



15 DECEMBER 2018

# GLOBAL AI BOOTCAMP

NEW INTELLIGENT WORLDS

LONDON  
2018

# Optimizing models with AutoML

Pablo Doval  
Data Pontifex @Plain Concepts



# Pablo Doval

DATA PONTIFEX @Plain Concepts

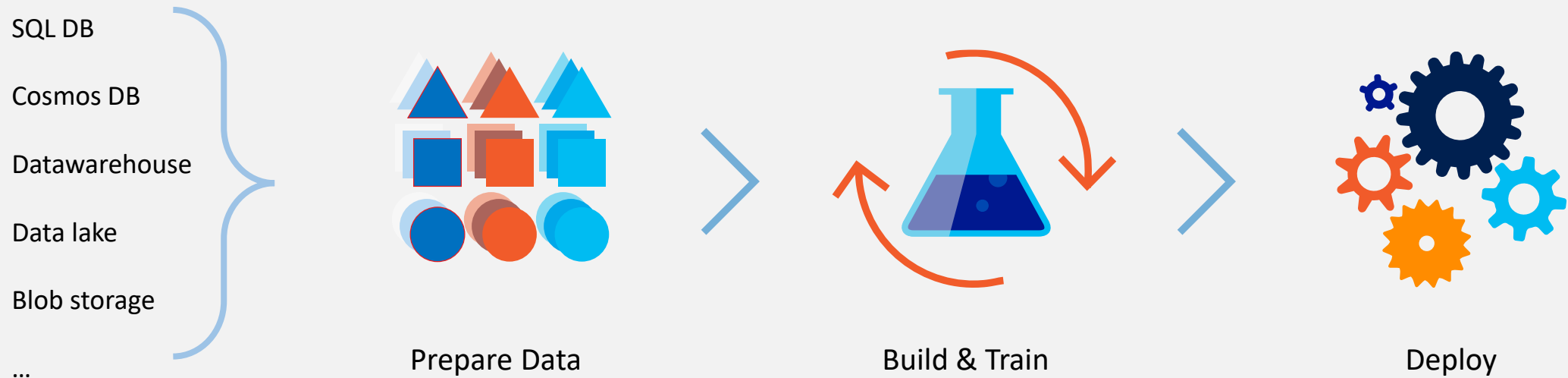
I work with code and data, but don't tell my mom; she thinks I'm a piano player in a whorehouse.

