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

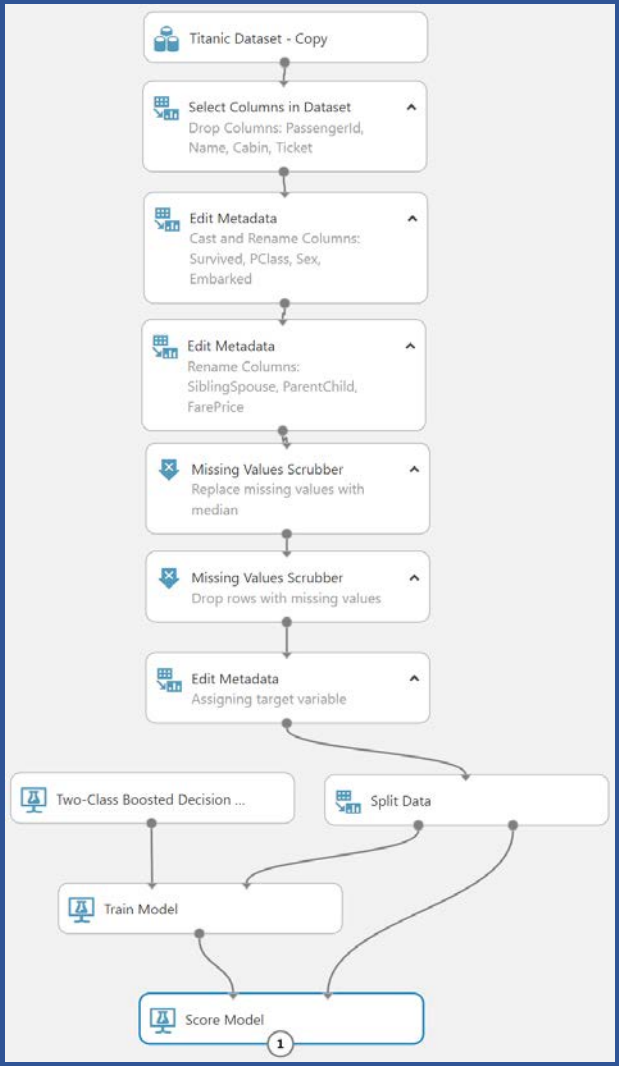
What to do

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

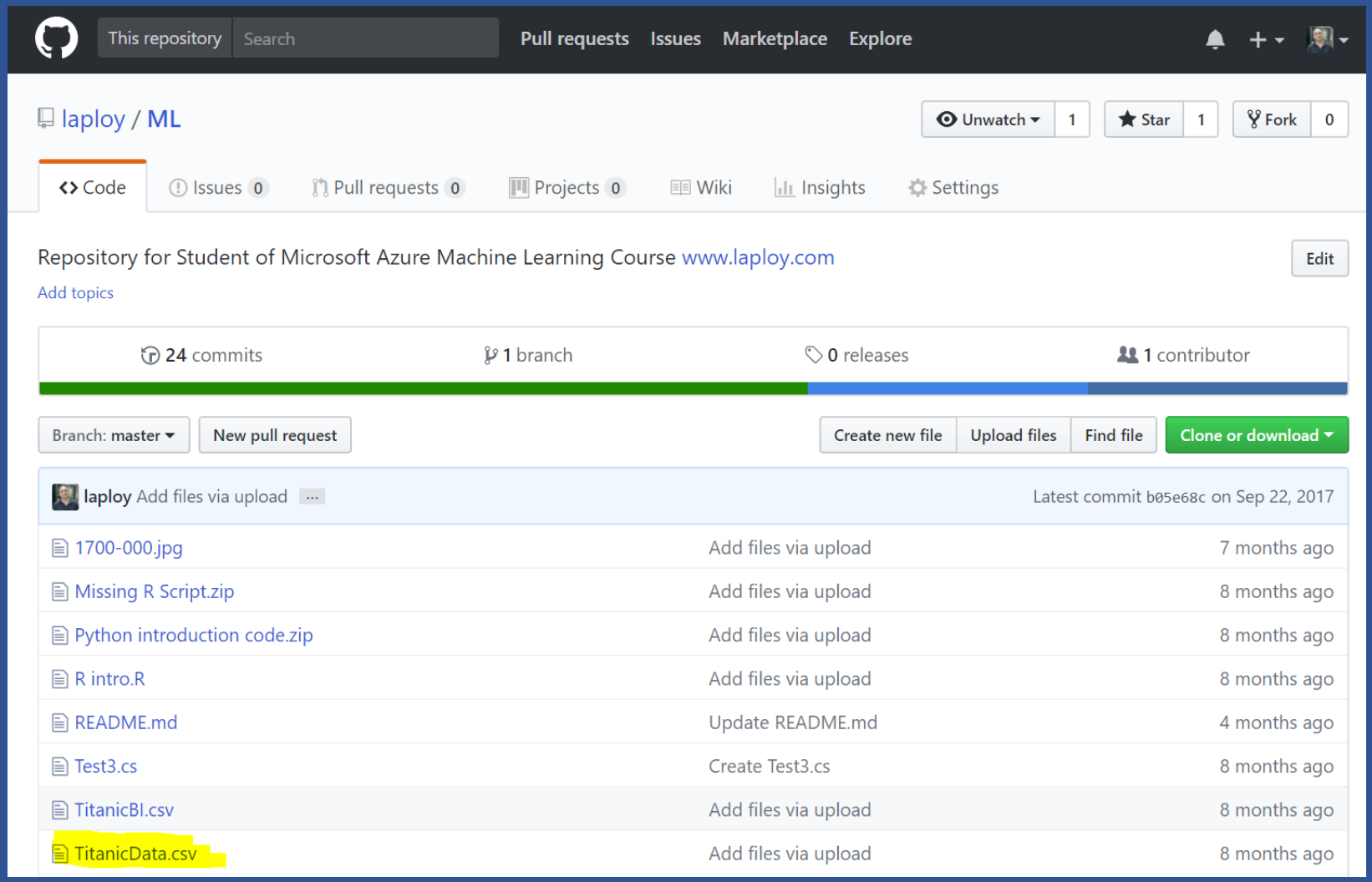
AML model development step

- 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

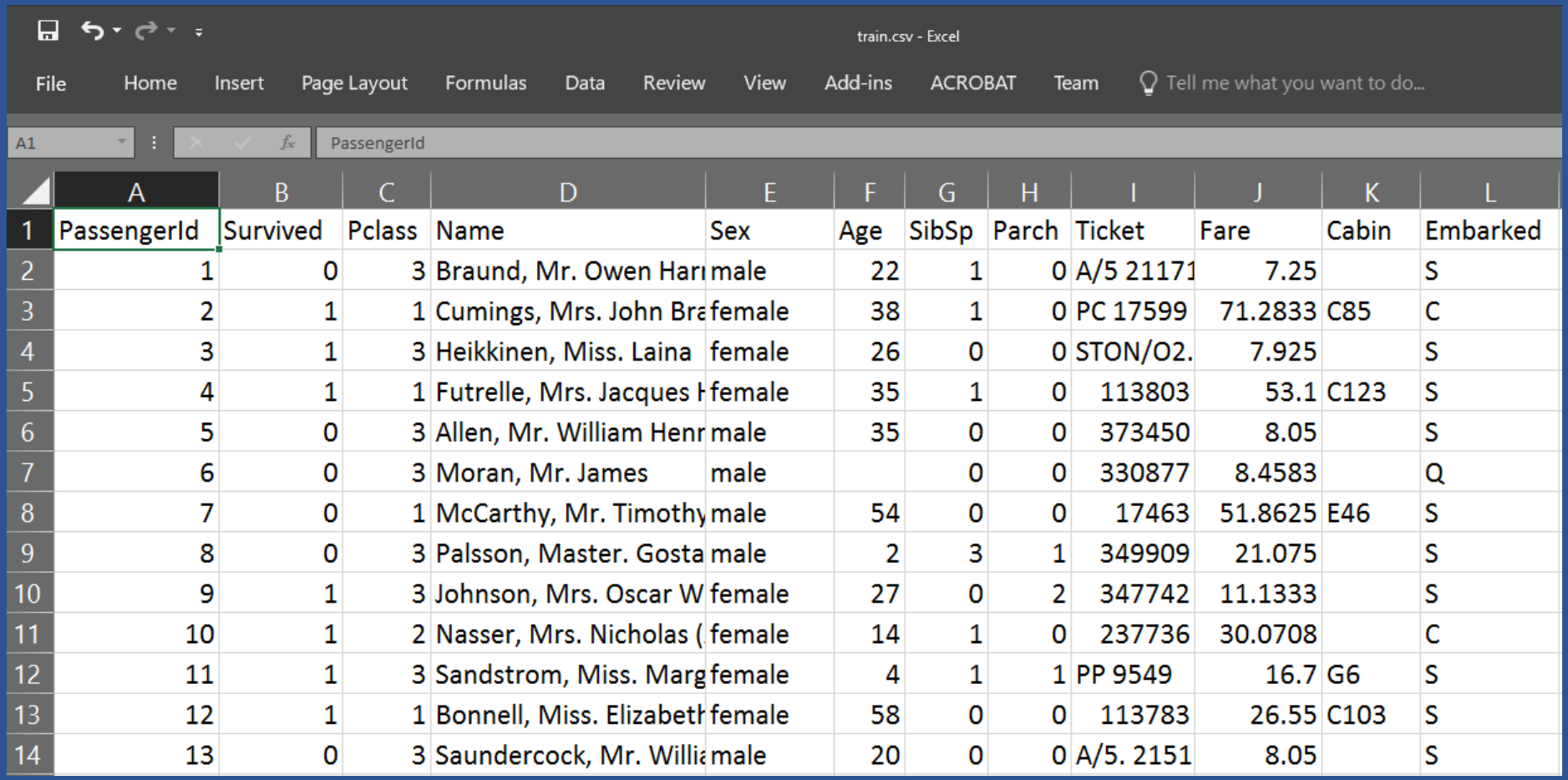
Building a classification model

Download Data set

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



Building a classification model



	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 Har	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. Willia	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 = 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

