Microsoft

Tutorial:
Building a machine learning model using Automated Machine Learning in Azure ML

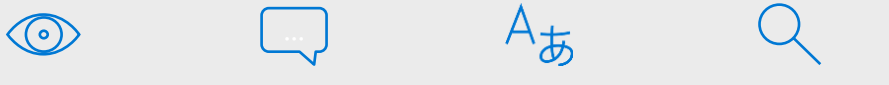
Ognian Dantchev
ALSO Group, AI team

Session goals

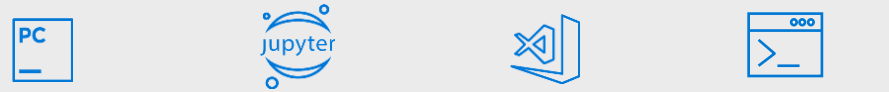
- Gain an **overview** of Azure AutoML and how to get started
- Understand the scenarios when AutoML makes sense
- Know the multiple **approaches to use** Azure AutoML
  - UI, Notebooks/SDK, Pipelines, local vs. remote compute and when to use each one of them.
- Learn additional Azure ML related features:
  - AML Workspace, Datasets, compute
  - Hyperparameter Tuning vs. AutoML

AI and Machine Learning in Azure

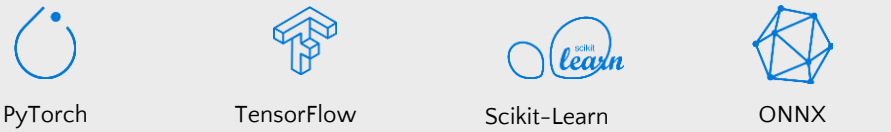
Domain Specific Pretrained Models
To reduce time to market



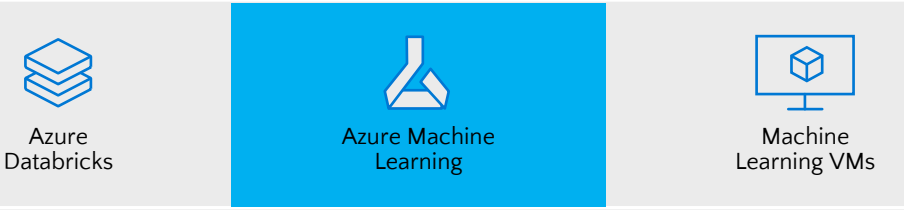
Familiar Data Science Tools
To simplify model development



Popular Frameworks
To build machine learning and deep learning solutions



Productive Services
To empower data science and development teams



Powerful Hardware
To accelerate deep learning

Azure Machine Learning service

Bring AI to everyone with an end-to-end, scalable, trusted platform

Boost your data science **productivity**

Increase your **rate** of experimentation

Deploy and **manage** your models everywhere

Built with your needs in mind

- Automated Machine Learning — Today's talk
- Managed compute
- DevOps/MLOps for machine learning
- Simple deployment
- Tool agnostic Python SDK
- Support for open source frameworks

Seamlessly integrated with the Azure Portfolio

Azure Machine Learning

Experience
SDK, Notebooks, Drag-n-drop, Automated ML *wizard*

MLOps
Reproducible, Reusable, Automatable, Git, CLI, REST

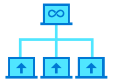
Today's talk

Dataset Management
Profiling, Drift, Labeling

Model Training
Experiments, Runs, **AutoML**

Model Management
Models

Model Serving
Batch, Realtime



Compute
Jobs, Clusters, Instances

Orchestration
Security, Mgmt, Deployment

Cloud
CPU, GPU, FPGA

Edge
CPU, GPU, NPU

Machine Learning Process

Files

Data lake

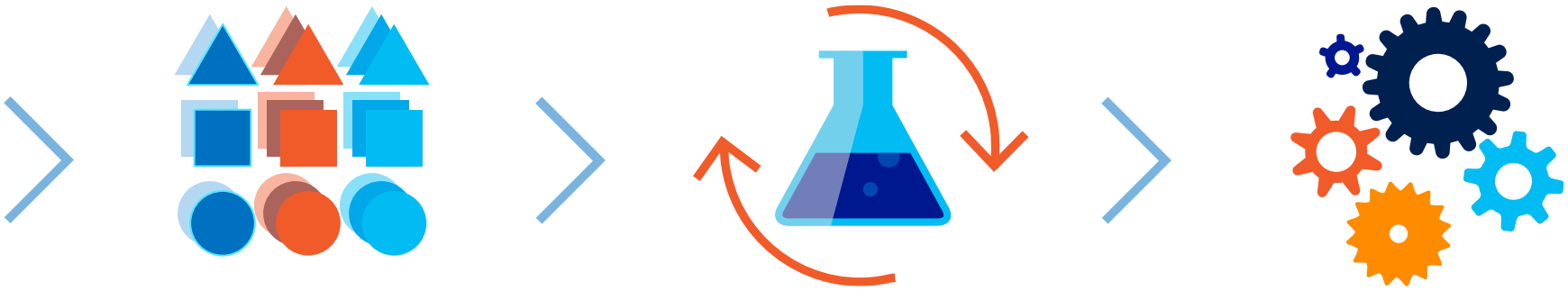
Blob storage

SQL DB

Cosmos DB

Datawarehouse

…

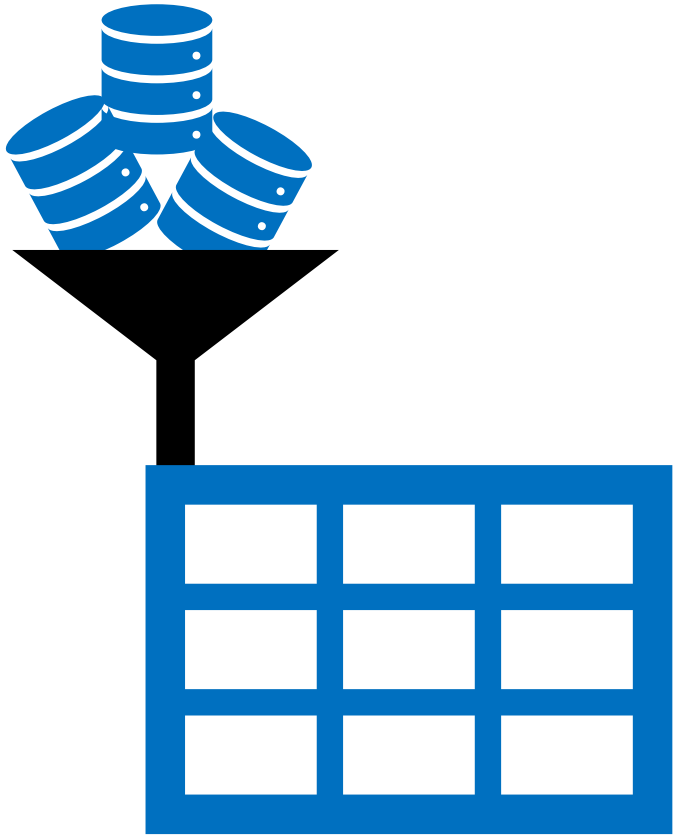


Prepare Data

Build & Train

Deploy

Automated ML Data Requirements



A single source for training:

- Tabular and flat data
- Sources: local file upload, blob, datalake, Azure SQL Database, web source (ex. Dataset files at GitHub), or anything in a Pandas dataframe

Value props:

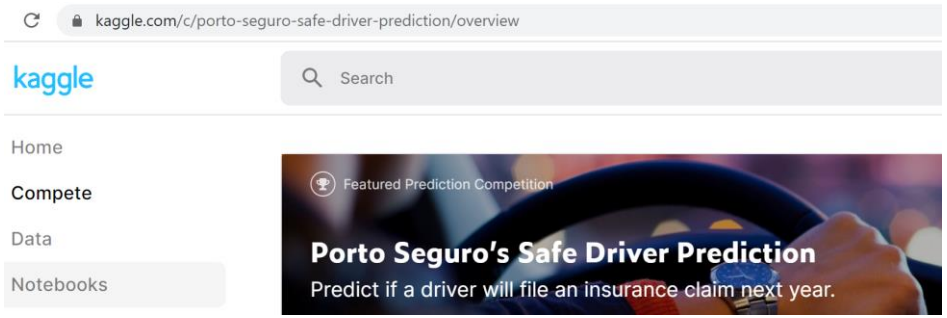
- Large data
- Azure Open Datasets
- Data Guardrails

Demo – Getting started with the Azure ML Workspace

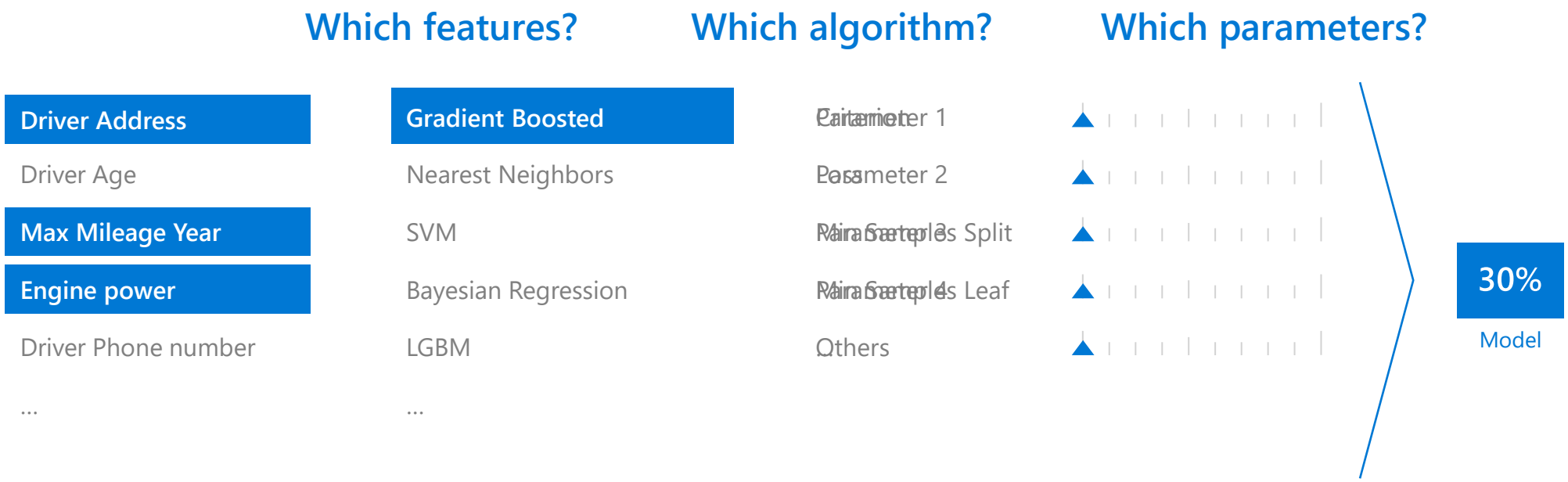
Machine Learning Problem Example

Will a customer/driver file an insurance claim next year?
('Safe Driver' Prediction)

Kaggle Competition: "Porto Seguro's Safe Driver Prediction"
https://www.kaggle.com/c/porto-seguro-safe-driver-prediction/overview



Model Creation Is Typically Time-Consuming



Model Creation Is Typically Time-Consuming

Which features?

- Max Mileage Year
- Driver Age
- Engine Power
- Car Color
- Driver Address
- ...

Which algorithm?

- Gradient Boosted
- Nearest Neighbors
- SVM
- Bayesian Regression
- LGBM
- ...

Which parameters?

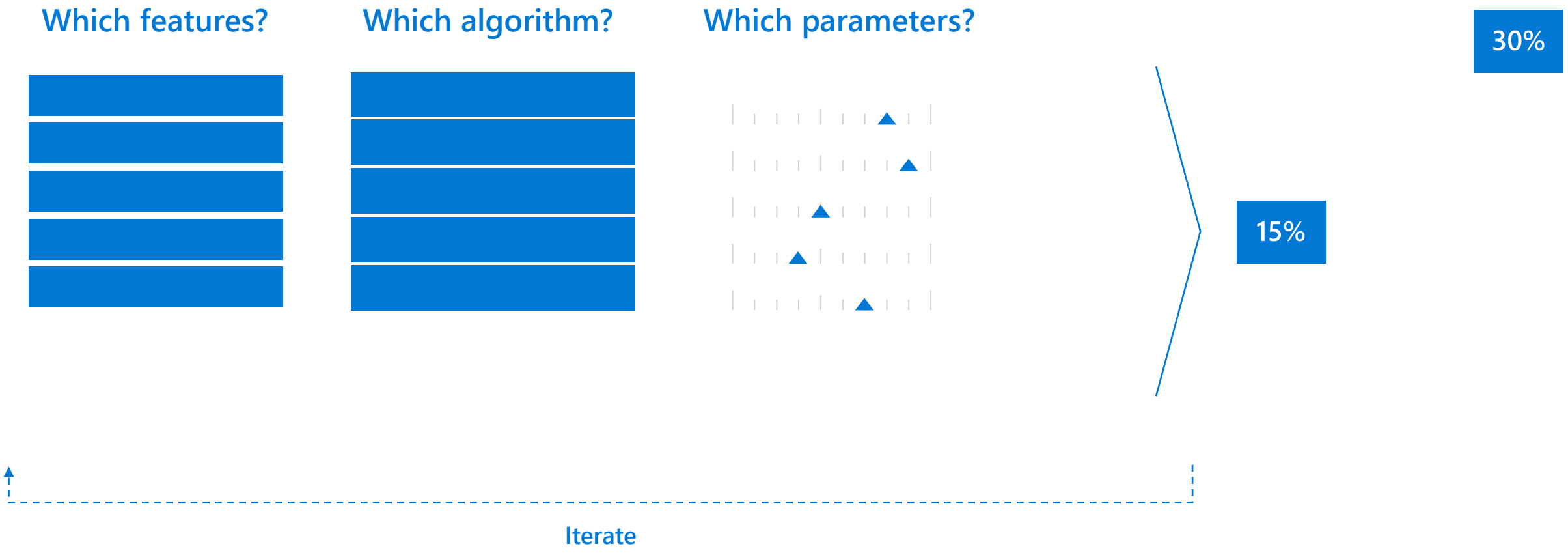
- Number of Neighbors
- Weights
- Min Samples Split
- Min Samples Leaf
- Others

30%

Model

Iterate

Model Creation Is Typically Time-Consuming



What if we can simplify machine learning?

