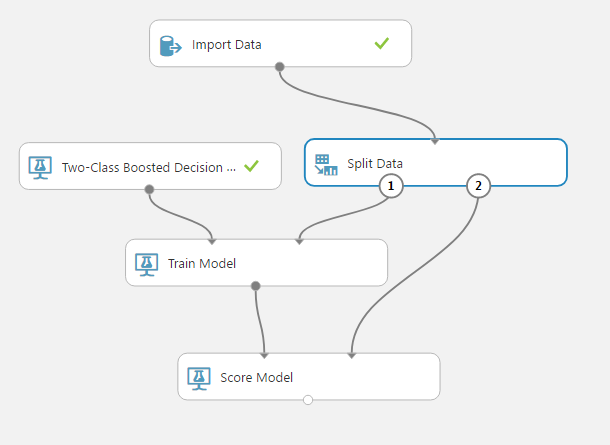


**FIRST EXPERIMENTAL**

In this session

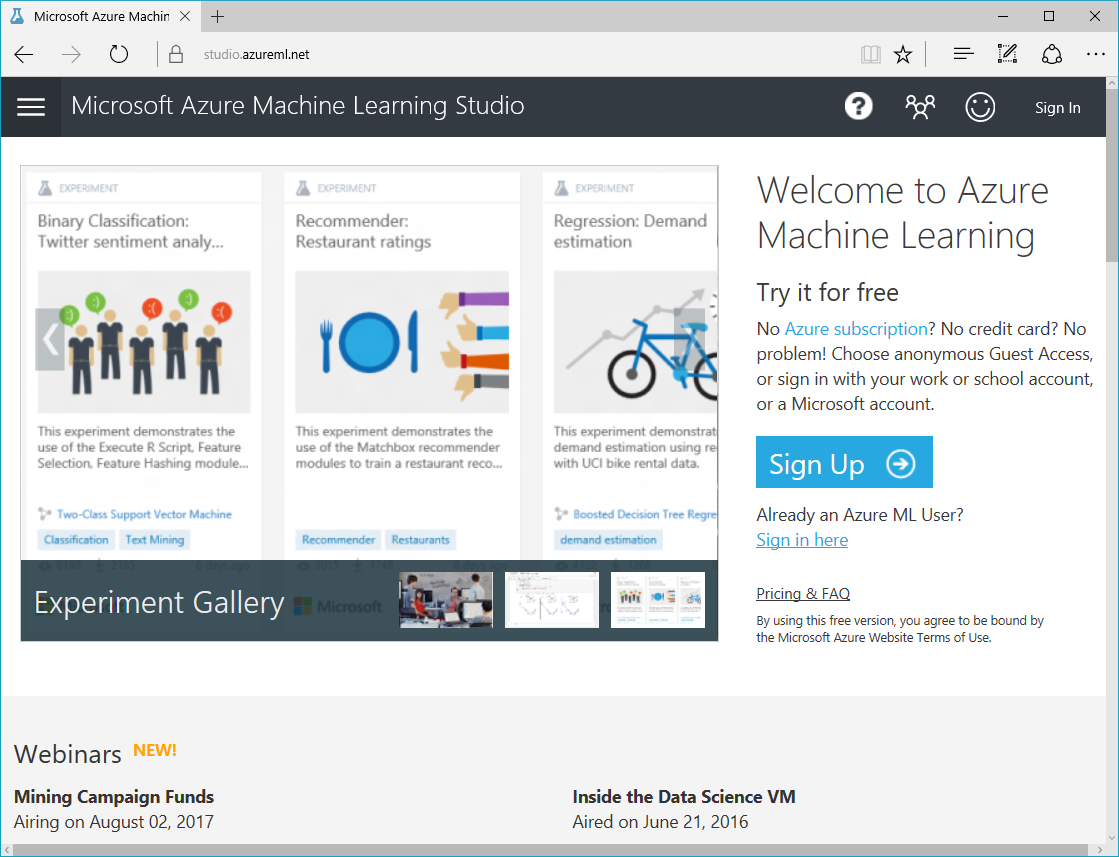
* Sing Up FREE Azure ML Studio Subscription
* Create Azure ML Studio workspace
* Train, Test, Evaluate for Two-Class Boosted Decision Tree
* Import census income dataset
* Create a new Azure Machine Learning experiment
* Train and evaluate a prediction model
* Type of datasets

