No easy access to

Large compute infrastructures (GPUs)
Trained ML Experts
Huge Datasets
Data science for the many, not only the mighty

We need to enable machine learning without large resources, infinite data, ivy-league education
Created @ Ínria around 2009
Top used machine learning library

Python implementation, well packaged and maintained, simple API, well documented and exemplified, efficient algorithms.
Python is born in Europe
European Machine Learning

Python + scikit-learn = European ML
We are actively looking for European partners to build the future of scikit-learn.
#7 Python project on GitHub
#117 project on GitHub

source: GitHub rankings based on stargazers
An industry standard

scikit-learn is frequently used by more than 70% of the data scientists

source: Kaggle survey 2022
(23,997 answers)
Automotive & Mobility Industry
Use case: Fleet Management

Leveraging scikit-learn's powerful histogram-based gradient boosted trees on a practical problem: **hourly forecasting the number of bike rentals** in a given city at various times of the week, month and year depending on weather conditions and business cycles.
Predictions by regression models

- Actual demand
- GBRT median
- GBRT mean (Poisson)
- GBRT 90% interval

0 20 40 60
0 200 400 600 800
Academics, Industrials, start-ups are solving mobility problems (forecasting, predictive maintenance...) using scikit-learn.
405,347 repositories & 9,847 packages rely on scikit-learn

source: GitHub
Human mobility analysis in Python. Represent trajectories and mobility flows with proper data structures. Extract human mobility metrics and patterns. Generate synthetic individual trajectories or synthetic mobility flows using standard migration models.
Understanding and simulating human mobility is of paramount importance for many present and future applications, such as traffic forecasting, urban planning, estimating migratory flows, and epidemic modeling, and hence for many actors, from urban planners to decision makers, industrial partner and mobility service provider...
It's not all about Deep Learning

When asked "Which of the following ML algorithms do you use on a regular basis?": The top 2 is about scikit-learn machine learning.

source: Kaggle survey 2022 (23,997 answers)
Which of the following ML algorithms do you use on a regular basis?

- Linear or Logistic Regression
- Decision Trees or Random Forests
- Convolutional Neural Networks
- Gradient Boosting Machines (xgboost, lightgbm, etc)
- Bayesian Approaches
- Dense Neural Networks (MLPs, etc)
- Recurrent Neural Networks
- Transformers Networks (BERT, gpt-3, etc)
- Neural Networks
- None
- Autoencoder Networks (DAE, VAE, etc)
- Generative Adversarial Networks
- Evolutionary Approaches
- Other

source: Kaggle survey 2022
(23,997 answers)
Let us move with European Machine Learning

Use a common basis for more interoperability of projects and better cooperation.
Industrial sponsors of the library

This Machine Learning is brought to you thanks to responsible actors fostering Open Source Software and Digital Commons.
Join our MOOC
Thank you for your attention

francois.goupil@inria.fr

scikit-learn.fondation-inria.fr