

Sentiment Analysis

binary classification

experimental in Azure ML

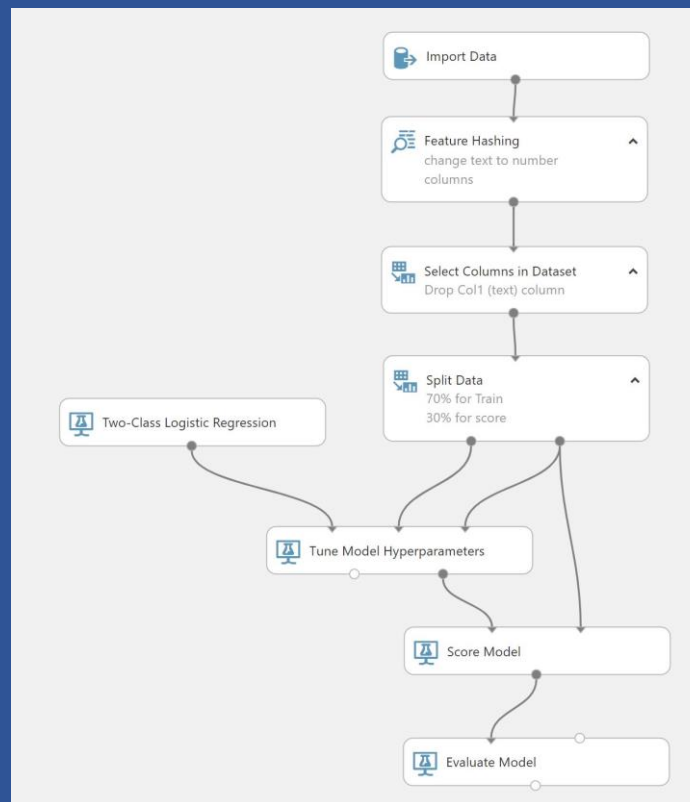
In this session

- Task and Data
- Create a new experiment in Azure ML Studio
- Import Data
- Run and Visualize
- Add Feature Hashing
- Run and Visualize

- Drop Col1
- Add Split Data
- Add Tune Model Hyperparameters
- Add Two-class Logistic Regression
- Add Score Model
- Add Evaluate Model

The finished model

<https://raw.githubusercontent.com/laploy/ML.NET/master/Sentiment/Azure-ML-Model.JPG>



Task and Data

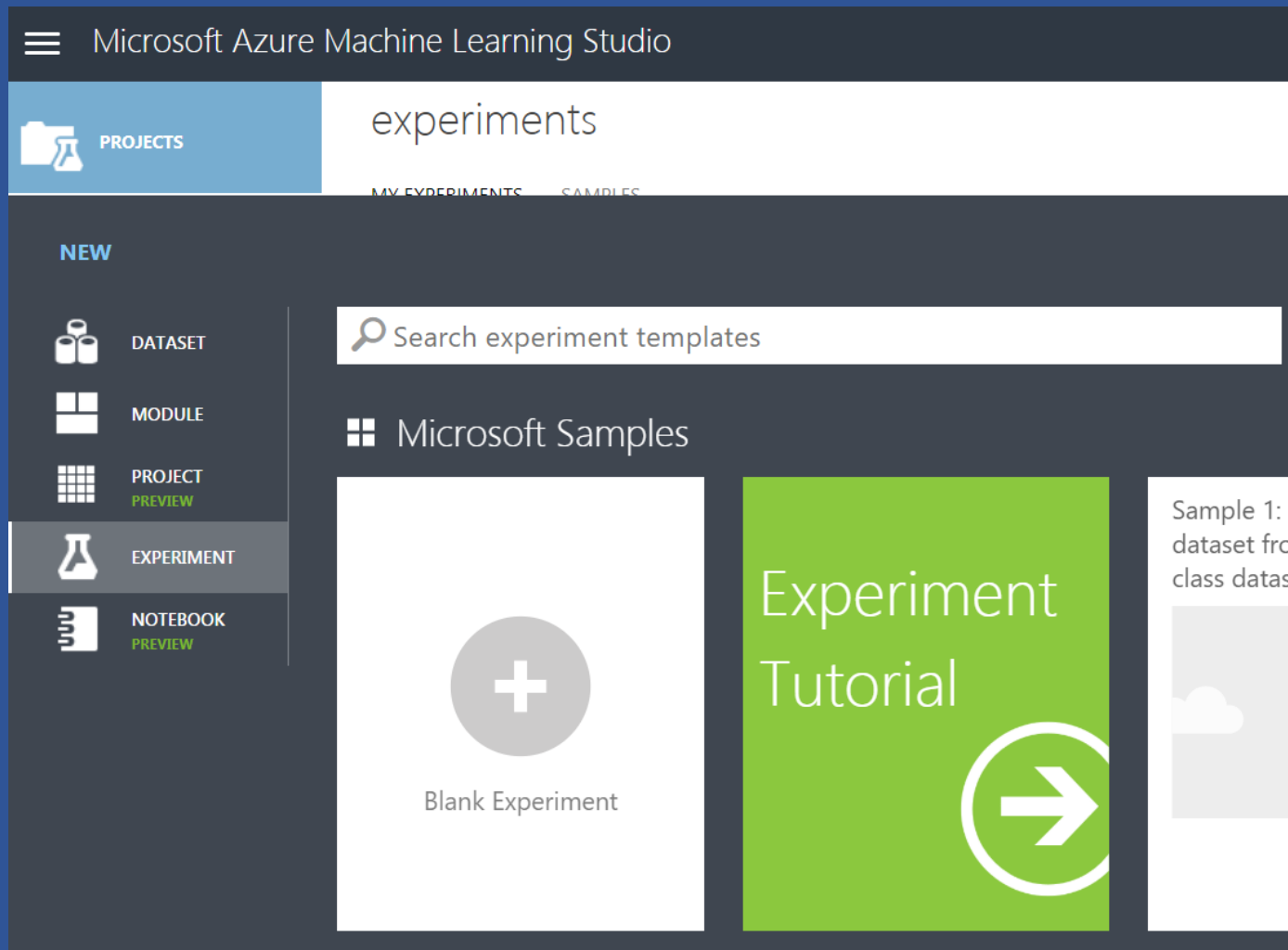
Task = Make a prediction if the user message is positive or negative