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

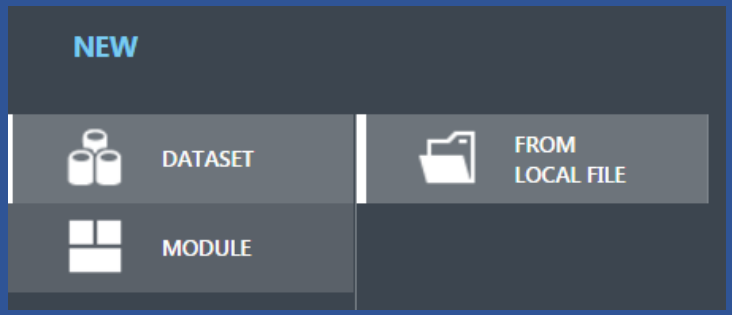
Building a classification model

Import Data set

1. Click DATASETS

2. Click NEW

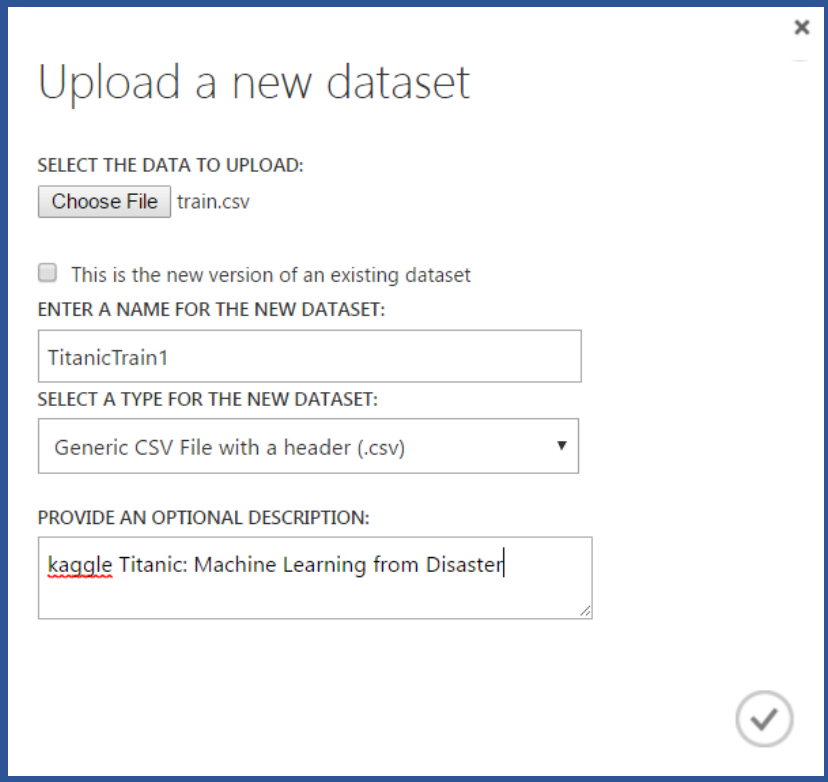
3. Click FROM LOCAL FILE



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

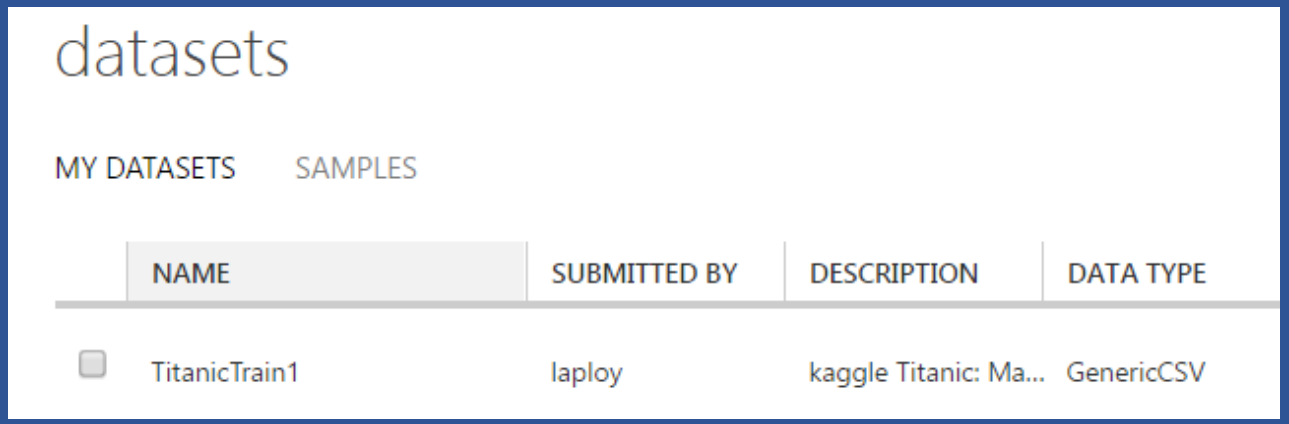


Building a classification model

Verify dataset uploaded

1. Click DATASETS

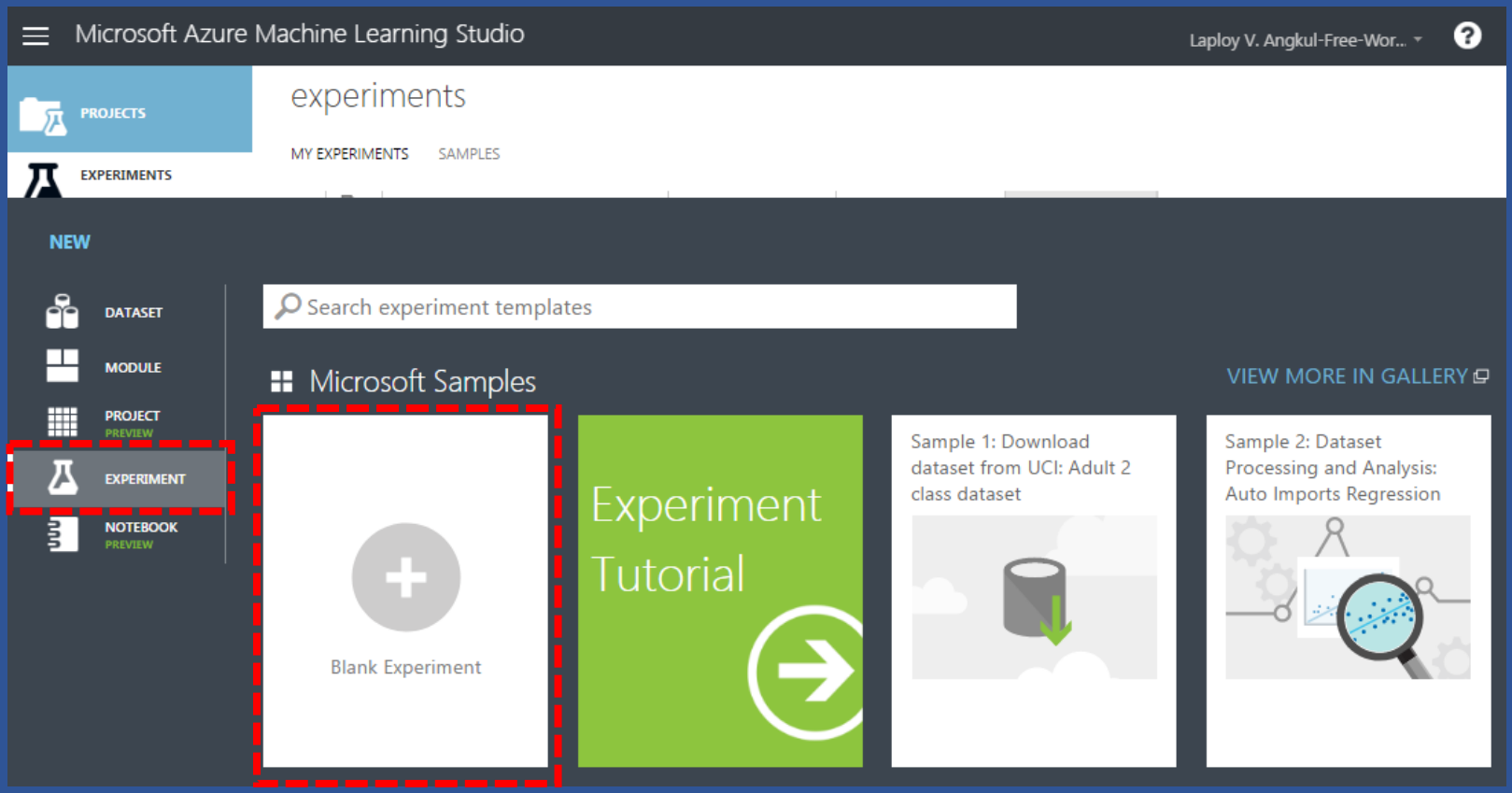
2. Make sure TitanicTrain1 is in MY DATASETS list