@PabloDoval

# Automated ML



# Simplifying the Machine Learning Process?

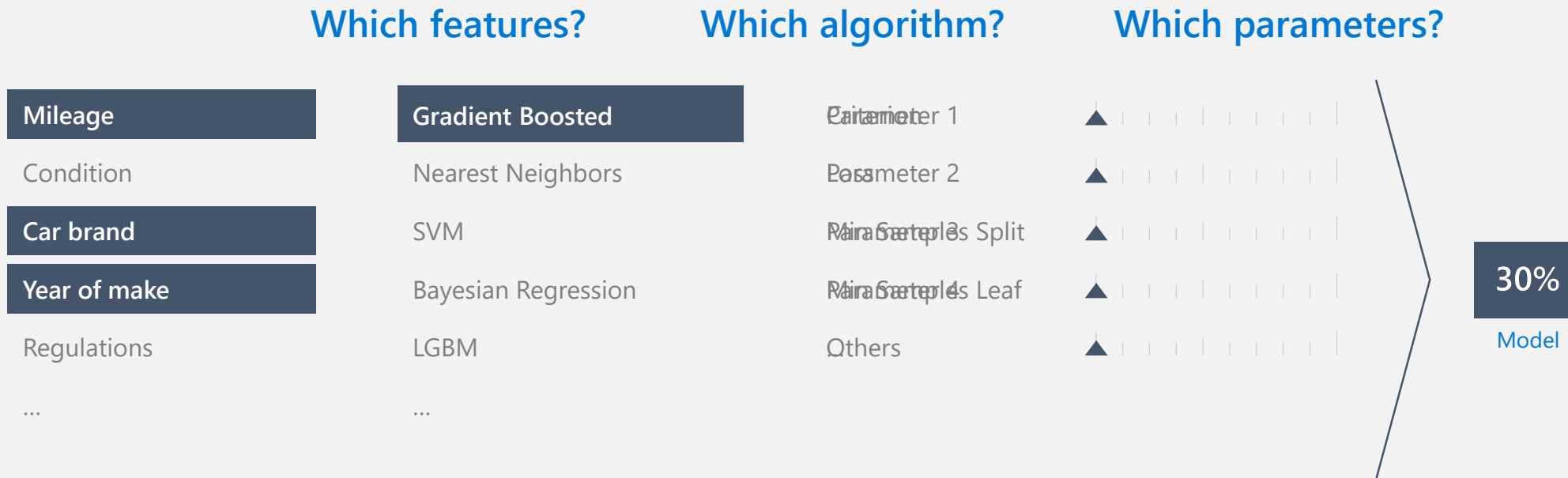


# ML Problem Example

How much is a car worth?



# Model Creation Is Typically Time-Consuming



# Model Creation Is Typically Time-Consuming

## Which features?

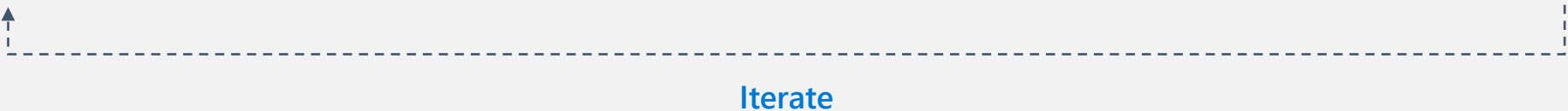
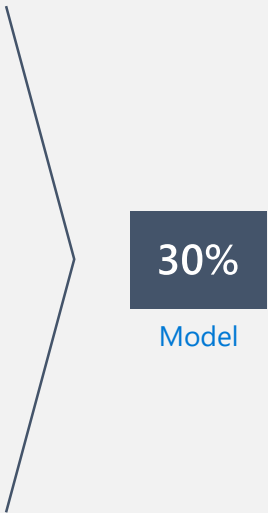
- Mileage
- Condition
- Car brand
- Year of make
- Regulations
- ...

## Which algorithm?

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

## Which parameters?

- Critereion
- Neighbors
- Weights
- Min Samples Split
- Min Samples Leaf
- Others

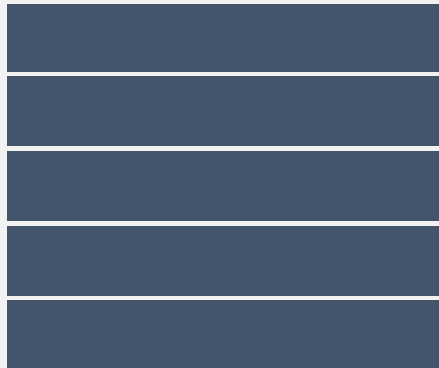


# Model Creation Is Typically Time-Consuming

Which features?



Which algorithm?



Which parameters?



