Python for Data Analytics and Machine Learning

Duration: 40 Hours (5 Days)

Course Overview

The Python for Data Analytics and Machine Learning course is designed to equip learners with the essential skills required to analyze data and build machine learning models using Python. This comprehensive program starts with a Course Introduction to set expectations and outline the learning journey. It progresses through a series of modules beginning with a Jupyter Overview, which is an essential tool for data science. Learners will get a solid foundation in Python before diving into specialized libraries like Numpy and Pandas for data analysis. Visualization techniques are covered extensively with libraries such as Matplotlib, Seaborn, Plotly, and more. As the course advances, it delves into machine learning topics, starting with an Introduction to Machine Learning and then exploring various algorithms and methods like Linear Regression, Logistic Regression, K Nearest Neighbors, and others. The course includes practical Data Capstone Projects for hands-on experience, and it concludes with cutting-edge topics such as Neural Nets and Deep Learning and Big Data and Spark with Python. This course is ideal for those looking to enhance their data analytics and machine learning capabilities, providing them with the knowledge and tools to tackle real-world data challenges.

Audience Profile

Koenig Solutions' Python for Data Analytics and Machine Learning course is designed for aspiring data professionals seeking practical Python skills.

- Data Analysts looking to enhance their Python proficiency
- Data Scientists aiming to streamline their workflows with Python
- Machine Learning Engineers building predictive models
- Software Developers interested in transitioning into data science
- Business Analysts leveraging Python for data analytics and insights
- Graduates pursuing careers in data-driven fields
- IT Professionals seeking to expand into data-centric roles
- Research Scientists applying machine learning techniques to their work
- Statisticians enhancing their analytical tools with Python
- Product Managers making informed, data-driven decisions
- Entrepreneurs harnessing data insights for business intelligence
- Marketing Professionals interpreting customer data and identifying trends

Course Syllabus

Module 01: Python Basics

- Introduction to Google Colaboratory
- Introduction to Python
- Variables
- Data Types: Lists, Tuples, Dictionaries, Sets, Strings
- Operators, Keywords, and Functions
- Loops and Conditional Statements: if, elif, while loop, for loop

Demo/Lab:

- Creating a Colab notebook
- Working with variables, operators, and predefined functions
- Creating custom functions
- Implementing loops and conditional statements

Module 02: Python for Data Analysis

- What is Data Analysis?
- NumPy Module Functions
- Pandas Module Functions

Demo/Lab:

- Performing data analysis with NumPy and Pandas functions

Module 03: Python for Data Cleaning and Data Preprocessing

- Handling Null Values
- Treating Outliers
- Encoding Techniques: Label Encoding and One-Hot Encoding
- Data Normalization: Min-Max Scaling and Standard Scaling

Demo/Lab:

- Hands-on data cleaning and preprocessing

Module 04: Exploratory Data Analysis

- Data Visualization Tools:
- Matplotlib
- Seaborn
- Plotly
- Choropleth Maps
- Pandas Visualization

Demo/Lab:

- Creating visualizations such as line plots, bar plots, histograms, pie charts, box plots, scatter plots, heatmaps, and 3D plots

Module 05: Machine Learning

- Introduction to Machine Learning
- Machine Learning Lifecycle
- Types of Machine Learning: Supervised and Unsupervised
- Supervised Learning: Regression and Classification
- Unsupervised Learning: Clustering, Natural Language Processing, and Recommendation Systems

Module 06: Regression Machine Learning

- Linear Regression
- Evaluation Metrics

- Linear Regression OLS Model
- Multicollinearity and Variance Inflation Factor (VIF)
- P-value, T-test, and F-test
- Lasso and Ridge Regression
- Recursive Feature Elimination
- KNN Regression
- Polynomial Regression
- Bias-Variance Tradeoff

Demo/Lab:

- Hands-on implementation of regression techniques

Module 07: Supervised Machine Learning - Classification

- Logistic Regression
- Evaluation Metrics
- Handling Imbalanced Datasets
- KNN Classification
- Decision Trees
- Ensemble Models
- Support Vector Classifier
- Naive Bayes Classifier
- Hyperparameter Tuning

Demo/Lab:

- Hands-on implementation of classification models
- Handling imbalanced datasets

Module 08: Unsupervised Machine Learning

- Clustering Techniques: K-Means Clustering and Hierarchical Clustering
- Principal Component Analysis (PCA)
- Recommendation Systems: Content-Based Filtering and Collaborative Filtering
- Natural Language Processing (NLP)

Demo/Lab:

- Implementing clustering and PCA
- Creating a recommender system
- Hands-on implementation of NLP