First experiment model



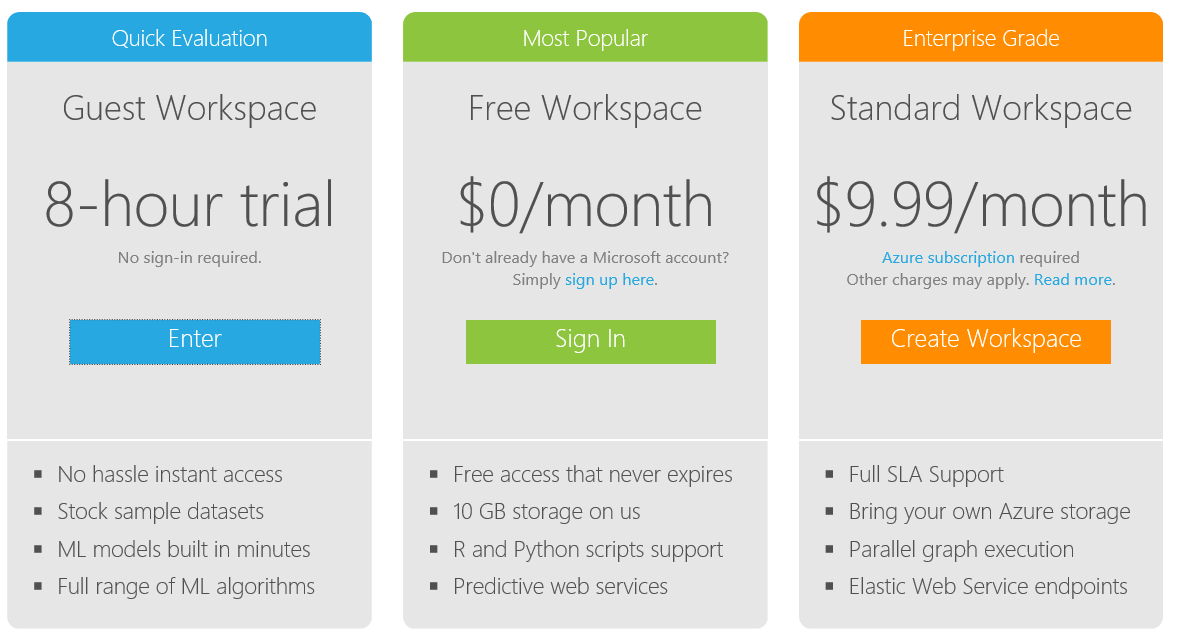
Sing Up FREE Azure ML Studio Subscription

