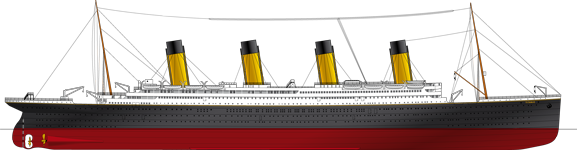
**BUILDING A CLASSIFICATION MODEL**



**In this session**

* Split data
* Select Algorithm
* Train
* Score
* Create web service
* Test web service
* Download Data set
* Data Dictionary
* View data set in Microsoft Excel
* Import Data set
* Create New Experiment
* Prepare Data
* Drop the columns
* Make categorical values
* Replace missing value with median
* Drop rows with missing data
* Create Label

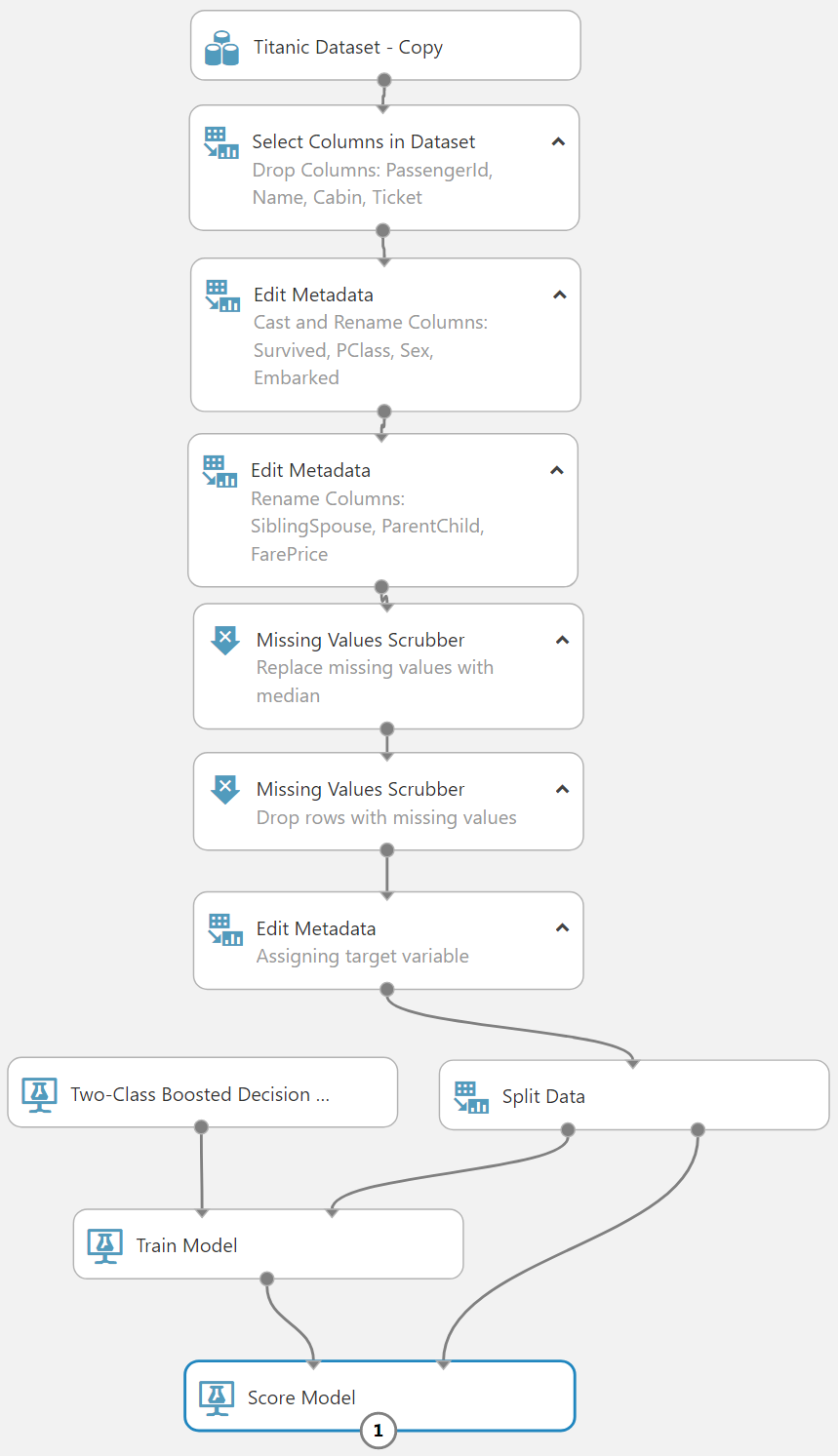
What to do

What to do;

* Create experiment
* Create Classification model
* Using Azure ML.
* Using the Titanic passenger data set
* Build a model for predicting the survival of a given passenger.



