Analytics Educator

Bansdroni, Kolkata
Mb: 9163223228
www.analyticseducator.com
analyticseducator@gmail.com

Machine Learning with Python

(33 hours class room + 30 hours of practice sessions)

About the Course

Every day, around the United States, more than 36,000 weather forecasts are calculated. They gather all 36,000 forecasts, put them in a database, and compare them to the actual conditions encountered in that location on that day. All that collection, analysis, and reporting take a lot of heavy analytical horsepower and it is done with one programming language: Python. Over 40% of all data scientists use Python in their day to day work. Python has long been known as a simple programming language to pick up, which has propelled it to be the most preferred tool for a Data Scientist. In this course you will learn how to use the power of Python to analyze data, create beautiful visualizations, and use powerful machine learning algorithms to formulate business strategies.

Overview of the course

Introduction to Python Programming Language
Python—The Programming Language
Installing Python
Anaconda
Spyder
Jupyter notebook
IDLE (Integrated DeveLopment Environment)
Implement the Code Using an IDE
Interact with Python
Writing Python Code
Make Calculations
Import New Libraries and Functions
Import additional libraries using pip install
Import msgpack to satisfy basic requirement
NumPy
Pandas
matplotlib

Bansdroni, Kolkata
Mb: 9163223228
www.analyticseducator.com
analyticseducator@gmail.com

The Pandas & NumPy Library

Pandas Data Structures
Introduction
Creating Your Own Data
Types of Data
The dtype Option
The Series
The list
The tupple
Difference between list & tupple
The DataFrame
Making Changes to Series and DataFrames
Exporting and Importing Data
CSV
Excel
Jason
Aggregate Functions
Indexing, Slicing, and Iterating
Indexing
Slicing
Iterating an Array
Conditions and Boolean Arrays
Shape Manipulation
The Index Objects
Other Functionalities on Indexes
Reindexing
Dropping
Arithmetic and Data Alignment
Operations between Data Structures
Flexible Arithmetic Methods
Operations between DataFrame and Series
Function Application and Mapping
Functions by Element
Functions by Row or Column
Statistics Functions
Sorting and Ranking
“Not a Number” Data
Assigning a NaN Value
Working With Missing Data
Filtering Out NaN Values
Filling in NaN Occurrences
Hierarchical Indexing and Leveling
Reordering and Sorting Levels
Summary Statistic by Level

Bansdroni, Kolkata
Mb: 9163223228
www.analyticseducator.com
analyticseducator@gmail.com

Pandas in Depth: Data Manipulation

Data Preparation
Merging
Merging Multiple Data Sets
Concatenating
Combining
Pivoting
Removing
Data Transformation
Tidy Data
Removing Duplicates
Mapping
Discretization and Binning
Detecting and Filtering Outliers
Permutation
String Manipulation
More String Methods
Built-in Methods for Manipulation of Strings
Regular Expressions
Data Aggregation
Group By
A Practical Example
Hierarchical Grouping
Group Iteration
Chain of Transformations
Functions on Groups
Advanced Data Aggregation

Groupby Operations

Introduction
Aggregate
Transform
Filter
The pandas.core.groupby .DataFrameGroupBy Object
Working With a MultiIndex

Bansdroni, Kolkata
Mb: 9163223228
www.analyticseducator.com
analyticseducator@gmail.com

The datetime Data Type
Introduction
Python's datetime Object
Converting to datetime
Loading Data That Include Dates
Extracting Date Components
Date Calculations and Timedeltas
Datetime Methods
Subsetting Data Based on Dates
Date Ranges
Shifting Values
Resampling

Data Visualization with matplotlib
The matplotlib Library
Installation
matplotlib Architecture
Backend Layer
Artist Layer
Scripting Layer (pyplot)
pyplot
Line chart
Scatter plot
Annotations: Add Text
Annotations: Properties
A Simple Interactive Chart
Set the Properties of the Plot
Working with Multiple Figures and Axes
Adding Further Elements to the Chart
Adding Text
Adding a Legend
Legends: Properties
Saving Your Charts
Saving the Code
Saving Your Chart Directly as an Image
Line Chart
Line Charts with pandas
Histogram

Bansdroni, Kolkata
Mb: 9163223228
www.analyticseducator.com
analyticseducator@gmail.com

Finding Your Center
Means: The Lure of Averages
The Average in Python: mean()
Medians: Caught in the Middle
The Median in Python: median()
Statistics à la Mode
The Mode in Python
Deviating from the Average
Measuring Variation
Back to the Roots: Standard Deviation
Standard Deviation in Python
Conditions, Conditions, Conditions …
Meeting Standards and Standings
Catching Some Z's

Summarizing It All
How Many?
The High and the Low
Living in the Moments
Tuning in the Frequency
Summarizing a Data Frame

What's Normal?
Hitting the Curve
Working with Normal Distributions
A Distinguished Member of the Family
Drawing Conclusions from Data

The Confidence Game: Estimation
Understanding Sampling Distributions
An EXTREMELY Important Idea: The Central Limit Theorem
Confidence: It Has Its Limits!
Fit to a t

Bansdroni, Kolkata
Mb: 9163223228
www.analyticseducator.com
analyticseducator@gmail.com