https://studio.azureml.net/

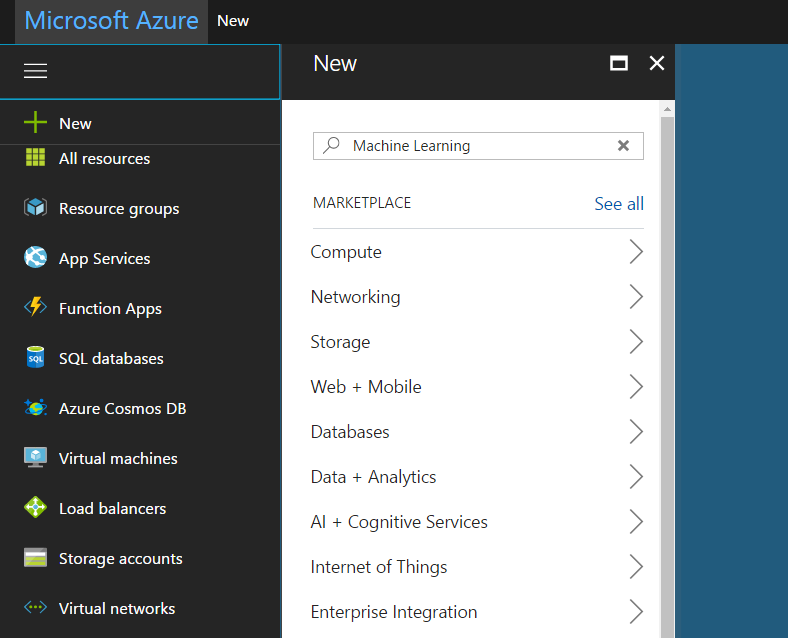


Sing Up FREE Azure ML Studio Subscription

Free Workspace -> sign up here

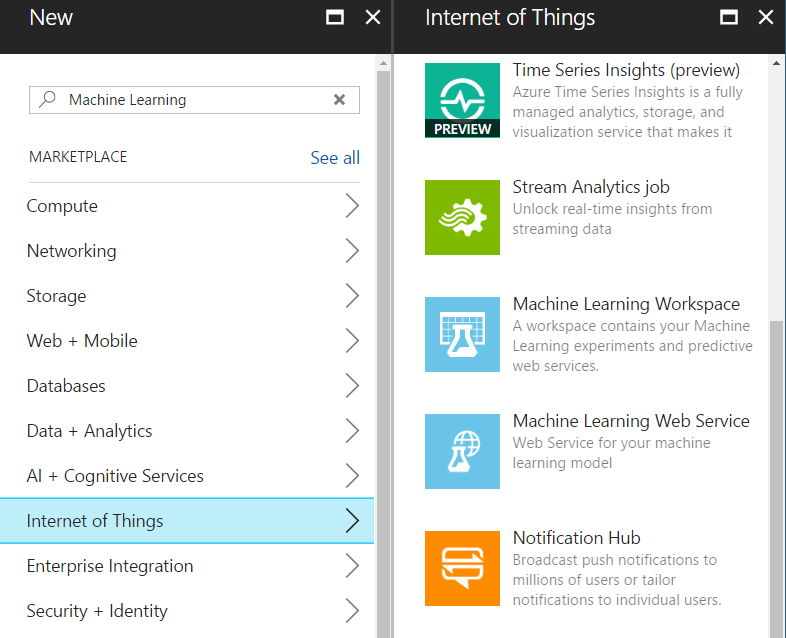


Create Azure ML Studio workspace

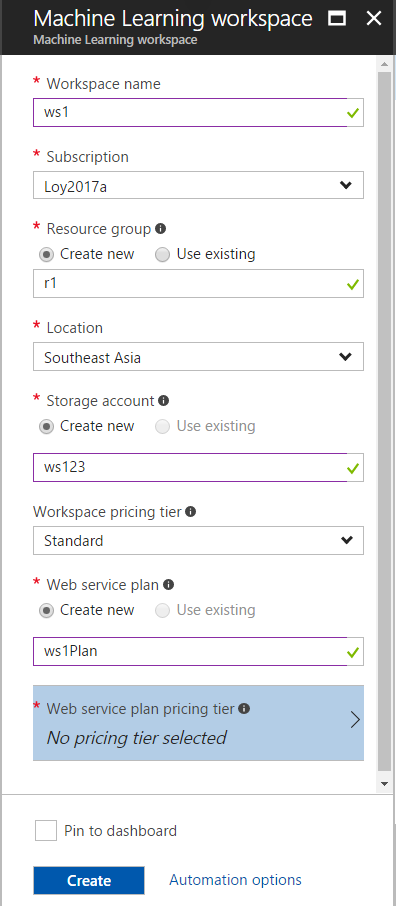
1. Go to the Azure portal https://portal.azure.com
2. Click +New

Create Azure ML Studio workspace

1. Select Internet of Things, click Machine Learning Workspace, then click Create

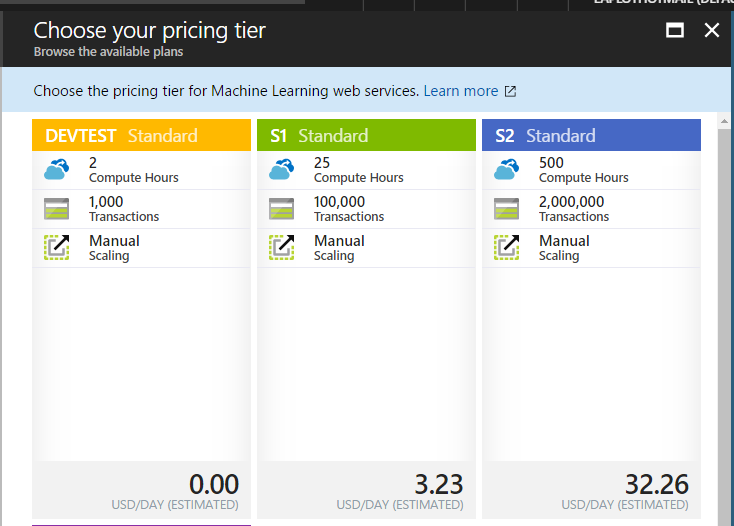


Create Azure ML Studio workspace



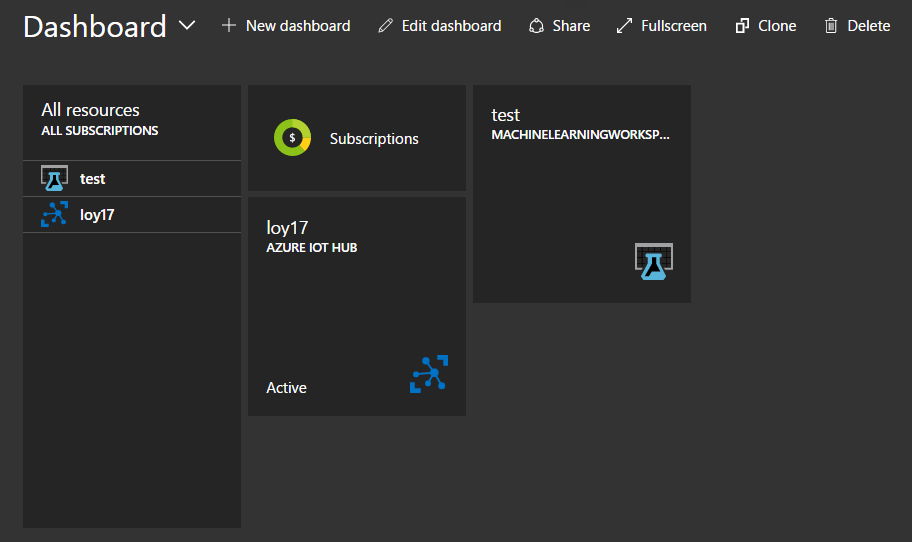
1. Workspace name = ws1
2. Subscription = defult
3. Resource group = Create new: rs1
4. Location = Southeast Asia
5. Storage account = Create new: names1
6. Workspace pricing tier = Standard
7. Web service plan = Create new: ws1Plan