Building a classification model

Create New Experiment

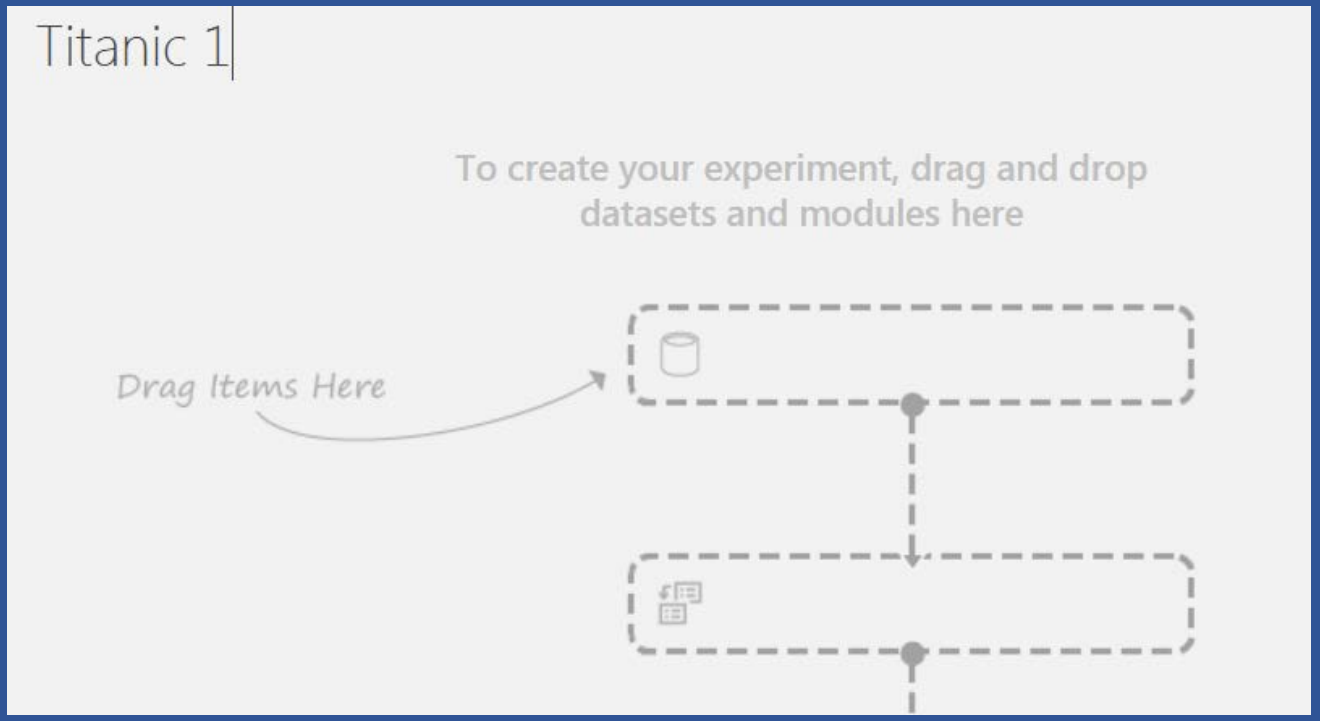
Create Blank Experiment



Building a classification model

Set experiment name

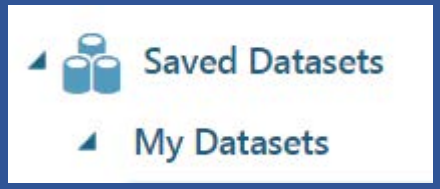
Type in name = Titanic 1



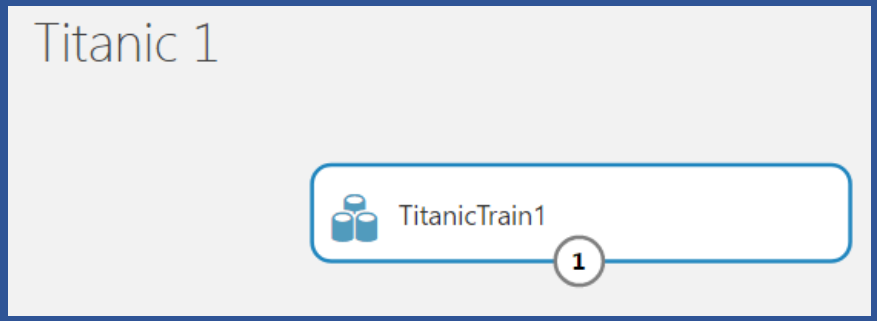
Building a classification model

Drag & drop dataset to canvas

1. Click Saved Datasets / My Datasets



2. Drag & drop TitanicTrain1 to canvas

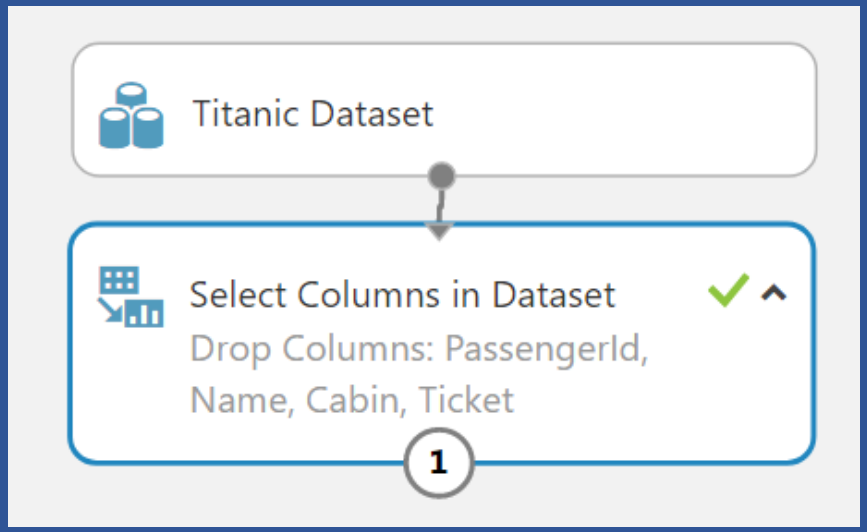


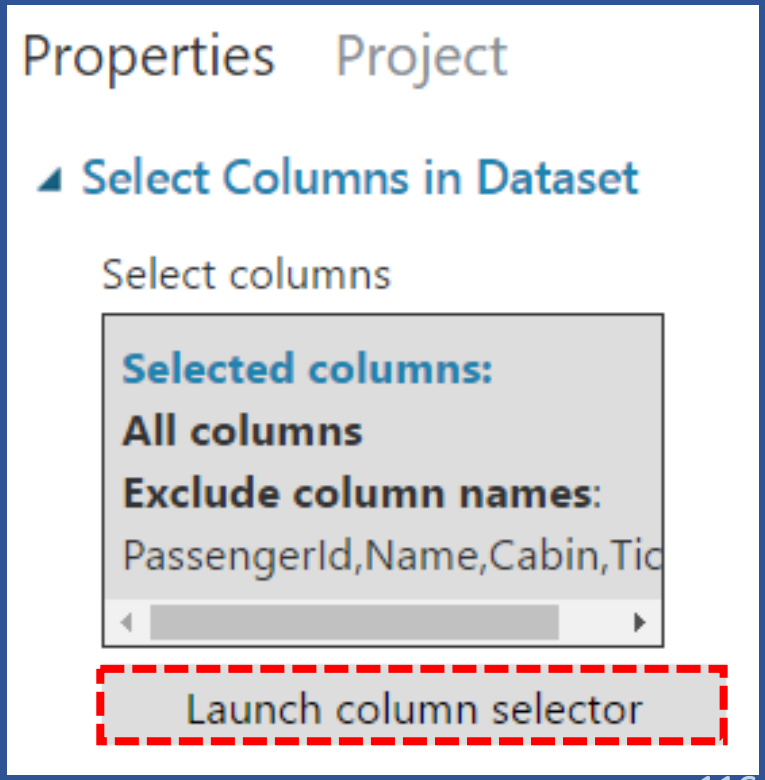
3. Visualize output

Building a classification model

Drop the columns PassengerID, Name, Ticket, Cabin

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

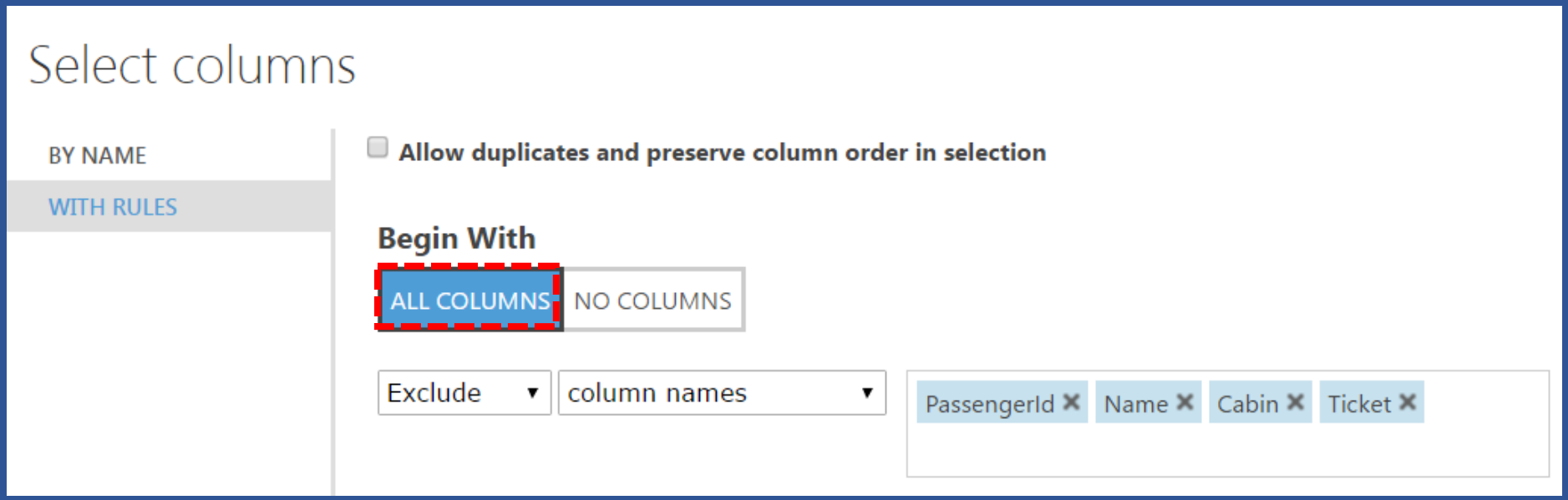




Building a classification model

Drop the columns PassengerID, Name, Ticket, Cabin

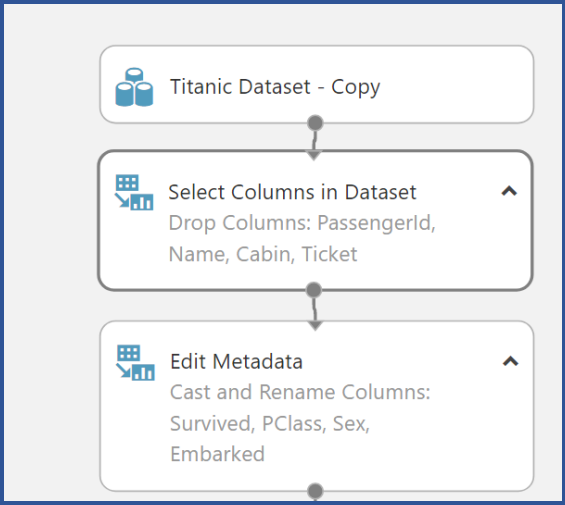
5. Begin With = ALL COLUMNS / Exclude / column name
6. Selected column PassengerID, Name, Ticket, Cabin
7. Click 
8. Visualize