30%

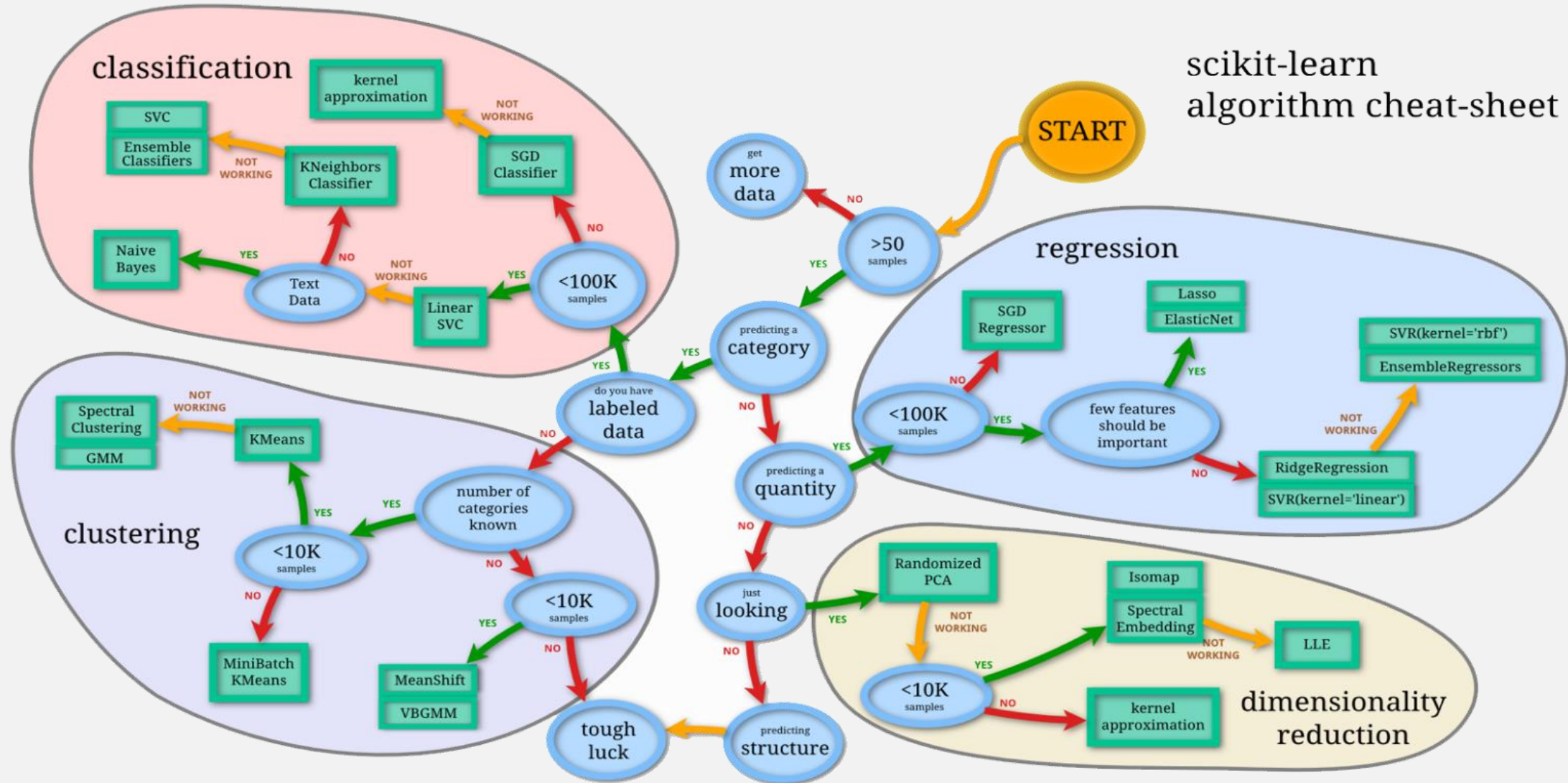
15%

Iterate

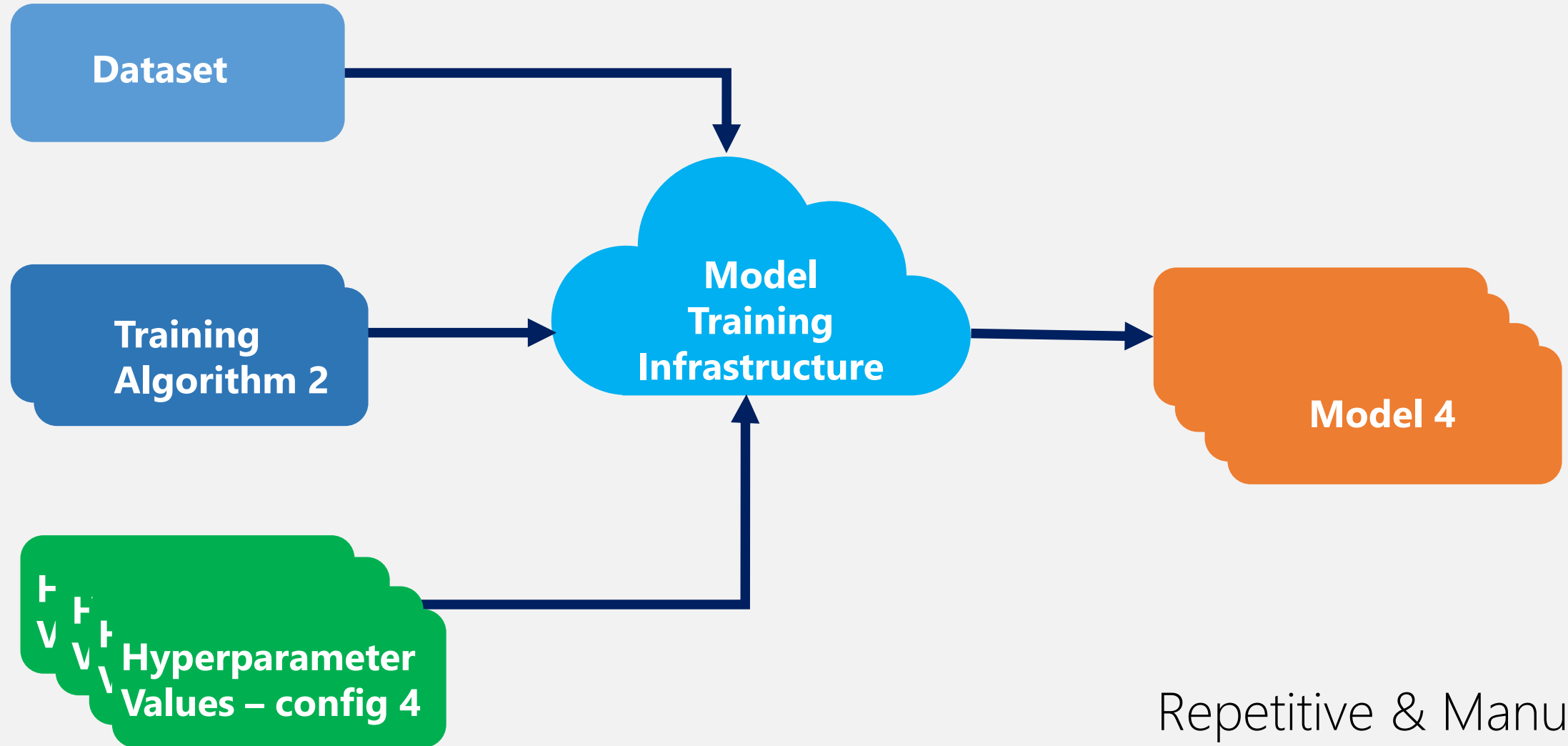


# Machine Learning Complexity

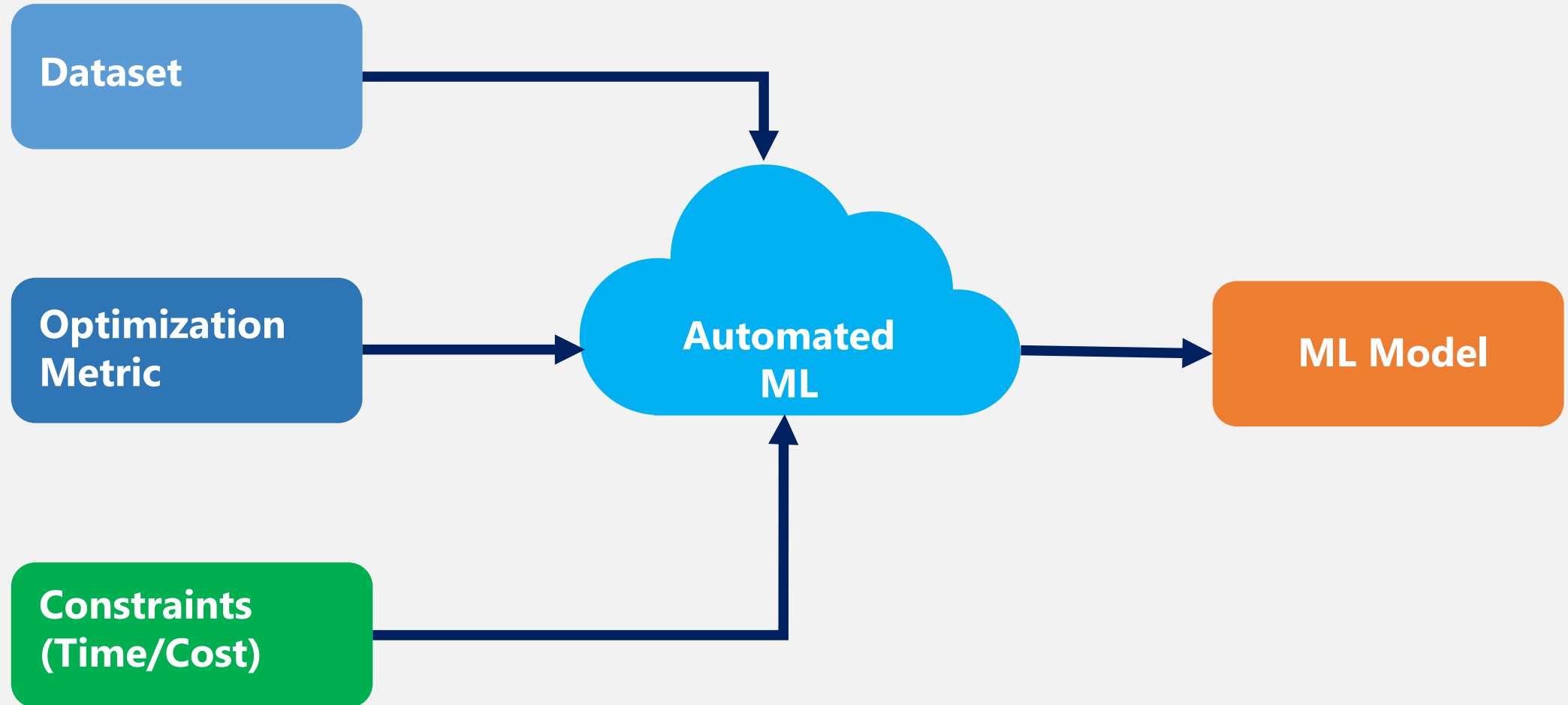