ML model when finished



AML model development step

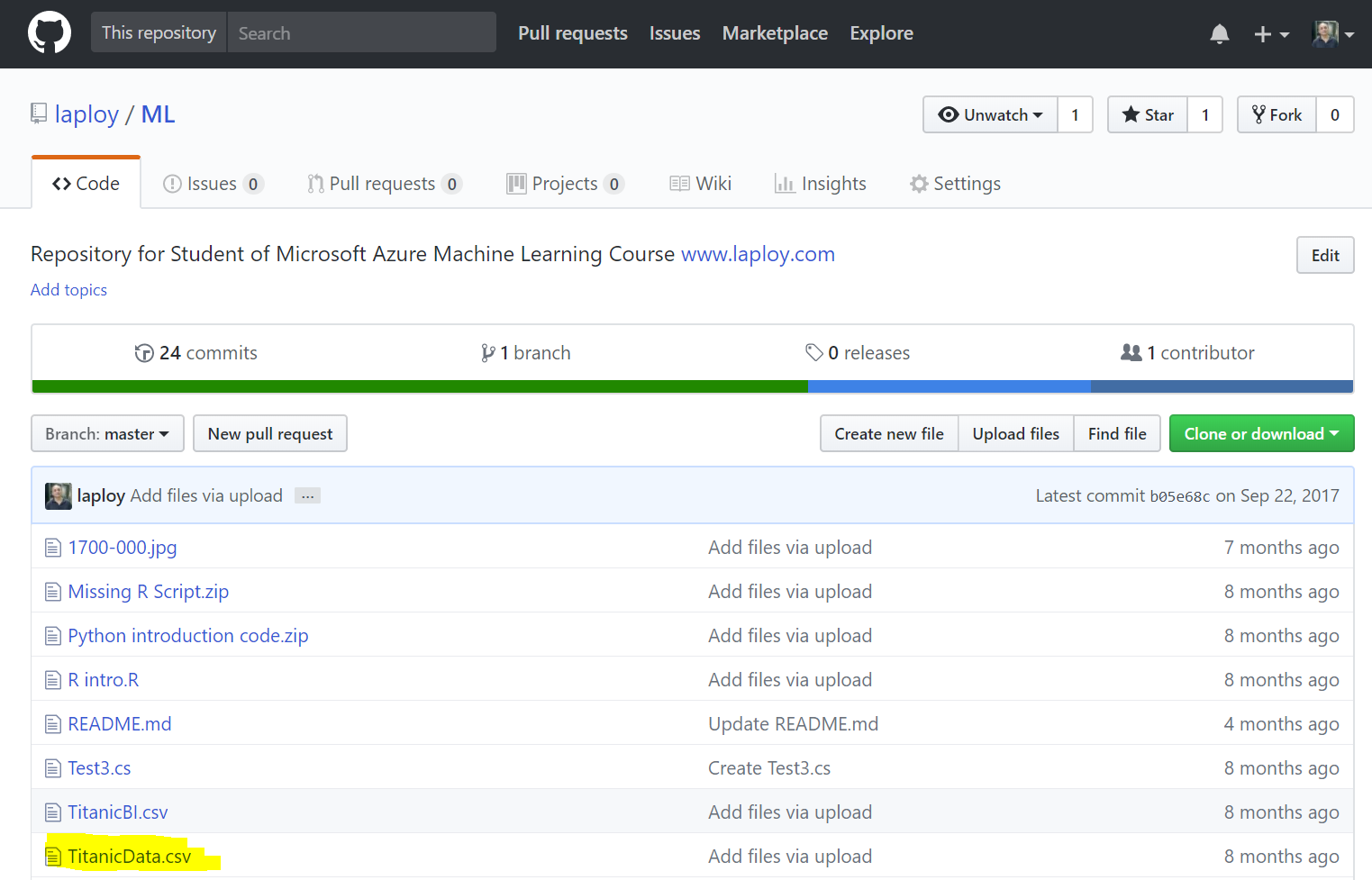


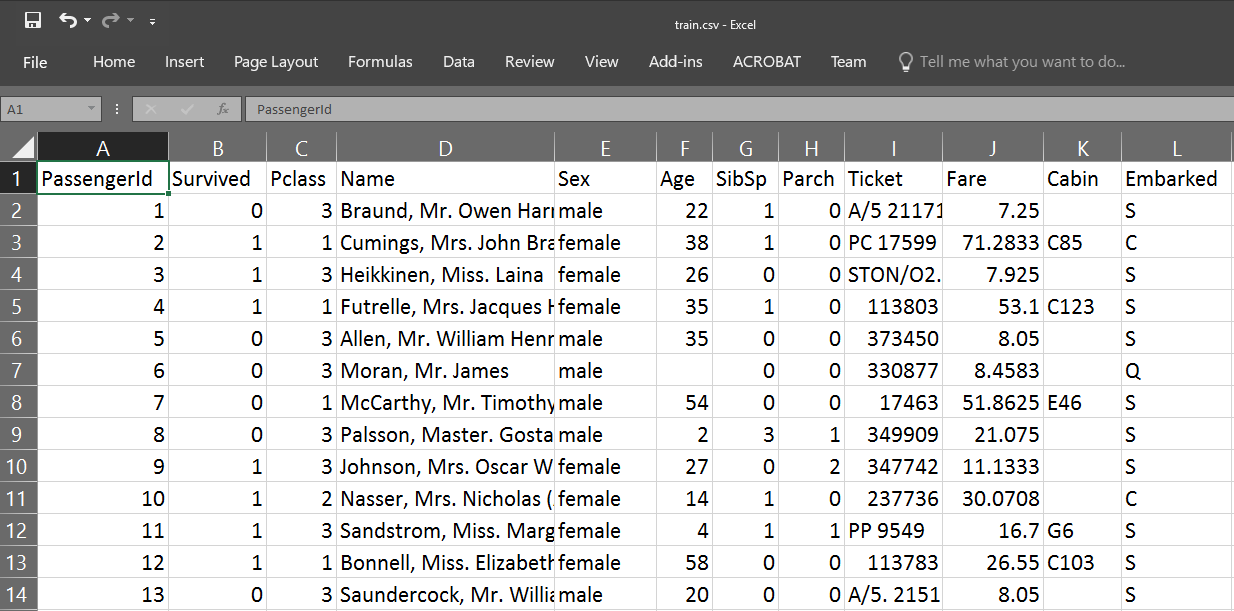
**AML model development step**

* Create: data preparation
* Train: teach the algorithm with data
* Score: see the performance
* Evaluate: compare performance of each algorithm
* Publish Web Service: production and re-train

Download Data set

<https://github.com/laploy/ML> TitanicData.csv





**Data Dictionary**

Variable Definition Key

PassengerID

survival Survival 0 = No, 1 = Yes

pclass Ticket class 1 = 1st, 2 = 2nd, 3 = 3rd

Name

sex Sex

Age Age in years

sibsp # of siblings / spouses aboard the Titanic

parch # of parents / children aboard the Titanic

ticket Ticket number

fare Passenger fare

cabin Cabin number

embarked Port of Embarkation C = Cherbourg, Q = Queenstown,

S = Southampton

Variable Notes

**Variable Notes**

pclass: A proxy for socio-economic status (SES)

* 1st = Upper
* 2nd = Middle
* 3rd = Lower

age: Age is fractional if less than 1. If the age is estimated, is it in the form of xx.5

sibsp: The dataset defines family relations in this way...

