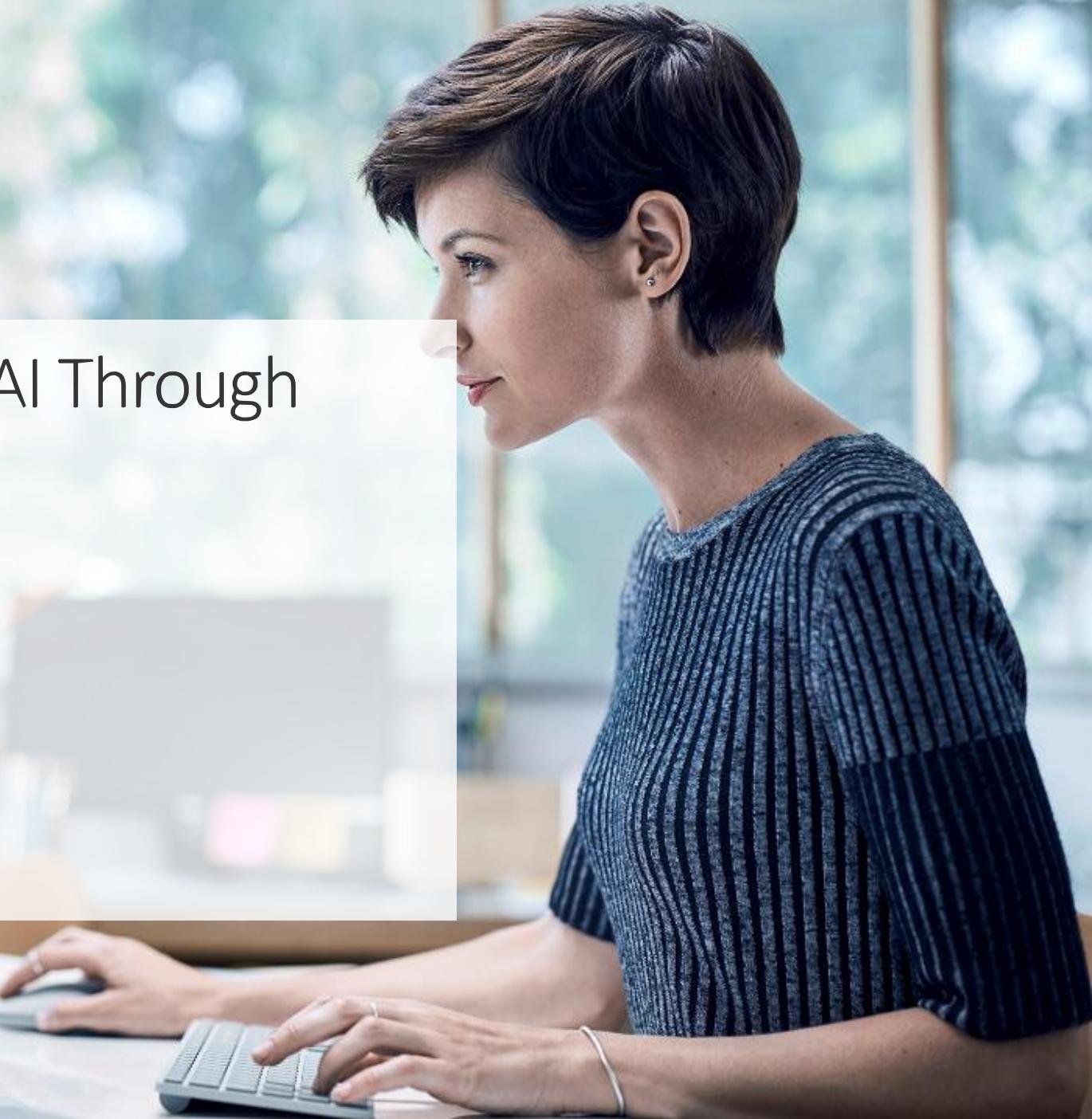


Democratizing & Accelerating AI Through Automated Machine Learning

AI Platform Team
[#AIGlobalNight](#)



global.ainights.com



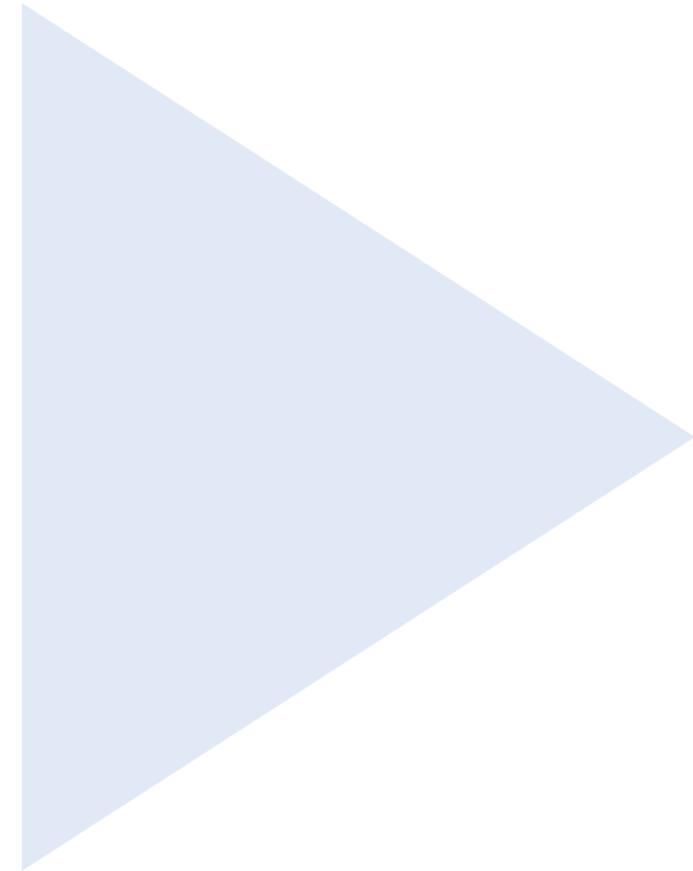
Agenda

- Welcome Video
- Why Automated Machine Learning
- Automated ML Capabilities
- How to Get Started





