amadeus

MLflow and beyond

Machine Learning in real life

Telecom Valley | SophiaConf
100% OPEN SOURCE - Conférences et Workshops

Generoso PAGANO
Davide SANAPO

The Speakers

Generoso

Davide

_ Data Engineers @ Amadeus

- Big-Data products
- Data Mesh architecture
- Machine Learning dev-ops

Agenda

_ Machine Learning in real life

_ MLflow

_ Pyntel

ML is everywhere!

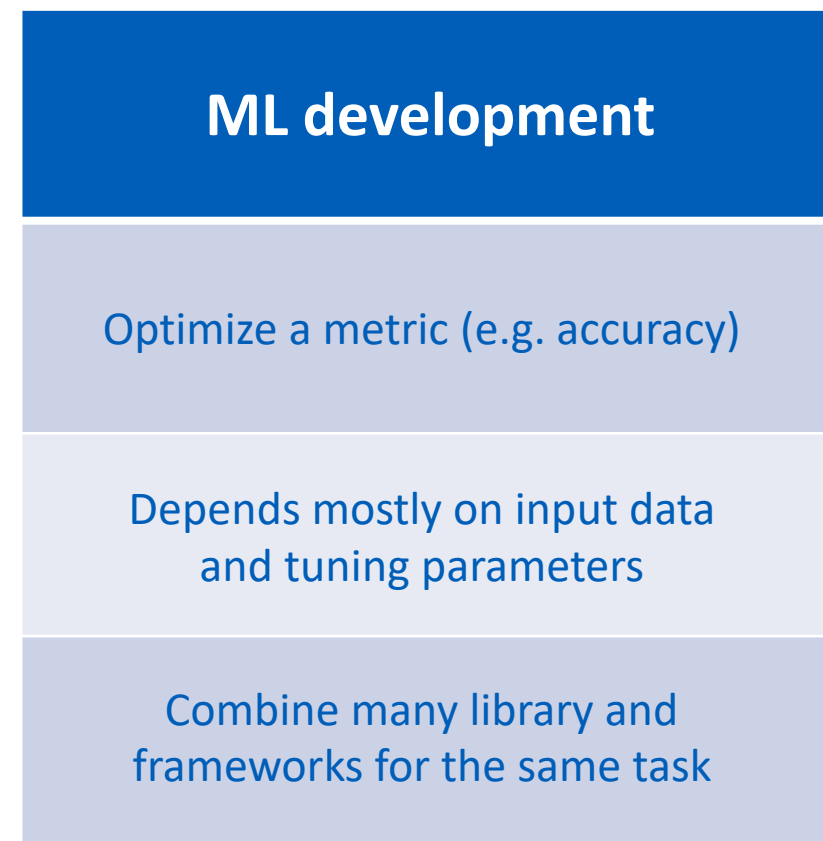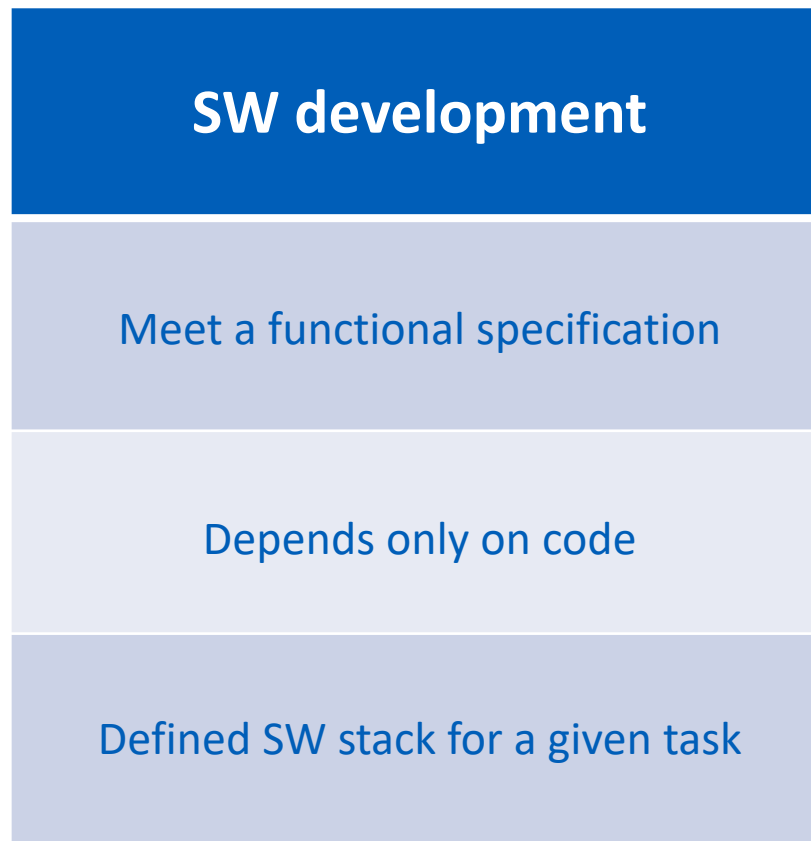_ ML is reshaping businesses and our daily life