scikit-learn  
algorithm cheat-sheet



# Model Selection & Hyperparameter Tuning

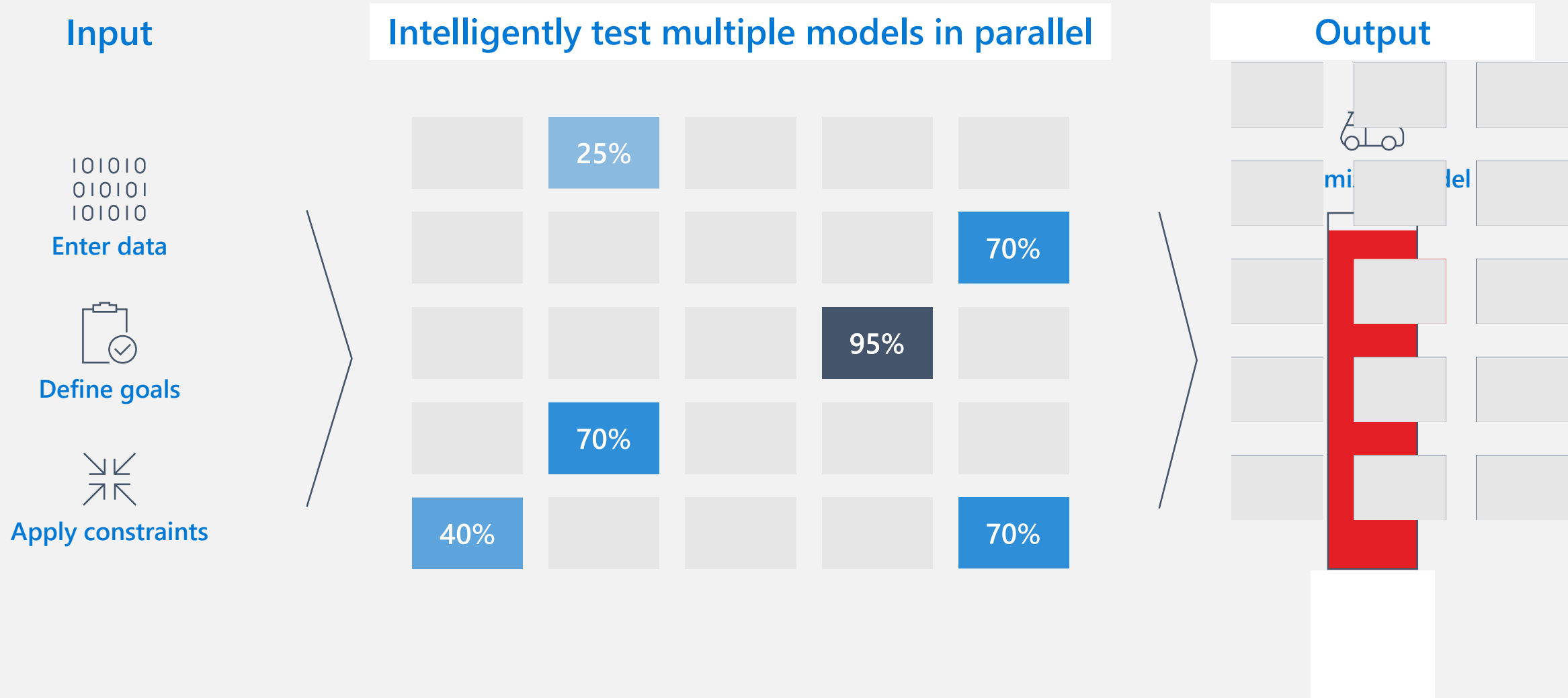


