictionaries

Control Flow & Conditional Statements

- Conditional, Arithmetic and Logical Operators
- If, Elif and Else Statements
- While Loops | For Loops
- Nested Loops and List/Dictionary Comprehensions

Functions

- Types of functions
- Lambda Functions
- Map, Filter and Reduce

File Handling

- Create, Read, Write files
- Errors and Exception Handling

MODULE 1: CLASS & OBJECTS | MODULE 2: DATA ANALYSIS IN PYTHON

Class and Objects

- Create a class | Create an object
- The `__init__()`
- Modifying Objects | Object Methods | Self
- Modify Object Properties | Delete Object
- Pass Statements

Numpy - Numerical Python

- Introduction to Array
- Creation and Printing of ndarray
- Basic Operations in Numpy
- Indexing
- Mathematical Functions of Numpy

Data Manipulation with Pandas

- Series and DataFrames
- Data Importing and Exporting through Excel, CSV Files
- Data Understanding Operations
- Indexing, Slicing and Conditional Slicing
- Groupby, Pivot table and Cross Tab
- Concatenating, Merging and Joining
- Descriptive Statistics | Removing Duplicates
- String Manipulation | Missing Data Handling

Data Visualization using Matplotlib & Pandas

- Introduction to Matplotlib | Basic Plotting | Properties
- Line plots | Pie chart | Bar Graph | Histograms
- Box, Violin Plots and Scatterplot

DATA VISUALIZATION (EDA) | MODULE 3: ADVANCED STATISTICS

Case Study on EDA and Visualizations

- What is EDA?
- Uni-Variate Analysis
- Bi-Variate Analysis
- Seaborn: Pair Plots, Catplot, Heat Maps, Count plot

Regular Expressions & Unstructured Data

- Structured Data and Unstructured Data
- Literals and Meta Characters
- How to use Regular Expressions using Pandas?
- Inbuilt Methods | Pattern Matching

Project: Web Scraping, Data Mining & EDA

Data Collection → Data Mining → Data Preprocessing → Data Visualization

Data Gathering Techniques

- Data types | Data Collection Techniques
- Sampling: Convenience, Simple Random, Stratified, Systematic and Cluster Sampling

Descriptive Statistics

- Univariate and Bi Variate Analysis
- Measures of Central Tendencies | Measures of Dispersion
- Skewness and Kurtosis | Box Plots and Outliers detection
- Covariance and Correlation

Probability Distribution

- Discrete: Bernoulli, Binomial, Poisson
- Continuous: Normal, Standard Normal

Inferential Statistics

- Z-test, t-test, Chi-Square, ANOVA
- Confidence Intervals
- Hypothesis Testing

MODULE 4: SQL FOR DATA SCIENCE

SQL for Data Science - Introduction to Databases

- Basics of SQL
- DML, DDL, DCL and Data Types
- Common SQL commands using SELECT, FROM and WHERE
- Logical Operators in SQL

SQL Joins

- INNER and OUTER joins to combine data from multiple tables
- RIGHT, LEFT joins to combine data from multiple tables

Filtering and Sorting

- Advanced filtering using IN, OR and NOT
- Sorting with GROUPBY and ORDER BY

SQL Aggregations

- Common Aggregations including COUNT, SUM, MIN and MAX
- CASE and DATE functions as well as work with NULL values

Subqueries and Temp Tables

- Subqueries to run multiple queries together
- Temp tables to access a table with more than one query

SQL Data Cleaning

- Perform Data Cleaning using SQL

MODULE 5: MACHINE LEARNING — REGRESSION TECHNIQUES

Introduction to Machine Learning

What Is Machine Learning? | Supervised vs Unsupervised Learning | Regression vs Classification | Assessing Model Accuracy

Linear Regression

- Simple Linear Regression: Estimating Coefficients
- Assessing Coefficient Estimates
- R Squared and Adjusted R Squared | MSE and RMSE

Multiple Linear Regression

- Estimating Regression Coefficients
- OLS Assumptions | Multicollinearity
- Feature Selection | Gradient Descent

Evaluating Regression Metrics

- Homoscedasticity and Heteroscedasticity of error terms
- Residual Analysis | Q-Q Plot
- Cook's distance and Shapiro-Wilk Test
- Identifying the line of best fit
- Qualitative Predictors | Interaction Terms
- Non-linear Transformations of the Predictors

Polynomial & Regularization

- Why Polynomial Regression | Creating polynomial linear regression
- Lasso, Ridge, ElasticNet Regularization