Welcome Video



Machine Learning on 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



Vision



Speech



Language



Search



PyCharm



Jupyter



Visual Studio Code



Command line



PyTorch



TensorFlow



Scikit-Learn



ONNX



Azure Databricks



Azure Machine Learning



Machine Learning VMs



CPU



GPU



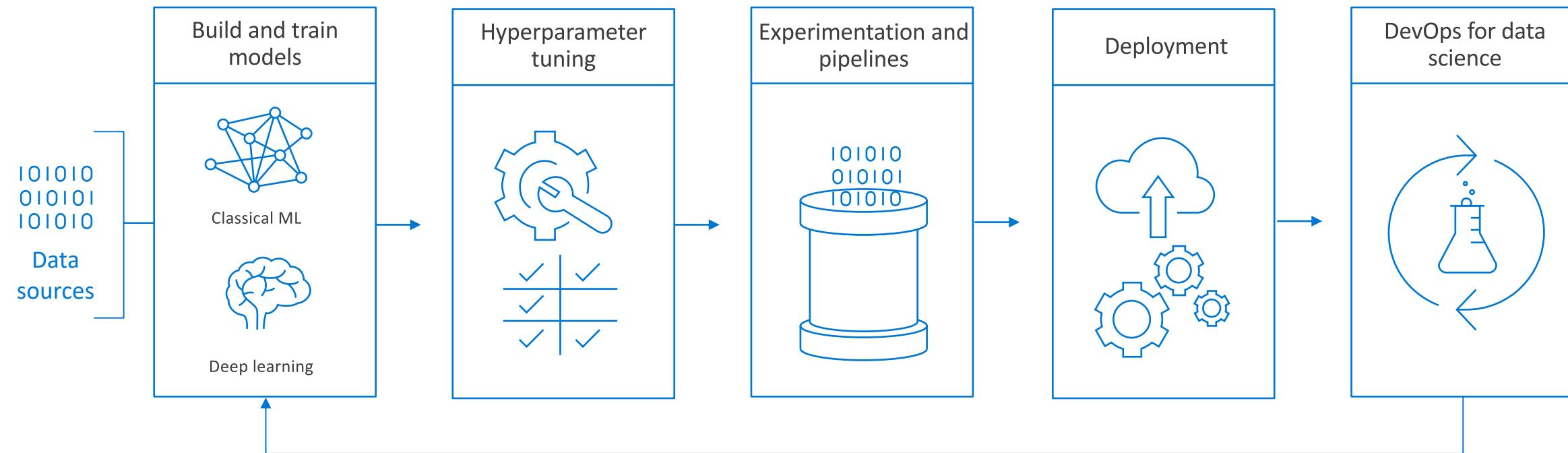
FPGA



From the Intelligent Cloud to the Intelligent Edge



Building blocks for a Data Science Project



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 Mission

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



Democratize AI

Enable Domain Experts & Developers to get rapidly build AI solutions

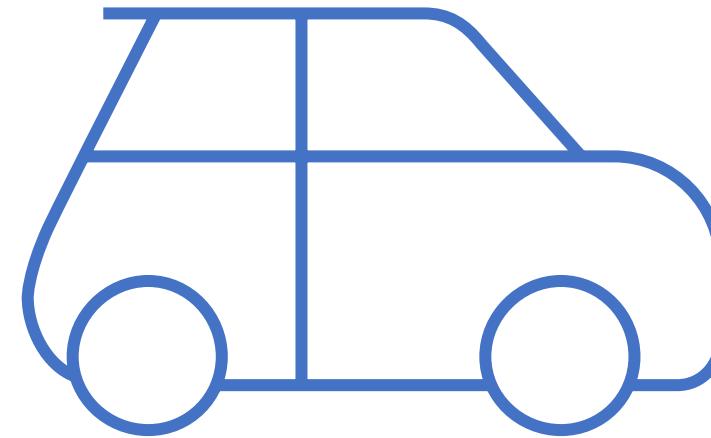
Accelerate AI

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

Scale AI

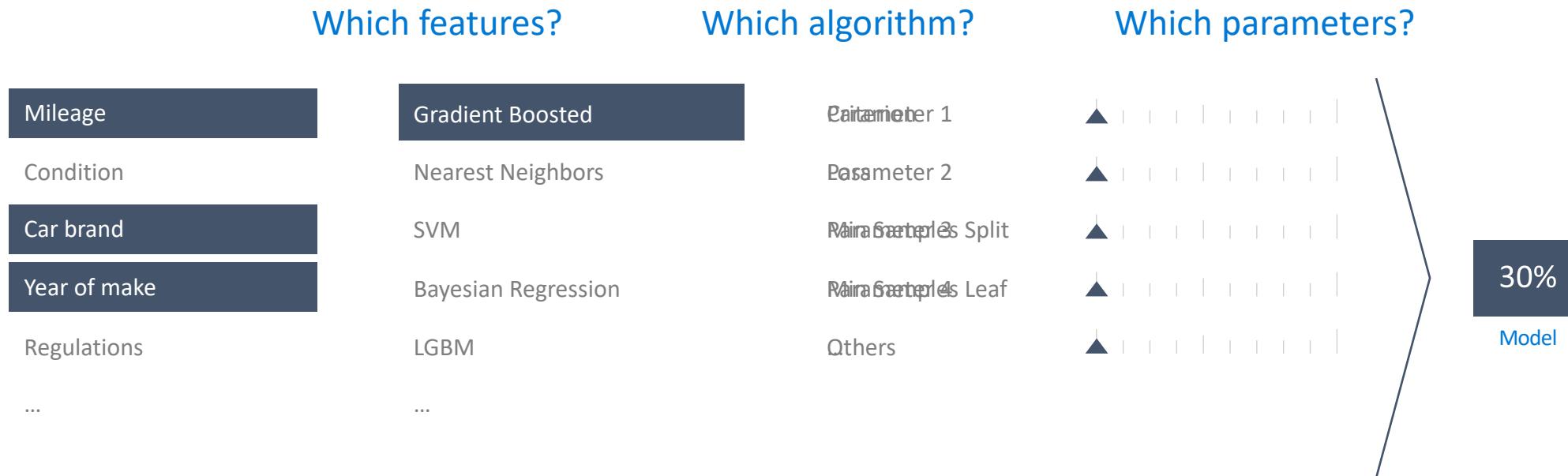
Build AI solutions at scale in an automated fashion

Machine Learning Problem Example



How much is this 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
...