Create Azure ML Studio workspace

1. Click No pricing tier selected
2. Click DEVTEST
3. Click Pin to dashboard
4. Click Create

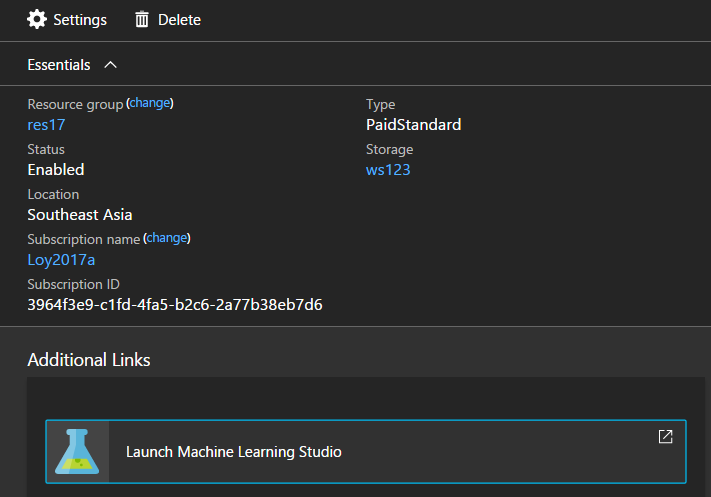
Create Azure ML Studio workspace

1. Click at Machine Learning workgroup on dashboard



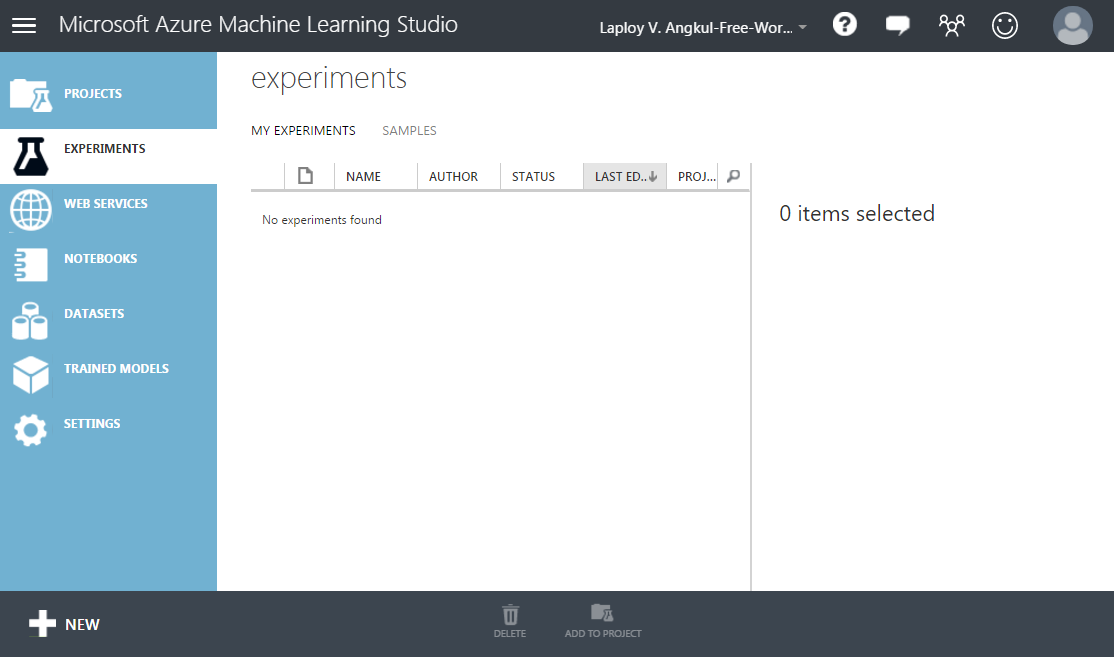
Create Azure ML Studio workspace

1. Click Launch Machine Learning Studio



Create Azure ML Studio workspace

