

Data Science with Python

Prepared by Mr. Biswapratap Chatterjee

OVERVIEW & PURPOSE

This short course is especially targeted for college graduates who do not hold a higher degree like Mtech, Ms, MSc or PhD in the field of Mathematics or Data Science. The instructor is a Btech in Computer Science but that didn't stop him from pursuing the dream to become a Data Scientist. The journey started 7 years ago and after a long struggle today the instructor has established himself as a prominent Data Scientist in elite organizations. The instructor clearly understands the challenges faced by someone not coming from a strong programming or math background and hence, has curated this course particularly for such students. The AI industry today either has excellent programmers or excellent theorists but not both. This leaves a huge opportunity for people like us as we can understand and deliver quality output using both programming and theory. This course is particularly designed for this purpose. It can serve as the golden gate from a general software developer/student to a full-fledged Data Scientist.

OBJECTIVES

1. Trigger the coder or programmer brain inside you.
2. Take away your statistics and math phobia associated with AI.
3. Turn you into an industry ready Data Scientist for a job or entrepreneurship.

PREREQUISITES

1. Basic idea about making logical decisions.
2. Basic familiarity about Calculus, Probability and Statistics.

SYLLABUS

- Basic - Total 8 months
 - Python - 1 month
 - Variable Types

- Operators
- Decision Making
- Loops
- Numbers
- Strings
- List
- Tuple
- Dictionary
- Date and Time
- Functions
- Modules
- Files I/O
- Exception Handling
- Python - OOPs Concepts
- Regular Expressions
- Python-MySQL Database Access
- Multithreaded and Multiprocessing
- Numpy - 1 month
 - Numpy - Array creation
 - Numpy - Indexing and Slicing
 - Numpy - Advanced Slicing
 - Numpy - Broadcasting
 - Numpy - Iteration
 - Numpy - Array Manipulations
 - Numpy - Binary Operators
 - Numpy - String Functions
 - Numpy - Mathematical Functions
 - Numpy - Arithmetic Operations
 - Numpy - Statistical Functions
 - Numpy – Sort, Search & Counting Functions
 - Numpy - Matrix Library
 - Numpy - Linear Algebra
- Pandas - 1 month
 - Python Pandas - Series
 - Python Pandas - DataFrame
 - Python Pandas - Statistics
 - Python Pandas - Function Application

- Python Pandas - Indexing, Slicing, Iteration, Sorting
- Python Pandas - Working with Text Data
- Python Pandas - Aggregations
- Python Pandas - Missing Data
- Python Pandas - GroupBy
- Python Pandas - Merging/Joining
- Python Pandas - Concatenation
- Python Pandas - Date Functionality
- Python Pandas - Timedelta
- Python Pandas - Categorical Data
- Python Pandas - Visualization
- Python Pandas - IO Tools
- Python Pandas - Sparse Data
- Python Pandas - Caveats & Gotchas
- Classical Machine Learning - 2.5 months
 - A Gentle Introduction To Machine Learning.
 - Cross Validation
 - The Confusion Matrix
 - Sensitivity and Specificity
 - Bias and Variance
 - Entropy
 - Linear Regression
 - Multiple Regression
 - T-Tests and Anova
 - ROU and AUC
 - Odds and Log(Odds)
 - Odds Ratios and Log(Odds Ratios)
 - Logistic Regression
 - Maximum Likelihood
 - R-Squared and P-Value
 - Ridge and Lasso Regularization
 - Principal Component Analysis
 - Linear Discriminant Analysis
 - t-SNE
 - K-Means
 - K-Nearest Neighbors
 - Naive Bayes

- Gaussian Naive Bayes
- Regression Trees
- Classification Trees
- Random Forest
- Gradient Descent, Chain Rule
- AdaBoost
- Gradient Boost
- Support Vector Machines
- XGBoost
- Deep Learning - 2.5 months
 - Neural Networks
 - Backpropagation
 - Activation Functions
 - Loss/Cost Functions
 - Tensors for Neural Networks
 - Deep Learning with PyTorch
- Advanced (Optional) - Total 8 months
 - Natural Language Processing - 1.5 months
 - Computer Vision - 2.5 months
 - Transformers - 1 month
 - Generative Networks (GANs) - 1.5 months
 - Reinforcement Learning - 1.5 months