

Automated Income Classification: AutoML in Azure

Step by step procedure for Automl in Azure

Creating a Dataset by uploading a csv file

The screenshot shows the Azure Machine Learning Studio interface for creating a new Automated ML job. The left sidebar contains navigation options: All workspaces, Home, Model catalog, Authoring (Notebooks, Automated ML, Designer), Assets (Data, Jobs, Components, Pipelines, Environments, Models, Endpoints), and Manage (Compute, Monitoring, Data Labeling). The main panel is titled 'Create a new Automated ML job' and shows a progress bar with five steps: Select data asset (active), Configure job, Select task and settings, Hyperparameter configuration (Computer Vision only), and Validate and test. The 'Select data asset' step is expanded, displaying a table of existing datasets. A success message at the top indicates that a new dataset 'Laptop_price' was created successfully. The table lists four datasets: 'new', 'Flight', 'Insurance', and 'Diabetes', all of which are tabular. The 'Created on' and 'Modified on' columns show the timestamps for each dataset. At the bottom of the main panel, there are 'Back', 'Next', and 'Cancel' buttons.

Name	Dataset type	Created on ↓	Modified on
new	Tabular	Jun 14, 2023 8:05 AM	Jun 14, 2023 8:05 A...
Flight	Tabular	Jun 14, 2023 12:47 AM	Jun 14, 2023 12:47 ...
Insurance	Tabular	Jun 11, 2023 5:34 PM	Jun 11, 2023 5:34 P...
Diabetes	Tabular	Jun 10, 2023 4:58 PM	Jun 10, 2023 4:58 P...

Selecting the target column and selecting the compute

Selecting the target column and selecting the compute

The screenshot shows the 'Configure job' step in the Azure AI Machine Learning Studio. The left sidebar contains a navigation menu with sections: All workspaces, Home, Model catalog, Authoring (Notebooks, Automated ML, Designer), Assets (Data, Jobs, Components, Pipelines, Environments, Models, Endpoints), and Manage (Compute, Monitoring, Data Labeling). The main area is titled 'Create a new Automated ML job' and shows a progress bar with steps: Select data asset, Configure job (current), Select task and settings, Hyperparameter configuration (Computer Vision only), and Validate and test. The 'Configure job' section includes: 'Data asset' set to 'Prediction (View data asset)'; 'Experiment name' with radio buttons for 'Select existing' (selected) and 'Create new'; 'Existing experiment' dropdown set to 'adult_set'; 'Target column' dropdown set to 'income (String)'; 'Select compute type' dropdown set to 'Compute cluster'; and 'Select Azure ML compute cluster' dropdown set to 'standard'. There are '+ New' and 'Refresh computes' links. At the bottom are 'Back', 'Next', and 'Cancel' buttons.

Selecting primary metric and exit criterion

The screenshot shows the 'Additional configurations' dialog box. The left sidebar is the same as the previous screenshot. The main area shows 'featurization settings' with options for 'Classification', 'Regression', 'Time series forecasting', 'Natural language processing', and 'Computer vision'. The 'Additional configurations' dialog includes: 'Primary metric' dropdown set to 'Accuracy'; 'Explain best model' checkbox checked; 'Use all supported models' checkbox checked; 'Blocked models' dropdown set to 'A list of models that Automated ML will not use during training.'; 'Exit criterion' section with 'Training job time (hours)' set to '0.5' and 'Metric score threshold' set to 'Metric score threshold'; and a 'Concurrency' section. At the bottom are 'Save' and 'Cancel' buttons.

Selecting the validation type

Default Directory > Plasma > Automated ML > Start job

Create a new Automated ML job

- Select data asset
- Configure job
- Select task and settings
- Hyperparameter configuration (Computer Vision only)
- Validate and test**

Select the validation and test type

You can choose a validation type and select a test data asset as an optional step. Providing your own validation and test data assets are currently preview features.