Blank, new ML Studio workspace

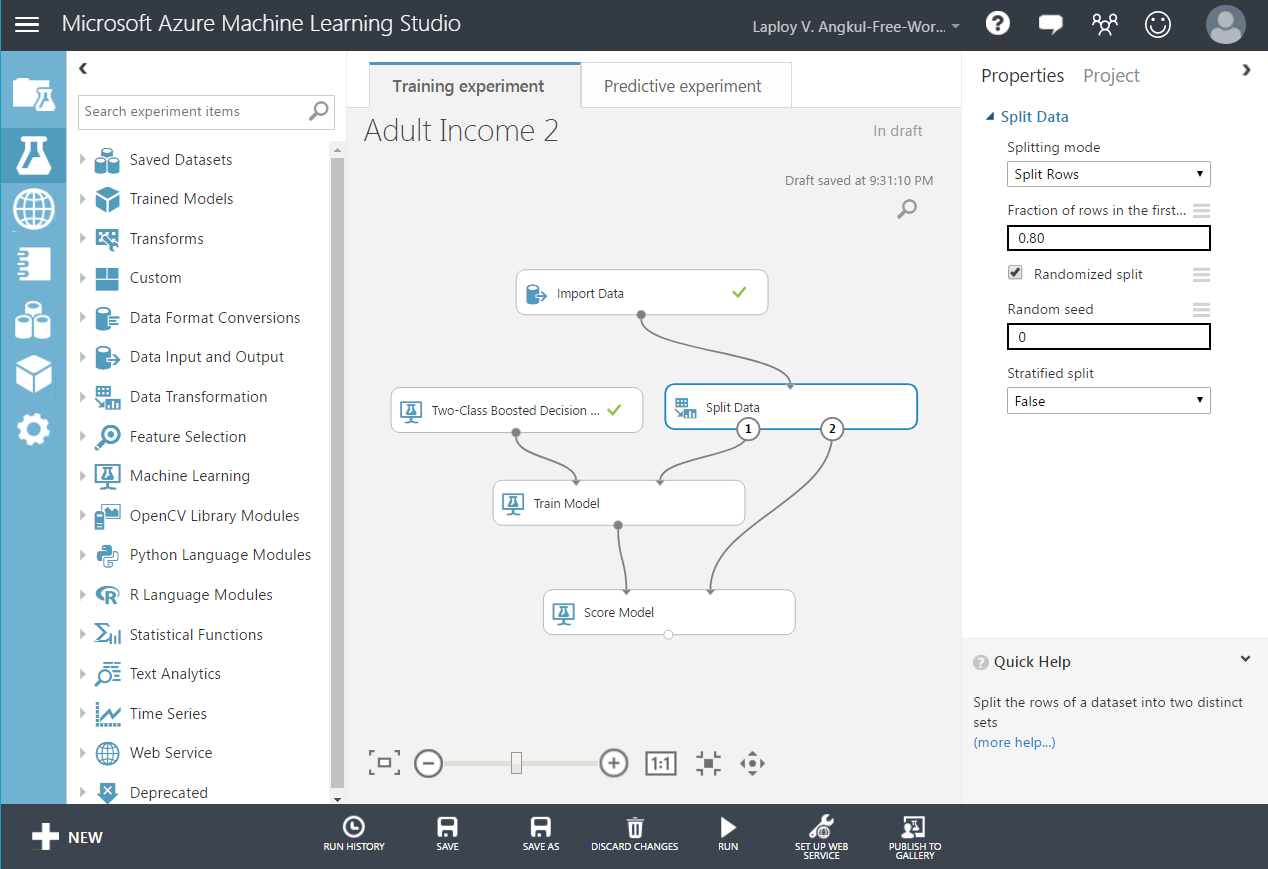


Train, Test, Evaluate for Binary Classification

Predicting whether a person’s income exceeds $50,000 per year based on his demographics or census data

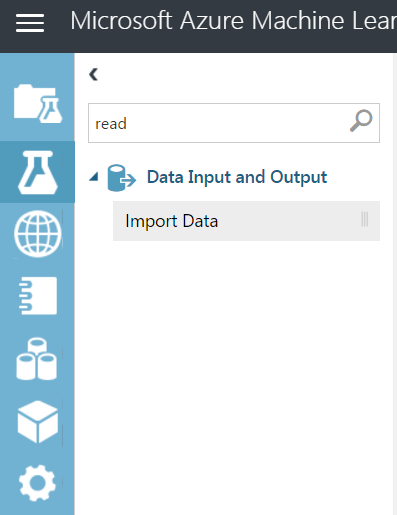
1. Download, prepare, and upload a census income dataset.
2. Create a new Azure Machine Learning experiment.
3. Train and evaluate a prediction model.

