

Building a classification model

BUILDING A CLASSIFICATION MODEL



Building a classification model

In this session

- 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
- Split data
- Select Algorithm
- Train
- Score
- Create web service
- Test web service

Building a classification model

What to do

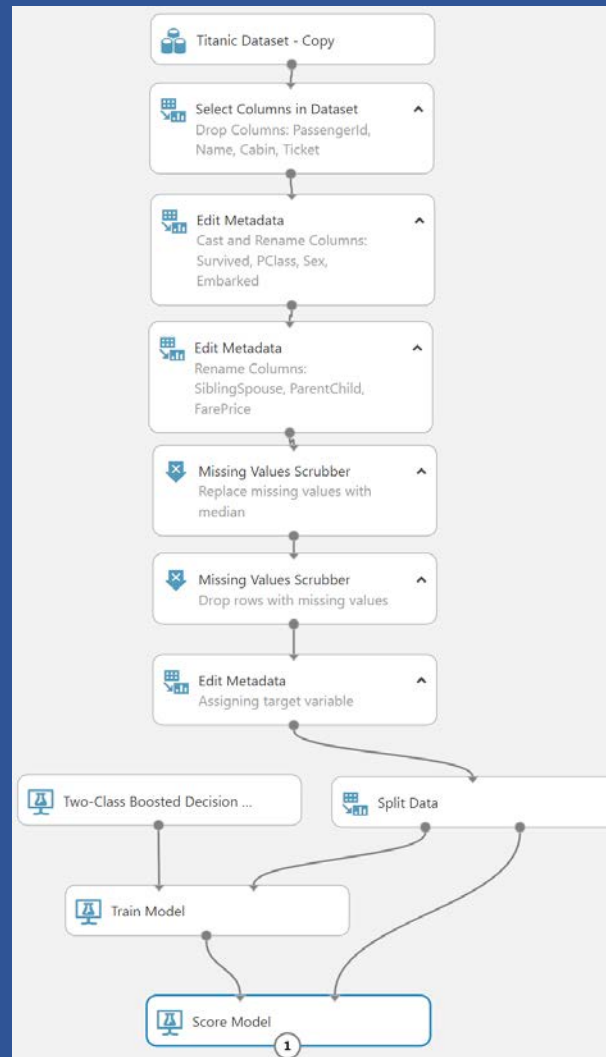
What to do;

- 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.



Building a classification model

ML model when finished



Building a classification model

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

Building a classification model

Download Data set

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

The screenshot shows the GitHub repository page for 'laploy / ML'. The repository is described as 'Repository for Student of Microsoft Azure Machine Learning Course www.laploy.com'. It has 24 commits, 1 branch, 0 releases, and 1 contributor. The 'Code' tab is selected, showing a list of files. The file 'TitanicData.csv' is highlighted in yellow. The repository is created by 'laploy' and the latest commit is 'b05e68c' on Sep 22, 2017.

Repository for Student of Microsoft Azure Machine Learning Course www.laploy.com

Add topics

24 commits 1 branch 0 releases 1 contributor

Branch: master New pull request Create new file Upload files Find file Clone or download

File	Action	Time
1700-000.jpg	Add files via upload	7 months ago
Missing R Script.zip	Add files via upload	8 months ago
Python introduction code.zip	Add files via upload	8 months ago
R intro.R	Add files via upload	8 months ago
README.md	Update README.md	4 months ago
Test3.cs	Create Test3.cs	8 months ago
TitanicBI.csv	Add files via upload	8 months ago
TitanicData.csv	Add files via upload	8 months ago

Building a classification model

train.csv - Excel

File Home Insert Page Layout Formulas Data Review View Add-ins ACROBAT Team Tell me what you want to do...

A1 :

	A	B	C	D	E	F	G	H	I	J	K	L
1	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
2	1	0	3	Braund, Mr. Owen Harri	male	22	1	0	A/5 21171	7.25		S
3	2	1	1	Cumings, Mrs. John Bra	female	38	1	0	PC 17599	71.2833	C85	C
4	3	1	3	Heikkinen, Miss. Laina	female	26	0	0	STON/O2.	7.925		S
5	4	1	1	Futrelle, Mrs. Jacques H	female	35	1	0	113803	53.1	C123	S
6	5	0	3	Allen, Mr. William Henr	male	35	0	0	373450	8.05		S
7	6	0	3	Moran, Mr. James	male		0	0	330877	8.4583		Q
8	7	0	1	McCarthy, Mr. Timothy	male	54	0	0	17463	51.8625	E46	S
9	8	0	3	Palsson, Master. Gosta	male	2	3	1	349909	21.075		S
10	9	1	3	Johnson, Mrs. Oscar W	female	27	0	2	347742	11.1333		S
11	10	1	2	Nasser, Mrs. Nicholas (female	14	1	0	237736	30.0708		C
12	11	1	3	Sandstrom, Miss. Marg	female	4	1	1	PP 9549	16.7	G6	S
13	12	1	1	Bonnell, Miss. Elizabeth	female	58	0	0	113783	26.55	C103	S
14	13	0	3	Saundercock, Mr. Willi	male	20	0	0	A/5. 2151	8.05		S

Building a classification model

Data Dictionary

Variable	Definition	Key
PassengerID		
survival	Survival	0 = No, 1 = Yes
pclass	Ticket class	1 = 1st, 2 = 2nd, 3 = 3 rd
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

Building a classification model

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.