One-Sample Hypothesis Testing
Hypotheses, Tests, and Errors
Hypothesis Tests and Sampling Distributions
Catching Some Z's Again
Z Testing in Python
t for One
t Testing in Python
Working with t-Distributions
Visualizing t-Distributions
Testing a Variance
Working with Chi-Square Distributions
Visualizing Chi-Square Distributions

Two-Sample Hypothesis Testing
Hypotheses Built for Two
Sampling Distributions Revisited
t for Two
Like Peas in a Pod: Equal Variances
t-Testing in Python
A Matched Set: Hypothesis Testing for Paired Samples
Paired Sample t-testing in Python
Testing Two Variances

Testing More than Two Samples
Testing More Than Two
ANOVA in Python
Another Kind of Hypothesis, Another Kind of Test
Getting Trendy
Trend Analysis in Python

More Complicated Testing
Cracking the Combinations
Two-Way ANOVA in Python
Two Kinds of Variables … at Once
After the Analysis

Introducing Machine Learning
Uses and abuses of machine learning
Machine learning successes
How machines learn
Machine learning in practice
Machine learning with Python

Bansdroni, Kolkata
Mb: 9163223228
www.analyticseducator.com
analyticseducator@gmail.com

Forecasting Numeric Data – Regression Methods

Understanding regression
Simple linear regression
Ordinary least squares estimation
Multiple Linear Regression
Regression: What a Line!
Linear Regression in Python
Juggling Many Relationships at Once: Multiple Regression
exploring and preparing the data
ANOVA: Another Look
Formulae and Linear Models
Model Building
training a model on the data
evaluating model performance
improving model performance
Goodness of Fit with Data—The Perils of Overfitting
Root-Mean-Square Error
Model Simplicity and Goodness of Fit
Assumption checking
Assumption checking using packages
Case studies of Linear Regression
Estimation the quality of wines
Price prediction of real estate
Movie popularity prediction
Retail sales prediction

When the Response Falls into Two Categories – Logistic Regression

Understanding logistic regression
The logit model
Generalized Linear Model
Simple logistic regression
Multiple logistic regression
Customer satisfaction analysis with the multiple logistic regression
Multiple logistic regression with categorical data
The Dataset and the Data Dictionary
Data Import in Python
EDD in Python
Outlier Treatment in Python
Missing Value treatment in Python
Variable transformation and Deletion in Python
Dummy variable creation in Python
Automatic dummy variable creation

Formulae and Logistic Models
Model Building
training a model on the data
evaluating model performance
improving model performance
Goodness of Fit with Data—The Perils of Overfitting
Confusion Matrix
Creating Confusion Matrix in Python

Time series models with Python

Introduction to Time Series Data
Notation for Time Series Data
Peculiarities of Time Series Data
Setting the Frequency
Treatment of missing values
White Noise
Stationarity
Seasonality
Correlation Between Past and Present Values
The Autocorrelation Function (ACF)
The Partial Autocorrelation Function (PACF)
Picking the Correct Model
The Autoregressive (AR) Model
ARMA
ARIMA
Automatic ARIMA

Cluster Analysis

Unsupervised Learning & Clustering: theory
K-Means Clustering: Theory
Example K-Means Clustering in Python
Visualize K-Means Results in Python
Model-based Unsupervised Clustering in Python
How to assess a Clustering Tendency of the dataset
Selecting the number of clusters for unsupervised Clustering methods (K-Means)
Assessing the performance of unsupervised learning (clustering) algorithms
How to compare the performance of different unsupervised clustering algoritms?

Bansdroni, Kolkata
Mb: 9163223228
www.analyticseducator.com
analyticseducator@gmail.com

Decision Tree & Random Forest

A Simple Tree Model
Deciding How to Split Trees
The stopping criteria for controlling tree growth
Tree Entropy and Information Gain
Pros and Cons of Decision Trees
Tree Overfitting
Pruning Trees
Decision Trees for Classification
Conditional Inference Trees
Conditional Inference Tree Classification
Building a decision tree in Python
Model Validation
Model Improvement
Model Interpretation
Ensemble technique
Random Forest Classification
Splitting Data into Test and Train Set in Python
Choose the number of trees
Model Validation
Model Improvement
Model Interpretation
Accuracy of the model
Decision Vs Random Forest

Important points:

1. After each class, assignments will be given as homework which are needed to be completed before the next class. The first 15 minutes of every class will be reserved to answer the participant's queries.
2. After every session, the discussed codes, presentations, handouts will be emailed to all the participants. Participants are advised to carry it either in soft copy or as print outs in the class.
3. Participants are advised to bring their own computers so that they can practice the codes along with the instructor.
4. Normally the class duration would be 3 hours, with a break of maximum 5-10 minutes depending of the requirement of the participants. In case all the queries of the participants are not answered with in the stipulated time of 3 hours then the instructor will extend the class by 15 minutes to 30 minutes.
5. After the completion of the module, there will be an option for all the participants to work on other case studies on real life data for further practice. (This is optional and will not be considered for calculating your final grade)
6. If a participant feels that he/she requires further help on certain topic, then they can attend the same session of some other batch.