- Price prediction
- SPAM detection
- Product recommendation
- Fraud detection
- Voice recognition
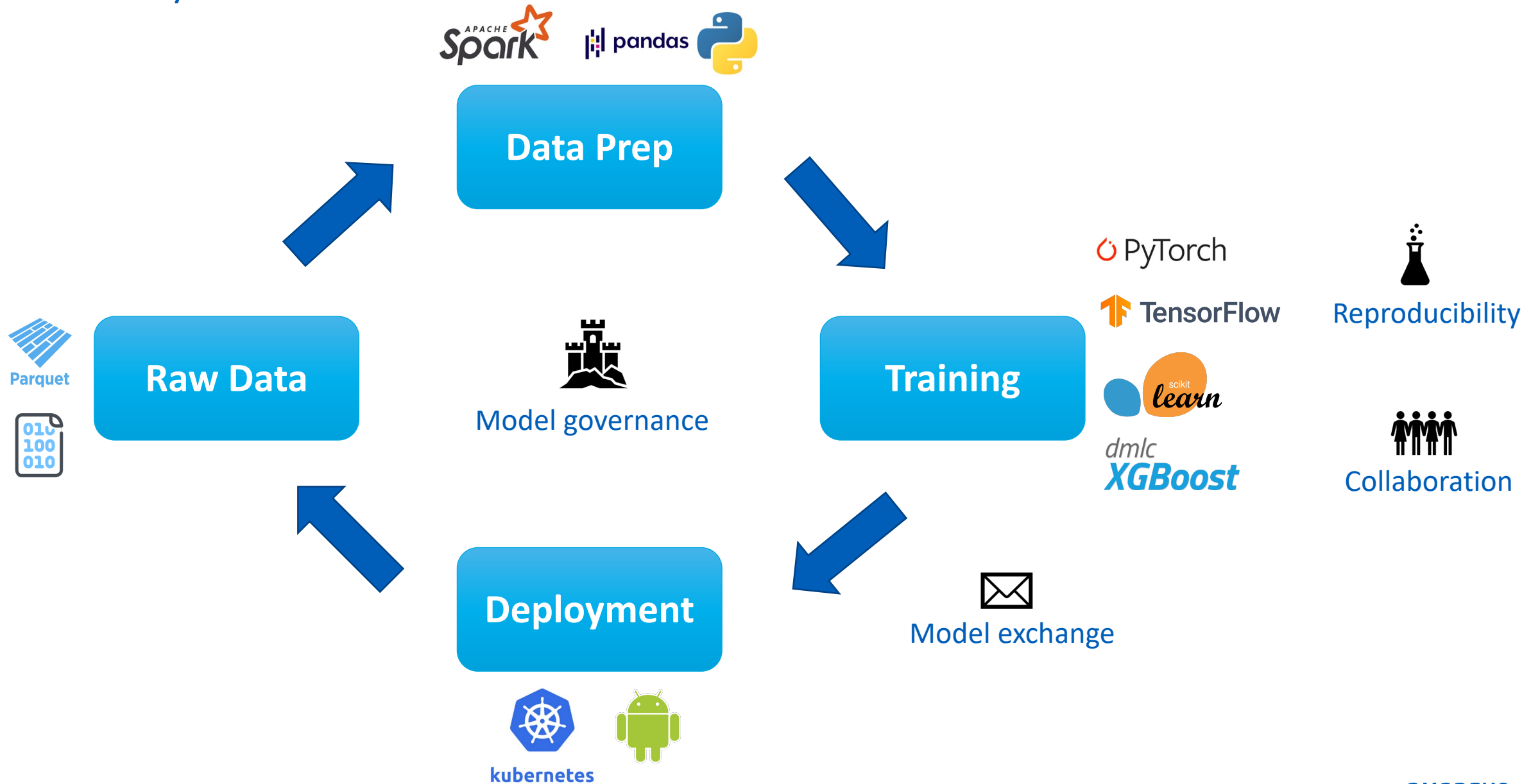- Chat bots
- Autonomous driving
- ...