Building a classification model

Machine Learning experiment creation working steps

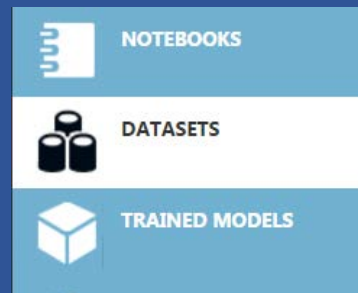
Working steps;

- 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

Building a classification model

Import Data set

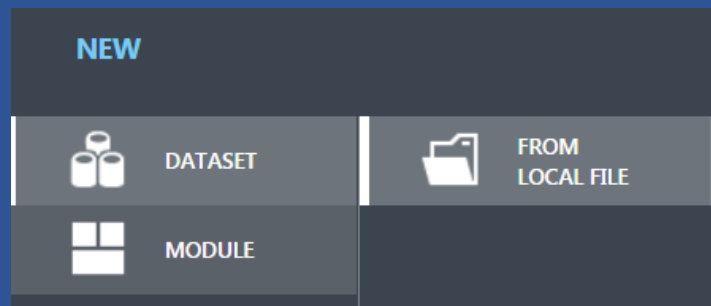
1. Click DATASETS



2. Click NEW



3. Click FROM LOCAL FILE



Building a classification model

Upload a new dataset

4. Click **Choose File**
5. Brows and select train.csv
6. ENTER A NAME FOR THE NEW DATASET
TitanicTrain1
7. SELECT A TYPE FOR THE NEW DATASET
Generic CSV File with a header (.csv)
8. PROVIDE AN OPTIONAL DESCRIPTION
kaggle Titanic: Machine Leering from disaster
9. Click



×

Upload a new dataset

SELECT THE DATA TO UPLOAD:

Choose File

train.csv

☐ This is the new version of an existing dataset

ENTER A NAME FOR THE NEW DATASET:

TitanicTrain1

SELECT A TYPE FOR THE NEW DATASET:

Generic CSV File with a header (.csv) ▼

PROVIDE AN OPTIONAL DESCRIPTION:

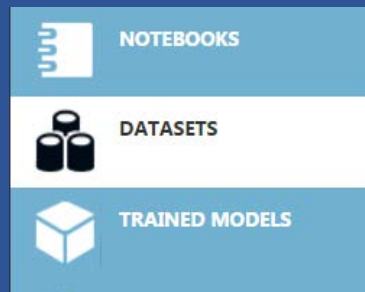
kaggle Titanic: Machine Learning from Disaster

✓

Building a classification model

Verify dataset uploaded

1. Click DATASETS



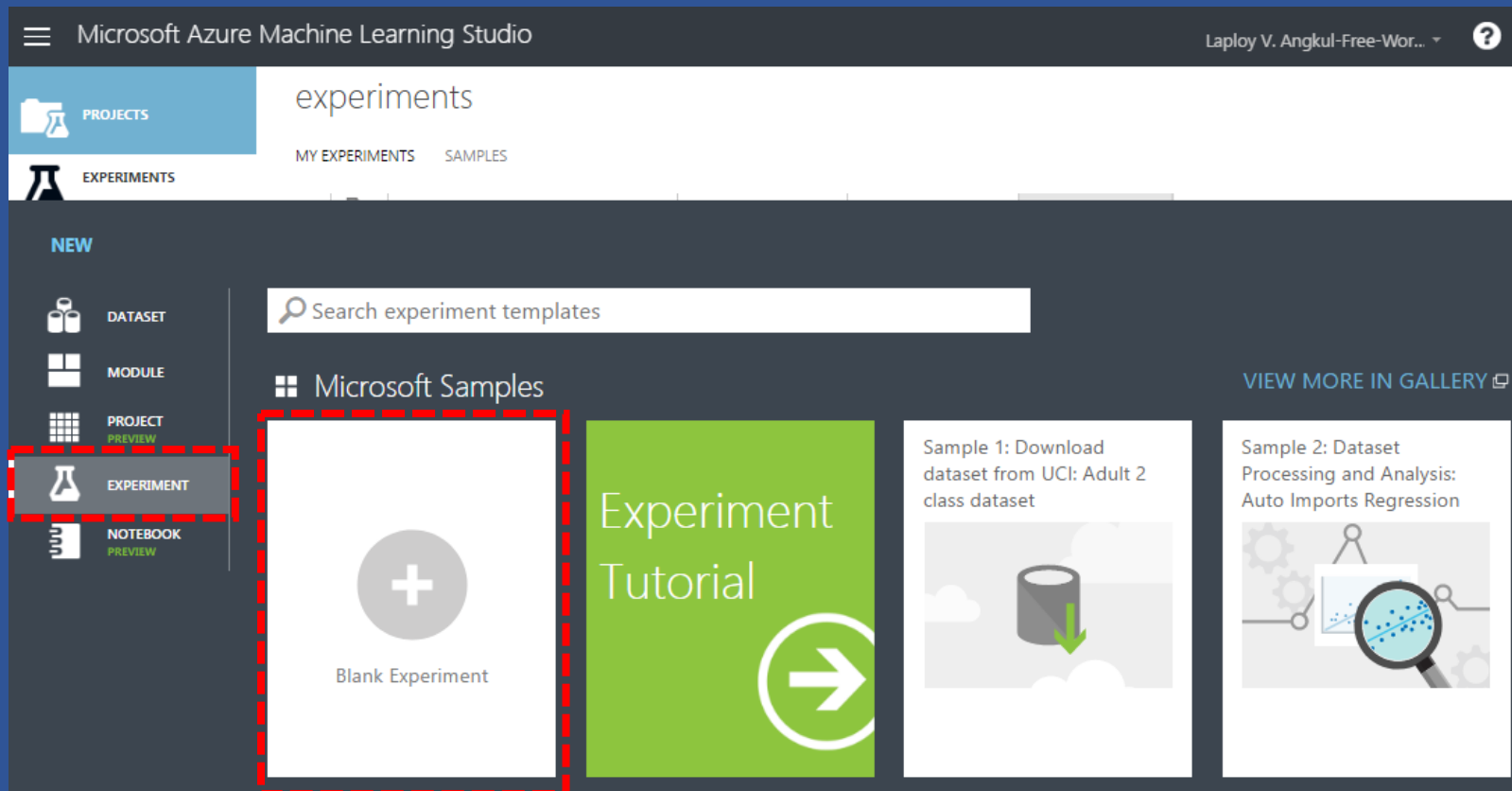
2. Make sure TitanicTrain1 is in MY DATASETS list

datasets				
MY DATASETS		SAMPLES		
	NAME	SUBMITTED BY	DESCRIPTION	DATA TYPE
<input type="checkbox"/>	TitanicTrain1	laploy	kaggle Titanic: Ma...	GenericCSV