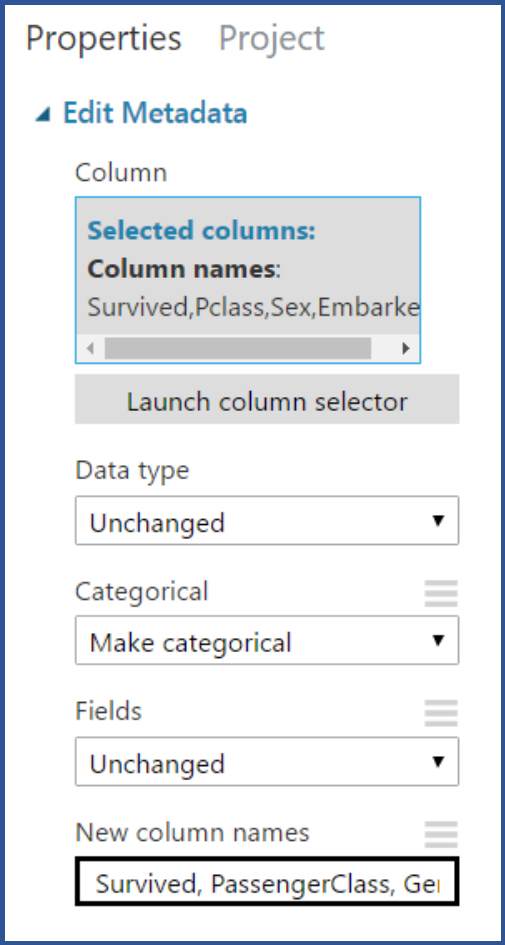


Building a classification model

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

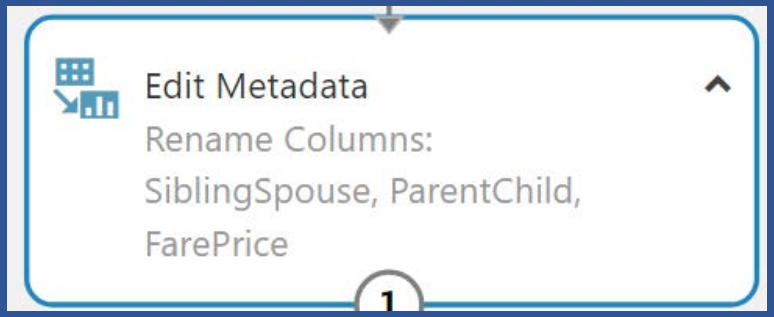


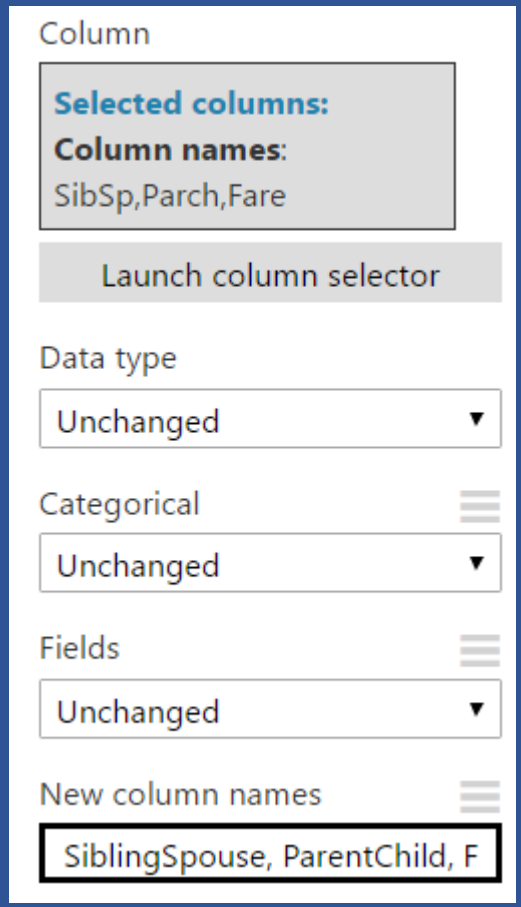


Building a classification model

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

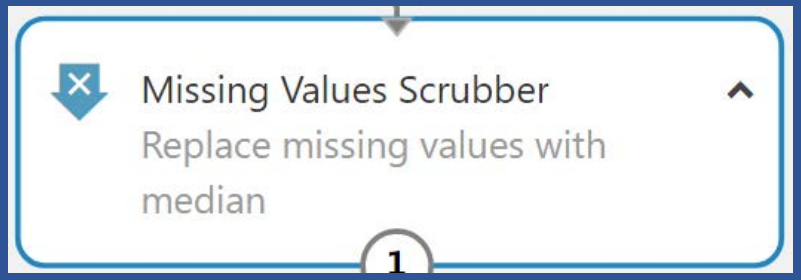


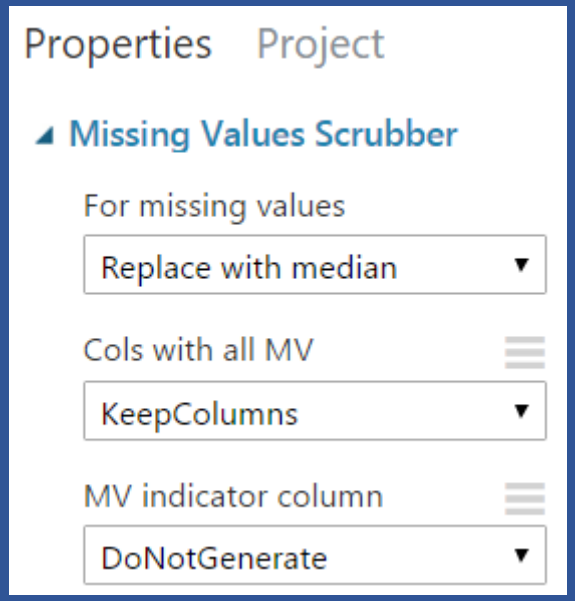


Building a classification model

Replace missing value with median

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

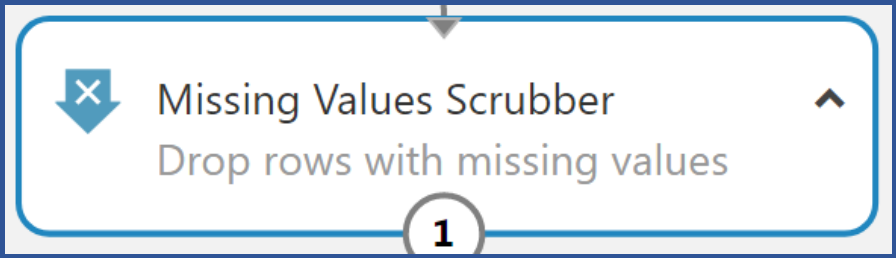


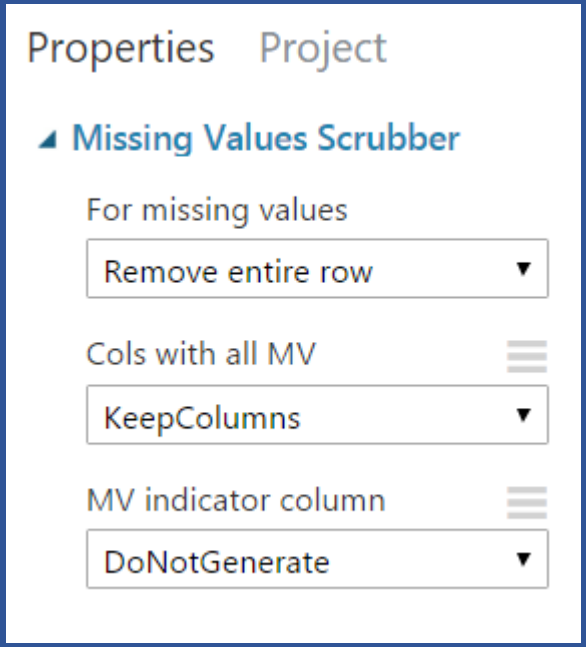


Building a classification model

Drop rows with missing data

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

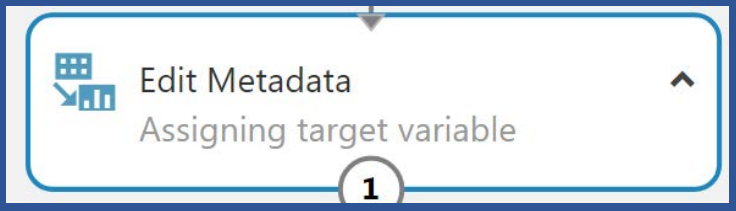


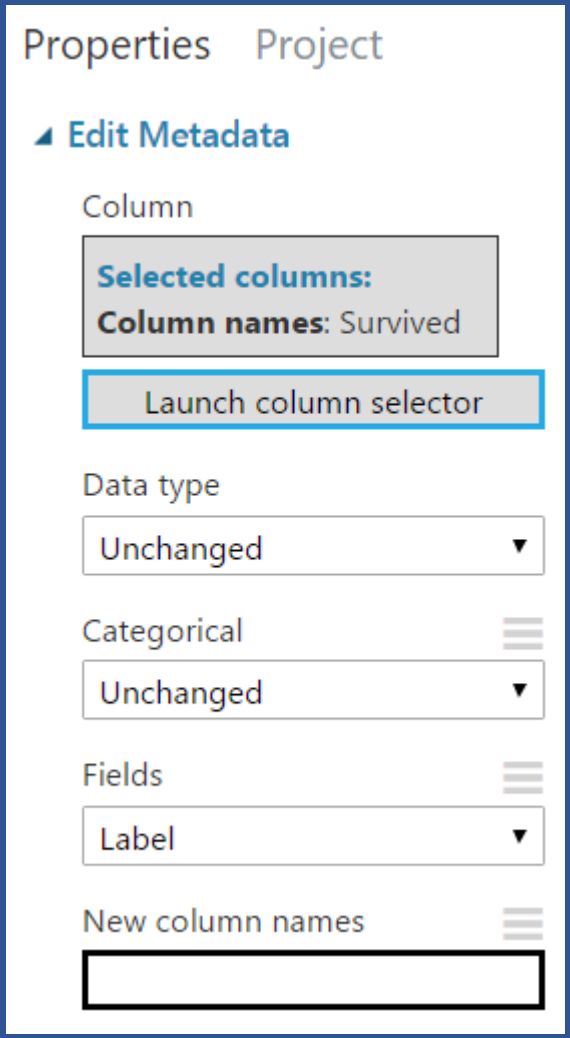


Building a classification model

Create Label

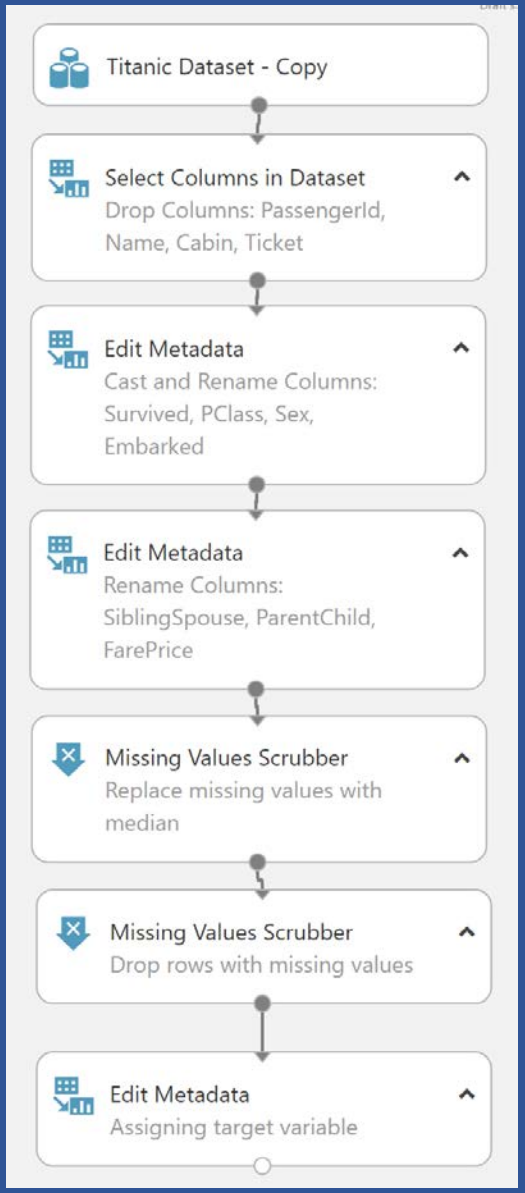
1. Drag & drop Edit Metadata
2. Comment = Assigning target variable
3. Selected column = Survived
4. Data type = Unchanged
5. Categorical = Unchanged
6. Fields = Label
7. New column name = -
8. Visualize





