

MLaaS-AutoML

Conf. dr. Cristian KEVORCHIAN
Computer Science Department
University of Bucharest

Machine Learning as a Service (MLaaS)

- ▶ Machine learning as a service or MLaaS is a **set of cloud services** that machine learning providers can offer as a part of cloud computing services. The services that MLaaS offers include face recognition, application programming interface (APIs), data visualization, predictive analytics, deep learning, and natural language processing.

Machine Learning Democratization

- Developing machine learning in a traditional way is a resource-intensive activity, requiring significant knowledge in the ML field but also time to develop and compare dozens of models.
- The lifecycle of production-ready machine learning models will be shortened when ML operations are automated.

CLOUD MACHINE LEARNING SERVICES COMPARISON

	Amazon ML and SageMaker	Microsoft Azure AI Platform	Google AI Platform (Unified)	IBM Watson Machine Learning
Classification	✓	✓	✓	✓
Regression	✓	✓	✓	✓
Clustering	✓	✓	✓	✗
Anomaly detection	✓	✓	✗	✗
Recommendation	✓	✓	✓	✗
Ranking	✓	✓	✗	✗
Data Labeling	✓	✓	✓	✓
MLOps pipeline support	✓	✓	✓	✓
Built-in algorithms	✓	✓	✓	✗
Supported frameworks	TensorFlow, MXNet, Keras, Gluon, Pytorch, Caffe2, Chainer, Torch	TensorFlow, scikit-learn, PyTorch, Microsoft Cognitive Toolkit, Spark ML	TensorFlow, scikit-learn, XGBoost, Keras	TensorFlow, Keras, Spark MLlib, scikit-learn, XGBoost, PyTorch, IBM SPSS, PMML