* Sibling = brother, sister, stepbrother, stepsister
* Spouse = husband, wife (mistresses and fiancés were ignored)

parch: The dataset defines family relations in this way...

* Parent = mother, father
* Child = daughter, son, stepdaughter, stepson
* Some children travelled only with a nanny, therefore parch=0 for them.

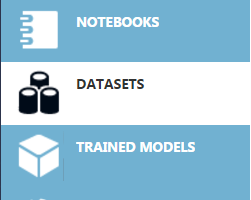
Machine Learning experiment creation working steps

**Working steps;**

* Import Data set
* Create New Experiment
* Prepare Data
  + - * + Drop the columns PassengerID, Name, Ticket, Cabin
        + Make categorical values: Survived, Pclass, Sex, Embarked
        + Replace missing value with median
        + Drop rows with missing data
        + Create Label
        + Split data 70% training and 30% scoring
* Select Algorithm : Two-Class Boosted Decision
* Train
* Score

Import Data set

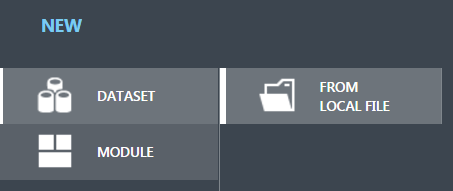
1. Click DATASETS



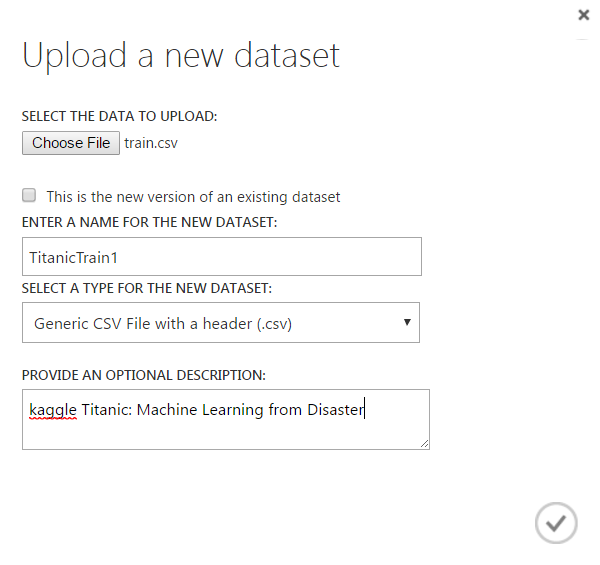
1. Click NEW



1. Click FROM LOCAL FILE



Upload a new dataset

1. Click Choose File
2. Brows and select train.csv
3. ENTER A NAME FOR THE NEW DATASET

TitanicTrain1

1. SELECT A TYPE FOR THE NEW DATASET

Generic CSV File with a header (.csv)

1. PROVIDE AN OPTIONAL DESCRIPTION

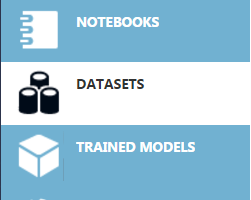
kaggle Titanic: Machine Leering from disaster