Automated ML & Hyperparameter tuning Mission

Enable automated building of machine learning with the goal of accelerating, democratizing and scaling AI

Accelerate AI

Improve Productivity for Data Scientists, Citizen Data Scientists, App Developers & Analysts

Democratize AI

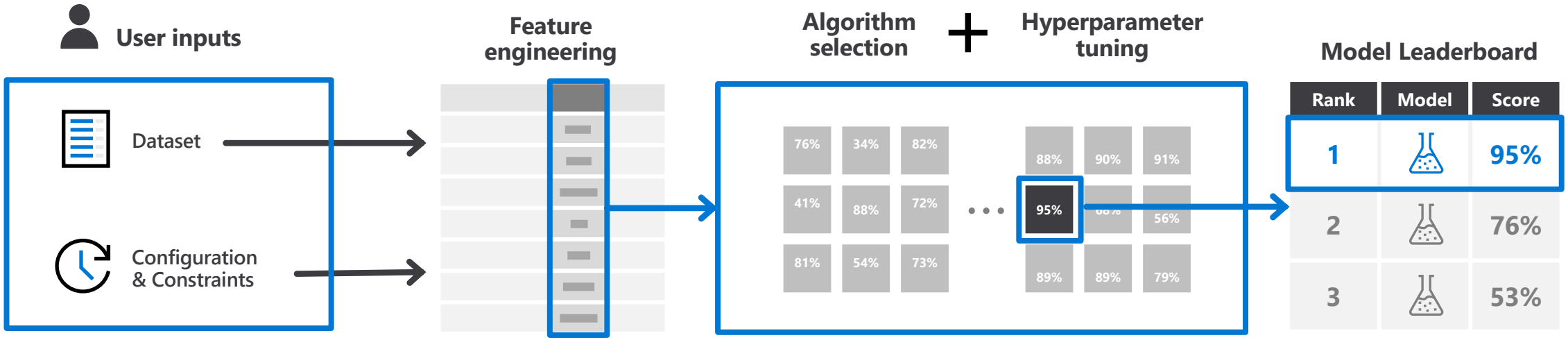
Enable developers and any data science practitioner to rapidly get started building Machine Learning based solutions

Scale AI

Build AI solutions at scale in the cloud out when using large data and multiple model training jobs, in an automated fashion

What is **Automated** Machine Learning?

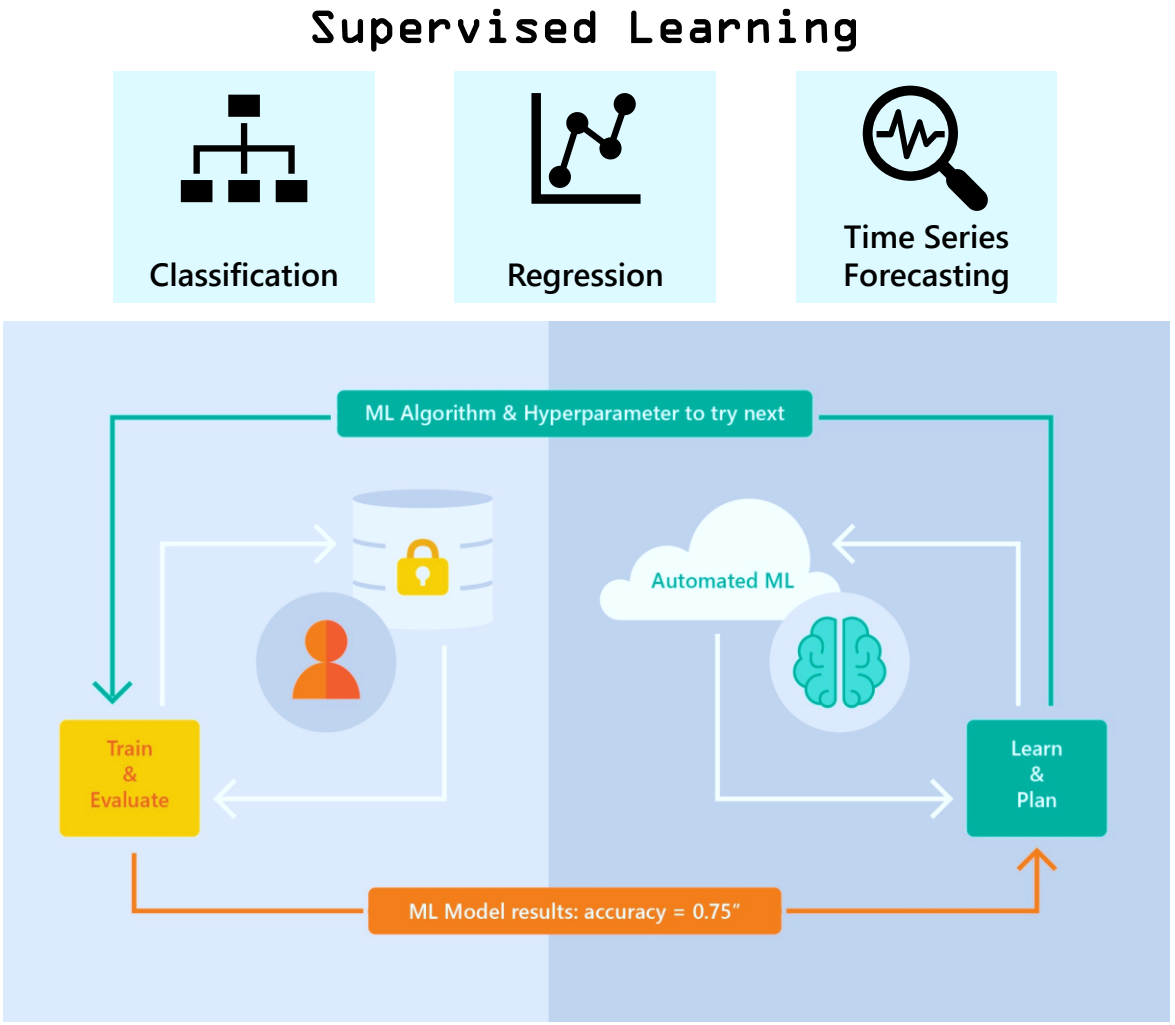
Automated machine learning (automated ML) automates feature engineering, algorithm and hyperparameter selection to find the 'best model' for your data.