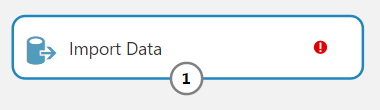
The overall workflow of the experiment



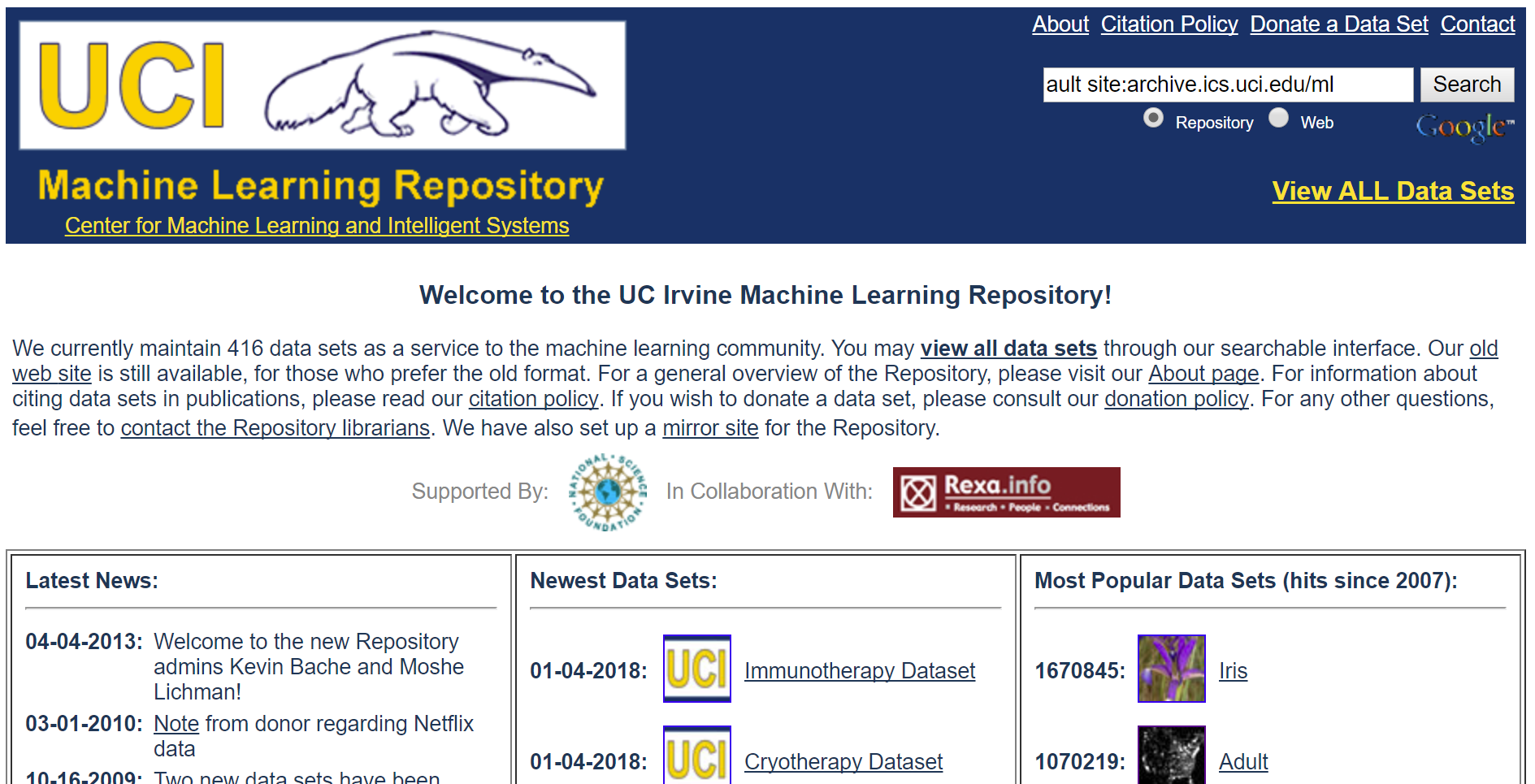
Train, Test, Evaluate for Binary Classification

* Create New blank experiment. Name = Adult Income 1
* Click Data Input and Output
* Drag & drop Import Data