1. Click

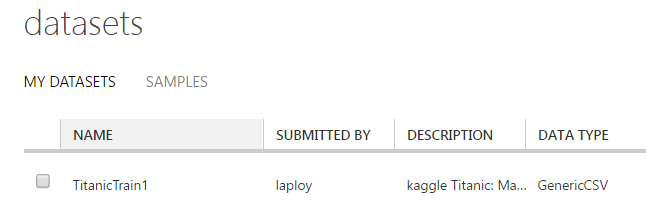


Verify dataset uploaded

1. Click DATASETS

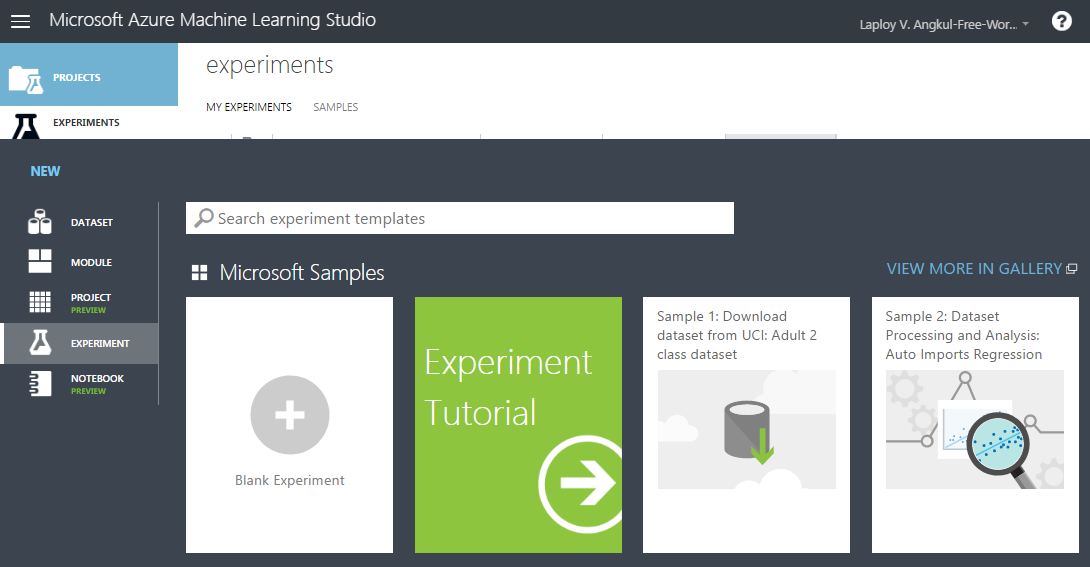


1. Make sure TitanicTrain1 is in MY DATASETS list



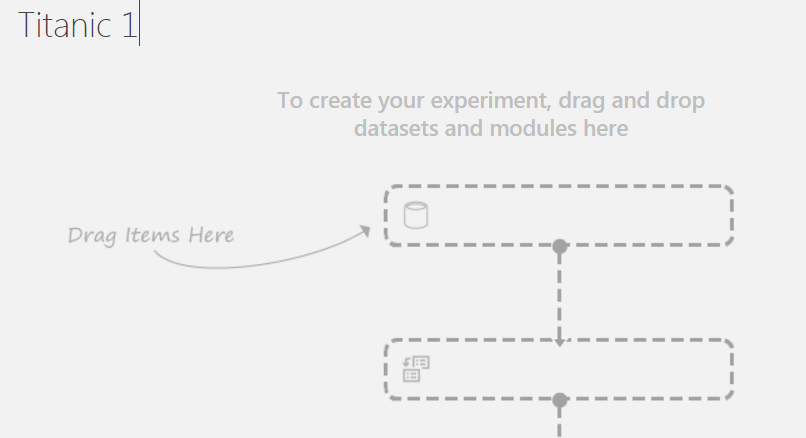
Create New Experiment

Create Blank Experiment



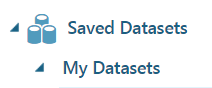
Set experiment name

Type in name = Titanic 1

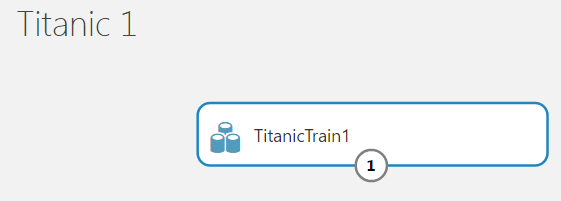


Drag & drop dataset to canvas

1. Click Saved Datasets / My Datasets



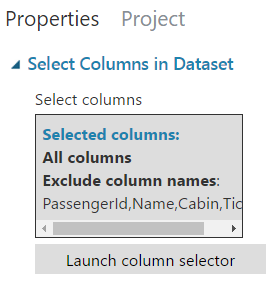
1. Drag & drop TitanicTrain1 to canvas

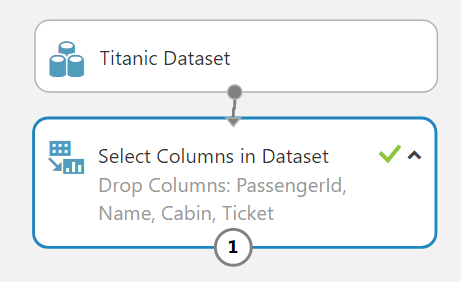


1. Visualize output

Drop the columns PassengerID, Name, Ticket, Cabin

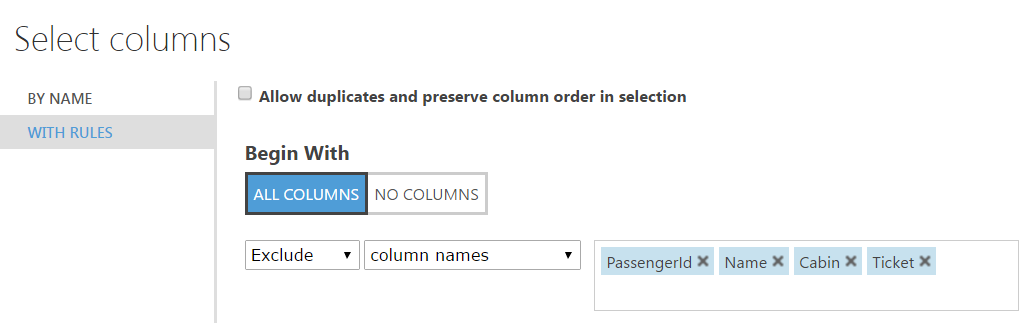
1. Drag & drop module Select Columns in Dataset
2. Selected columne = Drop Columns: PassengerId, Name, Cabin, Ticket
3. Click Launch column selector
4. Visualize



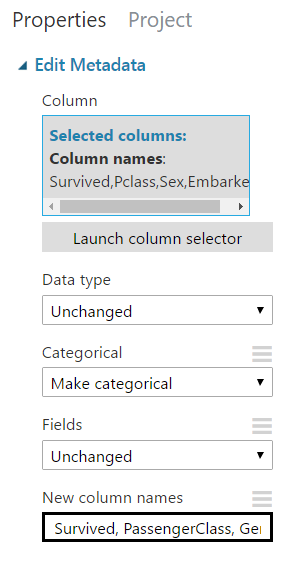
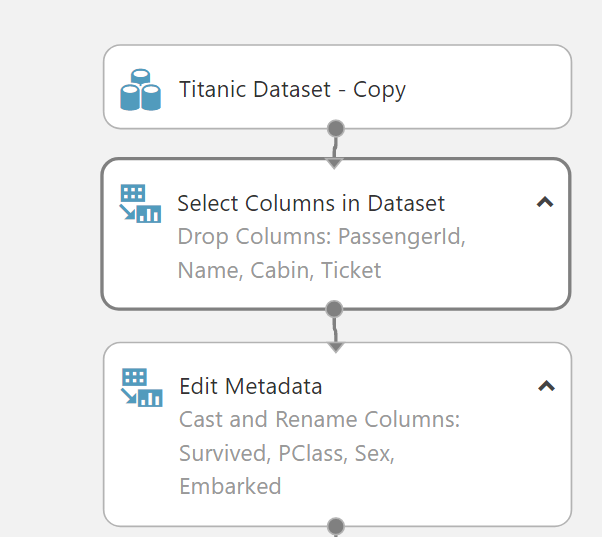


Drop the columns PassengerID, Name, Ticket, Cabin

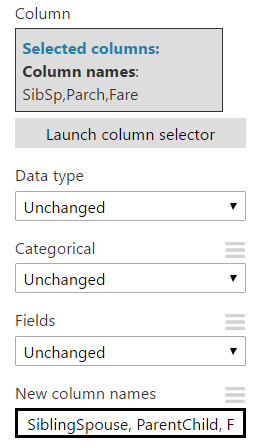
1. Begin With = ALL COLUMNS / Exclude / column name
2. Selected column PassengerID, Name, Ticket, Cabin
3. Click 
4. Visualize