Building a classification model

Create New Experiment

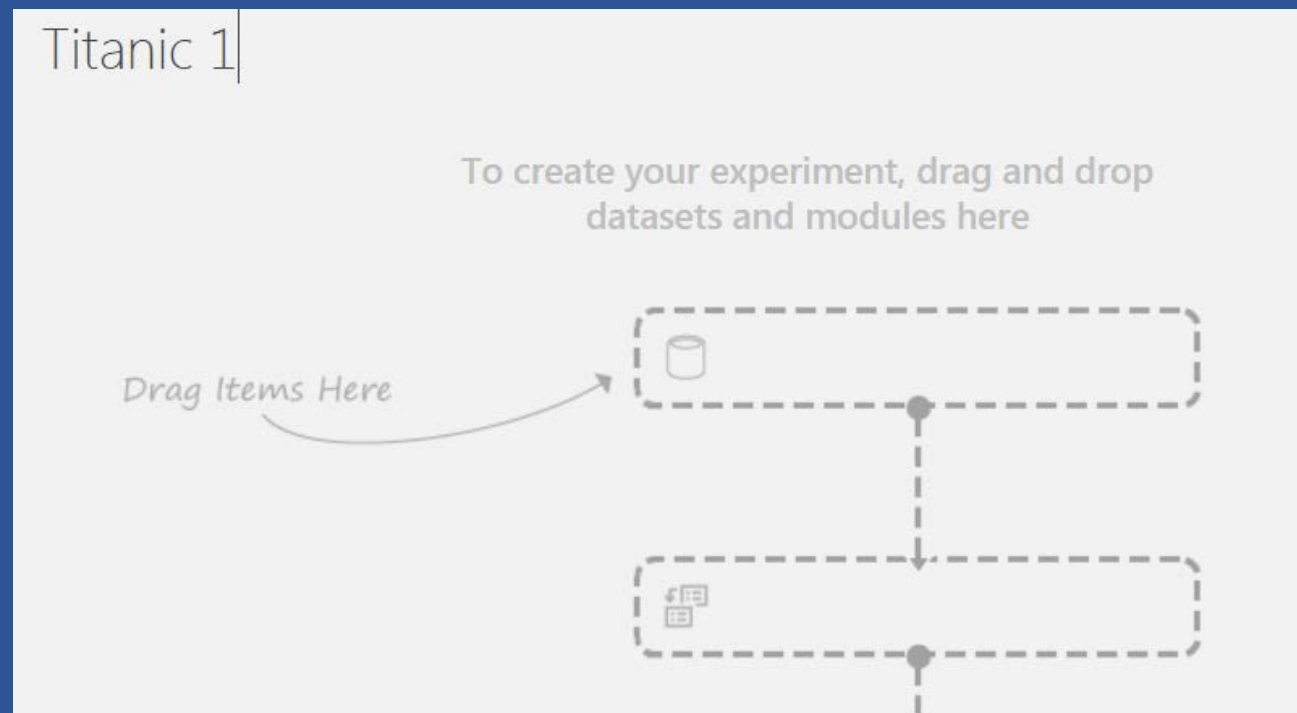
Create **Blank Experiment**



Building a classification model

Set experiment name

Type in name = **Titanic 1**



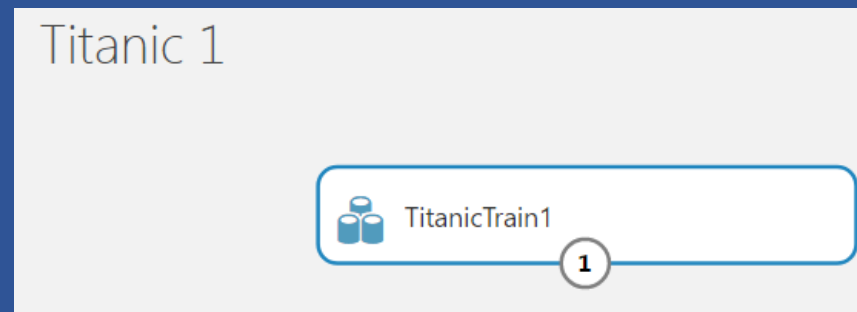
Building a classification model

Drag & drop dataset to canvas

1. Click **Saved Datasets / My Datasets**



2. Drag & drop **TitanicTrain1** to canvas

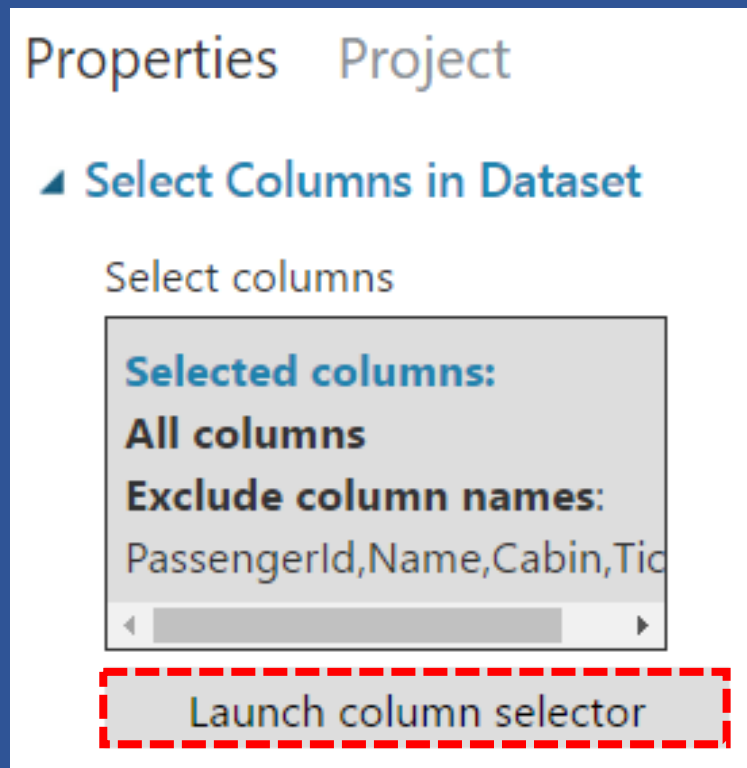
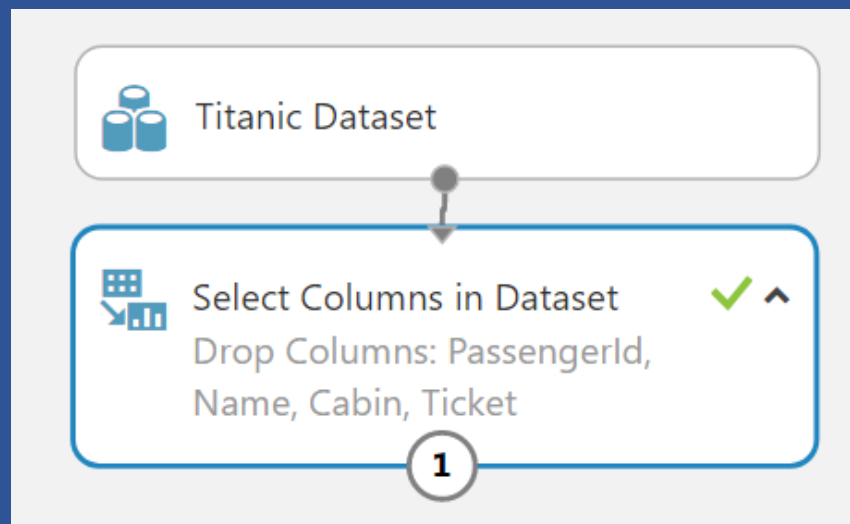


3. Visualize output

Building a classification model

Drop the columns PassengerId, Name, Ticket, Cabin

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



Building a classification model

Drop the columns PassengerID, Name, Ticket, Cabin

5. Begin With = ALL COLUMNS / Exclude / column name

6. Selected column **PassengerID, Name, Ticket, Cabin**

7. Click 

8. Visualize

Select columns

BY NAME