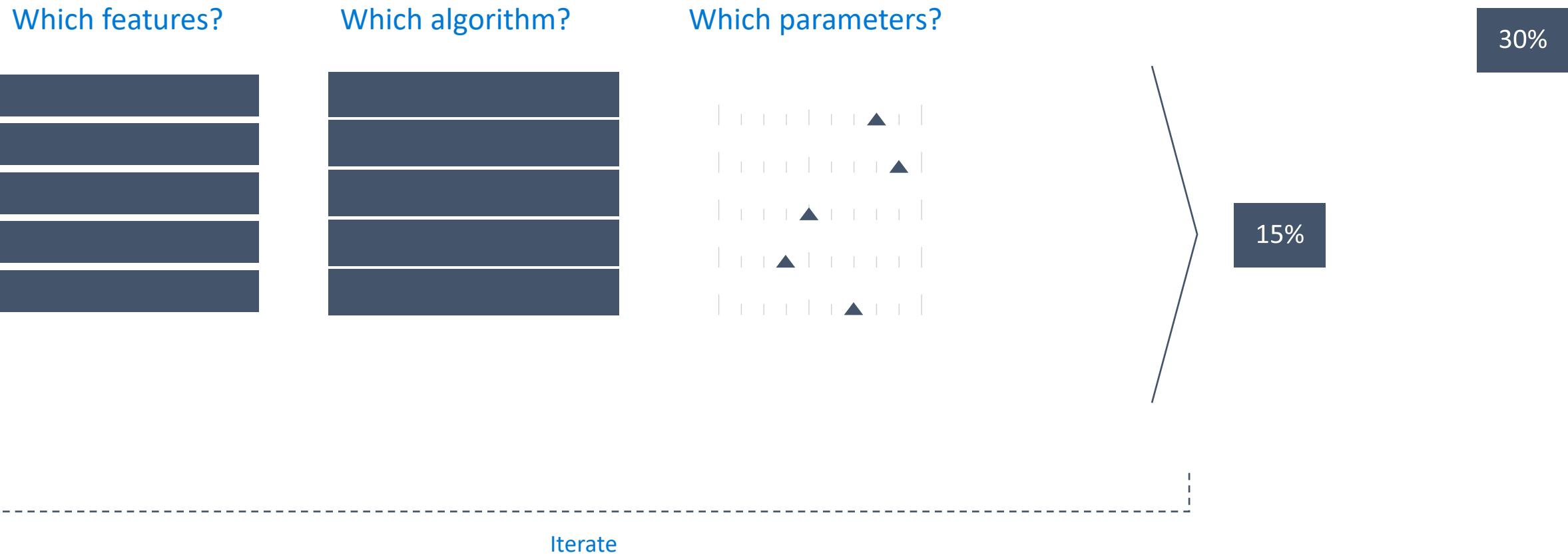
Which parameters?

Neighbors
Weights
Min Samples Split
Min Samples Leaf
Others

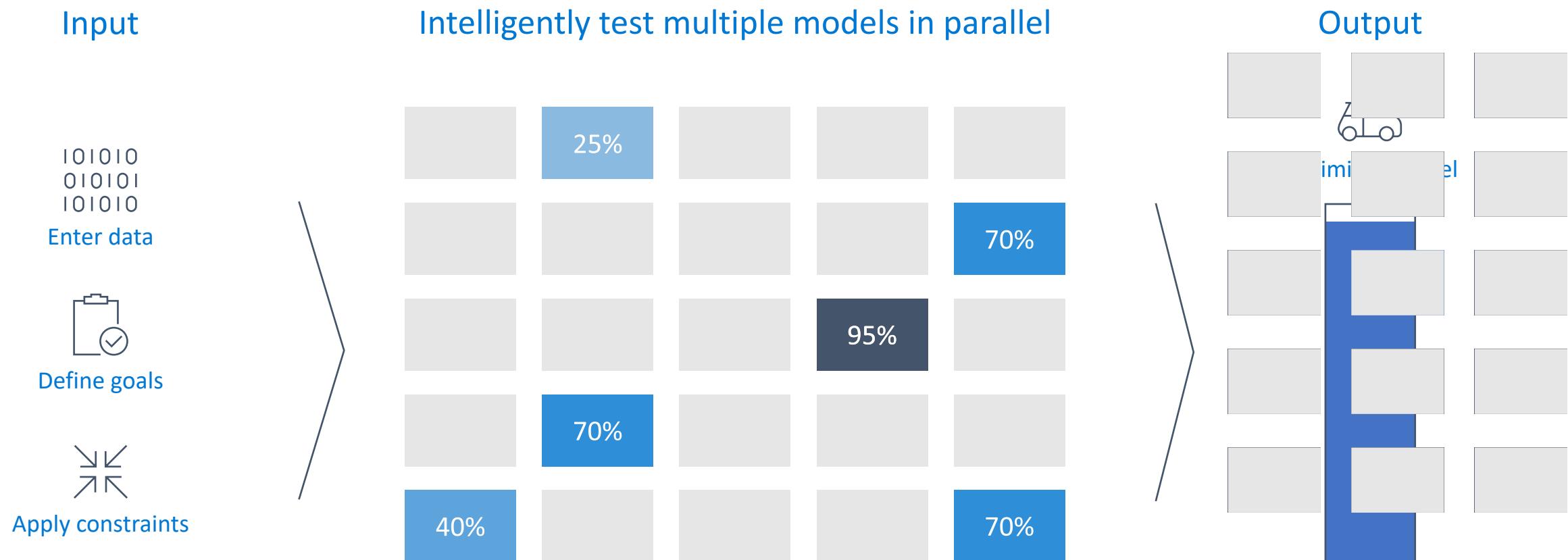
30%
Model



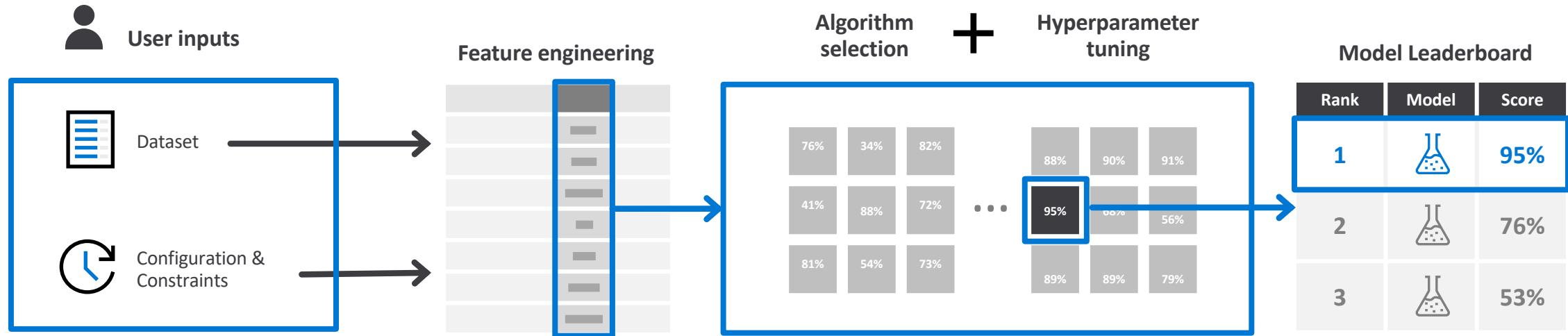
Model Creation Is Typically Time-Consuming