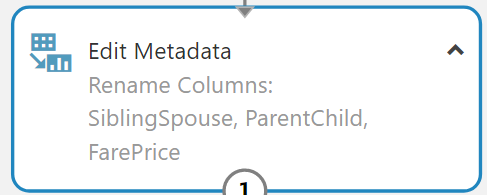


Make categorical values: Survived, Pclass, Sex, Embarked

1. Drag & drop Edit Metadata
2. Comment = Cast and Rename Columns: Survived, PClass, Sex, Embarked
3. Selected column Survived, Pclass, Sex, Embarked
4. Data type = Unchanged
5. Categorical = Make categorical
6. Fields = Unchanged
7. New column name = Survived, PassengerClass, Gender, PortEmbarkation
8. Visualize

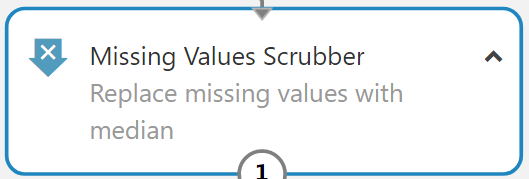
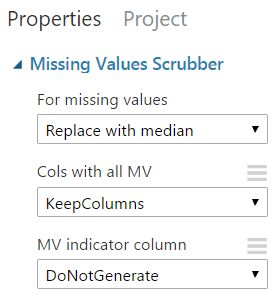
Rename columns

1. Drag & drop Edit Metadata
2. Comment = Rename Columns: SiblingSpouse, ParentChild, FarePrice
3. Selected column SibSp, Parch, Fare
4. Data type = Unchanged
5. Categorical = Unchanged
6. Fields = Unchanged
7. New column name = SiblingSpouse, ParentChild, FarePrice
8. Visualize

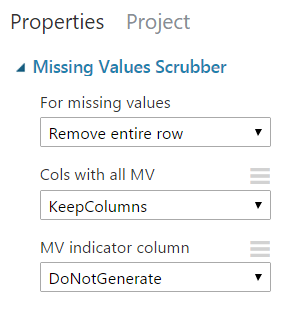


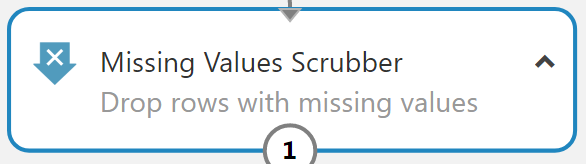
Replace missing value with median

1. Drag & drop Missing Values Scrubber
2. Comment = Replace missing value with median
3. Set properties
4. Visualize

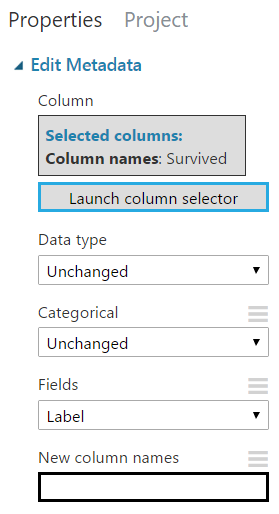


Drop rows with missing data

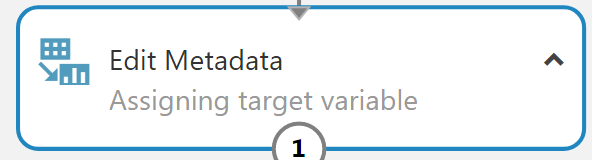
1. Drag & drop Missing Values Scrubber
2. Comment = Drop rows with missing values
3. Set properties
4. Visualize