Automated ML – How it works

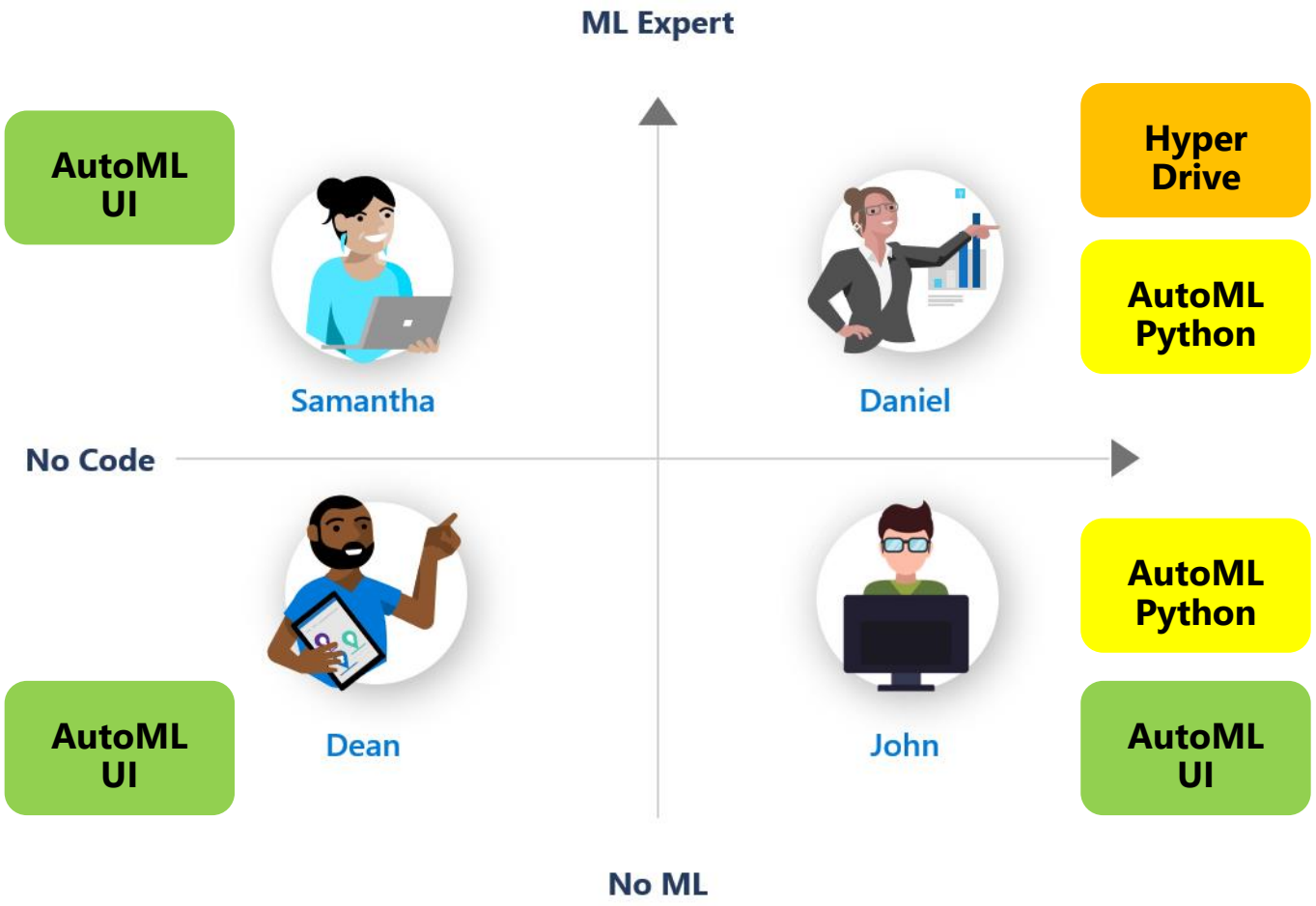
- ML Tasks supported:

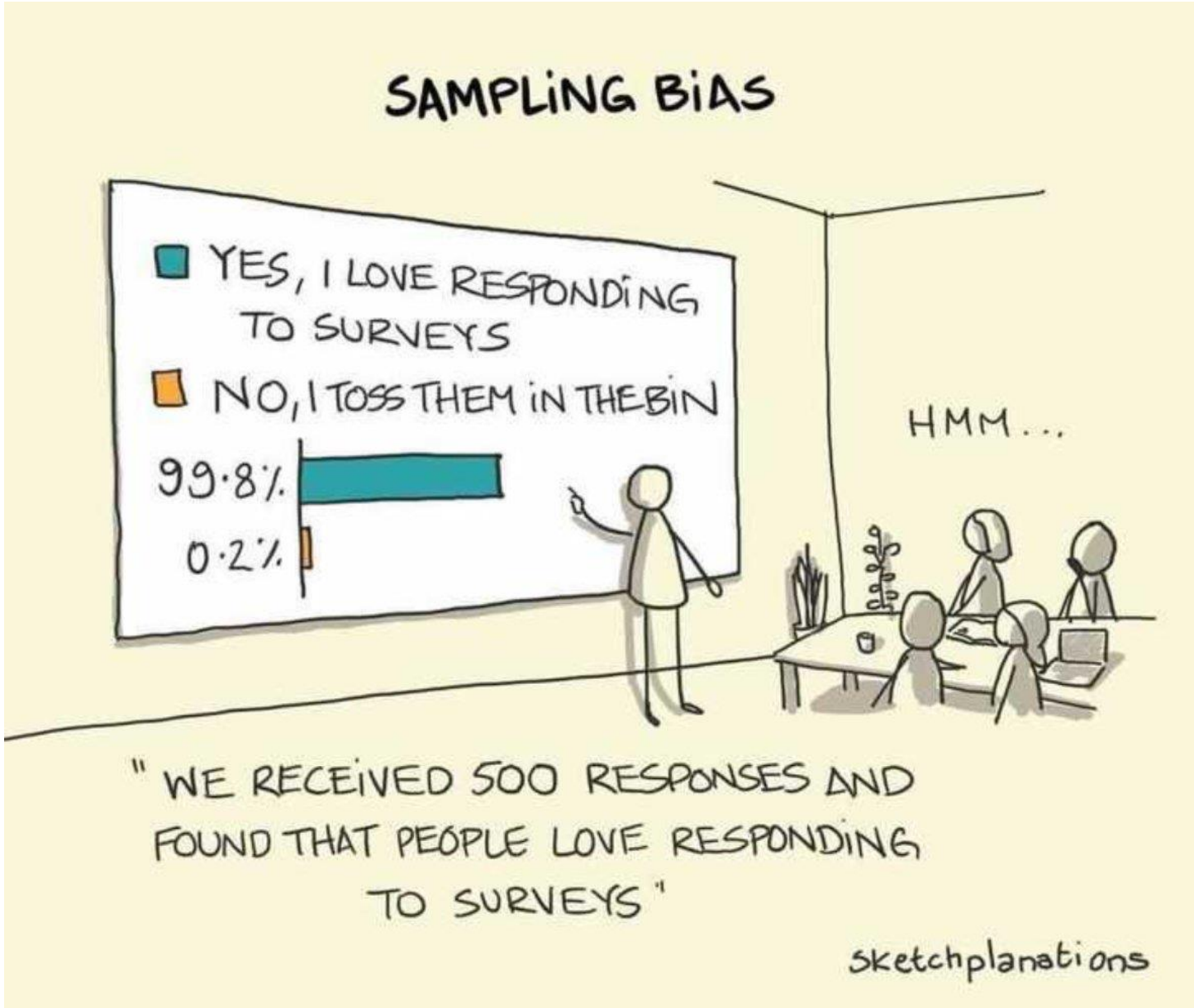
  Classification Regression and Time-Series Forecasting
  + DNN-based text featurization and algos (BERT).
- Based on Microsoft Research
- Brain trained with millions of experiments
- Privacy preserving: No need to "see" the data
- (*) Deep Learning – Computer Vision