MODULE 5: CLASSIFICATION TECHNIQUES

Logistic Regression

- An Overview of Classification
- Logistic Model, Logit and Sigmoid functions
- Estimating Coefficients and Making Predictions
- Setting threshold and decision boundary
- Logistic Regression for >2 Response Classes

Evaluation Metrics

- Confusion Matrix
- Accuracy and Error rate
- TPR and FPR
- Precision and Recall, F1 Score
- AUC – ROC | Kappa Score

Naive Bayes

- Principle of Naive Bayes Classifier
- Bayes Theorem and Terminology
- Posterior, Prior probability, Likelihood
- Multinomial, Bernoulli and Gaussian Naive Bayes

Tree Based Modules - Decision Trees

- Basic Terminology, Root Node and Terminal Node
- Regression Trees and Classification Trees
- Gini Index | Overfitting and Pruning
- Stopping Criteria | Accuracy Estimation

Ensemble Methods & Random Forest

- What is Ensemble Learning?
- Bootstrap Aggregation Classifiers
- What is Random Forest and how it works?
- Variable selection using Random Forest

Boosting & K-Nearest Neighbors

- AdaBoost | Gradient Boosting
- K-Nearest Neighbor Algorithm
- Curse of Dimensionality | Pros and Cons
- How to improve KNN performance

MODULE 6: UNSUPERVISED LEARNING

Support Vector Machines (SVM)

- Maximal Margin Classifier | HyperPlane
- Support Vector Classifiers and SVMs
- Hard and Soft Margin Classification
- Classification with Non-linear Decision Boundaries
- Kernel Trick – Polynomial and Radial
- Tuning Hyper parameters: Gamma, Cost and Epsilon
- SVMs with More than Two Classes

Principal Components Analysis (PCA)

- Introduction to Dimensionality Reduction
- What Are Principal Components? | 2D & 3D PCA
- EigenValues, EigenVectors and Orthogonality
- Proportion of variance explained in PCA

Recommendation Systems

- Content based & Collaborative filtering
- Cold start problem | Matrix factorization
- Building recommendation engine

K-Means Clustering

- Centroids and Medoids
- Deciding optimal value of 'k' using Elbow Method
- Linkage Methods

Hierarchical Clustering

- Divisive and Agglomerative Clustering
- Dendrograms and their interpretation
- Applications of Clustering | Practical Issues in Clustering

Association Rules & Apriori

- Market Basket Analysis
- Metric Support/Confidence/Lift
- Improving Supervised Learning with clustering

MODULE 7: DEEP LEARNING

Introduction to Neural Networks

- Introduction to Perceptron & History of Neural networks
- Activation functions: Sigmoid, Relu, Softmax, Leaky Relu, Tanh
- Gradient Descent | Learning Rate and tuning
- Optimization functions | Introduction to Tensorflow & Keras
- Back propagation and chain rule | Fully connected layer
- Cross entropy | Weight Initialization | Regularization

TensorFlow 2.0

- Introducing Google Colab
- Tensorflow basic syntax | Tensorflow Graphs | Tensorboard

Artificial Neural Network with Tensorflow

- Neural Network for Regression and Classification
- Evaluating, Improving and tuning the ANN
- Saving and Restoring Graphs

MODULE 8: CNN & COMPUTER VISION

- UNIT 1: Working with Images & CNN Building Blocks
- Working with Images – Reshaping, Filtering, Digitization
- Introduction to Convolutions | 2D convolutions for Images
- Convolution Backward | Transposed Convolution
- Pooling: Max Pooling and Other pooling options

CNN Architectures & Transfer Learning

- CNN Architectures and LeNet Case Study
- Case Studies: AlexNet, ZFNet, VGGNet, GoogleNet, ResNet
- GPU vs CPU | Transfer Learning Principles and Practice
- Hands-on Keras Demo: SVHN Transfer learning from MNIST

Object Detection & Semantic Segmentation

- Object Detection with Yolo and SSD
- U-Net Architecture for Semantic Segmentation
- Siamese Network for Metric Learning

MODULE 9: NATURAL LANGUAGE PROCESSING (NLP)

Unit 1: Introduction to Statistical NLP Techniques

- Introduction to NLP
- Preprocessing: NLP Tokenization, stop words, normalization, Stemming and lemmatization
- Preprocessing: Bag of words, TF-IDF as features
- Language model: probabilistic models, n-gram model and channel model
- Hands on NLTK