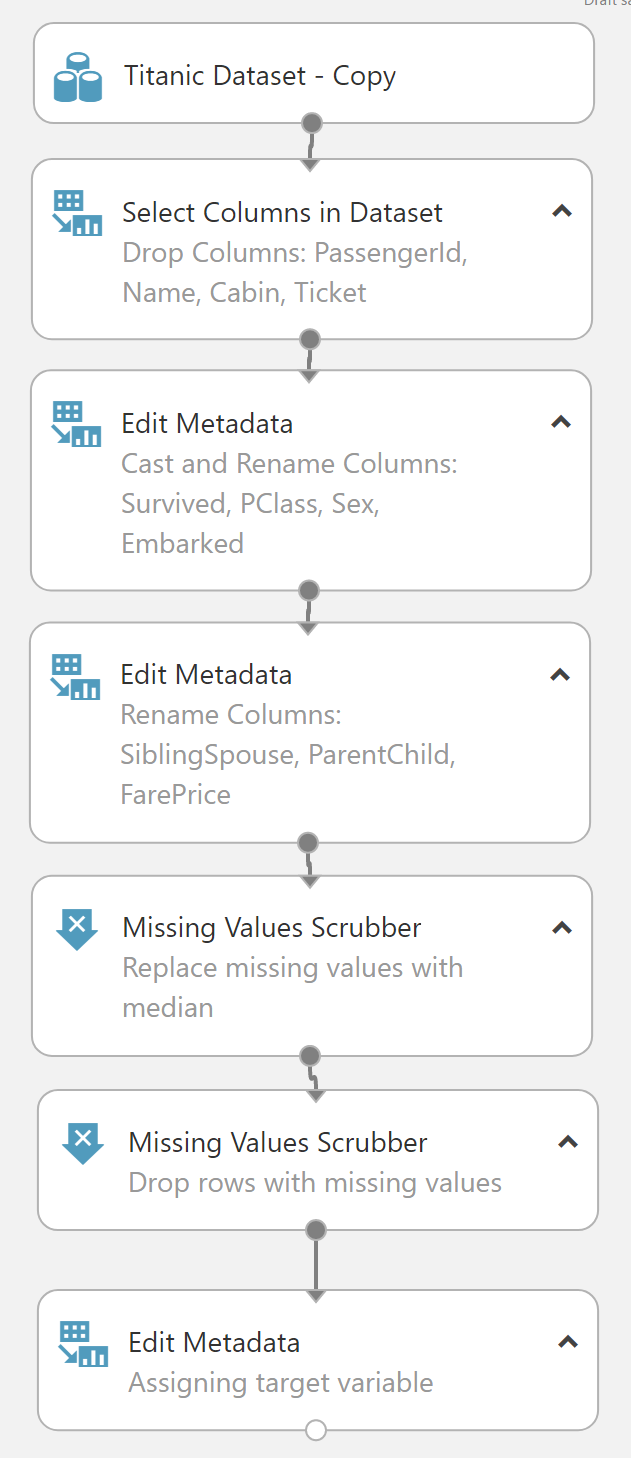
Create Label



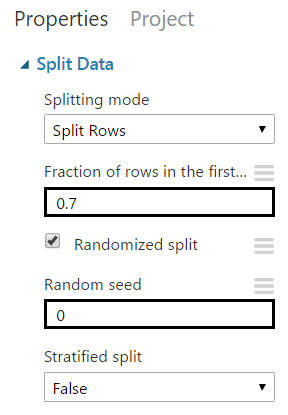
1. Drag & drop Edit Metadata
2. Comment = Assigning target variable
3. Selected column = Survived
4. Data type = Unchanged
5. Categorical = Unchanged
6. Fields = Label
7. New column name = -
8. Visualize

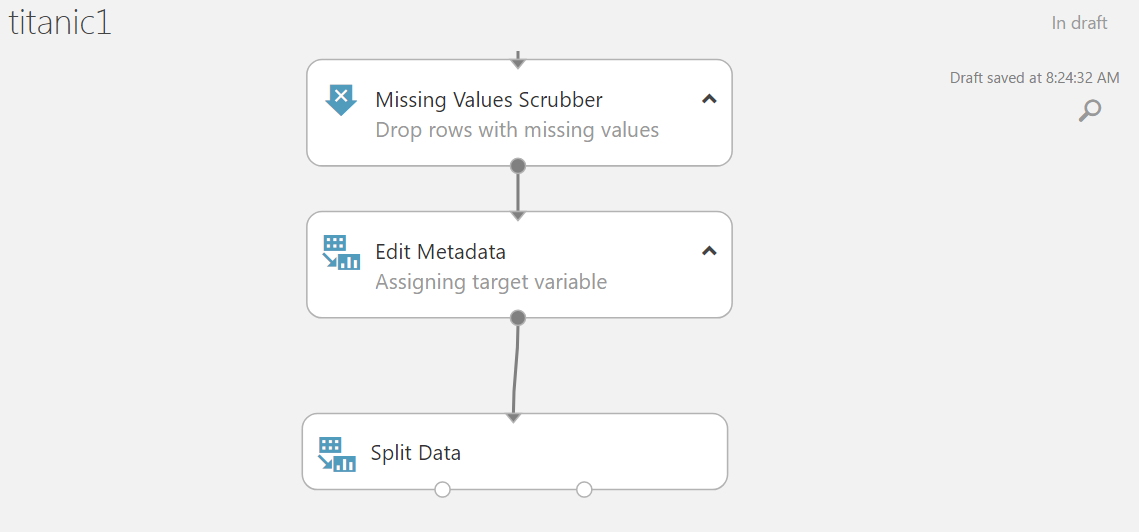


Import and Dataset preparation



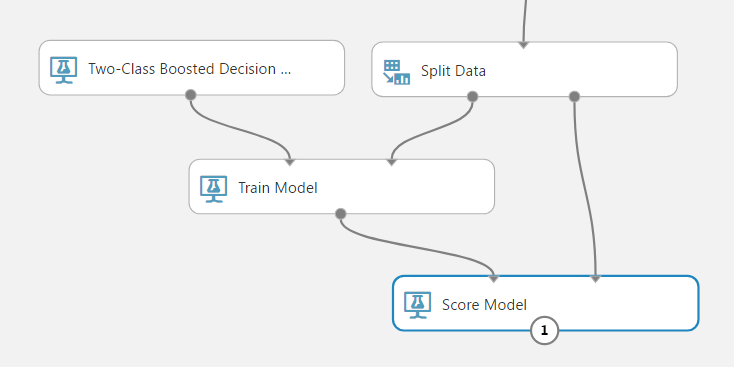
Split data 70% training and 30% scoring

1. Drag & drop Split data
2. Set properties

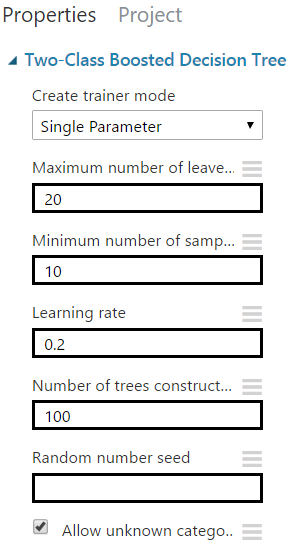


Add Algorithm, Train and Score

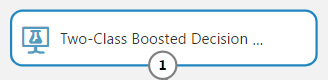
* Add Two-Class Boosted Decision tree
* Add Train Model
* Add Score Model



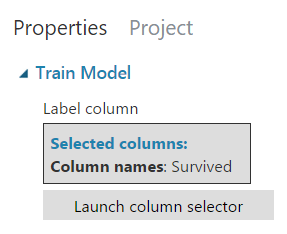
Add Two-Class Boosted Decision tree

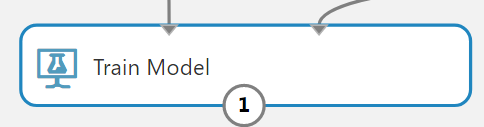


1. Drag & drop Two-Class Boosted Decision tree
2. Set properties



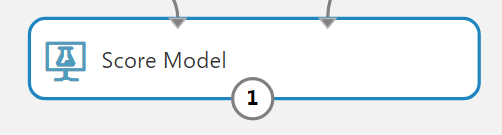
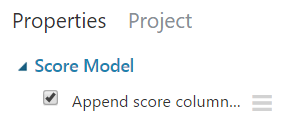
Add Train Model