Building a classification model

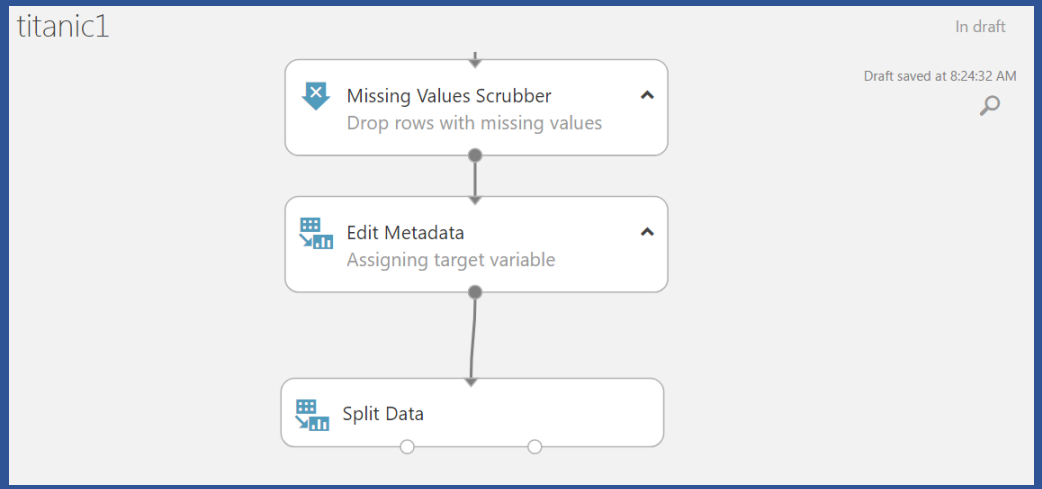
Import and Dataset preparation

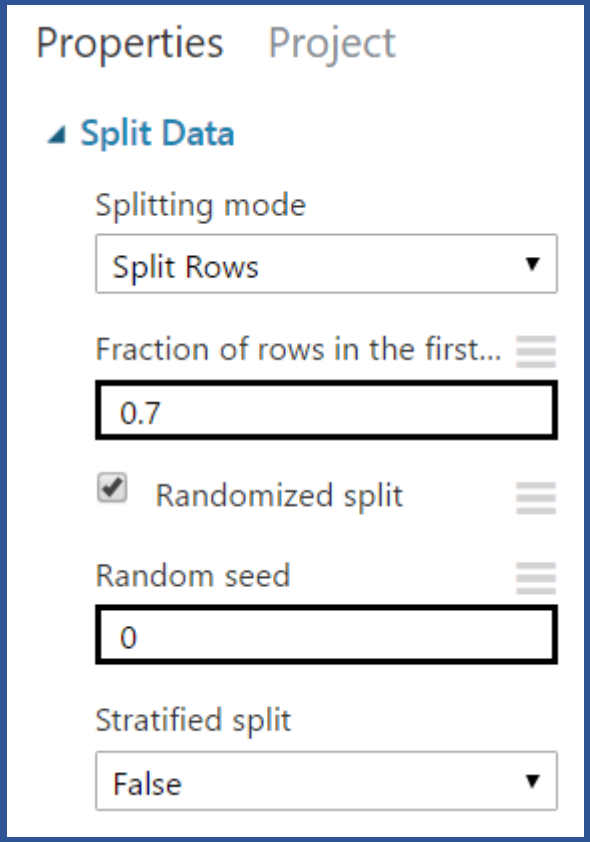


Building a classification model

Split data 70% training and 30% scoring

1. Drag & drop Split data
2. Set properties

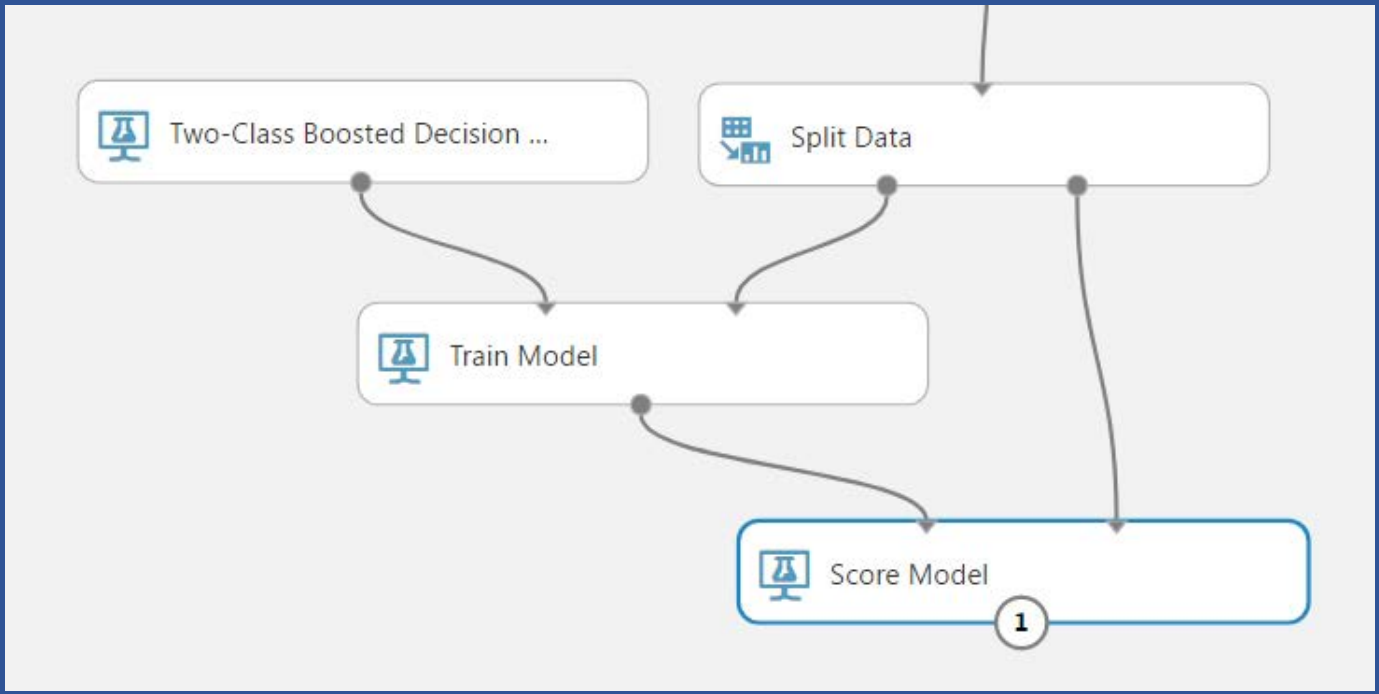




Building a classification model

Add Algorithm, Train and Score

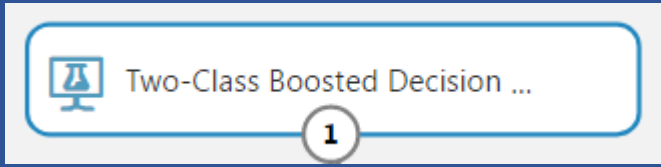
- Add Two-Class Boosted Decision tree
- Add Train Model
- Add Score Model

