

## DATA SCIENCE WITH PYTHON

### DATA SCIENCE / ARTIFICIAL INTELLIGENCE?

Data Science is a concept used to understand and analyze data to get insights into that topic and includes data cleansing, preparation and analysis. It is an umbrella term which encompasses several related fields like Mathematics, Statistics, Computer Science, Machine Learning, Data Mining, Visualization etc. It is the foremost problem-solving tool of the current industrial age.

Machine learning (ML) is said to be one of the most critical skills of current times. From autonomous vehicles to self-tuned databases, ML is found everywhere. Industry analysts often refer to ML-driven automation as the job killer. Almost every domain and industry vertical are getting impacted by ML. Platform companies with massive investments in Artificial Intelligence (AI) research are shipping new tools and frameworks at a rapid pace.

All the above factors have put an average developer under pressure to acquire machine learning skills. Despite the interest and a sense of urgency, students are struggling to learn the essential skills required to master ML. Most institutions offering to teach the course also just barely cover basics and without proper structured learning.

This is where Technocourses.com steps in with our 'Data Science with Python', program. We make learning Data Science, Machine Learning and Deep Learning fun, easy and highly engaging to people of all backgrounds.

### WHY TECHNOCOURSES.COM?

- **Systematic Learning:** Become expert in Data Science the right way quickly, instead of stumbling around online for days unnecessarily. Use that time by working on even more projects to add to your portfolio.
- **Amazing Projects:** Finish over 10 plus projects such as Sales Prediction, Stock Analysis, Image Classification, House Price Prediction, etc., and few capstone projects.
- **Stay Connected:** With Google Classrooms and WhatsApp groups, stay in touch with your tutor to clarify your doubts and receive a curated content of articles every day to stay up to date in this field.
- **Continuous Learning:** Receive assignments and reference notebooks every day to continue the learning at your own time. Refer books on Data Science present at the institute to further your understanding.
- **Add-on Courses:** Get Add-on courses on Business English and Soft Skills free of cost, to further your skills.

### CONTENT

#### PART I: Python Essentials for Data Science – 36 Hours

- Module 1: Python Fundamentals and Programming – 12 Hrs.
- Module 2: Data Handling with NumPy and Pandas – 12 Hrs.
- Module 3: Data Visualization with Matplotlib – 6 Hrs.
- Module 4: Advanced Plotting with Seaborn– 6 Hrs.

#### PART II: Statistics for Data Science – 30 Hours

- Module 5: Introduction to Statistics for Data Science – 4 Hrs.
- Module 6: Sampling methods in Statistics – 7 Hrs.
- Module 7: Exploratory Analysis and Distributions – 12 Hrs.
- Module 8: Advanced Statistics for Data Science – 7 Hrs.

#### PART III: Data Science & Applied Machine Learning – 54 Hours

- Module 9: Machine Learning (ML) Fundamentals – 3 Hrs.
- Module 10: ML Regression and Classification Algorithms – 12 Hrs.
- Module 11: ML Advanced Algorithms & Techniques– 16 Hrs.
- Module 12: Deep Learning & Neural Networks – 13 Hrs.
- Module 13: Applications of ML in Bigdata, Cloud Computing etc., – 10 Hrs.

#### PART IV: Practice Sessions – 30 Hours

### ADDONS

- Advance MS Excel
- CV preparation, mock interviews & Guidance
- Business English & Communication Skills

### **Module 1 – Python Fundamentals and Programming**

- What is Python?
- Why is Python essential for Data Science?
- Versions of Python, How to install Python
- Anaconda Distribution, Jupyter Notebooks
- Command line basics, GitHub overview
- Understand Python scripts, Data Types
- Programming Concepts, Python, Operators
- Conditional Statement, Loops, Lists, Tuples, Dictionaries, Sets, Methods and Functions
- Errors and Exception Handling
- Object Oriented Programming in Python
- Modules and Packages

