Machine Learning made easy with PyCaret

Moez Ali

October 5, 2022
This presentation and demo is available on:

https://www.github.com/pycaret/pycaret-gotocph
About me

- **Background**: Finance + Economics + Data Science
- **Industry**: Healthcare + Education + Consulting + Fintech + Retail
- **Current Work**: AI Product Strategy Director @ Antuit.AI
- **Academic / Research**: Queens University, Cornell University, University of Regina
- **Open-source**: PyCaret
By the end of this session:

• What is PyCaret

• Live demo: Use-case implementation in Python
AI and Data Landscape - 2022

Source: Linux Foundation
PyCaret

PyCaret is an open source, low-code machine learning library for building and deploying machine learning pipelines in Python. It is commonly used for rapid prototyping and primarily targeted towards Citizen Data Scientists.
Phases of Machine Learning

Training Phase
- Labeled input
- Algorithm

Prediction Phase
- New, unlabeled input
- Model

MATHS!!!
Machine Learning Life Cycle

- Business Problem
- Data Sourcing & ETL
- Exploratory Data Analysis (EDA)
- Data Prep.
- Model Training & Selection
- Deployment & Monitoring
Data Prep, Model Training and Selection

Data Prep
- Train Test Split
- Missing Values Imputation
- Feature Scaling
- Encodings
- Feature Engineering

Model Training
- Train
- Final Pipeline
- Deploy
- Monitor

Model Selection
- Cross Validation Environment
  - Model Training
  - Model Tuning
  - Model Selection
  - Model Ensemble

Final Pipeline
Deploy
Monitor

Optional

Final Pipeline
Deploy
Monitor

Optional
Features:

Data Preparation

Model Training

Hyperparameter Tuning

Analysis & Interpretability

Model Selection

Experiment Logging
Use-cases supported in PyCaret

SUPERVISED
- Classification
- Regression
- Time Series

UNSUPERVISED
- Clustering
- Anomaly Detection
- Association Rule Mining
- Natural Language Processing
PyCaret by numbers

- 6.5M+ Downloads
- 6,400+ Git Stars
- 90 Contributors
- 4,000+ Commits
Integrations

- `scikit-learn`
- RAPIDS
- RAY
- mlflow
- XGBoost
- LightGBM
- CatBoost
- OPTUNA
- Yellowbrick
- plotly
- GENSIM
- spaCy
Demo
PyCaret Masterclass: Machine Learning for Business Analysts

REGISTER

Course Syllabus

- Machine Learning Lifecycle (CRISP-DM)
- Time Series Forecasting
- Supervised Machine Learning
- Automated Hyperparameter Tuning
- Clustering Analysis
- Experiment Logging and MLOps
- Classification Models
- Model Ensembling
- Regression Models
- Exploratory Data Analysis

This 8-hour full-day Masterclass on PyCaret: Machine Learning for Business Analysts specifically targets data science enthusiasts and future experts. In this fully hands-on Masterclass, you will learn several important concepts and techniques in machine learning with the help of real-life business problems. You will have an opportunity to learn directly from Moez Ali – the creator of PyCaret himself. PyCaret is one of the most used and popular open-source, low-code, machine learning libraries in Python with over 5M downloads and a welcoming community. Moez is the founding developer of PyCaret and a senior data scientist who has led multimillion-dollar AI projects in Healthcare, Fintech, Retail, and
Don’t forget to rate this session in the GOTO Guide app.