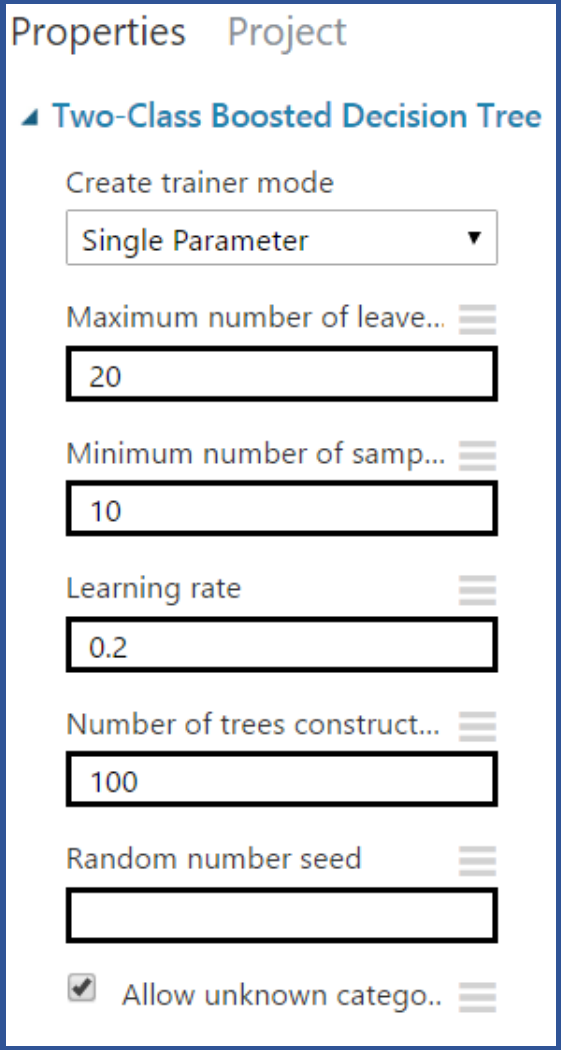


Building a classification model

Add Two-Class Boosted Decision tree

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

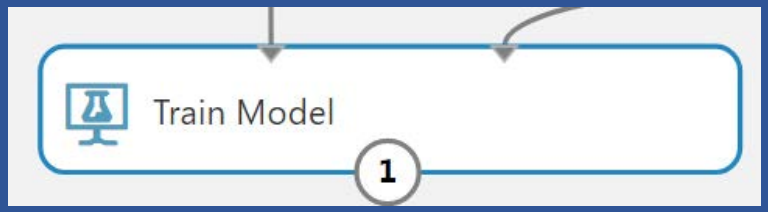


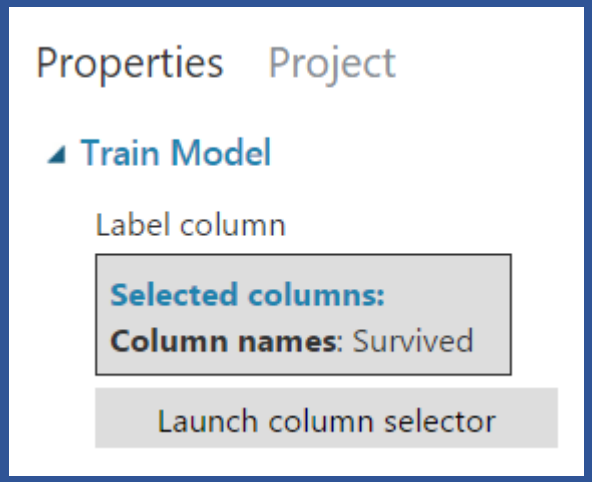


Building a classification model

Add Train Model

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

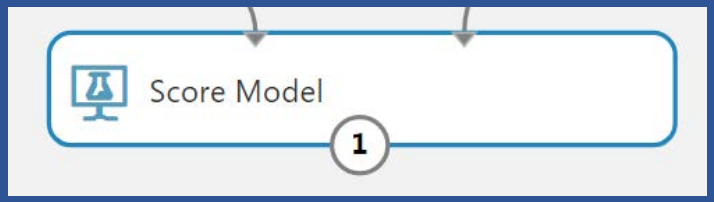


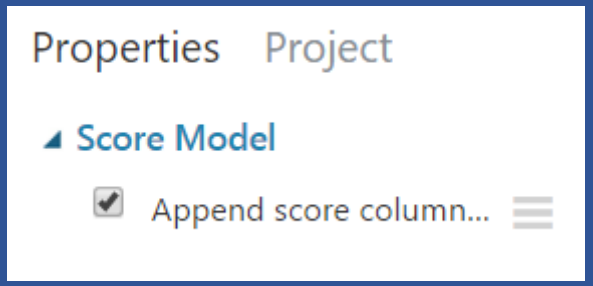


Building a classification model

Add Score Mode

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

