Implementing scikit-learn Models in Databricks

Janani Ravi
Co-founder, Loonycorn

www.loonycorn.com

Overview

An introduction to scikit-learn

Build ML models with scikit-learn on Databricks using MLflow

An Overview of scikit-learn

scikit-learn - easy-to-use, very comprehensive and efficient Python library for traditional ML models

Traditional ML Models

Have a fundamental algorithmic structure to solve problems

The algorithm is fed data which trains the algorithm's parameters

Called model parameters

Traditional ML Models

Regression models: Linear, Lasso, Ridge, SVR

Classification models: Naive Bayes, SVMs, Decision trees, Random forests

Dimensionality Reduction: Manifold learning, factor analysis

Clustering: K-means, DBSCAN, Spectral clustering

Support for Complete ML Workflow

All common families of models supported

Data pre-processing, cleaning, feature selection, and extraction

Model validation and evaluation

Demo

Building and training a regression model in Databricks using scikit-learn and MLflow

Demo

Building and training a classification model in Databricks using scikit-learn and MLflow

Summary

An introduction to scikit-learn

Build ML models with scikit-learn on Databricks using MLflow

Up Next:
Implementing XGBoost Models in Databricks