Data for train

- file name: yelp_labelled.txt
- Row: 1,000
- Column : 2
- Label: Col2

	A	B
1	Wow... Loved this place.	1
2	Crust is not good.	0
3	Not tasty and the texture was just nasty.	0
4	Stopped by during the late May bank holiday off Rick Steve recommen	1
5	The selection on the menu was great and so were the prices.	1
6	Now I am getting angry and I want my damn pho.	0
7	Honeslty it didn't taste THAT fresh.)	0
8	The potatoes were like rubber and you could tell they had been made	0
9	The fries were great too.	1
10	A great touch.	1

Create a new experiment in Azure ML Studio



Import Data

https://raw.githubusercontent.com/laploy/ML.NET/master/Sentiment/yelp_labelled.txt

Draft saved at 8:24:28 AM

Import Data

1

Import Data

Launch Import Data Wizard

Data source

Web URL via HTTP

Data source URL

https://raw.githubusercontent.com/laploy/ML.NET/master/Sentiment/yelp_labelled.txt

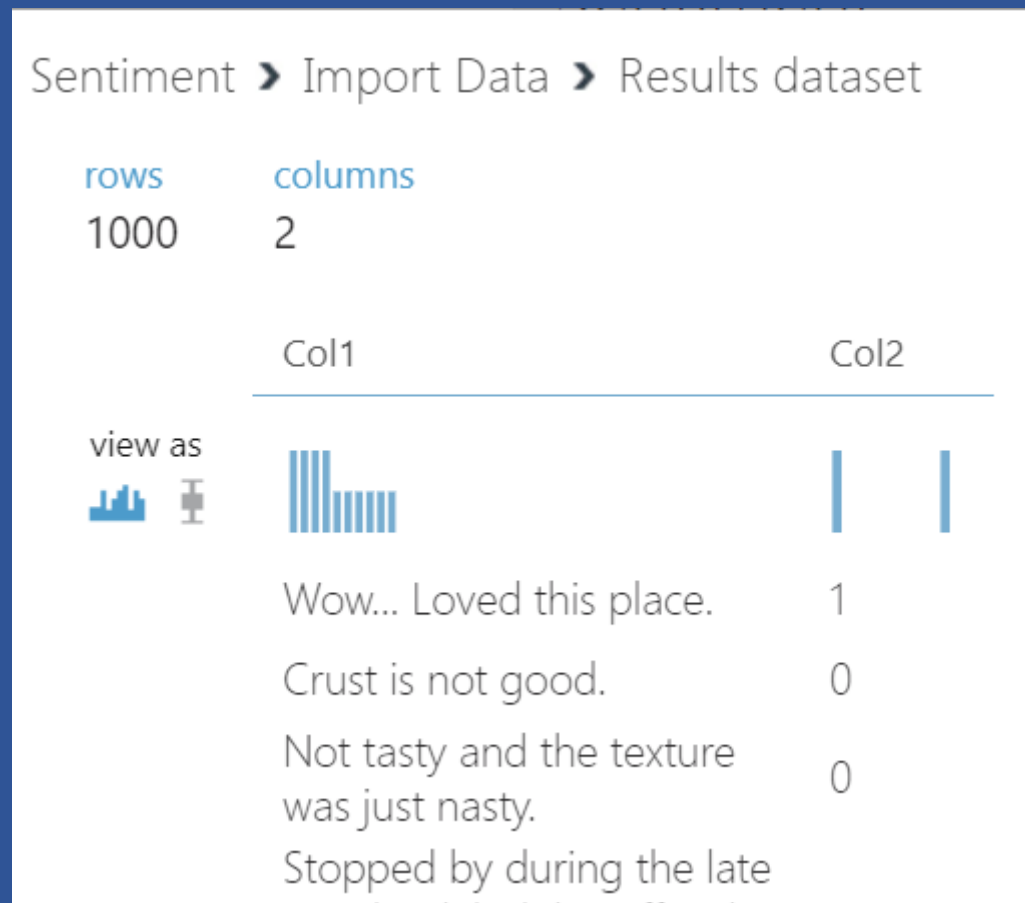
Data format

TSV

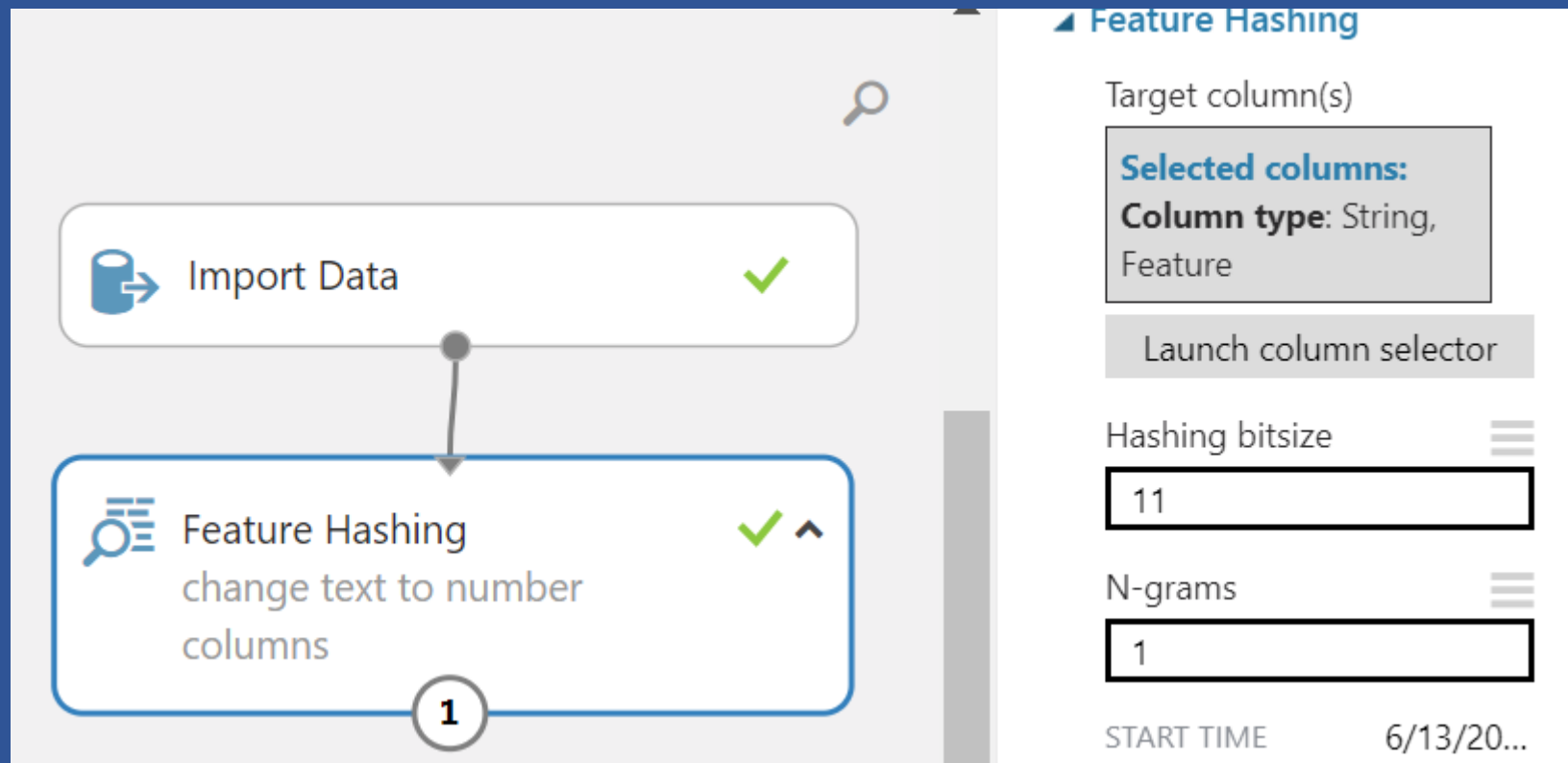
☐ CSV or TSV has header

☐ Use cached results

Run and Visualize



Add Feature Hashing



The screenshot displays the Azure ML Studio interface. On the left, a pipeline is shown with two steps: 'Import Data' and 'Feature Hashing'. The 'Feature Hashing' step is highlighted with a blue border and a circled '1' below it. The right-hand pane shows the configuration for the 'Feature Hashing' step.