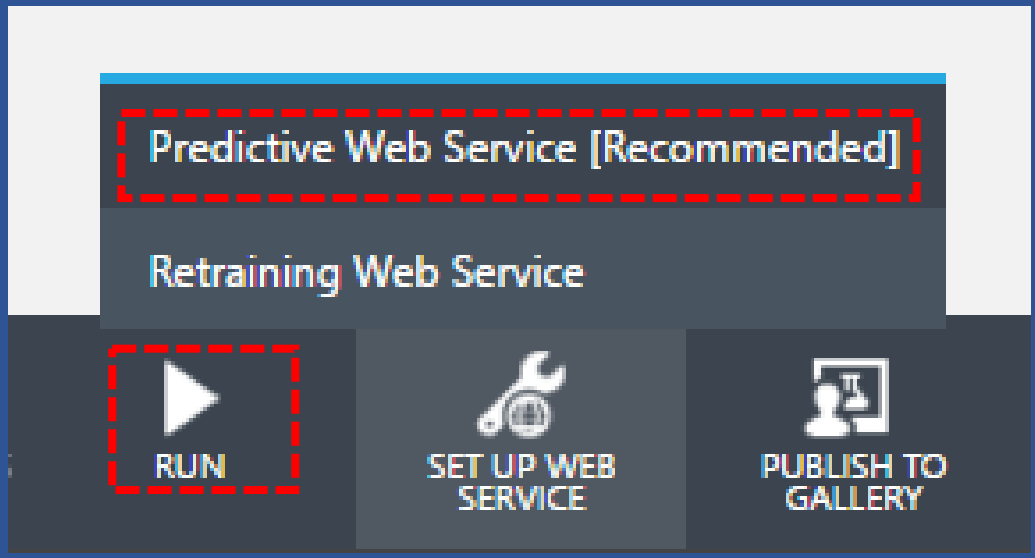




Building a classification model

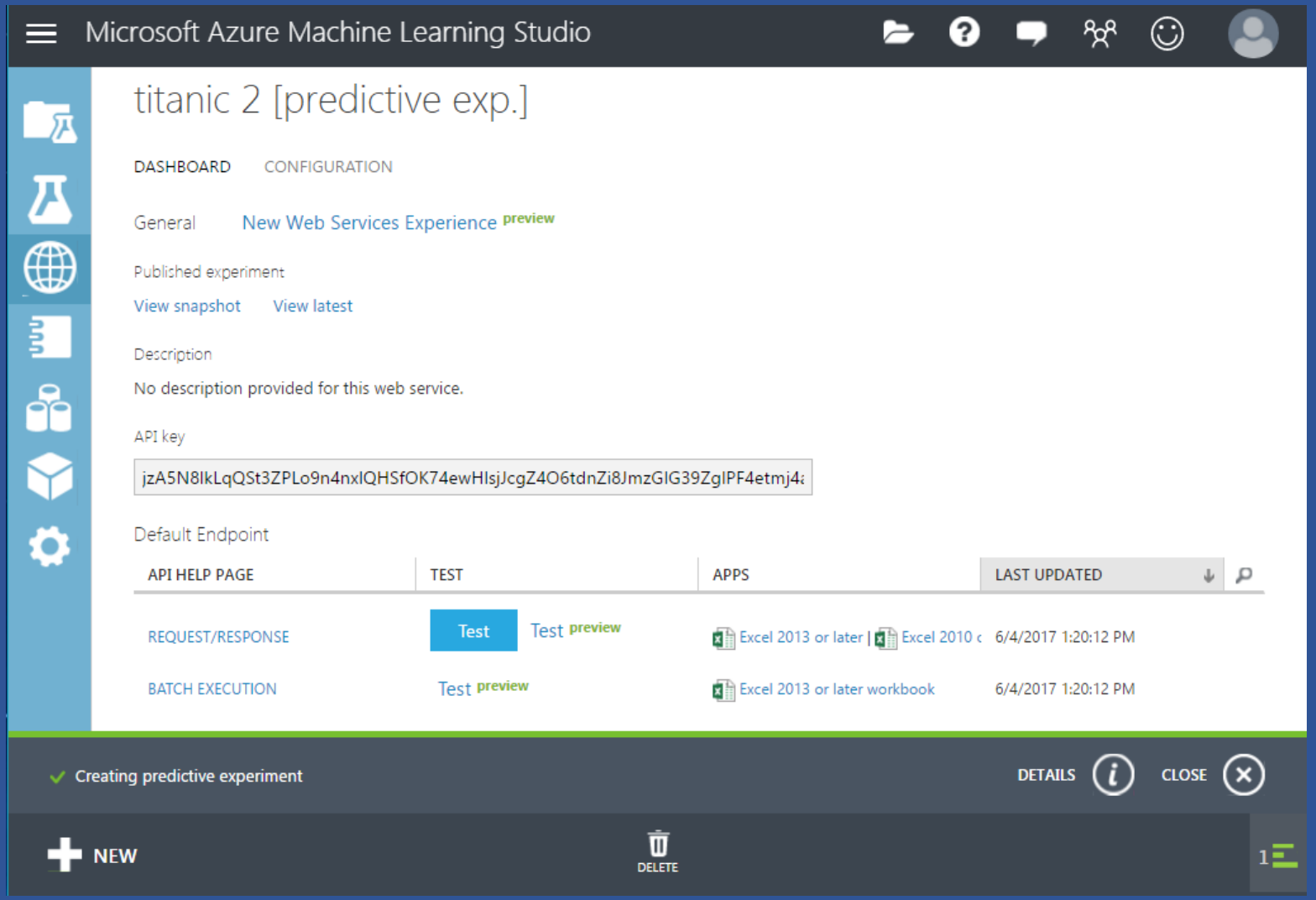
Create web service

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

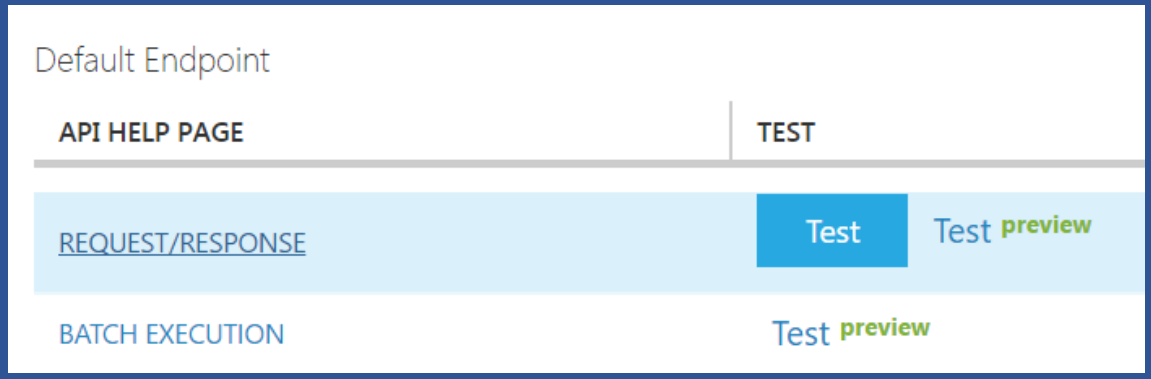


Building a classification model

Test web service

Web service testing

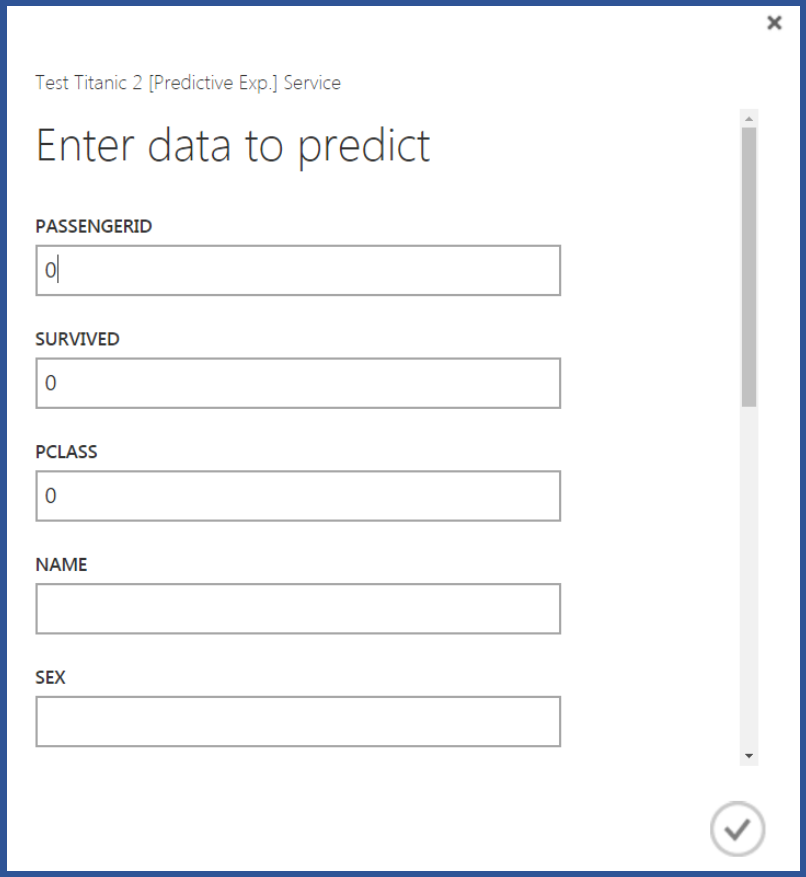
- REQUEST/RESPONSE Test
- REQUEST/RESPONSE Test preview
- REQUEST/RESPONSE Excel workbook test
- BATCH EXECUTION Test preview
- BATCH EXECUTION Excel workbook test