Validation type ⓘ
Train-validation split

Percentage validation of data ⓘ
30
Automated ML recommends that between 10 and 30 percent of data is held out for validation

Test data asset (preview) ⓘ
No test data asset required

Back Finish Cancel

After completion of automl the best model is voting ensemble

Default Directory > Plasma > Automated ML > adult_set > ivory_parsnip_16fn94vp

ivory_parsnip_16fn94vp ⚙️ ⭐️ ✅ Completed

Overview Data guardrails **Models** Outputs + logs Child jobs

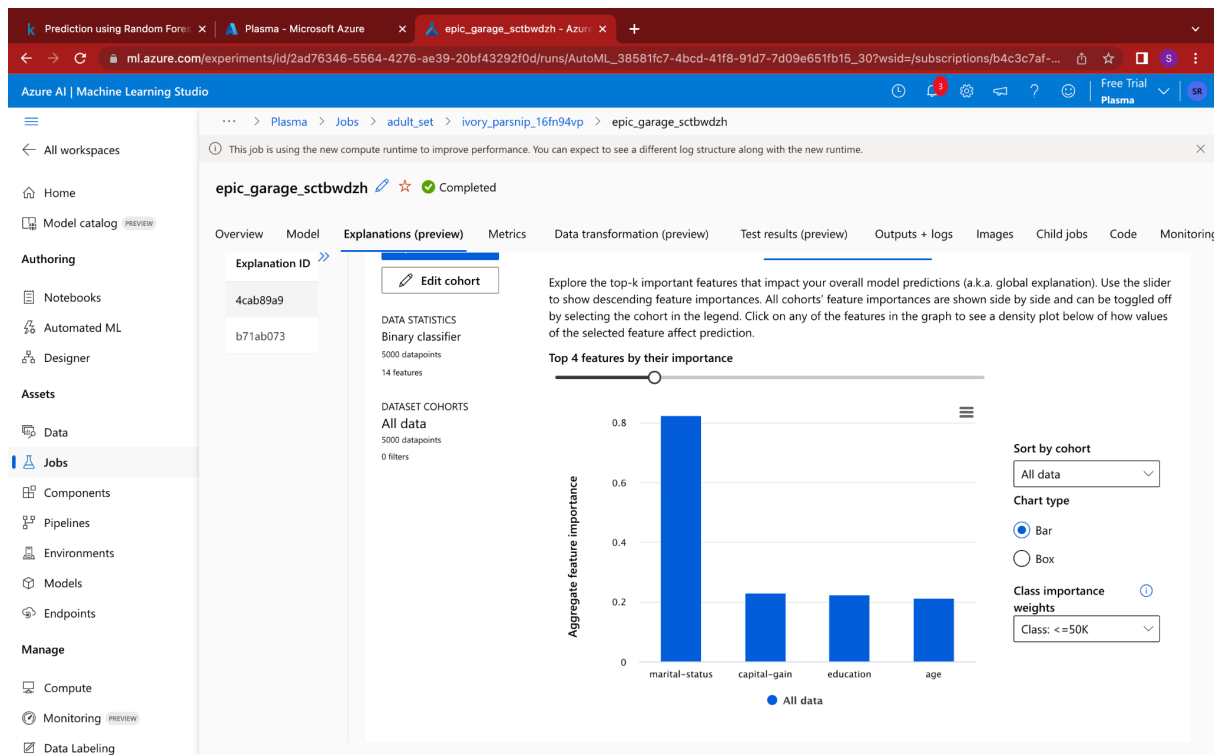
Refresh Deploy Download Explain model View generated code View options

Search Filter Columns

Algorithm name	Explained	Accuracy ↓	Sampling	Created on	Duration
VotingEnsemble	View explanation	0.87818	100.00 %	Jun 16, 2023 11:47 PM	1m
StackEnsemble		0.87675	100.00 %	Jun 16, 2023 11:48 PM	1m
MaxAbsScaler, LightGBM		0.87648	100.00 %	Jun 16, 2023 11:16 PM	1m
MaxAbsScaler, XGBoostClassifier		0.87566	100.00 %	Jun 16, 2023 11:16 PM	1m
StandardScalerWrapper, XGBoostClassifier		0.87218	100.00 %	Jun 16, 2023 11:38 PM	1m
StandardScalerWrapper, XGBoostClassifier		0.87033	100.00 %	Jun 16, 2023 11:16 PM	1m
StandardScalerWrapper, XGBoostClassifier		0.87033	100.00 %	Jun 16, 2023 11:16 PM	1m
StandardScalerWrapper, XGBoostClassifier		0.87013	100.00 %	Jun 16, 2023 11:16 PM	1m
StandardScalerWrapper, XGBoostClassifier		0.86945	100.00 %	Jun 16, 2023 11:16 PM	1m

<< < Page 1 of 2 > >> 25/Page

Top four features. Marital status is the important feature



Metrics of the completed job