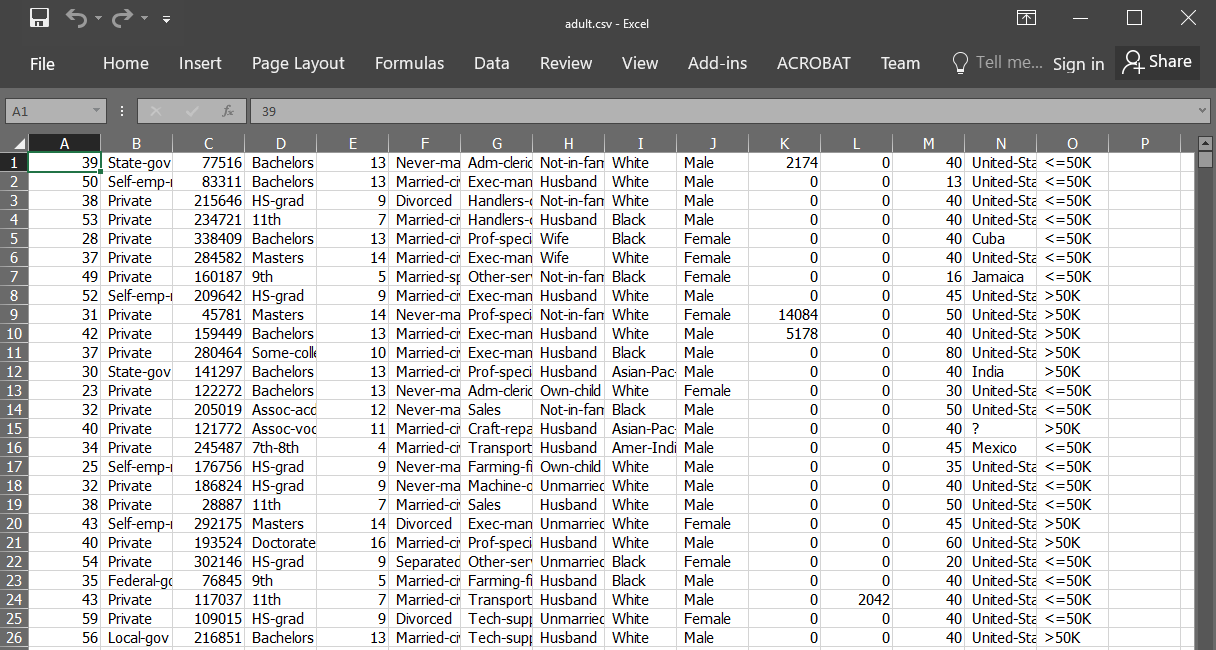


http://archive.ics.uci.edu/ml



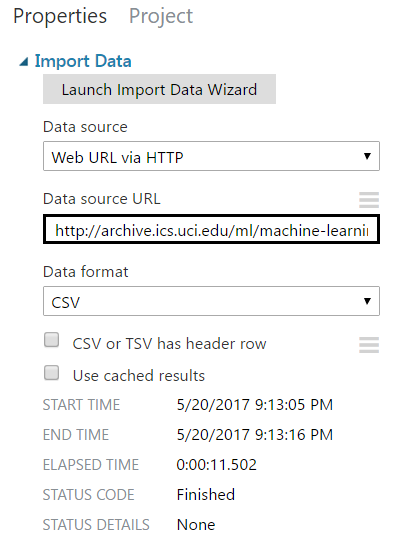
http://archive.ics.uci.edu/ml/datasets/Adult





Train, Test, Evaluate for Binary Classification

Configure Import data module:

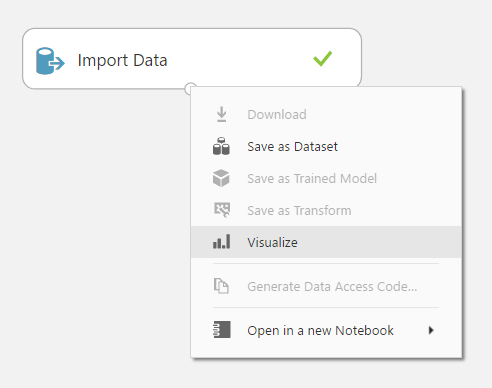
* Data source = Web URL via HTTP
* Data source URL =

http://archive.ics.uci.edu/ml/machine-learning-databases/adult/adult.data

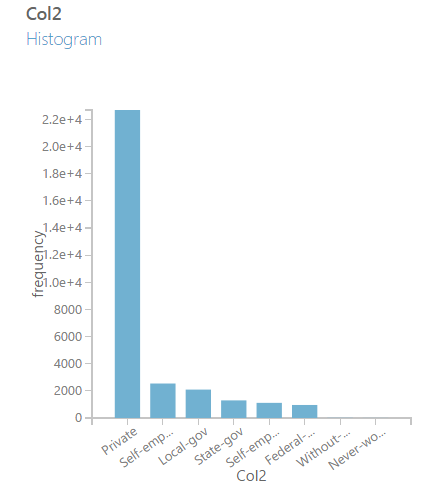
* Data format = CSV
* Run experiment

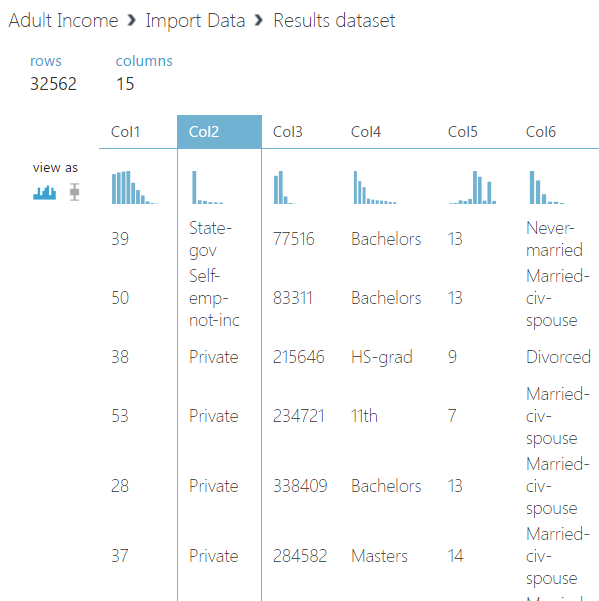
Train, Test, Evaluate for Binary Classification

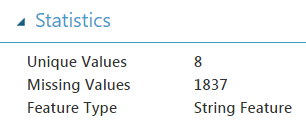
* Right click at the output of Import Data
* Click Visualize



Train, Test, Evaluate for Binary Classification

* Click on Col2
* Look at Statistics and Histogram





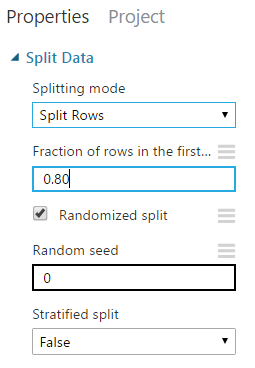
Train, Test, Evaluate for Binary Classification

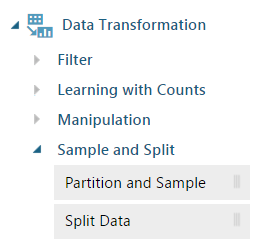
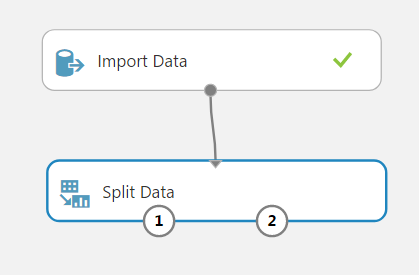
Split up the dataset

* Training data This grouping is used for creating our new predictive model based on the inherent patterns found in the historical data via the ML algorithm we use for the solution.
* Validation data This grouping is used for testing the new predictive model against known outcomes to determine accuracy and probabilities.

