INTRODUCTION TO SCIKIT-LEARN

MACHINE LEARNING IN PYTHON

with Corey Wade, Director & Founder of Berkeley Coding Academy

OVERVIEW

- Prerequisites - Python
- Requirements - Jupyter Notebook (anaconda.com)
- Code & Links - github.com/coreyjwade/odsc
- Introductory survey - forms.gle/d5Zc7XijanX5NuTS6
- Libraries - pandas, numpy, sklearn, xgboost

WORKSHOP INCLUDES

- Intro to Machine Learning in Python
- Focus on Tabular Data (rows & columns of mostly numeric text)
- A brief introduction to pandas
- Many models in sklearn
- Fit models, score models, make predictions, optimize parameters
- Coding tips + bonus features like feature_importances_
- 2 contests + opportunities to practice on your own data

WORKSHOP DOES NOT INCLUDE

- Unstructured datasets (with images and text)
- Neural networks
- Unsupervised learning

MEET YOUR INSTRUCTOR

Corey Wade

- Director & Founder of Berkeley Coding Academy
- Author of 2 Python Machine Learning Books
- Math/Programming Teacher at Berkeley Independent Study

SUPPORT STAFF

Mary Orozco

- Berkeley Coding Academy lead instructor
- Completing Data Science degree from UCB
- Competitive youth soccer coach in Bay Area

Yash Sharma

- Berkeley Coding Academy alumni
- Silicon Valley Science Fair Winner - Machine learning drone project
- Still in high school!

MODULES

- Introductory comments 10 min + 5 min code set-up
- Module 1 - Preparing Data for ML with pandas 15 min
- Module 2 - Supervised learning with sklearn 30 min + 15 min lab
- Module 3 - Cross-validation with sklearn 15 min + 15 min break
- XGBoost slides 10 min
- Module 4 - Fine-tuning models with sklearn 30 min + 20 min lab
- Module 5 - Most influential columns with sklearn 15 min
- Closing comments 10 min

GETTING STARTED

- If you have never used Jupyter Notebook, download Anaconda now
- Download files from GitHub - github.com/coreyjwade/odsc
- Open Jupyter Notebook **sklearn_intro_starter_code**
- Complete introductory survey - forms.gle/d5Zc7XijanX5NuTS6