### **Module 2 – Data Handling with NumPy and Pandas**

- NumPy overview, Arrays & Matrices
- NumPy basic operations, functions
- NumPy for Data Analysis , Importing Pandas
- Pandas overview, Pandas Series and Data Frames, Dealing with missing data
- GroupBy, Merging, Concatenating and joining
- Data Input & Output

### **Module 3 – Data Visualization with Matplotlib**

- Why visualize data?, Importing Matplotlib
- Chart: Line Chart, Bar Charts and Pie Charts
- Plotting from Pandas object
- Object Oriented Plotting: Setting axes limits and ticks
- Plot Formatting: Custom Lines, Markers, Labels, Annotations, Colors

### **Module 4 –Advanced Data Visualization with Seaborn**

- Importing Seaborn, Seaborn overview
- Distribution and Categorical Plotting
- Matrix plots & Grids, Regression Plots
- Style & Color , Review Session

### **Module 5: Introduction to Statistics for Data Science**

- Applied statistics in business
- Descriptive & Inferential Statistics
- Statistics Terms and definitions
- Types of Data, Data Measurement Scales

### **Module 6: Sampling methods in Statistics**

- Sampling Data, with and without replacement
- Sampling Methods, Random vs Non-Random
- Measurement on Samples
- Random Sampling methods
- Simple random, Stratified, Cluster, Systematic sampling., Biased vs unbiased sampling
- Sampling Error, Data Collection methods

### **Module 7: Exploratory Analysis and Distributions**

- Measures of Central Tendencies
- Mean, Median and Mode
- Data Variability: Range, Quartiles, Standard Deviation
- Calculating Standard Deviation
- Z-Score/Standard Score
- Empirical Rule,
- Calculating Percentiles
- Outliers, Distributions Introduction
- Normal Distribution, Central Limit Theorem
- Histogram - Normalization
- Other Distributions: Poisson, Binomial etc.,
- Normality Testing
- Skewness , Kurtosis, Measure of Distance
- Euclidean, Manhattan and Murkowski Distance

### **Module 8: Advanced Statistics for Data Science**

- Hypothesis Testing
- Null Hypothesis, P-Value
- Need for Hypothesis Testing in Business
- Two tailed, Left tailed & Right tailed test
- Hypothesis Testing Outcomes: Type I & II errors
- Parametric vs Non-Parametric Testing
- Parametric Tests, T - Tests: One sample, two sample, Paired
- One Way ANOVA
- Importance of Parametric Test
- Non-Parametric Tests: Chi-Square, Mann-Whitney, Kruskal-Wallis etc.,
- Which Test to Choose?
- Asserting accuracy of Data
- Correlation & Regression

### **Module 9: Machine Learning (ML) Fundamentals**

- Machine Learning Introduction
- Traditional Programming vs ML
- AI vs ML vs DL
- Applications of Machine Learning
- Machine Learning vs Deep Learning vs Artificial Intelligence
- Languages and platforms
- Machining learning Tools
- Exploratory Data Analysis
- Train-Test-Split
- Fixing missing Values
- Label and One-Hot Encoding
- Evaluation Metrics
- Over-fitting & Underfitting

### **Module 10: ML Regression and Classification Algorithms**

- Linear Regression
- Logistic Regression
- SVM Classifiers
- K – Nearest Neighbors
- K-Means Clustering
- Principle Component Analysis

### **Module 11: ML Advanced Algorithms & Techniques**

- Multiple Linear & Logistic Regression, Polynomial Regression
- Trees & Forest Classifiers
- XG Boosting
- Model Selection & Cross-Bias Variance
- K-Fold Cross Validation
- Natural Language Processing

### **Module 12: Deep Learning and Neural Networks**

- Deep Learning Fundamentals
- Working of Neural Networks
- Gradient Descent and Back Propagation
- TensorFlow Overview
- Building Artificial Neural Networks (ANN) with Python
- Crafting Advanced ANNs
- OpenCV Overview
- Building Convolutional Neural Networks for Image Classification

### **Module 13: Applications of ML in Bigdata and Cloud Computing**

- Introduction to Tableau and building Dashboards
- Introduction to Hadoop and Spark