MACHINE LEARNING ÎN AZUREML

Domain Specific Pretrained Models

To reduce time to market



Vision



Speech



Language



Search

Familiar Data Science Tools

To simplify model development



PyCharm



Jupyter



Visual Studio Code



Command line

Popular Frameworks

To build machine learning and deep learning solutions



PyTorch



TensorFlow



Scikit-Learn



ONNX

Productive Services

To empower data science and development teams



Azure
Databricks



Azure Machine Learning



Machine
Learning VMs

Powerful Hardware

To accelerate deep learning



CPU



GPU



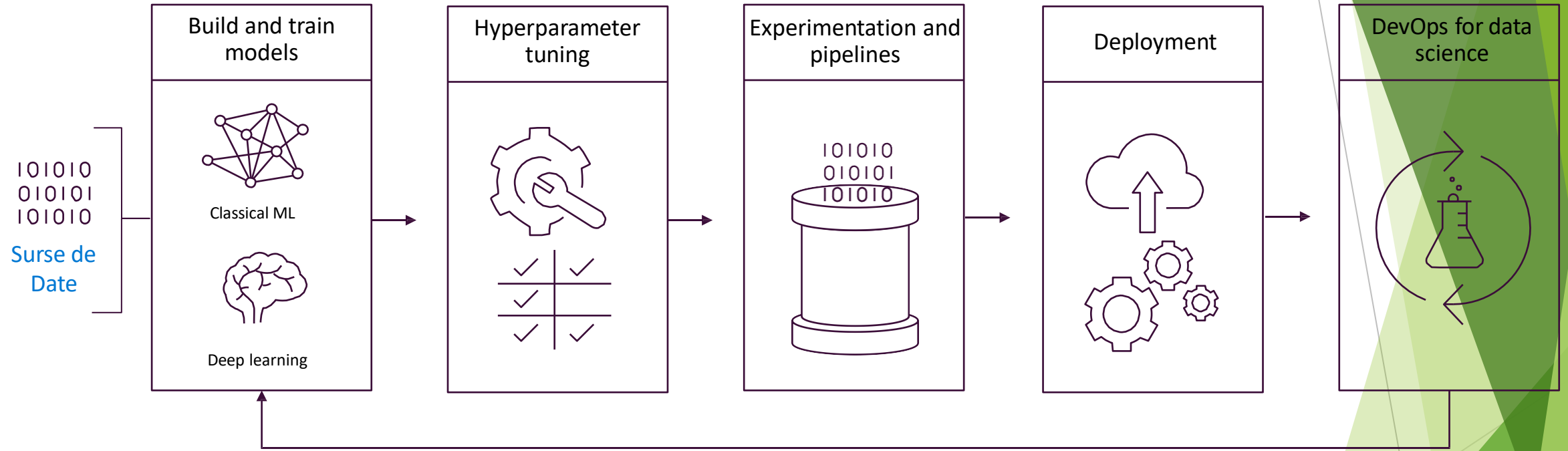
FPGA



From the Intelligent Cloud to the Intelligent Edge



DATA SCIENCE Project Lifecycle



Azure Machine Learning

- ▶ The core environment for dataset administration, model training, and deployment is Azure Machine Learning.
- ▶ Machine Learning Studio, a web-based and low-code environment for quickly configuring machine learning operations and pipelines, is included in the platform.
- ▶ In general, Azure Studio provides tools for data exploration, preprocessing, method selection, and modeling result validation.
- ▶ Around 100 methods are supported by the Studio, including classification (binary and multiclass), anomaly detection, regression, recommendation, and text analysis. The platform only offers one clustering algorithm, which is worth noticing (K-means).

AUTOML - DEFINITION

- ▶ Automated Machine Learning (AutoML) is a method for developing machine learning models that involves automating iterative and time-consuming operations.
- ▶ It enables data scientists, analysts and developers to build high-complexity, efficient and productive ML models while maintaining the quality of the model.
- ▶ AUTOML is part of Azure Machine Learning and is based on Microsoft Research.

Some similarities with standard approach

- ▶ Practitioners in a typical machine learning application have a collection of input data points to train with.
- ▶ The raw data may not be in a format that can be used by all algorithms.
- ▶ An expert may need to use proper data pre-processing, feature engineering, feature extraction, and feature selection procedures to make the data suitable for machine learning.
- ▶ Following these stages, practitioners must choose an algorithm and optimize hyperparameters to improve their model's prediction performance.
- ▶ Each of these phases might be difficult, making machine learning difficult to implement.
- ▶ For non-experts, AutoML seeks to make these tasks easier.

AutoML: Classification, Regression, Forecasting, Time Series and Computer Vision

- ▶ When the user want to train and optimize a model regarding a data science problem based it must to established a target measure, when use automated ML.
- ▶ Automated ML democratizes the machine learning model creation process, empowering users to find an end-to-end machine learning pipeline for any problem, regardless of their data science knowledge.
- ▶ Automated machine learning can be used by data scientists, analysts, and developers in a variety of industries to:
 - ▶ ML solutions can be implemented without substantial programming experience.
 - ▶ Make the most of data science best practices.
 - ▶ Provide quick problem-solving solutions.

Classification

- ▶ A common machine learning job is classification.
- ▶ Classification is a form of **supervised learning** in which models learn from training data and then apply what they've learned to fresh data.
- ▶ Deep neural network text featurizers for classification, for example, are available in Azure Machine Learning particularly for these purposes.
- ▶ Classification models' major purpose is to anticipate which categories incoming data will fall into based on what they've learned from their training data. **Fraud detection, handwriting recognition, and object detection** are all examples of categorization.

Regression

- ▶ Regression problems, like classification, are a frequent supervised learning task.
- ▶ Regression models predict numerical output values based on independent predictors, as opposed to classification, which predicts categorical output values.
- ▶ The goal of regression is to estimate how one variable affects the others in order to assist build a relationship between those independent predictor variables.
- ▶ For instance, a car's price is determined by factors such as gas mileage, safety rating, and so on.
- ▶ For certain jobs, Azure Machine Learning provides customizations.

Time-series forecasting

- ▶ A multivariate regression problem is used to analyze an automated time-series experiment.
- ▶ Time-series values from the past are "pivoted" to become additional dimensions for the regressor, along with other predictors. Unlike traditional time series approaches, this methodology has the benefit of intuitively including various contextual variables and their relationships during training.
- ▶ For all items in the dataset and prediction horizons, automated ML learns a single, but frequently internally branched model. As a result, more data is available to estimate model parameters, and generalization to previously unknown series is possible.
- ▶ Whether it's income, inventory, sales, or client demand, forecasting is an important aspect of any organization. You can combine strategies and approaches using automated machine learning to generate a recommended, high-quality time-series forecast.