- List of supported algorithms in Azure AutoML
- List of Featurization areas
- FeaturizationConfig class (SDK doc) - (Sample notebook)
- List of algorithms and hyperparameters used in AutoML-created model (Sample notebook)

Personas and skills using AutoML & HyperDrive




SAMPLING BIAS
YES, I LOVE RESPONDING TO SURVEYS
NO, I TOSS THEM IN THE BIN
99·8%
0·2%
HMM...
" WE RECEIVED 500 RESPONSES AND FOUND THAT PEOPLE LOVE RESPONDING TO SURVEYS "
sketchplanations

Guardrails


Class imbalance


Train-Test split, CV, rolling CV


Missing value imputation


Detect high cardinality features


Detect leaky features

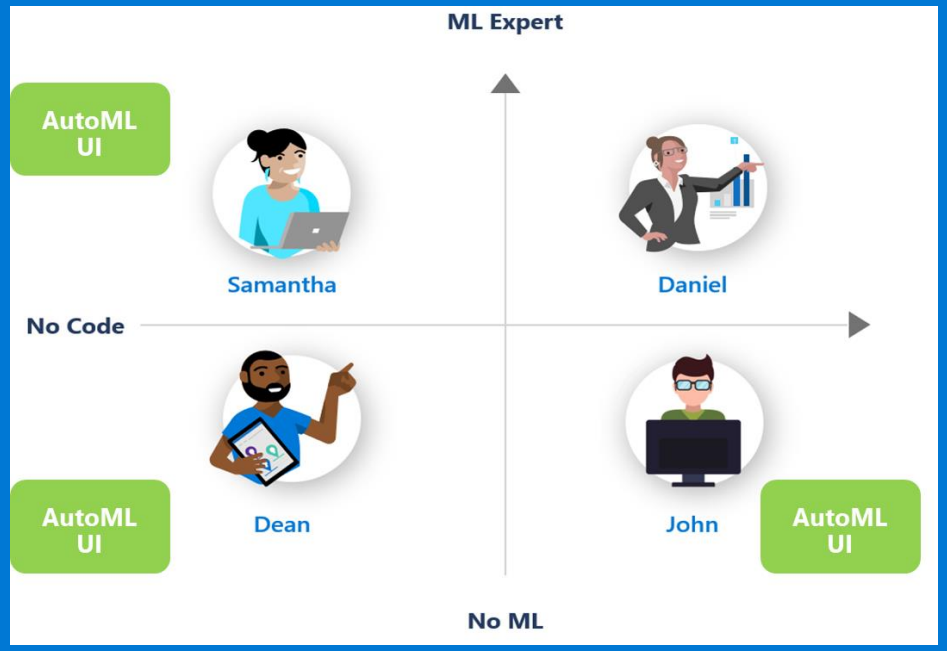

Detect overfitting


Model Interpretability / Feature Importance

Featurization

- [New] Deep Learning based (BERT) featurization
- [New] Custom Featurization
- Automatically engineer features
- Automated pre-processing of data

Demo – Using AutoML UI to create a model



Using AutoML code (AML SDK) in notebooks

AML SDK, packages, install, etc.

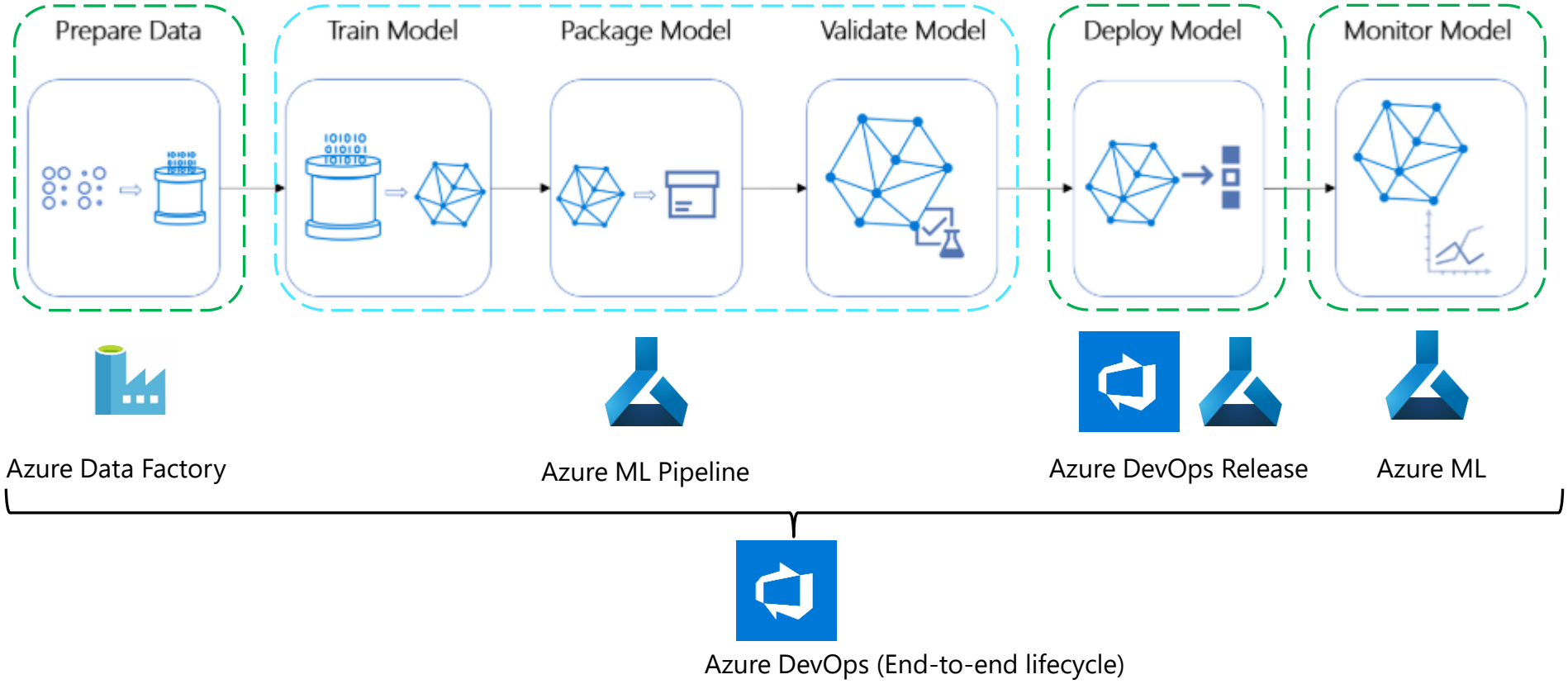
Demo – Using AutoML local runs, with SDK and notebooks, for training the *Safe Driver Prediction model*