Automated ML Accelerates Model Development



Automated Machine Learning Under the Hood



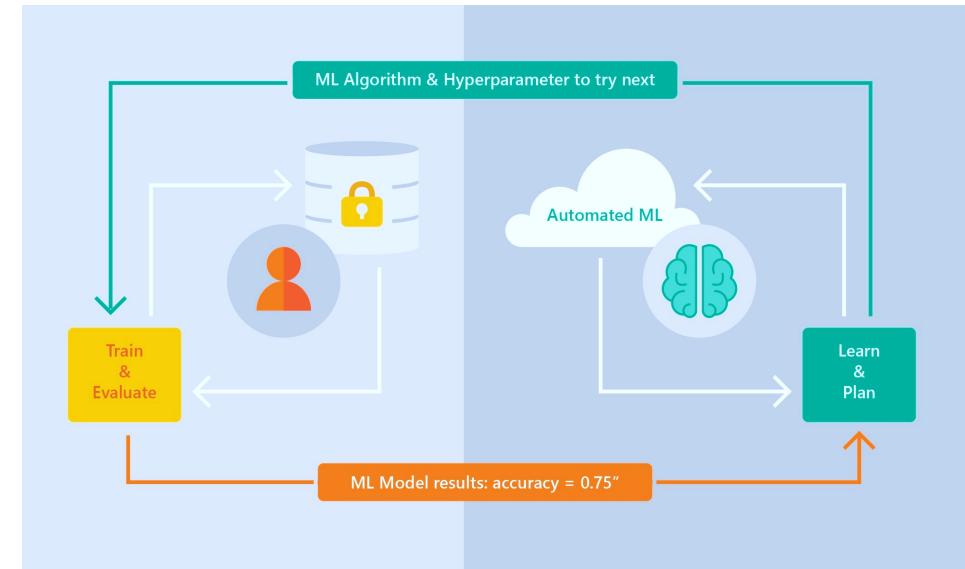
Automated Machine Learning Capabilities

Based on Microsoft Research

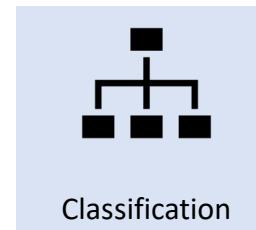
Brain trained with several million experiments

Personalized recommendation approach:
Collaborative filtering and Bayesian optimization

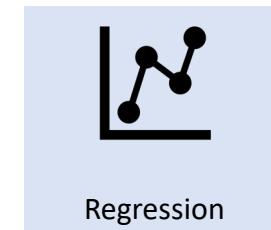
Privacy preserving: No need to “see” the data



Supervised Learning



Classification



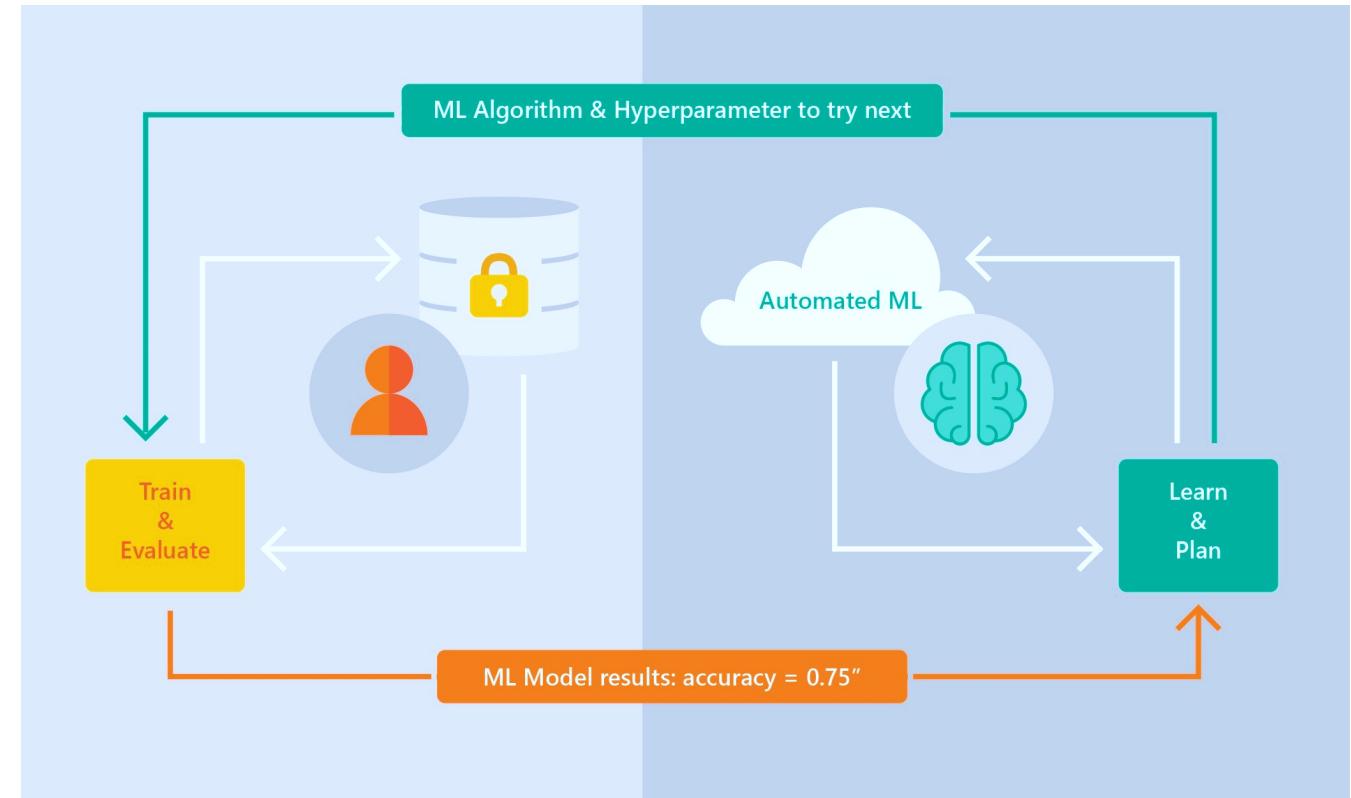
Regression



Time Series
Forecasting

Automated ML Capabilities

- ML Scenarios: Classification & Regression, Forecasting
- Languages: Python SDK for deployment and hosting for inference – Jupyter notebooks
- Training Compute: Local Machine, AML Compute, Data Science Virtual Machine (DSVM), Azure Databricks*
- Transparency: View run history, model metrics, explainability*
- Scale: Faster model training using multiple cores and parallel experiments



* In Preview

Guardrails:

Detection and auto-correction of data issues



Class imbalance



Train-Test split, CV, rolling CV



Missing value imputation



Detect high cardinality features

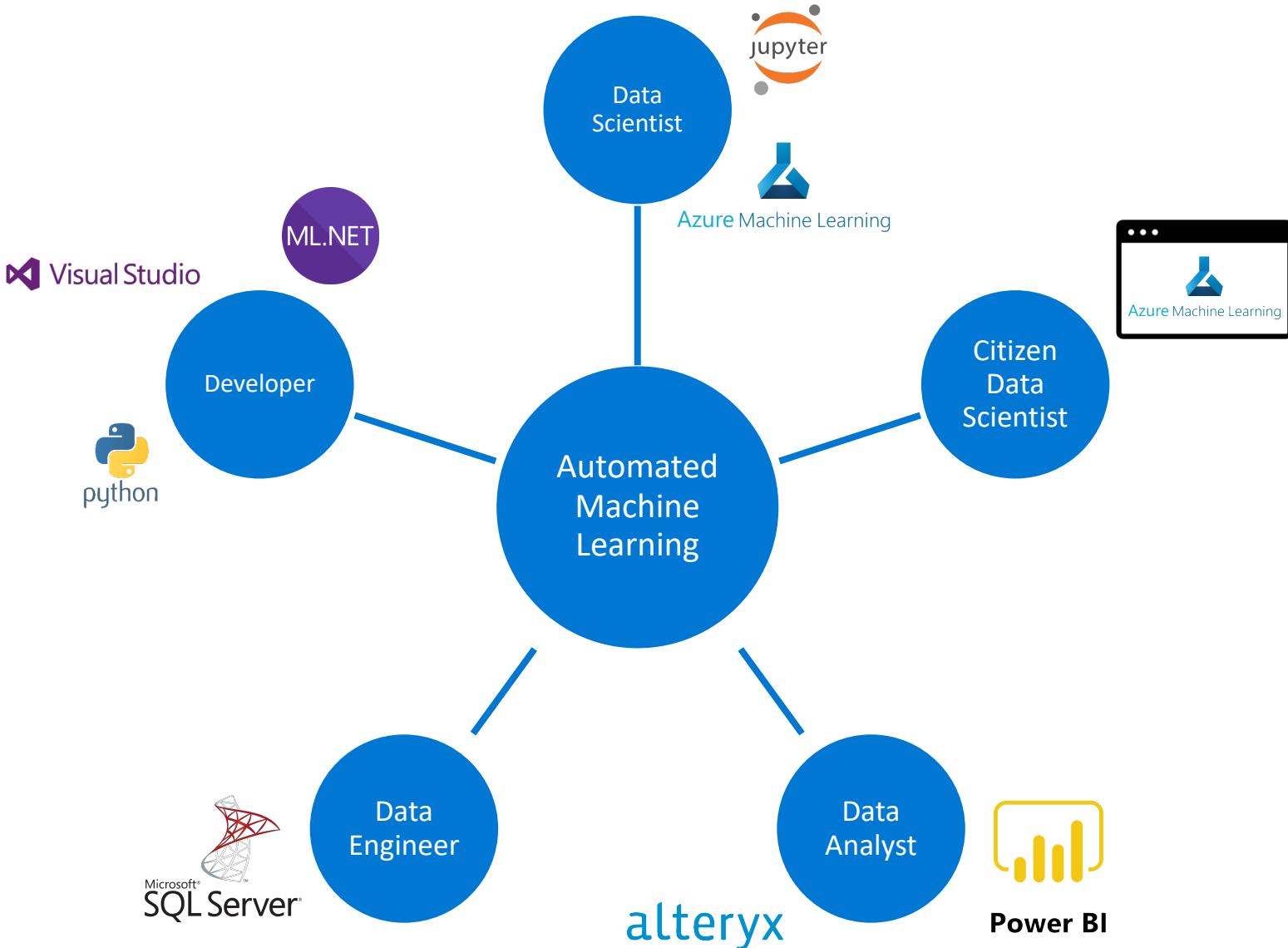


Detect leaky features



Detect overfitting

Machine Learning for Everyone



Customer Testimonials



With one line of code, it runs through different algorithms within the prediction family and the different parameter (or variable) combos that previously were manually tested by the scientists. The power of the cloud comes in here. The results are comparable to what the data scientists produced.

Manish Naik, BP, Digital Innovation



Auto ML's execution of different models was an impressive that enabled data scientists to work **iteratively** on machine learning experiments to increase auction sales by 10% and optimize the time auction cars are kept in the showroom to less than 30 days.

Farika Maharani, PT. Serasi Autoraya, Data Platform Supervisor



The CBRE AI and Data Engineering Team have successfully deployed a complete Azure Machine Learning model to their new API gateway leveraging the Azure AutoML solution in Azure Databricks. The API Gateway plus the model deployment goes into production this March.

Francis Dogbey, Microsoft CSA



In evaluating Azure Automated ML we discovered real potential in shortening the time to market for producing predictive models. The availability of the Automated ML UI also holds the promise of opening the ML space to non data science trained resources which in turn allows the democratizing of the predictive work without the pain of hiring expensive/ hard to retain staff.

Bogdan Rosca, Senior Director, Principal Information Architect



We see advantages moving over to Azure AutoML because we think we will be able to increase our speed to create models significantly and do more with less in terms of labor hours.

Dan Metzendorf, Data Science Manager, The Sherwin Williams Company



AutoML resulted in a significant improvement in model performance (1) Consistently produced better models than other automated ml libraries (TPOT) (2) Also outperformed hand-tuned models. AutoML explored a solution space larger than what was plausible to do manually

David Robinson, Devon Energy Data Scientist



Walgreens Boots Alliance

The reason we see the sharp uplift in sales is the customers are getting content that really connects with them, and they're getting offers for things that are truly relevant and relevant at that moment in time.... Microsoft—they are really wanting to be our partners and were really going to help us on this journey, which was very differentiating

Daniel Humble, Chief Data and Analytics Officer, Walgreens Boots Alliance

If I have 200 models to train—I can just do this all at once. It can be farmed out to a huge computer cluster, and it can be done in minutes so I'm not waiting for days or setting experiments to run over the weekend anymore.

Dean Riddlesden, Senior Data Scientist, Walgreens Boots Alliance



With automated machine learning in Azure Machine Learning service, we can focus our testing on the most accurate models and avoid testing a large range of less valuable models, because it retains only the ones we want. That saves months of time for us.

Matthieu Boujonnier: Analytics Application Architect and Data Scientist, Schneider Electric

Models evolve over time. And we use automated machine learning to speed that process, from four months for our first-generation models to a day for our newest models.

Loryne Bissuel-Beauvais: Data Scientist, Schneider Electric

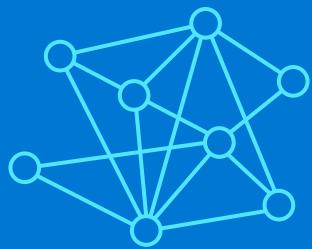


We tried AutoML for aspect ratio model and pleased to see AutoML produced much better model than our baseline. We need to build almost 50 models and are looking forward to the productivity boost we will get by not hand tuning each one of them!