WITH RULES

☐ Allow duplicates and preserve column order in selection

Begin With

ALL COLUMNS NO COLUMNS

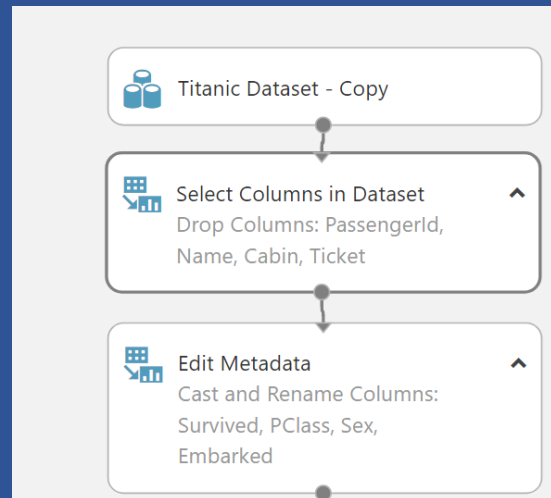
Exclude column names

PassengerId X Name X Cabin X Ticket X

Building a classification model

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



Properties Project

Edit Metadata

Column

Selected columns:
Column names:
Survived,Pclass,Sex,Embarke

Launch column selector

Data type
Unchanged

Categorical
Make categorical

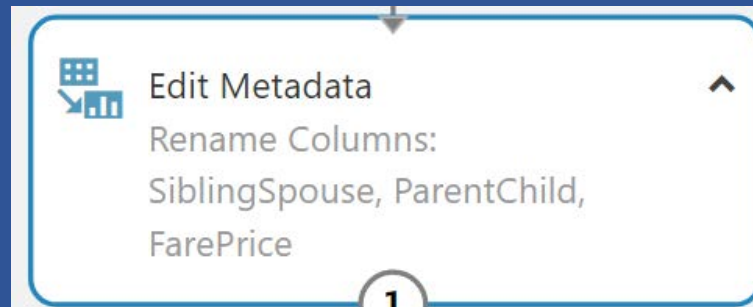
Fields
Unchanged

New column names
Survived, PassengerClass, Gei

Building a classification model

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



Column

Selected columns:
Column names:
SibSp,Parch,Fare

Launch column selector

Data type
Unchanged ▼

Categorical
Unchanged ▼

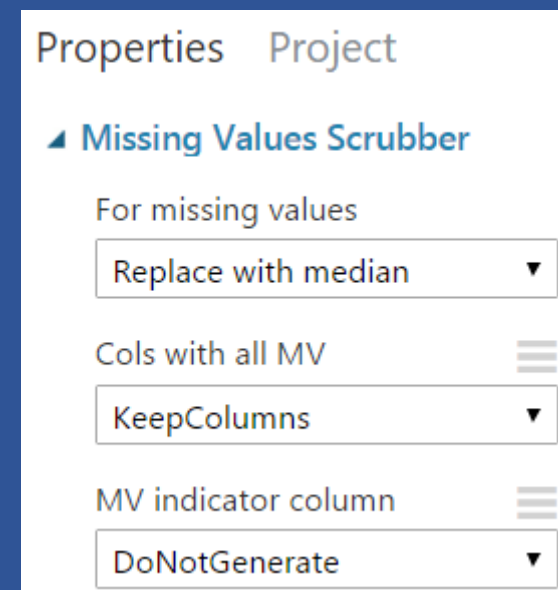
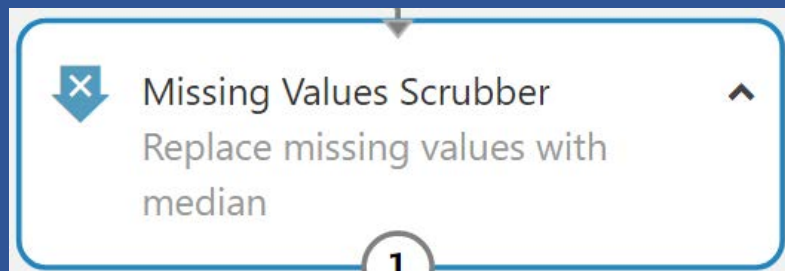
Fields
Unchanged ▼