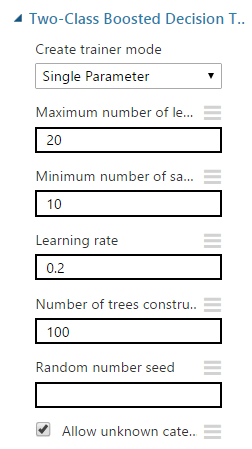
Train, Test, Evaluate for Binary Classification

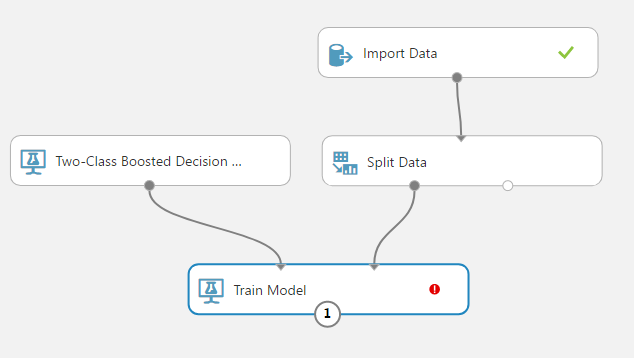
Add Split Data:

* Click Data Transformation
* Click Sample and Split
* Drag & drop Split Data module into canvas
* Connect Import Data to Split Data
* Set properties Fraction of row to 0.80



Train, Test, Evaluate for Binary Classification

* Add Two-Class Boosted Decision Tree and Train Model
* Connect Two-Class Boosted Decision Tree to Train Model
* Connect Split Data to Train Model

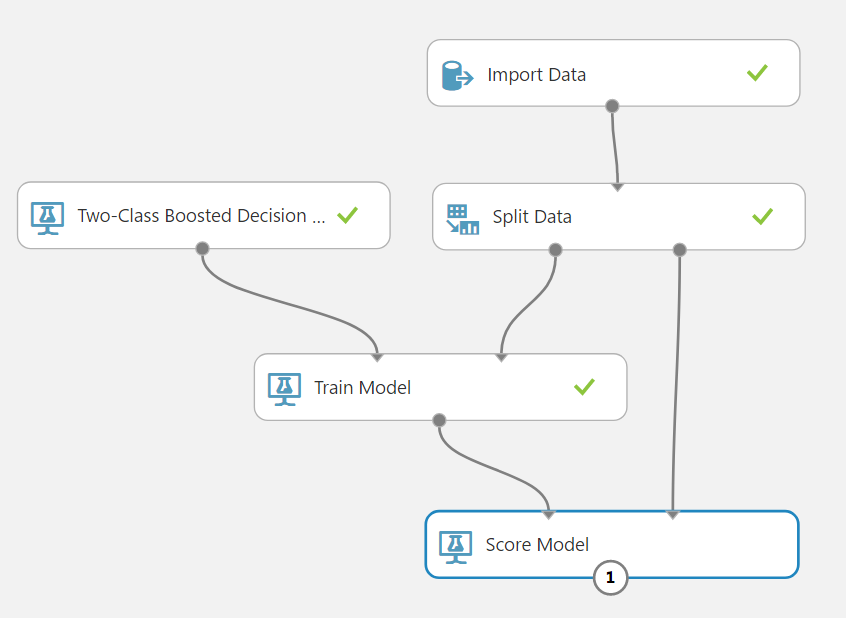


Label column specification

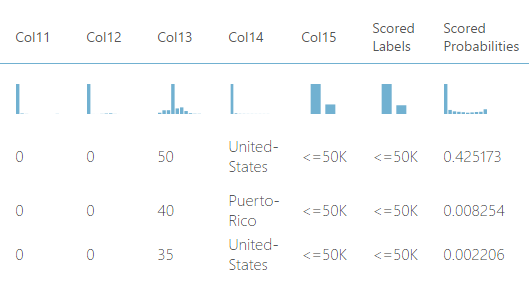
* Click Train Model
* Click Launch column selector
* Include col15
* Click
* Save
* Run

Score the model

Score the model:

* Add Score Model to canvas
* Connect Score Model to Train and Split model
* Run

Visualize the model results



Visualize the model results:

* Visualize output of Score Model
* Scored Labels This column denotes the model’s prediction for this row of the dataset.
* Scored Probabilities This column denotes the numerical probability (or the likelihood) of whether the income level for this row exceeds $50,000.

Type of datasets

**Training set**

* A set of examples used for learning
* Where the answer value is known.

**Validation set**

* A set of examples data
* Used to tune the architecture of a classifier
* And estimate the error

**Test set**

* Use to test the performances of a classifier
* Never used during the training process
* Give estimate of error

More Information

Two-Class Boosted Decision Tree

https://msdn.microsoft.com/en-us/library/azure/dn906025.aspx

Score Model

https://msdn.microsoft.com/en-us/library/azure/dn905995.aspx

Published Experiment

https://gallery.cortanaintelligence.com/Experiment/Adult-Income-1