Saurabh Naik, Sr. Software Engineer



Teams uses machine learning to analyze, gain insights and improve the quality of calls. We use AutoML to significantly scale up the application of ML solutions by semi-automating model train tasks



What's new?

Latest announcements ([Blog post with all the announcements](#))

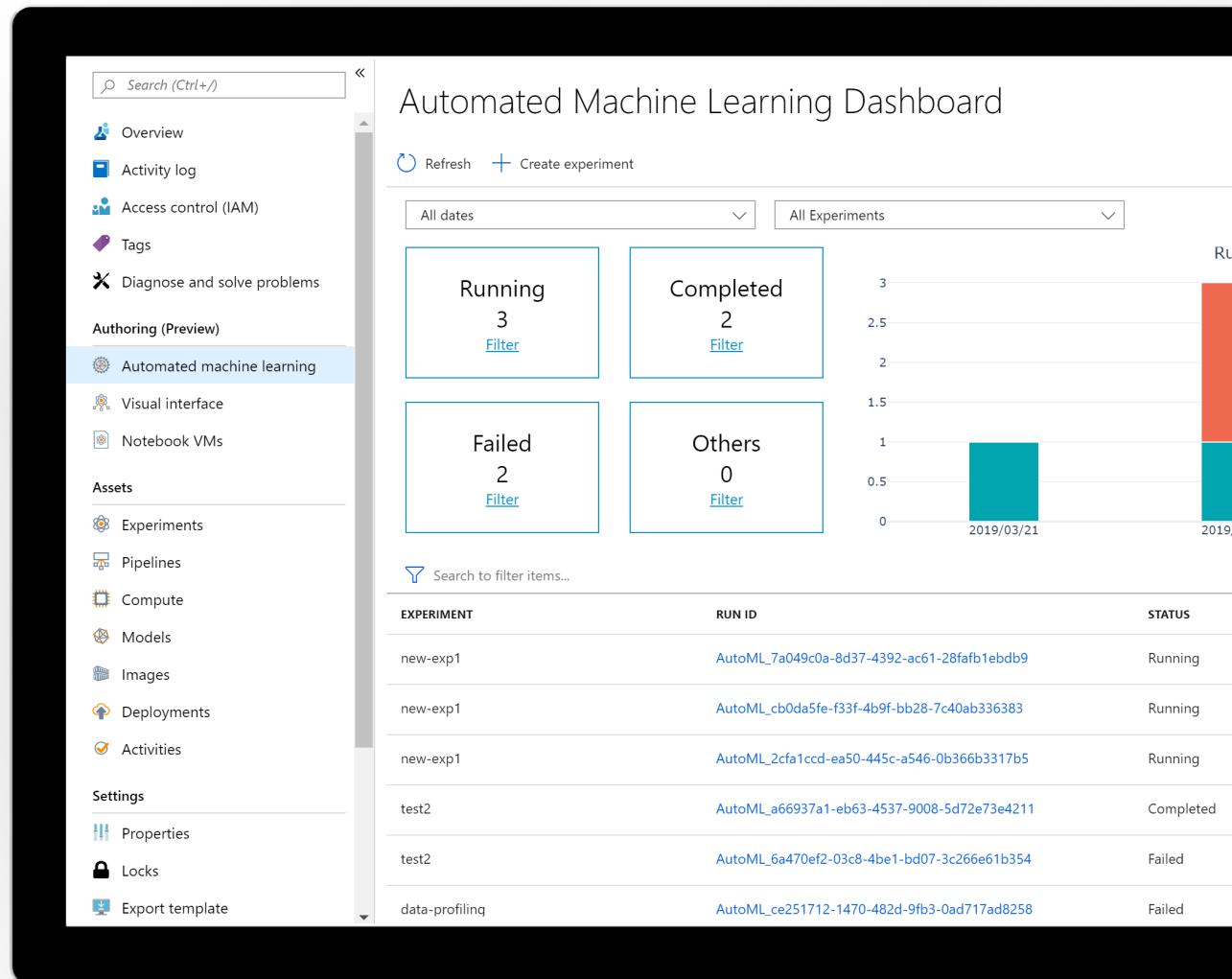
Automated ML UI in Azure portal (Preview)

- End-to-end no-code experience for non-data scientists to train ML models
- Classification, Regression, Forecasting
- Deploy models easily and quickly
- Advanced settings for power users to tune the training job

[Blog post](#)

Coming up next

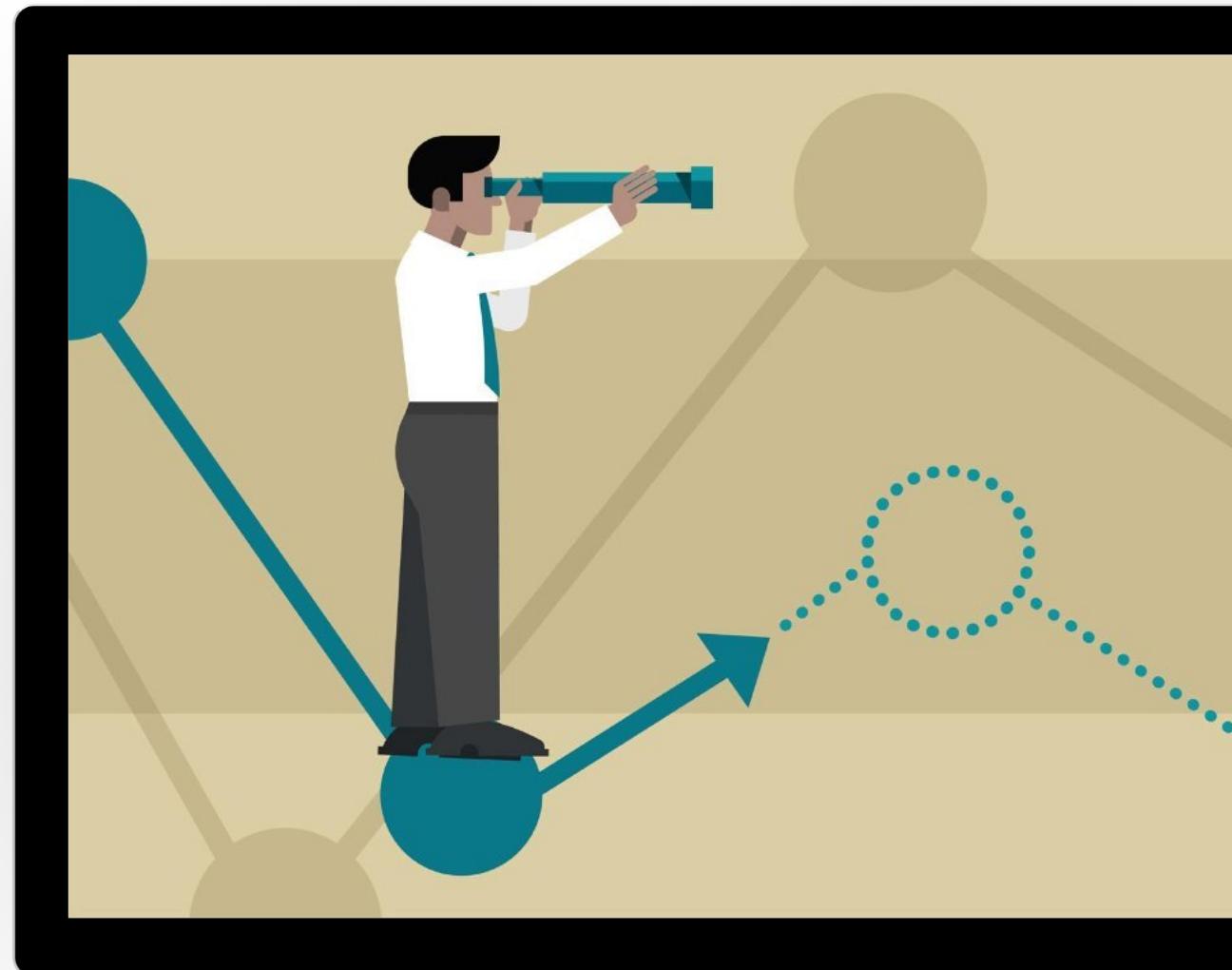
- Model explainability
- Additional data sources (with Datasets)
- Re-run experiments



