

Python for data science and ecommerce

FA-0502

Software Development · Data & Analytics

Introduction

This course has been designed to cater to the needs of learners who are comfortable with python and picking up a specialization of python as an additional concentrated skill. This course will also delve into the various usages of python which and its functions in terms of e-commerce. This course aims to alleviate the unnecessary jargon and cover the parts that are indeed needed.

This course shall also enable the user to have the necessary foundation to take their training to a more advanced level should they wish to do so.

This course is designed for those who are experienced in programming with python. For absolute beginners, a Python beginners course is recommended prior to attempting this course.

Learning outcome:

1. Overview of python basics.
2. IDEs of python.
3. Understand the basic syntax structures.
4. Get updated on recent syntax changes in python 3.8 and 3.9 respectively.
5. Advanced Data handling in python.
6. Working with external libraries in python.
7. Working with files and encoding.
8. Understand different forms of data acquisition.
9. Data Extraction and conversion.
10. Learn statistical analytics of data.
11. Use Python to get the basics of Data Analytics.
12. Use simple scripting to present data.

13. Have the ability to visualize data and manipulate them using simple python.
14. Understand how data science delves into Machine Learning.
15. Machine learning and coding in python using Scikit Learn
16. Machine learning and coding in python using XGBoost
17. Using machine learning for predictions and analytics.
18. Data Acquisition and scraping.
19. Spark
20. Applications of machine learning using Spark

Training outline

1. Day 1: Statistics and Probability and tools to use for Data Analytics

2. Python Revision

3. A Crash Course in matplotlib.

4. Advanced Visualization with Seaborn.

5. Covariance and Correlation.

6. Conditional Probability.

7. Bayes' Theorem.

8. Day 2: Real hands on Data Analytics

9. Basic Machine Learning (using Python).

10. Supervised vs. Unsupervised Learning, and Train/Test.

11. Using Train/Test to Prevent Overfitting a Polynomial Regression.

12. Bayesian Methods: Concepts.

13. Practice with:

14. Implementing a Spam Classifier with Naive Bayes.

15. K-Means Clustering.

16. Entropy

17. Decision Trees

18. Assessment,

19. Day 3: Dealing with Real-World Data

20. Bias/Variance Tradeoff

21. K-Fold Cross-Validation to avoid overfitting

22. Data Cleaning and Normalization

23. Cleaning web log data

24. Normalizing numerical data

25. Detecting outliers

26. Feature Engineering

27. Imputation Techniques for Missing Data

28. Handling Unbalanced Data: Oversampling,

29. Undersampling, and SMOTE

30. Binning, Transforming, Encoding, Scaling, and Shuffling

31. Day 4: Apache Spark: Machine Learning on Big Data

32. Installing Spark

33. Spark Introduction

34. Spark and the Resilient Distributed Dataset (RDD)

35. Introducing MLlib

36. Introduction to Decision Trees in Spark

37. K-Means Clustering in Spark

38. TF / IDF

39. Searching Wikipedia with Spark

40. Using the Spark 2.0 DataFrame API for MLlib

41. Day 5: E-Commerce and ML in the Real World

42. Deploying Models to Real-Time Systems

43. Testing Concepts

44. T-Tests and P-Values

45. Hands-on With T-Tests

46. Determining How Long to Run an Experiment

47. A/B Test Gotchas

© 2026 Faysal Aziz · Kuala Lumpur

<https://faysalaziz.com/courses/python-for-data-science-and-ecommerce/>