Demo – Using AutoML remote runs, with SDK and notebooks, for training the *Safe Driver Prediction model*

Azure ML Pipelines and *"the path towards DevOps for ML"*

Azure ML Pipelines within the e2e model lifecycle

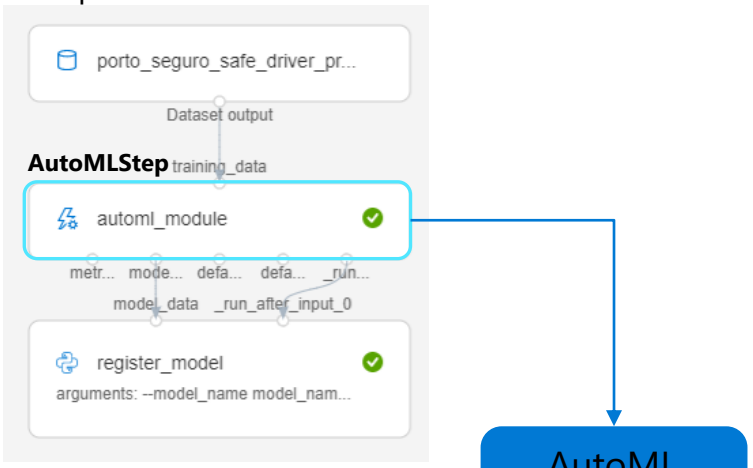
- Azure Machine Learning pipelines allow you to create workflows in your ML projects.
- Those pipelines can later be triggered from external operationalization orchestrators such as **Azure DevOps** CI/CD



AutoMLStep within Azure ML Pipelines

AML Pipelines

Example 1:



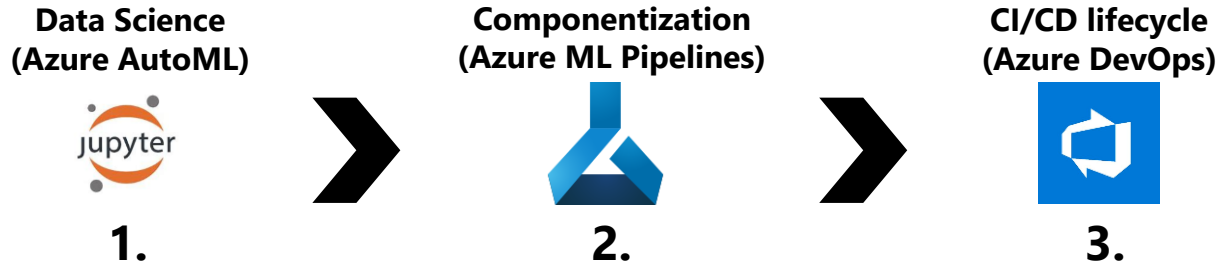
AutoML training

Example 2:

AutoMLStep provides an easy way to encapsulate AutoML configuration and training code as an **AML Pipeline step.**

Typical code refactoring process:

1. **Initial code:** Explore/define your regular AutoML code in **Jupyter notebooks**
2. **AML Pipelines**: Evolve/refactor code into decoupled pieces of code:
   - **PythonScriptStep** with .py script for each data transforms & AML SDK code
   - **AutoMLStep** for encapsulating AutoML as a pipeline step
3. **CI/CD for ML**: Integrate with higher level model lifecycle in Azure DevOps



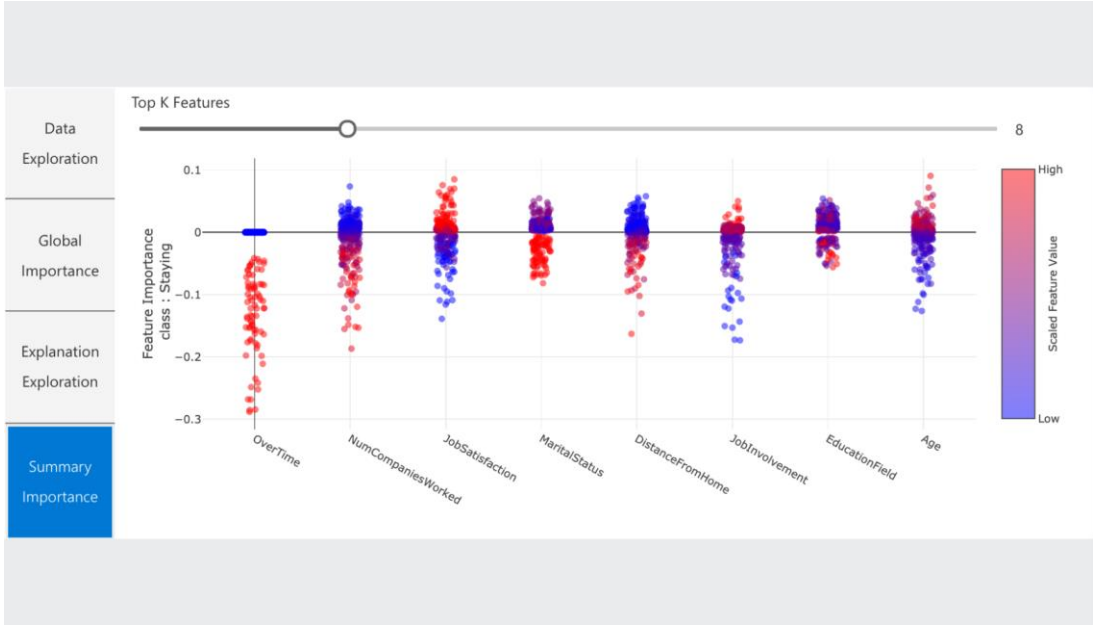
Demo – Using AutoML in AML Pipelines
(Pipeline implementation with SDK/Code in Python)

