

KOENIG SOLUTIONS · COURSE OUTLINE

Fundamentals of Artificial Intelligence (AI) and Machine Learning (ML) (Python Institute)

Learn to build, evaluate and deploy machine learning models in Python, from statistics and regression to Spark, recommender systems and REST APIs.

12

MODULES

66

TOPICS

COURSE OVERVIEW

This course sets up a Python data science environment and reviews the statistics and probability concepts needed for machine learning, including distributions, covariance, correlation and Bayes' theorem. It then works through core algorithm types and predictive models such as linear, polynomial and multivariate regression, before applying machine learning techniques including train/test validation, Naive Bayes classification, K-Means clustering, recommender systems, K-nearest neighbors and dimensionality reduction with PCA. It covers handling real-world data issues such as bias/variance trade-off, cross-validation, cleaning and normalisation, and outlier detection, then extends machine learning to big data with Apache Spark and MLlib. It closes with experimental design through A/B testing and building a UI and REST API for models.

WHO THIS COURSE IS FOR

Data Analysts

Machine Learning Engineers

Python Developers

Data Scientists

Software Engineers

LEARNING OBJECTIVES

- Install and configure a Python data science environment
- Apply statistics and probability concepts to data analysis
- Build predictive models using linear and polynomial regression
- Implement classification with Naive Bayes and K-Means clustering
- Develop recommender systems using item-based collaborative filtering
- Apply K-nearest neighbors and PCA for dimensionality reduction
- Evaluate models using cross-validation and A/B testing
- Deploy machine learning models with Apache Spark and REST APIs

12 Modules, 66 Topics

01

Getting Started

TOPICS

- Installing a Python Data Science Environment
- Using and understanding iPython (Jupyter) Notebooks
- **Python basics:** Part 1
- Understanding Python code
- Importing modules
- **Python basics:** Part 2
- Running Python scripts

02

Statistics and Probability Refresher and Python Practice

TOPICS

- Types of data
- Mean, median, and mode
- Using mean, median, and mode in Python
- Standard deviation and variance
- Probability density function and probability mass function
- Types of data distributions
- Percentiles and moments

03

Matplotlib and Advanced Probability Concepts

TOPICS

A crash course in Matplotlib · Covariance and correlation · Conditional probability · Bayes' theorem

04

Algorithm Overview

TOPICS

Data Prep · Linear Algorithms · Non-Linear Algorithms · Ensembles

05

Predictive Models

TOPICS

- Linear regression
- Polynomial regression
- Multivariate regression and predicting car prices
- Multi-level models

06

Applied Machine Learning with Python

TOPICS

- Machine learning and train/test
- Using train/test to prevent overfitting of a polynomial regression
- Bayesian methods: Concepts
- Implementing a spam classifier with Naïve Bayes
- K-Means clustering

07

Recommender Systems

TOPICS

- What are recommender systems?
- Item-based collaborative filtering
- How item-based collaborative filtering works?
- Finding movie similarities
- Improving the results of movie similarities
- Making movie recommendations to people
- Improving the recommendation results

08

More Applied Machine Learning Techniques

TOPICS

- K-nearest neighbors - concepts
- Using KNN to predict a rating for a movie
- Dimensionality reduction and principal component analysis
- A PCA example with the Iris dataset
- Data warehousing overview
- Reinforcement learning

09

Dealing with Data in the Real World

TOPICS

Bias/variance trade-off · K-fold cross-validation to avoid overfitting · Data cleaning and normalization · Cleaning web log data · Normalizing numerical data · Detecting outliers

10

Apache Spark: Machine Learning on Big Data

TOPICS

- Installing Spark
- Spark introduction
- Spark and Resilient Distributed Datasets (RDD)
- Introducing MLlib
- Decision Trees in Spark with MLlib
- K-Means Clustering in Spark
- TF-IDF
- Searching wikipedia with Spark MLlib
- Using the Spark 2.0 DataFrame API for MLlib

11

Testing and Experimental Design

TOPICS

- A/B testing concepts
- T-test and p-value
- Measuring t-statistics and p-values using Python
- Determining how long to run an experiment for
- A/B test gotchas

12

GUIs and REST

TOPICS

Build a UI for your Models · Build a REST API for your Models

WHAT IS INCLUDED

Everything You Need to Succeed

- Koenig course material for every module
- Delivery by certified instructors and subject matter experts
- Hands-on exercises throughout the course
- Course completion certificate from Koenig

ABOUT KOENIG

Who We Are

Established in 1993, Koenig Solutions is a globally recognised training provider with 35+ technology partnerships, 5,000+ training programmes and a network of certified instructors and subject matter experts, supporting individuals and enterprise teams worldwide.

FLEXIBLE DELIVERY

Learn the Way That Suits You

Classroom Training

Live Online Classes (ILO)

Fly-Me-A-Trainer (FMAT)

Flexi (Self-Paced)

1-on-1 Training

Ready to Get Started?

Talk to a learning advisor about schedules, group training and customised delivery.

Visit koenig-solutions.com

✉ info@koenig-solutions.com