Unit 2: Word Embedding

- Word2vec | GloVe
- POS Tagger | Named Entity Recognition (NER)
- POS with NLTK | TF-IDF with NLTK

Unit 3: Sequential Models

- Introduction to sequential models
- Introduction to RNN | Intro to LSTM
- LSTM forward pass | LSTM backprop through time
- Hands on keras LSTM

Unit 4: Applications

- Sentiment Analysis
- Sentence generation | Machine translation
- Advanced LSTM structures
- Keras-machine translation
- ChatBot

MODULE 10: TABLEAU | TOOLS & TECHNOLOGIES

Tableau for Data Science

- Install Tableau for Desktop 10
- Connect Tableau to a variety of datasets
- Analyze, Blend, Join and Calculate Data
- Visualize Data in the form of Various Charts, Plots and Maps
- Work with Data Blending in Tableau | Work with Parameters
- Create Calculated Fields | Adding Filters and Quick Filters
- Create Interactive Dashboards | Adding Actions to Dashboards

Tools & Technologies You Will Master

Python

NumPy + Pandas

Matplotlib

Seaborn

Plotly

SQL / MySQL

NLTK

RegEx

Jupyter

Scikit-learn

TensorFlow

Keras

OpenCV

Tableau

IBM Watson

MODULE 11: AI TOOLS FOR DATA SCIENTISTS

Introduction to AI Tools

- What are AI Tools and why every Data Scientist needs them
- How AI is transforming the Data Science workflow
- Difference between Traditional ML and AI-powered tools

AI Coding Assistants

- GitHub Copilot for Data Science projects Amazon CodeWhisperer
- Google Gemini for code and analysis
- Auto-complete and debugging with AI

AI for Automation & Workflow

- Make (Integromat) for automating Data pipelines
- Zapier AI for workflow automation
- Notion AI for documentation of projects
- Auto-generating project reports with AI

ChatGPT & Large Language Models (LLMs)

- Using ChatGPT for Data Analysis and Code Generation
- Prompt Engineering for Data Science tasks
- Auto-generating Python code, SQL queries and reports
- Summarizing datasets and generating insights

AI for Data Visualization

- AI-powered chart generation tools
- Julius AI — natural language to charts
- Tableau AI and Power BI Copilot features
- Auto-generating dashboards from raw data

Generative AI in Data Science

- Introduction to Generative AI concepts
- Using Midjourney / DALL-E for data presentations
- AI-generated synthetic datasets for training models
- Ethical use of AI tools in Data Science

100% JOB PLACEMENT ASSISTANCE

Our Placement Process

- Dedicated Placement Cell with experienced team
- Resume building and LinkedIn profile optimization
- Portfolio project guidance for GitHub
- Mock interviews with industry experts
- Soft skills and communication training

Top Roles, We Place You In

- Data Scientist — ₹6–20 LPA
- Machine Learning Engineer — ₹8–22 LPA
- Data Analyst — ₹4–10 LPA
- Data Engineer — ₹6–18 LPA
- AI/ML Researcher — ₹10–25 LPA
- Business Intelligence Analyst — ₹5–12 LPA
- NLP Engineer — ₹8–20 LPA
- Computer Vision Engineer — ₹8–22 LPA

Our Hiring Partners

- 200+ companies in our placement network
- Top MNCs, Product companies and Startups
- IT firms: TCS, Infosys, Wipro, HCL, Tech Mahindra
- Product companies: Amazon, Flipkart, Paytm, Zomato
- Analytics firms: Mu Sigma, Fractal Analytics, Tiger Analytics

Placement Statistics

- 50+ topic-wise mock interview sessions
- Data Science interview Q&A preparation
- Live coding rounds practice (HackerRank / LeetCode)
- System design interview guidance
- Company-specific preparation for top firms Resume shortlisting support and job referrals

Interview Preparation Support

- 95% Placement Rate
- Average Salary: ₹6.5 LPA
- Highest Package: ₹24 LPA
- 5000+ Students Placed
- 200+ Hiring Partners
- Placement within 3–6 months of course completion



START YOUR DATA SCIENCE JOURNEY TODAY!



Call / WhatsApp

+91 9911144059, 9911144051




Website

www.airoman.in



Email

info@airoman.in
airoman.del@gmail.com

 Tilak Nagar Metro Station Gate No. 3, Double Story Panther Brother Building 29/3, 1st Floor,
Ashok Nagar, Next to KFC, New Delhi, Delhi 110018