# Introducing Automated Machine Learning



Accessible & Faster

# Automated ML Accelerates Model Development



# Automated ML Capabilities

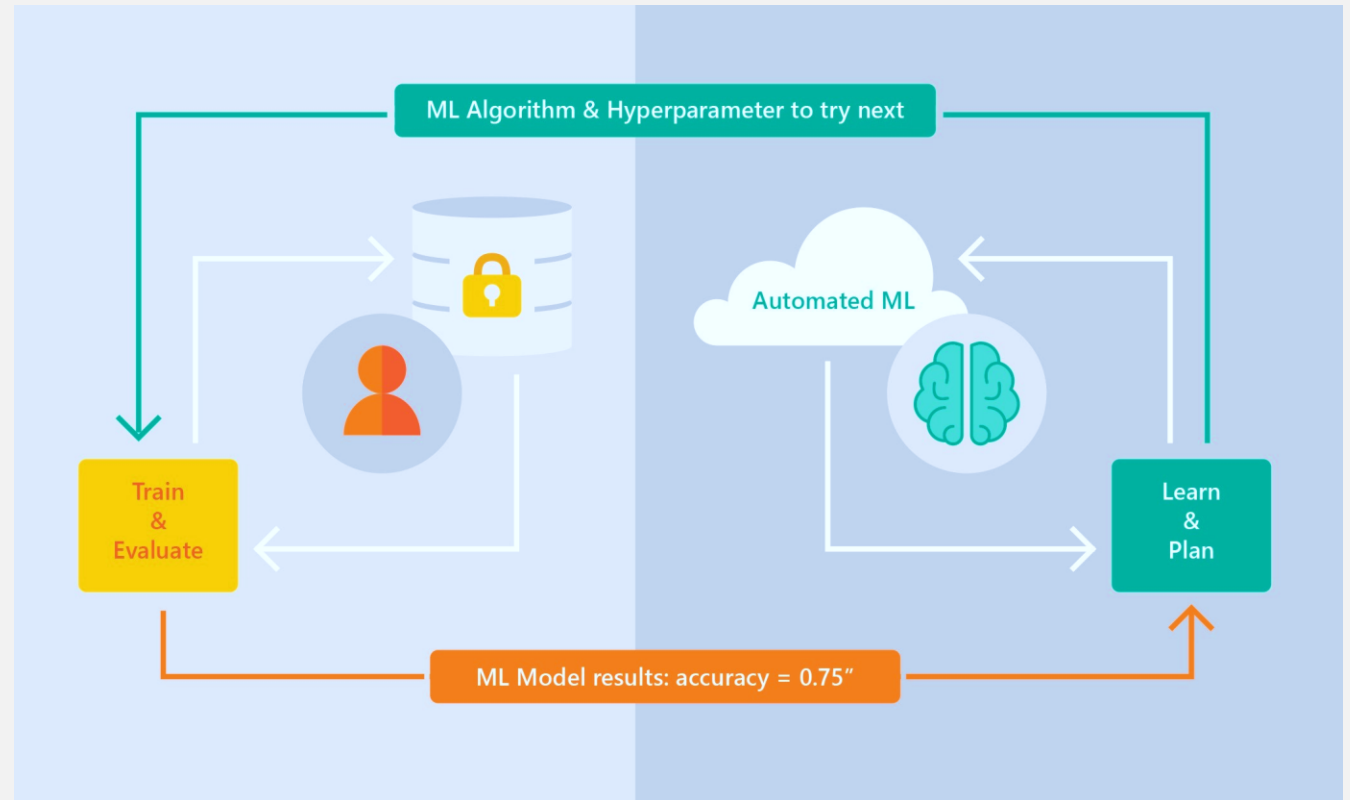


**GLOBAL AI  
LONDON 2018**



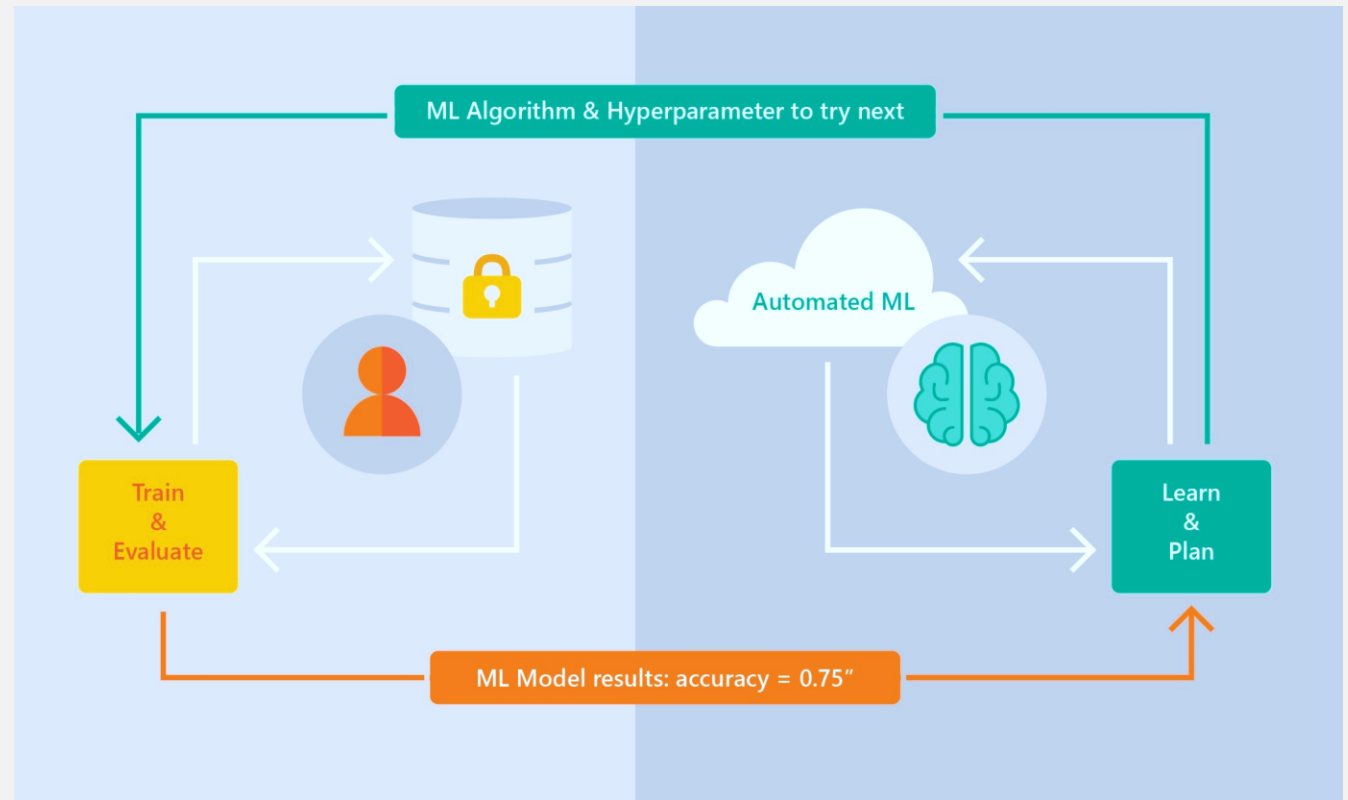
# Automated ML Capabilities

- Based on Microsoft Research
- Brain trained with several million experiments
- Collaborative filtering and Bayesian optimization
- Privacy preserving: No need to "see" the data