1. Drag & drop train model
2. Set column to Survived

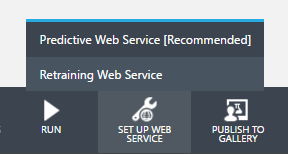


Add Score Mode

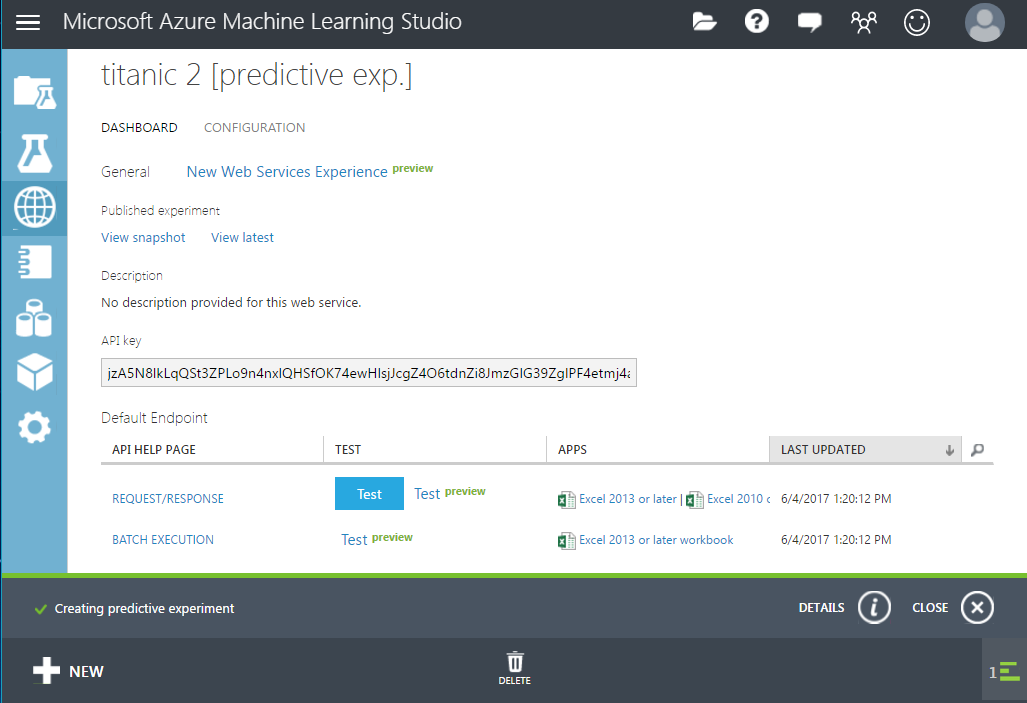
1. Drag & drop Score Model
2. Set property = Append score column
3. Save
4. Run experiment
5. Visualize



Create web service

1. Save as Titanic 2
2. Run Titanic 2
3. Click SET UP WEB SERVICE
4. Click Predictive Web Service
5. Click RUN
6. Click SET UP WEB SERVICE

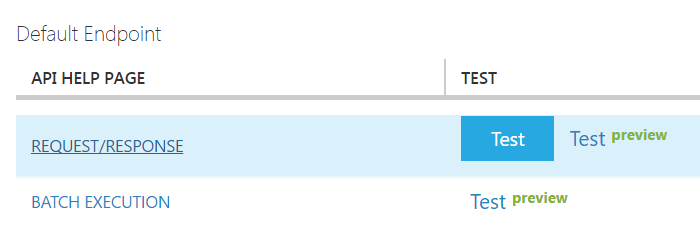
Create web service Titanic 2 [predictive exp.] page