Advanced forecasting configuration

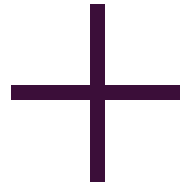
- ▶ Many models support through grouping rolling-origin cross validation configurable lags rolling window aggregate features holiday detection and featurization time-series and DNN learners (Auto-ARIMA, Prophet, ForecastTCN)
- ▶ Many models support through grouping rolling-origin cross validation configurable lags rolling window aggregate features

Computer Vision

- ▶ Support for computer vision tasks has been added to Automated ML for images (preview), allowing you to quickly construct models trained on image data for situations such as image categorization and object recognition.
- ▶ With this capability you can:
 - Seamlessly integrate with the Azure Machine Learning data labeling capability
 - Use labeled data for generating image models
 - Optimize model performance by specifying the model algorithm and tuning the hyperparameters.
 - Download or deploy the resulting model as a web service in Azure Machine Learning.
 - Operationalize at scale, leveraging Azure Machine Learning MLOps and ML Pipelines capabilities.

AZURE ML SERVICE

Azure Cloud
Services



Python
SDK



Allow us to:

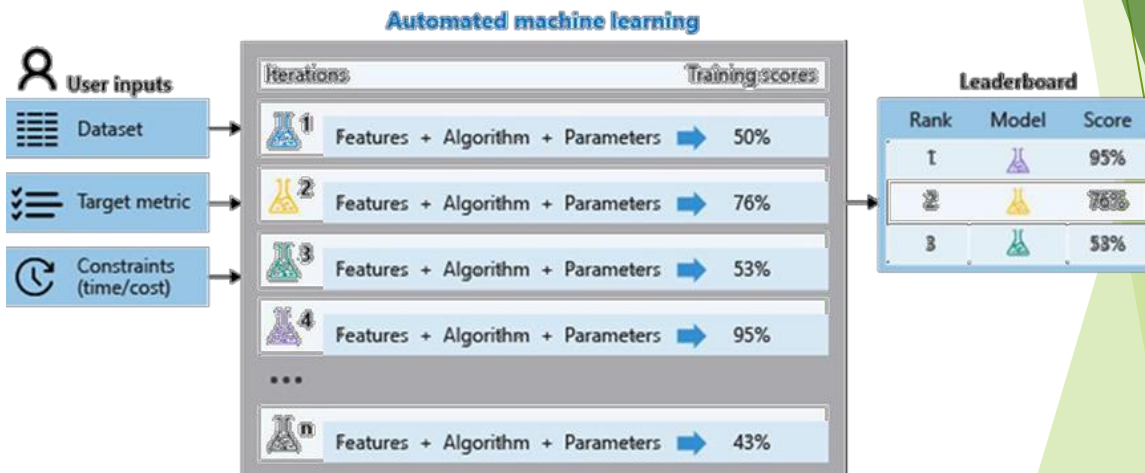
- ✓ Preparing Data Source
- ✓ Design Models
- ✓ Train Models

- ✓ Managementul Models
- ✓ Monitorizarea Experimentelor
- ✓ Deployment-ul Modelelor

ML Tasks

- **ML Task** = {data} + {problem type} + {loss function}
 - **ML project effort and budget**
 - 80% data preparation, 20% models design and evaluation
 - Effort that is repeated (A less elasticity to change)
 - **AutoML as a tool**
 - In the creation of ML pipelines, is a recommendation system for short-term accuracy
- ## Objective
- Keeping scientists safe from tedious duties
 - Finding a solution over data with minimal loss is automated.

AUTOML - GENEREAZĂ ML PIPELIN-URI



- ▶ **Input:** Restrictions (CPU, RAM, Time), Dataset, and Performance Goals Auto ML
- ▶ **Results:** With minimum loss, automatically determines the structure of a pipeline.

▶ AutoML Steps

- ▶ Establish pipeline structure
- ▶ Algorithm selection
- ▶ Establish hyper-parameters
- ▶ Performance evaluation

All three steps must be completed.;
Iterate until the goal is achieved.

DEMO