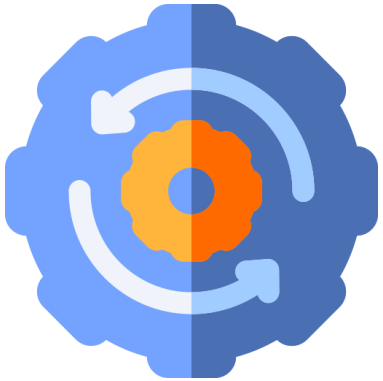
The Niche of scikit-learn in Machine Learning

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

scikit-learn



Developed as a Google summer of code project in 2007

Currently has 30+ active contributors

Sponsored by INRIA, Google, Tinyclues, and the Python Software Foundation

Attraction of scikit-learn

Easy-to-use

Comprehensive

Efficient

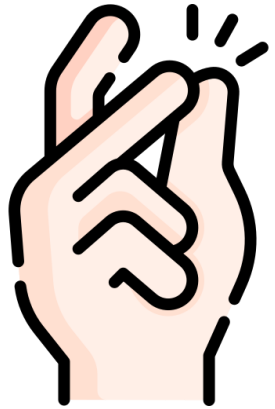
Attraction of scikit-learn

Easy-to-use

Comprehensive

Efficient

Ease of Use



- Estimator API for consistent interface
- Estimator for all kinds of models
- Create a model project
- Fit to training data
- Predict for new data
- Pipelines for complex operations

Attraction of scikit-learn

Easy-to-use

Comprehensive

Efficient

Support for Complete ML Workflow

All common families of models supported

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

Model validation and evaluation

Completeness

Regression, classification clustering, dimensionality reduction

Feature extraction and selection using statistical and dimensionality reduction

Data pre-processing

Data generation

- Swiss rolls, S-curves

Cross-validation to evaluate models

Hyperparameter tuning

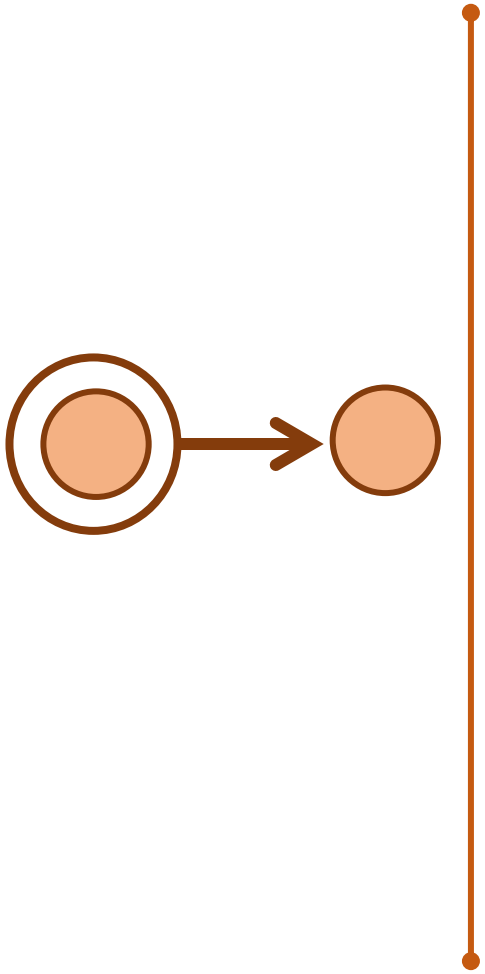
Attraction of scikit-learn

Easy-to-use

Comprehensive

Efficient

Efficiency

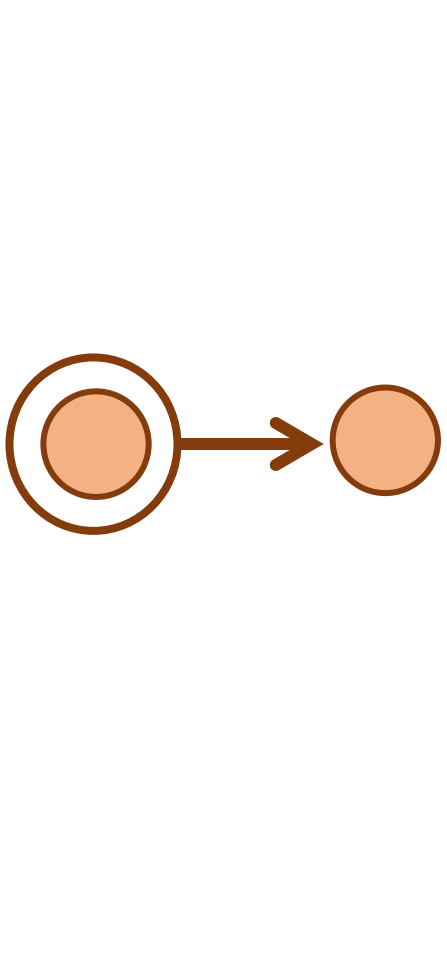


Highly optimized implementation

Built on SciPy, hence scikit prefix

Interoperates with all common Python libraries for data science

Efficiency



NumPy: Base n-dimensional array package

SciPy: Fundamental library for scientific computing

Matplotlib: Comprehensive 2D/3D plotting

Sympy: Symbolic mathematics

Pandas: Data structures and analysis

Foundational Libraries for scikit-learn

NumPy: Base n-dimensional array package

SciPy: Fundamental library for scientific computing

Matplotlib: Comprehensive 2D/3D plotting

Sympy: Symbolic mathematics

Pandas: Data structures and analysis