New column names
SiblingSpouse, ParentChild, F

Building a classification model

Replace missing value with median

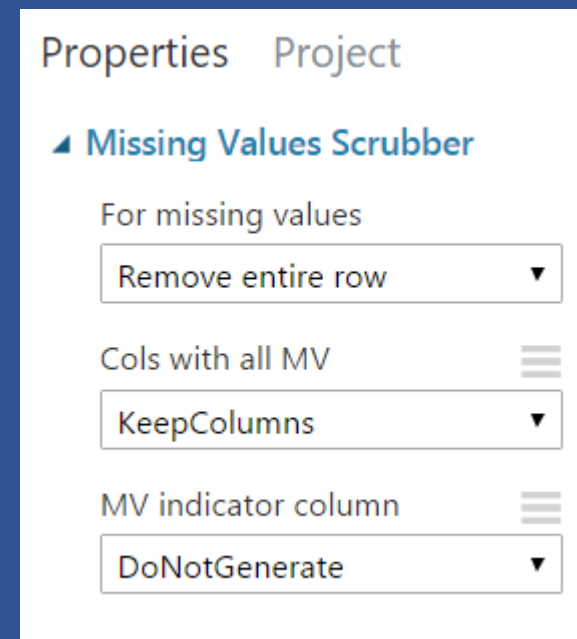
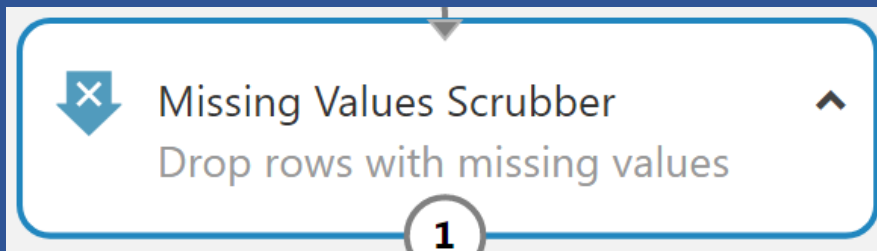
1. Drag & drop **Missing Values Scrubber**
2. Comment = **Replace missing value with median**
3. Set properties
4. Visualize



Building a classification model

Drop rows with missing data

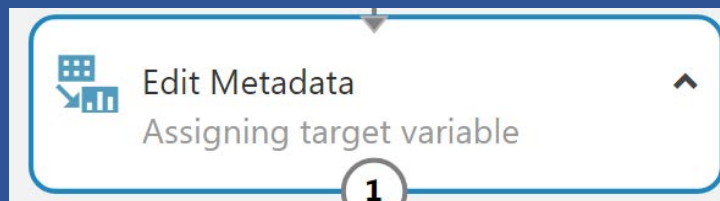
1. Drag & drop **Missing Values Scrubber**
2. Comment = **Drop rows with missing values**
3. Set properties
4. Visualize