Building a classification model

REQUEST/RESPONSE Test

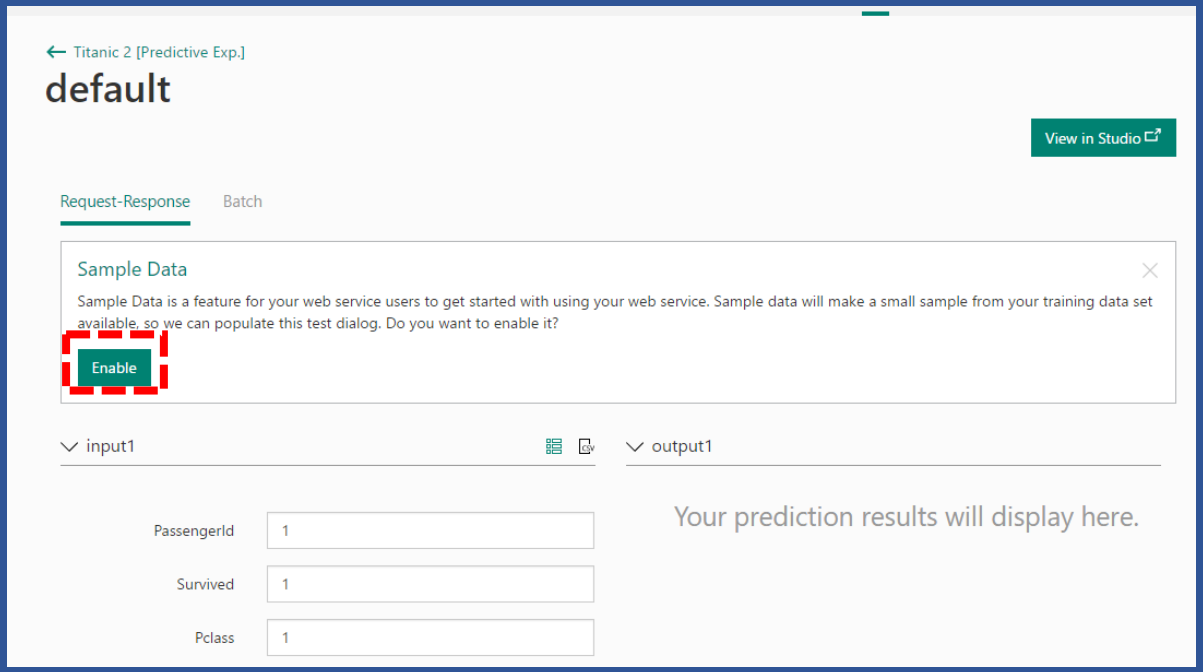
1. Test with know result
2. Open file kaggle test.csv
3. Take one passenger
4. Click REQUEST/RESPONSE Test
5. Fill in the form



Building a classification model

REQUEST/RESPONSE Test preview

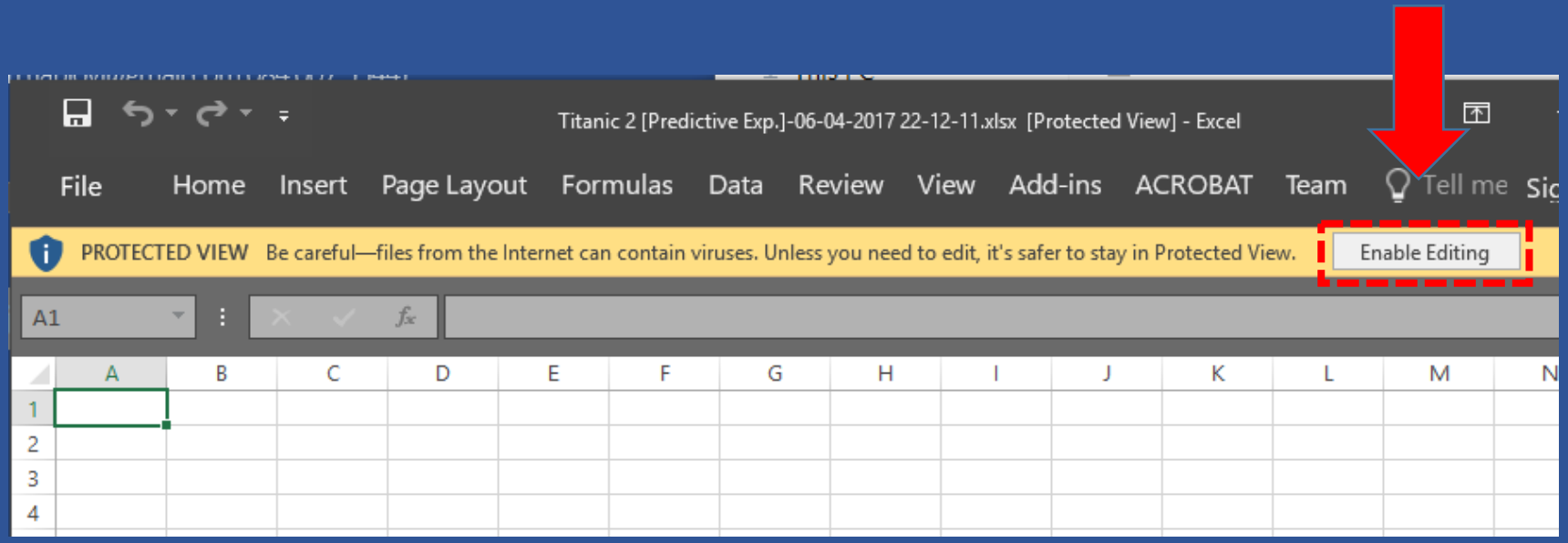
1. Test with know result
2. Open file kaggle test.csv
3. Take one passenger
4. Click REQUEST/RESPONSE Test preview
5. Click Enable (Sample Data)
6. Fill in the form
7. Click Test Request-Response



Building a classification model

REQUEST/RESPONSE Excel workbook test

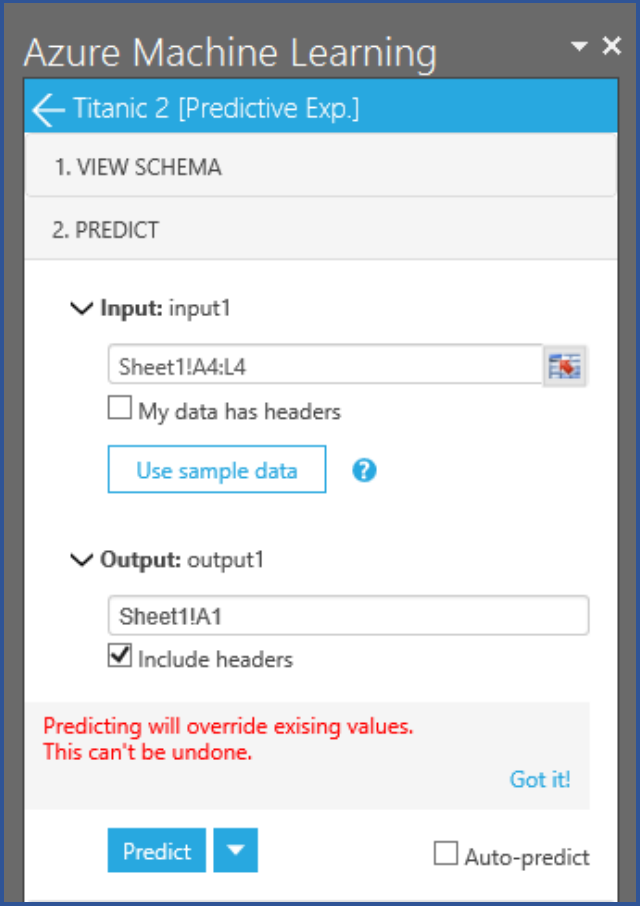
1. Test with know result
2. Open file kaggle test.csv
3. Take one passenger
4. Click REQUEST/RESPONSE Excel 2013 or later
5. Open file Titanic 2 [Predictive Exp.] on Desktop
6. Click Enable Editing



Building a classification model

REQUEST/RESPONSE Excel workbook test

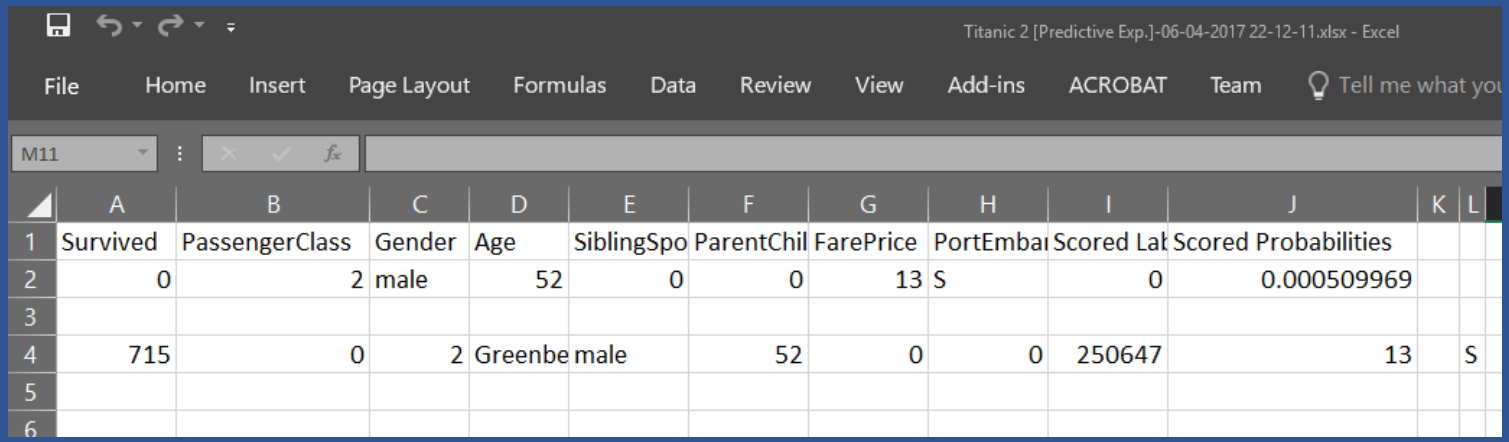
1. Input = Sheet1!A4L4
2. My data has headers = uncheck
3. Output = Sheet1!A1
4. Include headers = check
5. Copy a line from file kaggle test.csv to A4
6. Click Predict



Building a classification model

REQUEST/RESPONSE Excel workbook test

Test result



	A	B	C	D	E	F	G	H	I	J	K	L
1	Survived	PassengerClass	Gender	Age	SiblingSpo	ParentChil	FarePrice	PortEmbai	Scored Lal	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