# Automated ML Capabilities

- ML Scenarios: Classification & Regression, Forecasting
- Integration: Azure Machine Learning, Azure Notebooks, Jupyter Notebooks
- Data Type: Numeric, Text
- Languages: Python SDK for deployment and hosting for inference
- Training Compute: Local Machine, Remote Azure DSVM (Linux), Azure Batch AI, Databricks
- Transparency: View run history, model metrics
- Scale: Faster model training using multiple cores and parallel experiments



# Feature Engineering

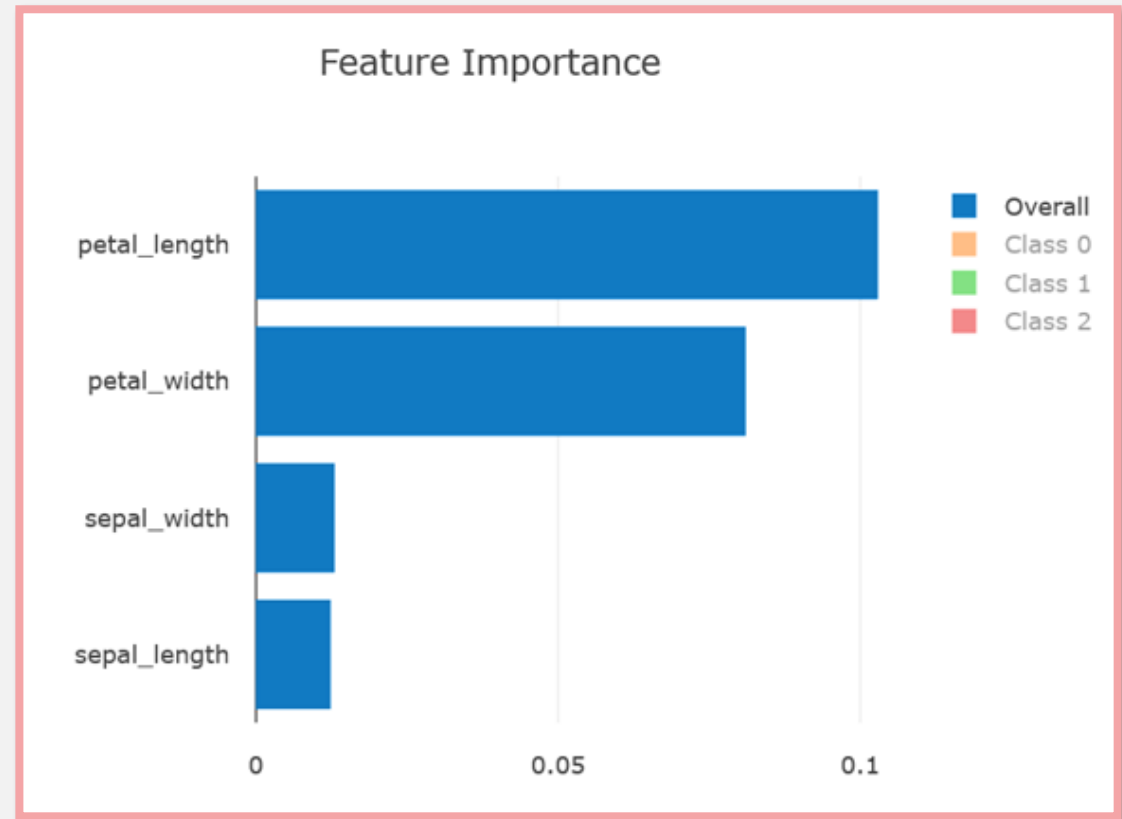
- Dropping high cardinality or no variance features
- Missing value imputation
- Generating additional features
- Transformations and encodings





# Model Explain-ability

- Feature importance as part of training
- Local feature importance for a given sample



# Learn more!

## Check out the lab for this session!

[aka.ms/GlobalAINotebook](https://aka.ms/GlobalAINotebook)

You'll learn:

- Use Azure Notebooks
- Build a basic Model
- Scale that model to the cloud
- Exploring and Testing the model



# Thanks!

Next up...

## AI strategy for leaders architects and businesses

Sherin Mathews

**GLOBAL AI**  
**LONDON 2018**



# **GLOBAL AI LONDON 2018**