Building a classification model

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



Properties Project

▲ **Edit Metadata**

Column

Selected columns:
Column names: Survived

Launch column selector

Data type

Unchanged ▼

Categorical

Unchanged ▼

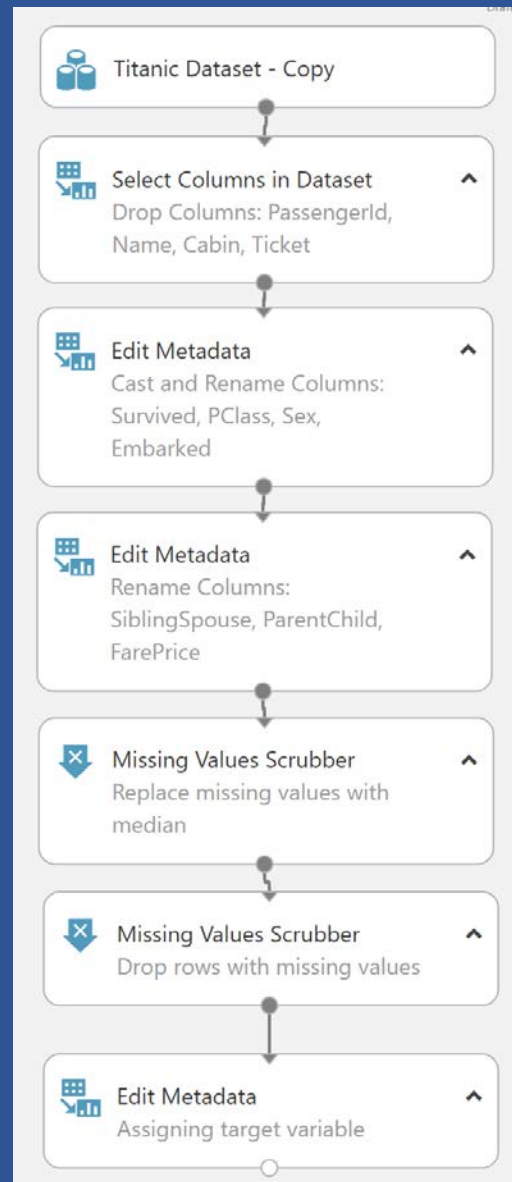
Fields

Label ▼

New column names

Building a classification model

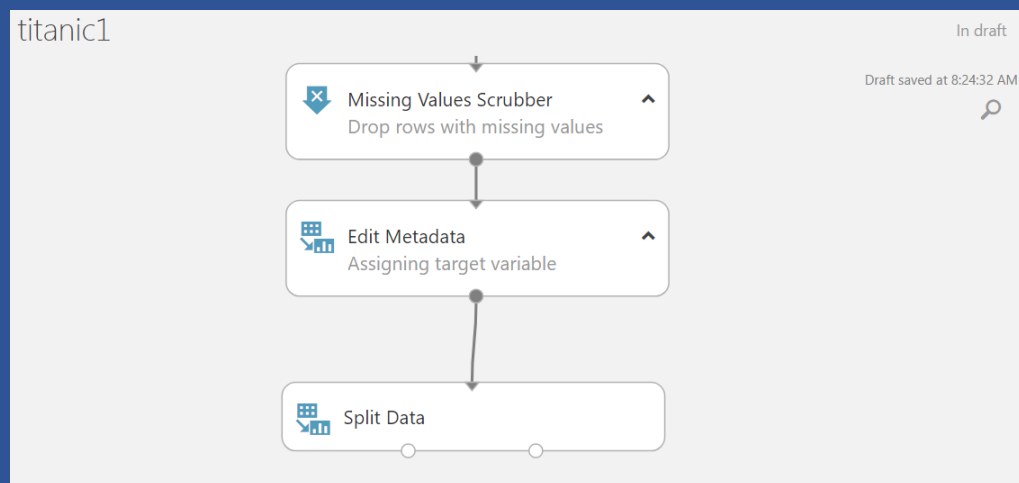
Import and Dataset preparation



Building a classification model

Split data 70% training and 30% scoring

1. Drag & drop **Split data**
2. Set properties



Properties Project

Split Data

Splitting mode
Split Rows ▼

Fraction of rows in the first...
0.7

☒ Randomized split

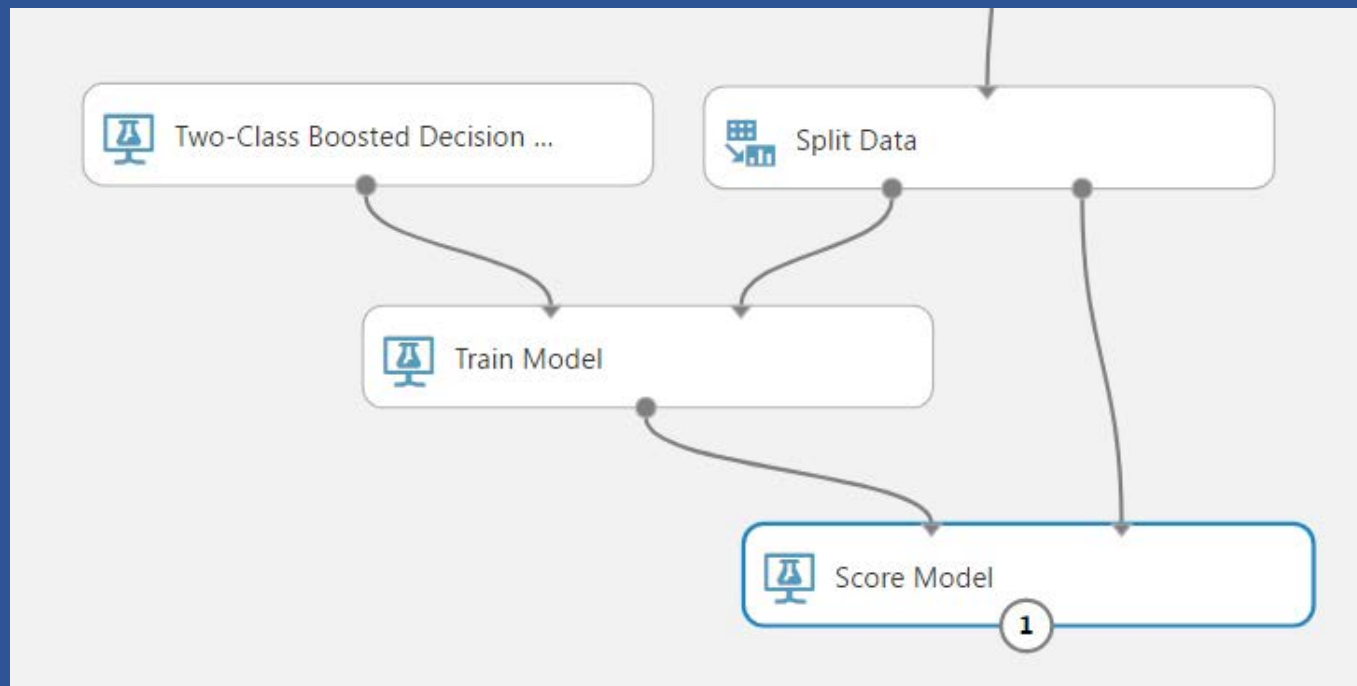
Random seed
0

Stratified split
False ▼

Building a classification model

Add Algorithm, Train and Score

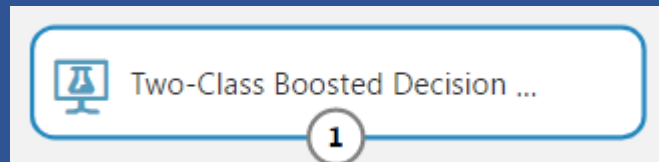
- Add Two-Class Boosted Decision tree
 - Add Train Model
 - Add Score Model



Building a classification model

Add Two-Class Boosted Decision tree

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



Properties Project

Two-Class Boosted Decision Tree

Create trainer mode
Single Parameter ▼

Maximum number of leave...
20

Minimum number of samp...
10

Learning rate
0.2

Number of trees construct...
100

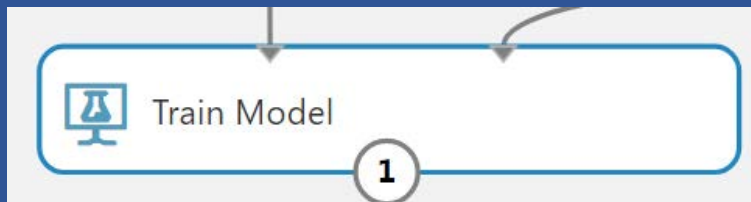
Random number seed

☒ Allow unknown catego..