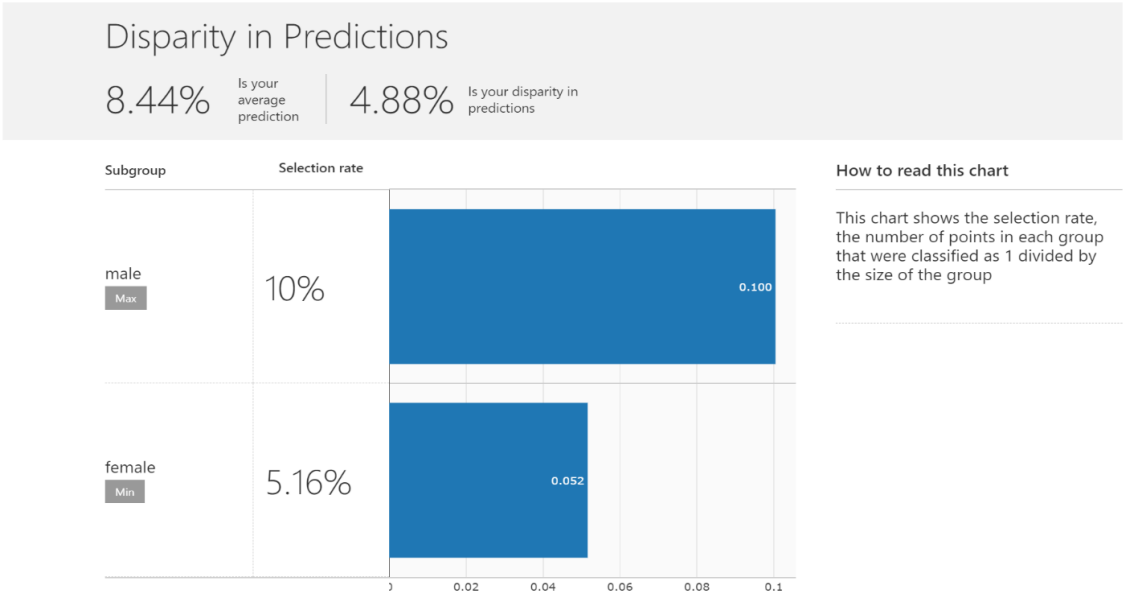
Model Deployment and inferencing in QA/Production

Demo – Deploying AutoML models and inferencing in QA/Production

UI + SDK demo

Model Explainability in Azure AutoML





Demo – Model Explainability in Azure Automated ML

What is Hyper Parameter Tuning (HyperDrive) in Azure ML?

…And when to use it versus Azure AutoML.

Azure **AutoML** vs. **HyperDrive**

- **Azure AutoML** supports *Regression*, *Classification* and *Time Series Forecasting* using tabular data.

- What if I want to optimize a different ML problem/tasks such as based on deep learning (image classification), or any other problem where **I also have my initial code to optimize**? (Scikit-Learn, TensorFlow, PyTorch, etc.)

→ Then, use **AML Hyperparameter Tuning (HyperDrive)**

Azure ML Hyperparameter Tuning Capabilities

1. 

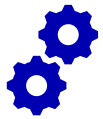
Automated Exploration

Automates the process of exploring the hyperparameter space, saving data scientists significant time and effort

2. 

Efficient Resource Usage

Intelligently early terminates poor performing configurations to save compute costs

3. 

User Controls

Users can specify a primary metric to optimize and a maximum budget in terms of number of jobs or duration

4. 

Deep Learning & Traditional ML

Applies to different models and learning domains. Users bring their own training algorithms and use in different scenarios

5. 

Framework agnostic

Use with Tensorflow, PyTorch, Scikit Learn or any other training framework of your choice

6. 

Easily find best performing config

Visualize all configs together and easily identify hyperparameter values that result in optimal performance. Deploy model from this best config and infer using AML