" ML development is harder than traditional SW development "

Matei Zaharia

Author of Apache Spark and MLflow

	SW development	ML development
Goal	Meet a functional specification	Optimize a metric (e.g. accuracy)
Quality	Depends only on code	Depends mostly on input data and tuning parameters
Stack	Defined SW stack for a given task	Combine many library and frameworks for the same task

ML lifecycle

Data Prep

Raw Data

Training

Deployment

Model governance

Model exchange

Reproducibility

Collaboration

Parquet

Apache Spark | pandas | python

PyTorch
TensorFlow
scikit learn
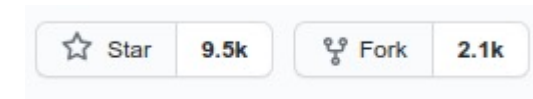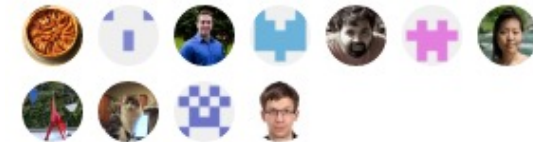dmlc XGBoost

kubernetes

ml*flow*

Open source platform for the ML lifecycle

- Works with most ML framework
- Works with any programming languages
- Scales to Big Data with Apache Spark

Some of the users…

☆ Star 9.5k   Fork 2.1k

Contributors 300

Tracking

Projects
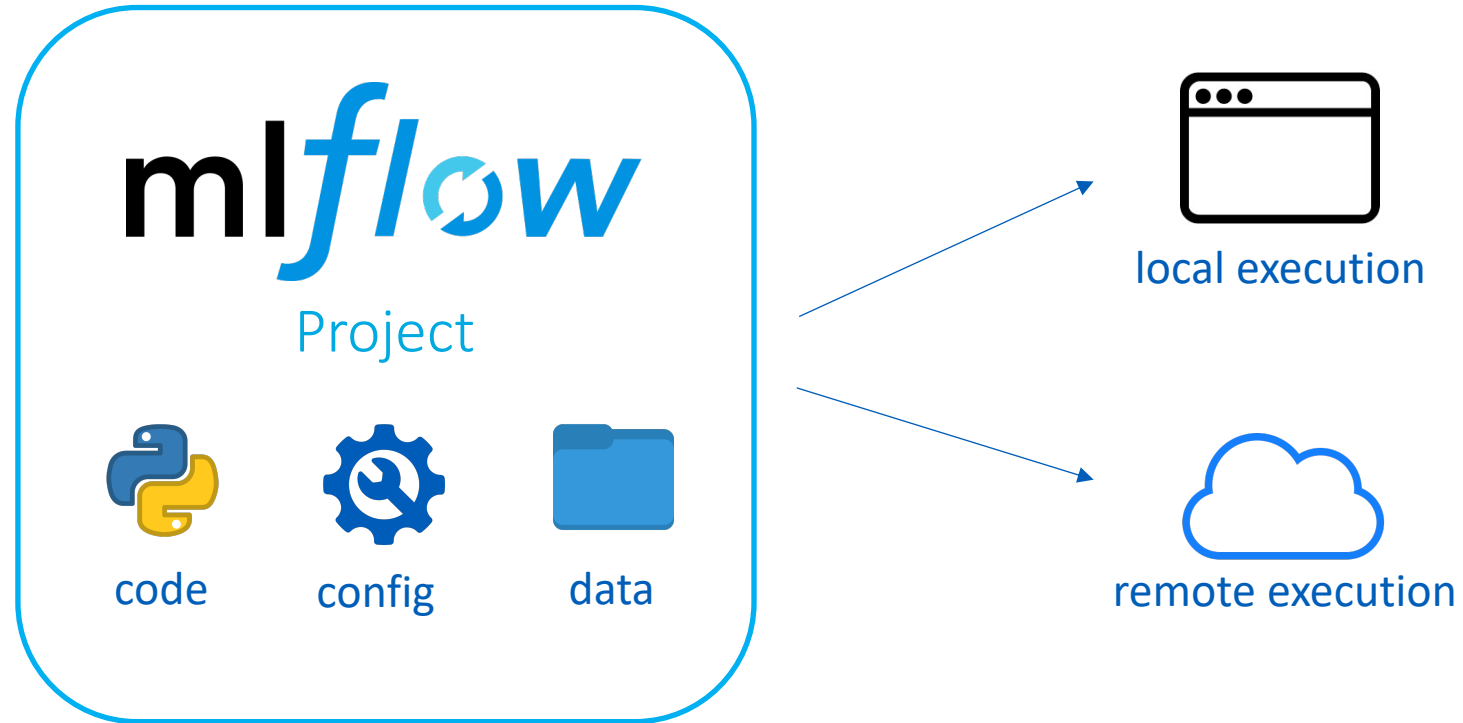
Registry

Models

MLflow
Tracking

_ API (Python, REST, R, Java) and UI for logging and querying

- parameters
- code version
- metrics

_ Analyse and compare results

MLflow
Projects

_ packaging data science code in a reusable and reproducible way
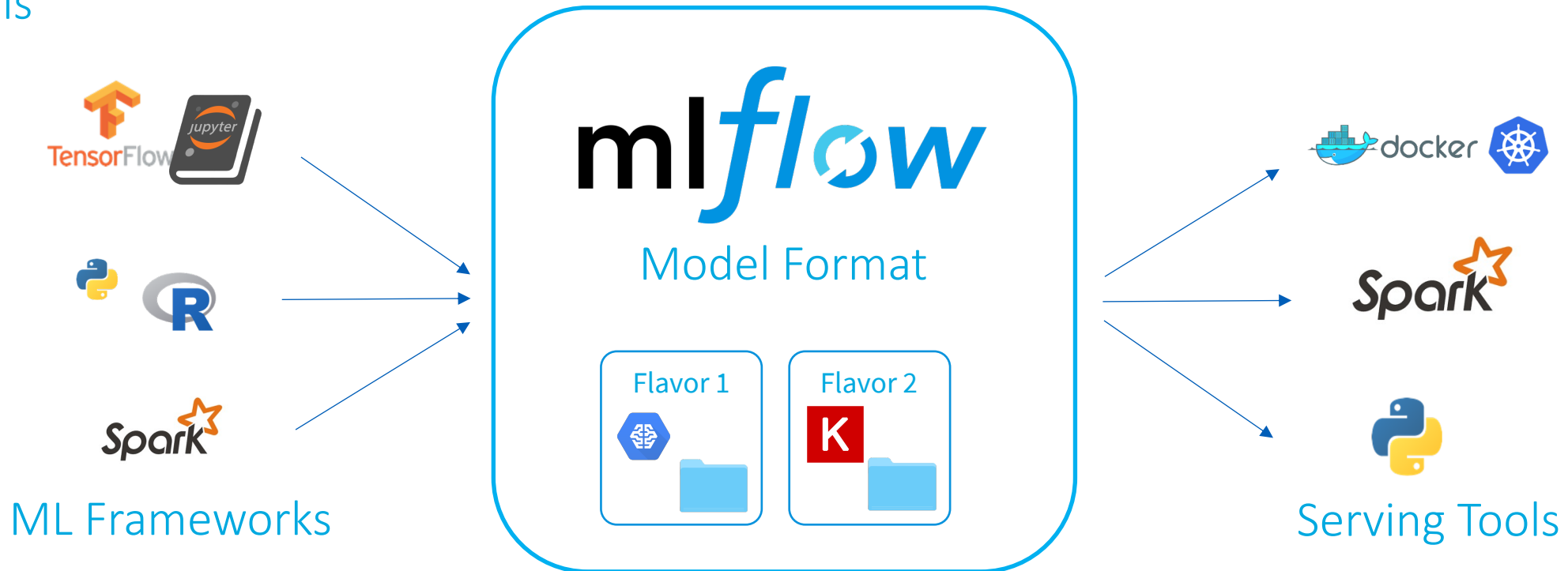
_ composing projects