Building a classification model

Add Train Model

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



Properties Project

Train Model

Label column

Selected columns:

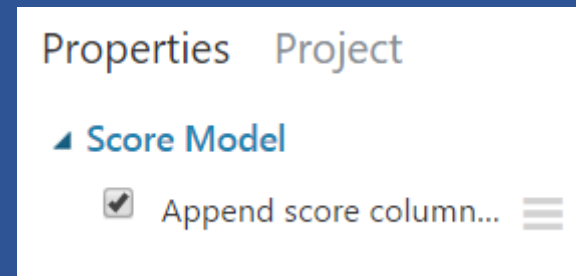
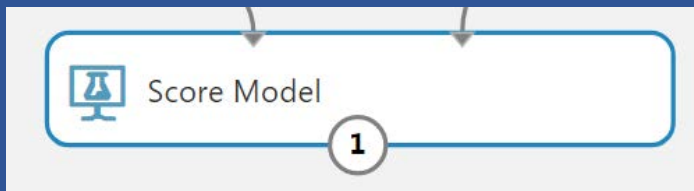
Column names: Survived

Launch column selector

Building a classification model

Add Score Mode

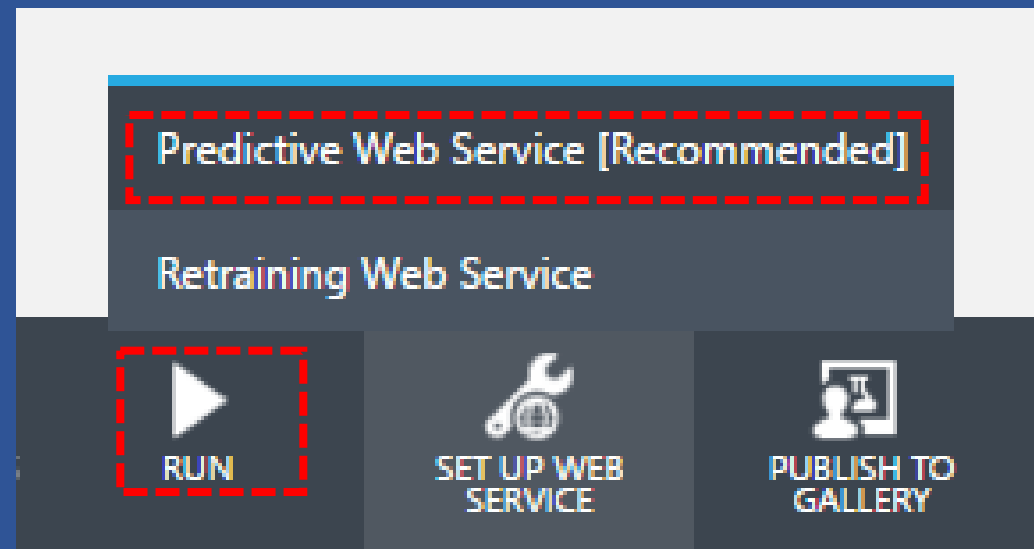
1. Drag & drop **Score Model**
2. Set property = **Append score column**
3. Save
4. Run experiment
5. Visualize



Building a classification model

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**



Building a classification model

Create web service Titanic 2 [predictive exp.] page

The screenshot shows the Microsoft Azure Machine Learning Studio interface. The top navigation bar includes a hamburger menu, the title 'Microsoft Azure Machine Learning Studio', and icons for file explorer, help, chat, collaboration, and a user profile. The left sidebar contains icons for workspace, experiment, published experiment, notebook, data, and settings.

The main content area displays the details for the 'titanic 2 [predictive exp.]' experiment. It includes tabs for 'DASHBOARD' and 'CONFIGURATION'. Under 'General', there is a link to 'New Web Services Experience' with a 'preview' tag. Below this, there are links for 'View snapshot' and 'View latest'. A 'Description' section states 'No description provided for this web service.' An 'API key' is displayed in a text box: 'jzA5N8lkLqQSt3ZPLo9n4nxiQHsfOK74ewHlsjJcgZ4O6tdnZi8JmzGIG39ZglPF4etmj4:'. A 'Default Endpoint' section is also present.

Below the configuration section is a table with columns: 'API HELP PAGE', 'TEST', 'APPS', and 'LAST UPDATED'. The table contains two rows of data.