How hyperparameter tuning in Azure ML works

Input

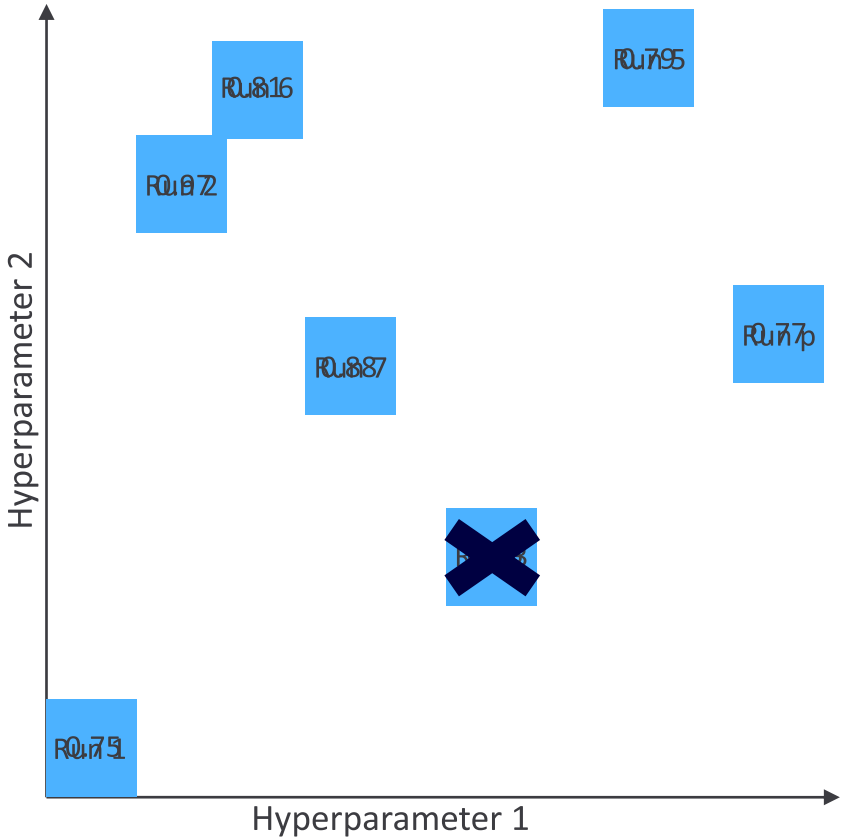
Training data

Training script

Define goals

Apply constraints

Launch multiple parallel training runs



Output

Hyperparameter config with best metric

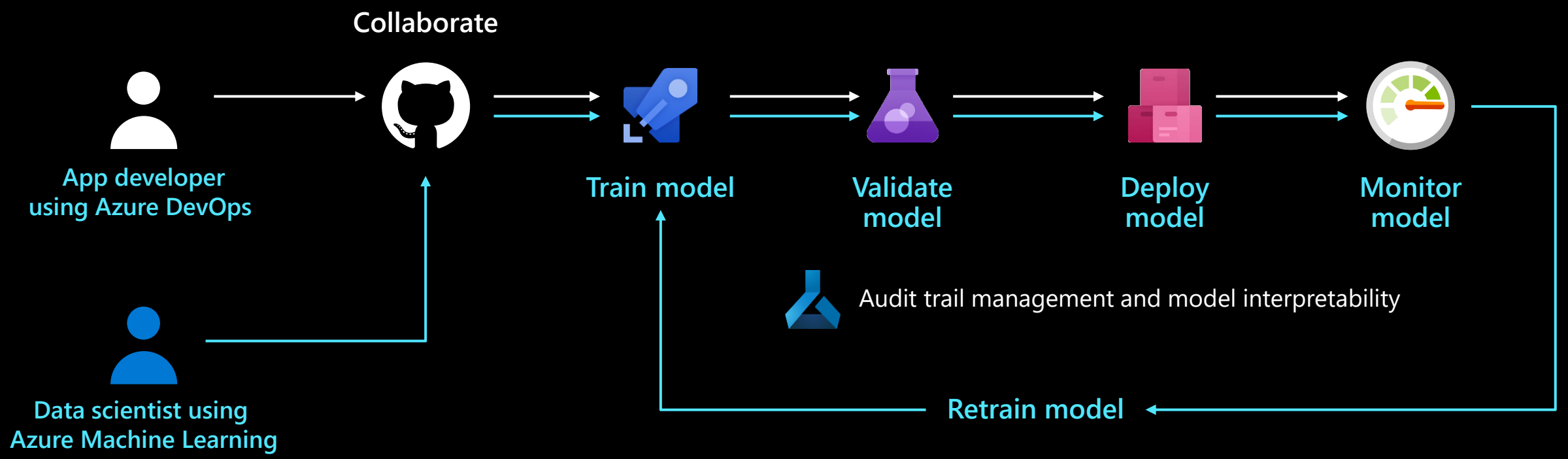
HyperDrive sample notebook

Demo – Using **Azure ML Hyperparameter Tuning** to optimize a deep learning Tensorflow model code in a notebook

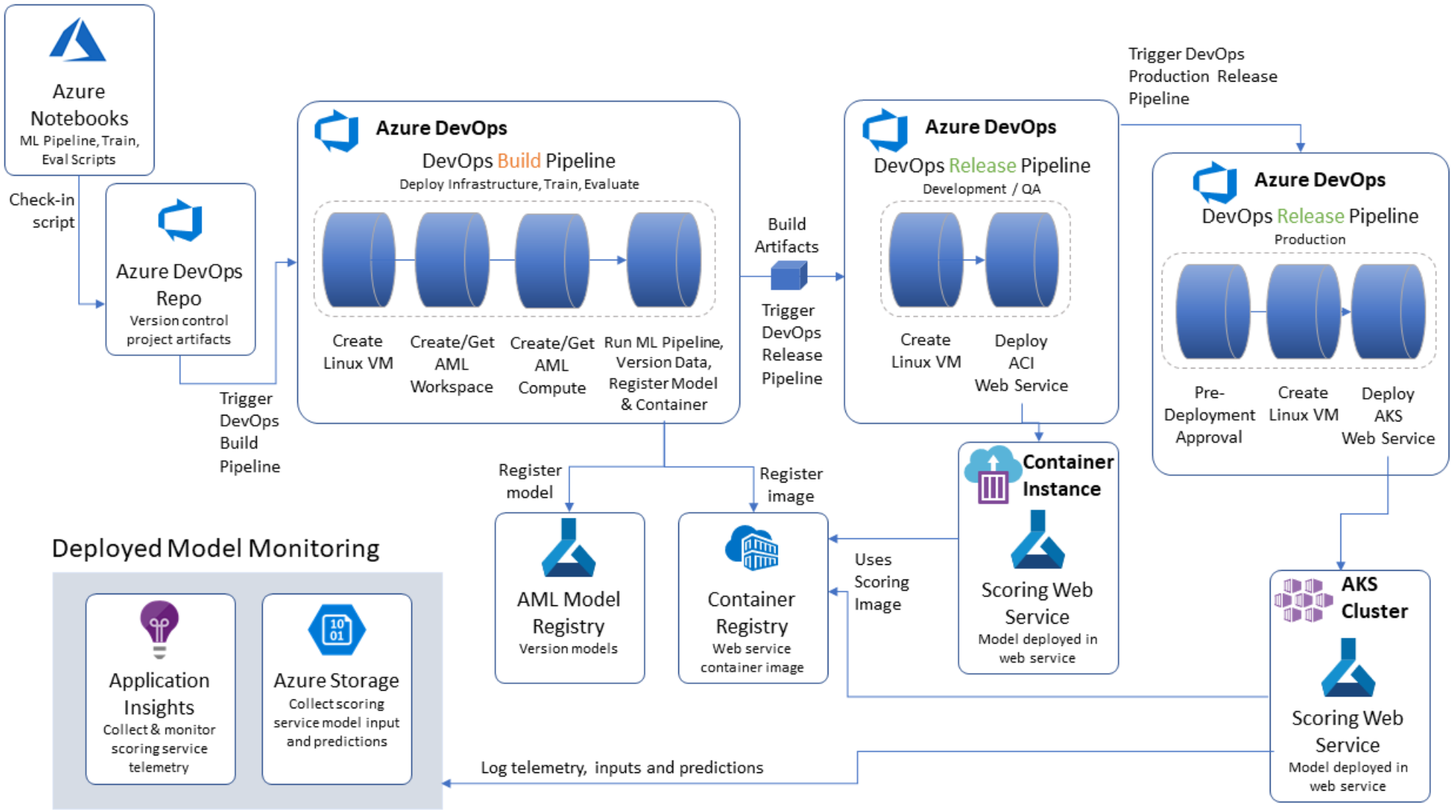
Next steps…

DevOps? MLOps? Operationalization?

MLOps with Azure Machine Learning



- Model reproducibility
- Model validation
- Model deployment
- Model retraining



- ☑ Model reproducibility
- ☑ Model validation
- ☑ Model deployment
- ☑ Model retraining

Additional topics not covered today in detail:

- AML designer (Visual designer UI)
- AML Dataset data-drift
- MLOps and Model Lifecycle (AML pipelines integration to Azure DevOps)
- NLP/DNN Featurization (Deep-Learning based featurization for Text)
- AML infrastructure (Workspace, Datasets, Environments, etc.)
- Other ML Tasks (Regression, Time-Series Forecast, Classification)
- Model interoperability with ONNX
- ‘Many Models’ (Hundreds of models trained in parallel)

Try it for free

http://aka.ms/amlfree

Learn more : https://aka.ms/automatedmldocs
Notebook Samples : https://aka.ms/automatedmlsamples
Blog: https://aka.ms/AutomatedML
The studio: https://ml.azure.com
Try Automated ML: https://aka.ms/tryautomatedml
Product Feedback : AskAutomatedML@microsoft.com

Microsoft

Microsoft

Appendix