MLflow
Model Registry

_ manage the full lifecycle of a model

_ model lineage

_ model versioning

_ stage transitioning

MLflow
Models

- serialization and packaging format for ML Models

- defines model dependencies

- save models from any framework in MLflow format

- deployment APIs

Mlflow Model

Limitations
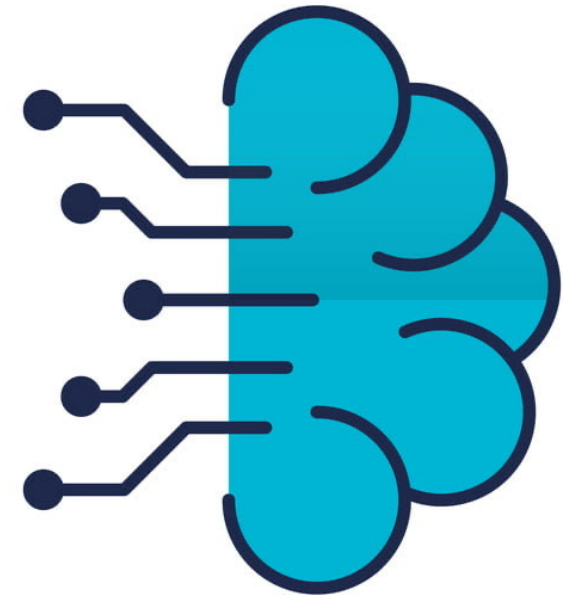
_ Model serving through REST API

- No swagger grammar

_ Open model serving

- No gRPC grammar

Pyntel
Overview

_ Python library

_ Extends MLflow Models

_ Model signature generation

  • Swagger grammar

  • gRPC schema

_ Enrich the model with cloud provider metadata

  • experiment id, subscription, resource group, runId

_ Validators

  • prediction validation

Conclusions

_ Machine Learning in production

_ Overview on Mlflow components

_ Pyntel
  • Contribution to MLflow planned
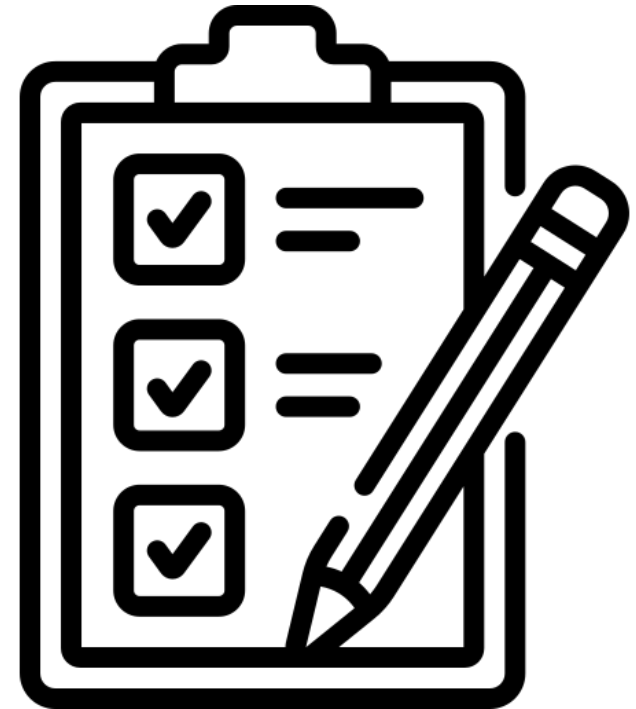
MLflow and beyond

Machine Learning in real life

Generoso PAGANO
Davide SANAPO