API HELP PAGE	TEST	APPS	LAST UPDATED
REQUEST/RESPONSE	Test Test preview	Excel 2013 or later Excel 2010 c	6/4/2017 1:20:12 PM
BATCH EXECUTION	Test preview	Excel 2013 or later workbook	6/4/2017 1:20:12 PM

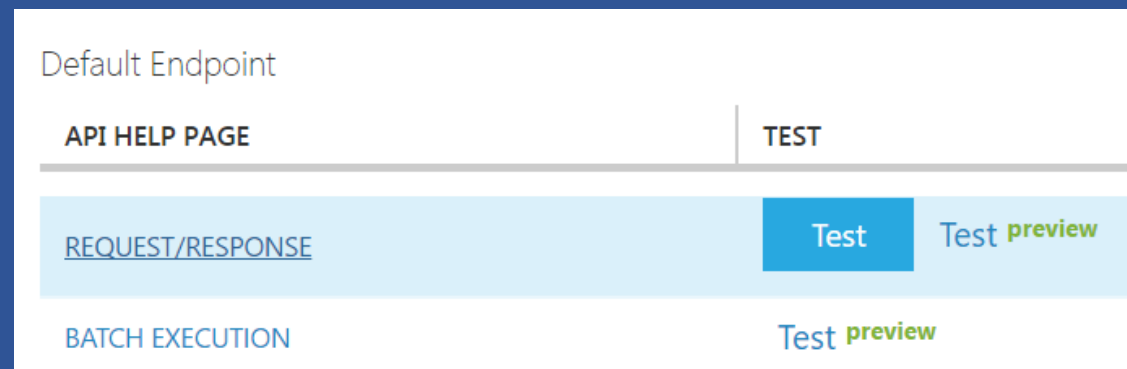
At the bottom of the interface, there is a status bar showing a green checkmark and the text 'Creating predictive experiment'. To the right of this are links for 'DETAILS', 'CLOSE', and a close button. At the very bottom, there is a 'NEW' button, a 'DELETE' button, and a tab indicator showing '1'.

Building a classification model

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



Building a classification model

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

Test Titanic 2 [Predictive Exp.] Service

Enter data to predict

PASSENGERID

SURVIVED

PCLASS

NAME

SEX

Building a classification model

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**

← Titanic 2 [Predictive Exp.]

default

[View in Studio](#)

Request-Response Batch

Sample Data

Sample Data is a feature for your web service users to get started with using your web service. Sample data will make a small sample from your training data set available, so we can populate this test dialog. Do you want to enable it?

Enable

input1

PassengerId 1

Survived 1

Pclass 1

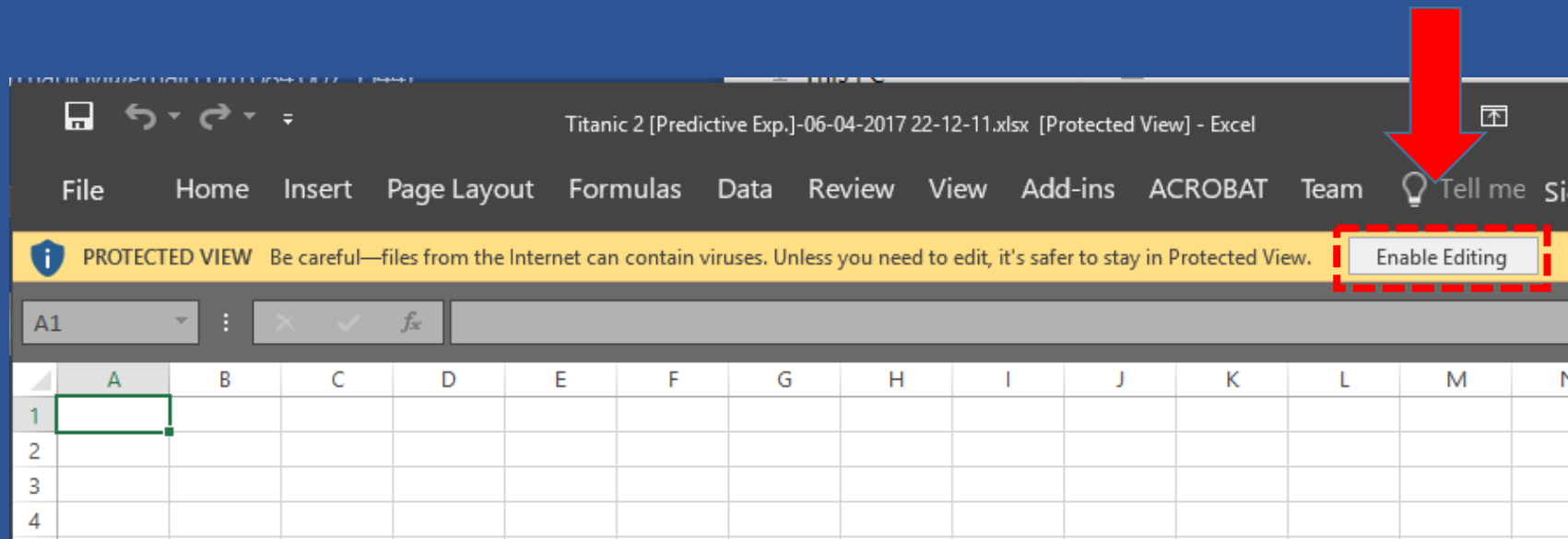
output1

Your prediction results will display here.

Building a classification model

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**



Building a classification model

REQUEST/RESPONSE Excel workbook test

1. Input = **Sheet1!A4:L4**
2. My data has headers = **unchecked**
3. Output = **Sheet1!A1**
4. Include headers = **checked**
5. Copy a line from file kaggle **test.csv** to **A4**
6. Click **Predict**

Azure Machine Learning

← Titanic 2 [Predictive Exp.]

1. VIEW SCHEMA

2. PREDICT

▼ Input: input1

Sheet1!A4:L4

☐ My data has headers

Use sample data ?

▼ Output: output1

Sheet1!A1

☒ Include headers

Predicting will override existing values.
This can't be undone.

Got it!

Predict ▼ ☐ Auto-predict

Building a classification model

REQUEST/RESPONSE Excel workbook test

Test result

Titanic 2 [Predictive Exp.]-06-04-2017 22-12-11.xlsx - Excel

	A	B	C	D	E	F	G	H	I	J	K	L
1	Survived	PassengerClass	Gender	Age	SiblingSpo	ParentChil	FarePrice	PortEmbar	Scored Lat	Scored Probabilities		
2	0	2	male	52	0	0	13	S	0	0.000509969		
3												
4	715	0	2	Greenbe	male	52	0	0	250647	13	S	
5												
6												

Building a classification model

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>