Latest announcements ([Blog post with all the announcements](#))

Time Series Forecasting Generally Available

- Rolling cross validation splits for time series data
- Configurable lags
- Window aggregation
- Holiday featurizer



Latest announcements ([Blog post with all the announcements](#))

Feature engineering updates

- Additional data guardrails and synthetic features
- Added XGBoost algorithm
- Improved transparency retrieving the engineered features

Coming up next

- Improved feature sweeping, text featurization
- Transparency: Get auto-featurized data

Preprocessing steps	Description
Drop high cardinality or no variance features	Drop these from training and validation sets, including features with all values missing, same value across all rows or with extremely high cardinality (for example, hashes, IDs, or GUIDs).
Impute missing values	For numerical features, impute with average of values in the column. For categorical features, impute with most frequent value.
Generate additional features	For DateTime features: Year, Month, Day, Day of week, Day of year, Quarter, Week of the year, Hour, Minute, Second. For Text features: Term frequency based on unigrams, bi-grams, and tri-character-grams.
Transform and encode	Numeric features with few unique values are transformed into categorical features. One-hot encoding is performed for low cardinality categorical; for high cardinality, one-hot-hash encoding.
Word embeddings	Text featurizer that converts vectors of text tokens into sentence vectors using a pre-trained model. Each word's embedding vector in a document is aggregated together to produce a document feature vector.
Target encodings	For categorical features, maps each category with averaged target value for regression problems, and to the class probability for each class for classification problems. Frequency-based weighting and k-fold cross validation is applied to reduce over fitting of the mapping and noise caused by sparse data categories.
Text target encoding	For text input, a stacked linear model with bag-of-words is used to generate the probability of each class.
Weight of Evidence (WoE)	Calculates WoE as a measure of correlation of categorical columns to the target column. It is calculated as the log of the ratio of in-class vs out-of-class probabilities. This step outputs one numerical feature column per class and removes the need to explicitly impute missing values and outlier treatment.
Cluster Distance	Trains a k-means clustering model on all numerical columns. Outputs k new features, one new numerical feature per cluster, containing the distance of each sample to the centroid of each cluster.



Workshop

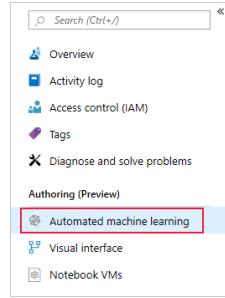
Prerequisites

1. Azure Subscription
2. Azure ML workspace



Automated ML using Azure Portal UI: Car price prediction

Follow instructions: <https://docs.microsoft.com/en-us/azure/machine-learning/service/how-to-create-portal-experiments>



Navigate to the left pane of your workspace. Select Automated Machine Learning under the Authoring (Preview) section.

Enter your experiment name, then select a compute from the list of your existing computes or create a new compute

Create a new automated machine learning experiment

[← Back](#)

Experiment name *

Select a compute * [?](#)

[Create a new compute](#) [Refresh compute](#)

Select a data file from your storage container, or upload a file from your local computer to the container

Car price dataset downloadable from:

[https://automlpmdemows6960037818.blob.core.windows.net/sampledata/Automobile price data.csv](https://automlpmdemows6960037818.blob.core.windows.net/sampledata/Automobile%20price%20data.csv)

Create a new automated machine learning experiment

[Back](#)

Experiment name *

Select a compute *

Storage Account

Storage Container

Select a CSV/TSV data file, or upload from your local computer

Preview data and keep all columns selected for training

Select the training job type: **regression**

Select target column: **price**

Open “Advanced settings”, set **training job time** to **10** minutes (for the sake of the workshop)

Hit

Start

And wait for the training job to start. You'll be able to see the models which are created during the run

Once the run is completed, click **deploy the best model**

Deploy Best Model

VotingEnsemble

Once deployed, follow [instructions](#) to consume from Power BI

Power BI | AI test | Hotel Reviews Dataflow

Edit queries

Get data Refresh Options Manage columns Transform table Reduce rows Add column All insights Map to standard Combine tables

Navigation[!\$schema="dbo", \$layer="Hotel_Reviews"]([Data])