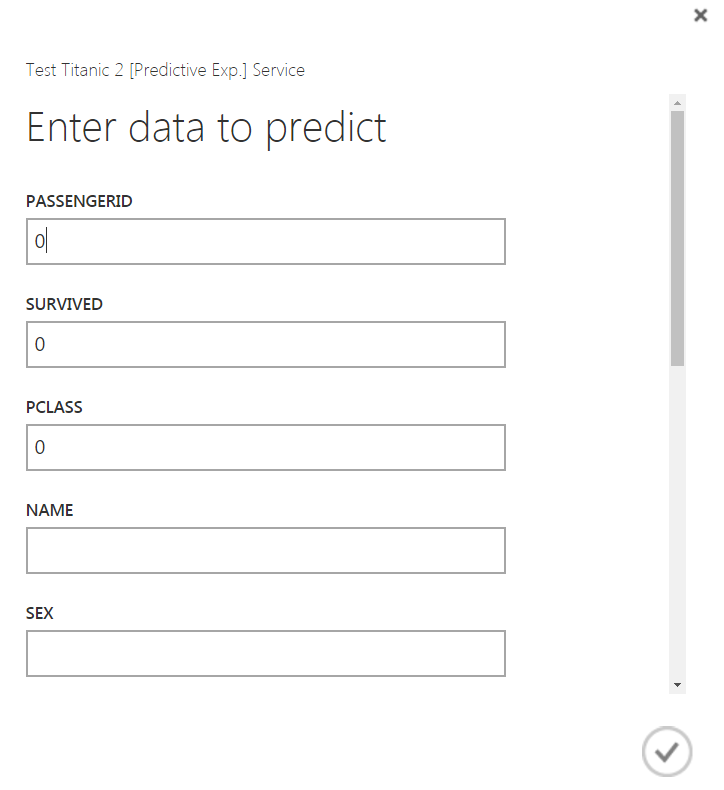
Test web service

**Web service testing**

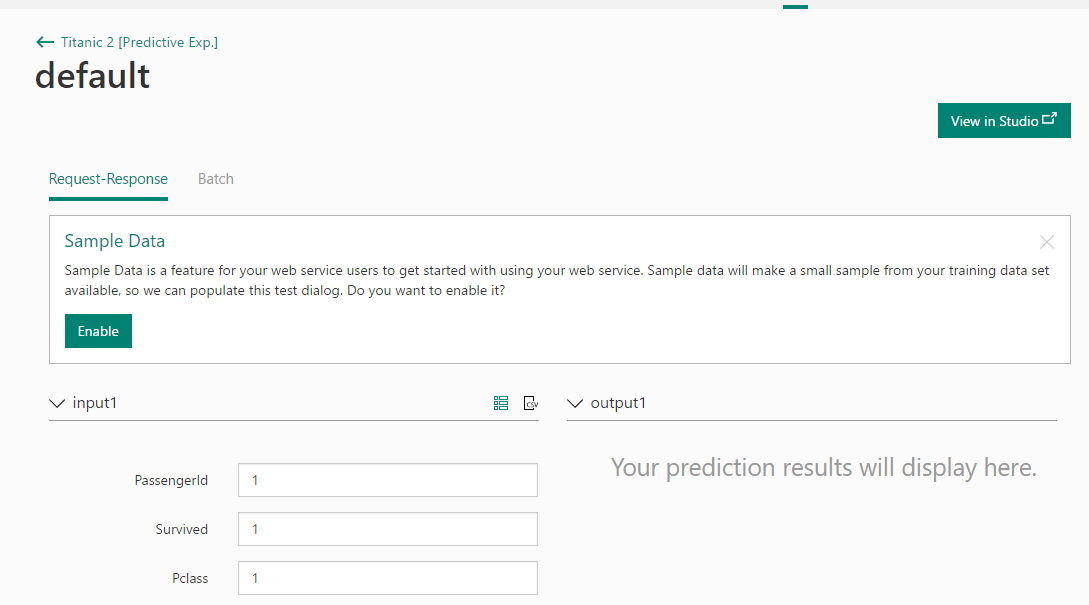
* REQUEST/RESPONSE Test
* REQUEST/RESPONSE Test preview
* REQUEST/RESPONSE Excel workbook test
* BATCH EXECUTION Test preview
* BATCH EXECUTION Excel workbook test



REQUEST/RESPONSE Test

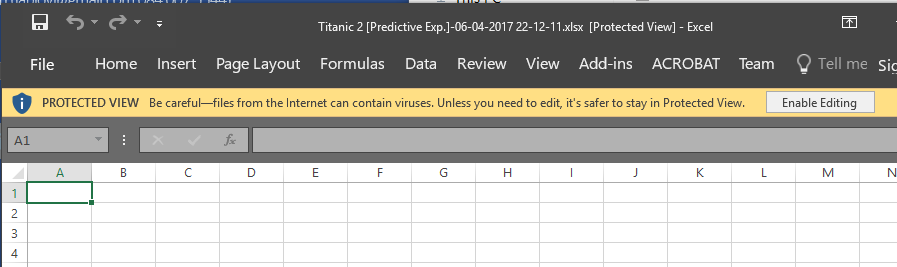
1. Test with know result
2. Open file kaggle test.csv
3. Take one passenger
4. Click REQUEST/RESPONSE Test
5. Fill in the form

REQUEST/RESPONSE Test preview

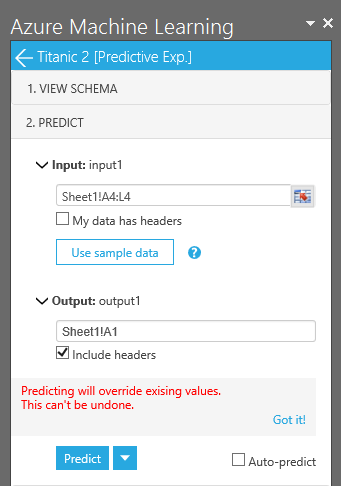
1. Test with know result
2. Open file kaggle test.csv
3. Take one passenger
4. Click REQUEST/RESPONSE Test preview
5. Click Enable (Sample Data)
6. Fill in the form
7. Click Test Request-Response

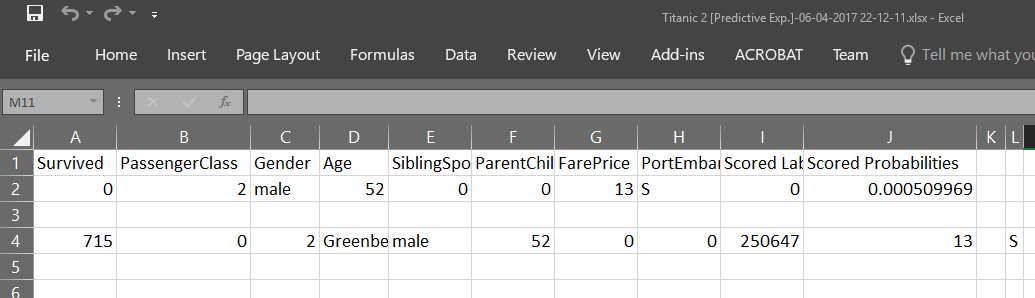
REQUEST/RESPONSE Excel workbook test

1. Test with know result
2. Open file kaggle test.csv
3. Take one passenger
4. Click REQUEST/RESPONSE Excel 2013 or later
5. Open file Titanic 2 [Predictive Exp.] on Desktop
6. Click Enable Editing

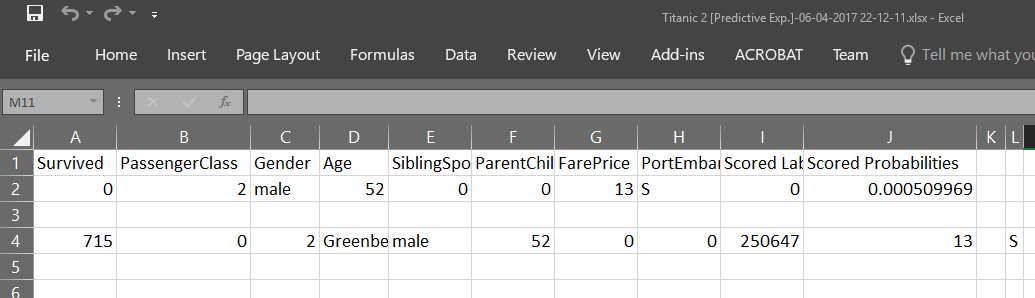


REQUEST/RESPONSE Excel workbook test

1. Input = Sheet1!A4L4
2. My data has headers = uncheck
3. Output = Sheet1!A1
4. Include headers = check
5. Copy a line from file kaggle test.csv to A4
6. Click Predict

REQUEST/RESPONSE Excel workbook test

Test result



More information

**More information on Classification Model**

Two-Class Boosted Decision Tree

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

Machine learning algorithm cheat sheet for Microsoft Azure Machine Learning Studio

https://docs.microsoft.com/en-us/azure/machine-learning/machine-learning-algorithm-cheat-sheet