	#	city	country	latitude	longitude	name	province	reviewer_id	reviewer_name	reviewer_state
1	Homes	Princetown	US	27.270	-129.481	Hotel 1	HI	5/20/2015, 5:00:00		6/1/2015, 5:00:00
2	Homes	Princetown	US	27.220	-129.481	Hotel 2	HI	6/1/2015, 5:00:00		6/1/2015, 5:00:00
3	Vacation_Rentals_Resorts_B...	US	27.280	-129.481	Hotel 3	HI	6/1/2015, 5:00:00		6/1/2015, 5:00:00	
4	Homes	Princetown	US	27.220	-129.481	Hotel 4	HI	6/23/2015, 5:00:00		6/23/2015, 5:00:00
5	Vacation_Rentals_Resorts_B...	Honolulu	US	21.282	-157.831	Hotel 5	HI	8/16/2015, 5:00:00		5/31/2015, 5:00:00
6	Vacation_Rentals_Resorts_B...	Kapaa	US	22.043	-159.481	Hotel 6	HI	8/16/2015, 5:00:00		5/31/2015, 5:00:00
7	Homes	Princetown	US	27.220	-129.481	Hotel 7	HI	6/1/2015, 5:00:00		6/1/2015, 5:00:00
8	Homes	Princetown	US	27.270	-129.481	Hotel 8	HI	6/1/2015, 5:00:00		6/1/2015, 5:00:00
9	Homes	Princetown	US	27.220	-129.481	Hotel 9	HI	6/1/2015, 5:00:00		6/1/2015, 5:00:00
10	Homes	Princetown	US	27.220	-129.481	Hotel 10	HI	7/29/2015, 5:00:00		7/29/2015, 5:00:00
11	Homes	Princetown	US	27.220	-129.481	Hotel 11	HI	6/1/2015, 5:00:00		6/1/2015, 5:00:00
12	Vacation_Rentals_Resorts_B...	Honolulu	US	21.282	-157.831	Hotel 12	HI	7/1/2015, 5:00:00		7/1/2015, 5:00:00
13	Homes	Princetown	US	27.220	-129.481	Hotel 13	HI	7/1/2015, 5:00:00		7/1/2015, 5:00:00
14	Vacation_Rentals_Resorts_B...	Honolulu	US	21.282	-157.831	Hotel 14	HI	8/2/2015, 5:00:00		8/2/2015, 5:00:00
15	Vacation_Rentals_Resorts_B...	Honolulu	US	21.282	-157.831	Hotel 15	HI	8/16/2015, 5:00:00		8/16/2015, 5:00:00
16	Vacation_Rentals_Resorts_B...	Honolulu	US	21.282	-157.831	Hotel 16	HI	12/10/2015, 4:00:00		12/10/2015, 4:00:00
17	Vacation_Rentals_Resorts_B...	Honolulu	US	21.282	-157.831	Hotel 17	HI	7/29/2015, 5:00:00		7/29/2015, 5:00:00
18	Vacation_Rentals_Resorts_B...	Honolulu	US	21.282	-157.831	Hotel 18	HI	7/15/2015, 4:00:00		7/15/2015, 4:00:00
19	Vacation_Rentals_Resorts_B...	Honolulu	US	21.282	-157.831	Hotel 19	HI	12/20/2015, 4:00:00		12/20/2015, 4:00:00
20	Vacation_Rentals_Resorts_B...	Honolulu	US	21.282	-157.831	Hotel 20	HI	3/26/2016, 5:00:00		3/26/2016, 5:00:00
21	Homes	Kapaa	US	22.043	-159.481	Hotel 21	HI	7/30/2015, 5:00:00		7/30/2015, 5:00:00
22	Vacation_Rentals_Resorts_B...	Honolulu	US	21.282	-157.831	Hotel 22	HI	10/4/2015, 5:00:00		10/4/2015, 5:00:00
23	Vacation_Rentals_Resorts_B...	Honolulu	US	21.282	-157.831	Hotel 23	HI	12/20/2015, 4:00:00		12/20/2015, 4:00:00
24	Vacation_Rentals_Resorts_B...	Honolulu	US	21.282	-157.831	Hotel 24	HI	3/1/2016, 4:00:00		3/1/2016, 4:00:00
25	Vacation_Rentals_Resorts_B...	Honolulu	US	21.282	-157.831	Hotel 25	HI	8/16/2015, 5:00:00		8/16/2015, 5:00:00
26	Vacation_Rentals_Resorts_B...	Honolulu	US	21.282	-157.831	Hotel 26	HI	10/25/2015, 1:00:00		10/25/2015, 1:00:00
27	Vacation_Rentals_Resorts_B...	Princetown	US	27.220	-129.481	Hotel 27	HI	5/20/2016, 5:00:00		5/20/2016, 5:00:00
28	Homes	Princetown	US	27.220	-129.481	Hotel 28	HI	4/19/2018, 11:00:00		

Get Data Cancel Done

Automated ML using Notebook VM : Energy Demand

1. Login into Azure Portal : <https://ms.portal.azure.com/#home>
2. Open your Machine Learning Service Workspace.

The screenshot shows the Azure Portal interface for a workspace named 'myws_nilesha2'. The left sidebar includes sections for Overview, Activity log, Access control (IAM), Tags, Diagnose and solve problems, Authoring (Preview), Notebook VMs (which is selected and highlighted in blue), Visual interface, Assets, Experiments, Pipelines, Compute, Models, Images, Deployments, and Activities. The main content area displays details about the workspace, including Resource group: 'myws_nilesha2', Location: 'East US 2', Subscription: 'AutoML Demo', and Application Insights: 'mywsnileshabf0fb0y'. Below this, there's a 'Getting Started' section with links for creating notebooks, automated machine learning models, building models using a visual interface, viewing documentation, and more.

3. Click on Notebook VM

This screenshot shows the 'Notebook VMs' section within the 'myws_nilesha2' workspace. The left sidebar is identical to the previous screenshot. The main area lists a single notebook VM named 'automlnilesha' with a status of 'Stopped'. At the top of the list, there is a red box around the 'New' button, which is used to create a new VM.

4. Click "New". If you have an existing VM go to step 6

This screenshot shows the 'Notebook VMs (Preview)' list. The 'New' button at the top left is highlighted with a red box. The table below lists one existing VM: 'automlnilesha' with a status of 'Stopped'. The columns include Name, Status, URI, Created By, VM Size, and Creation Time.

Name	Status	URI	Created By	VM Size	Creation Time
automlnilesha	Stopped	N/A	nilesha@microsoft.com	STANDARD_D4	May 28, 2019 11:59 AM

5. Name your VM and select machine size and click on create

The screenshot shows the 'New Notebook VM' dialog. It has fields for 'Name' (set to 'AutoMLVM') and 'Virtual machine size' (set to 'STANDARD_D3_V2'). There are 'Create' and 'Cancel' buttons at the bottom. The 'Create' button is highlighted with a red box.

6. After VM has started (~10mins) click on Jupyter link

This screenshot shows the 'Notebook VMs (Preview)' list again. The VM 'automl' is listed with a status of 'Running'. The 'Jupyter' link next to it is highlighted with a red box. Other columns include Name, Status, Created By, VM Size, and Creation Time.

7. Click on root folder > Samples-x.x.xx > how-to-use-azureml > automated-machine-learning > forecasting-energy-demand > /auto-ml-forecasting-energy-demand.ipynb

8. Follow instructions in notebook executing each cell.

Resources

<http://aka.ms/amlfree>

Learn more : <https://aka.ms/automatedmldocs>

Notebook Samples : <https://aka.ms/automatedmlsamples>

Blog Post : <https://aka.ms/AutomatedML>

Product Feedback : AskAutomatedML@microsoft.com



Thank You!



AI Platform Team