Feature Hashing

Target column(s)
Selected columns:
Column type: String, Feature

Launch column selector

Hashing bitsize
11

N-grams
1

START TIME 6/13/20...

Run and Visualize

Sentiment > Feature Hashing > Transformed dataset

rows

1000

columns

2050

view as



Col1

Col2

Col1_HashingFeature_1

Col1_HashingFeature_2

Col1



Wow... Loved this place.

1

0

0

0

Crust is not good.

0

0

0

0

Not tasty and the texture was just nasty.

0

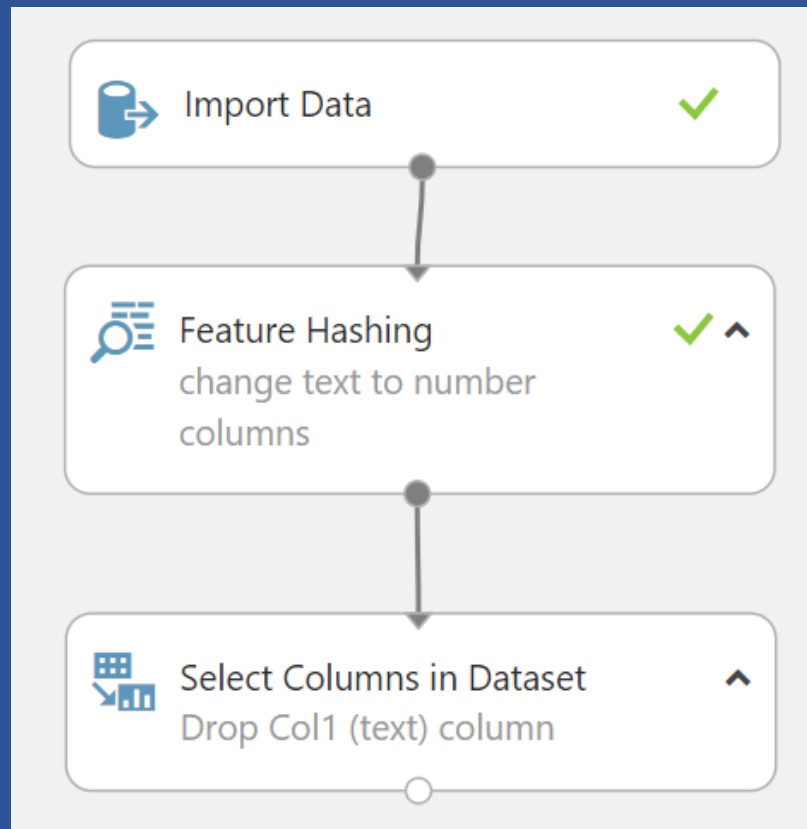
0

0

0

Stopped by during the late May bank

Add **Select Columns in Dataset** module to Drop Col1

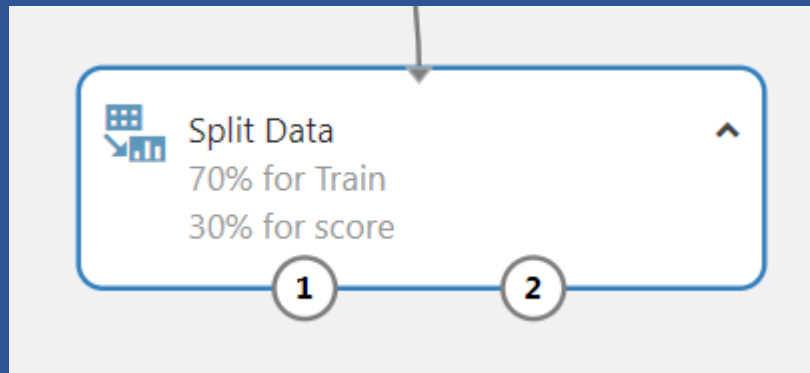


Begin With

ALL COLUMNS **NO COLUMNS**

Exclude **column names** **Col1** **x**

Add Split Data



Properties Project

Split Data

Splitting mode
Split Rows ▼

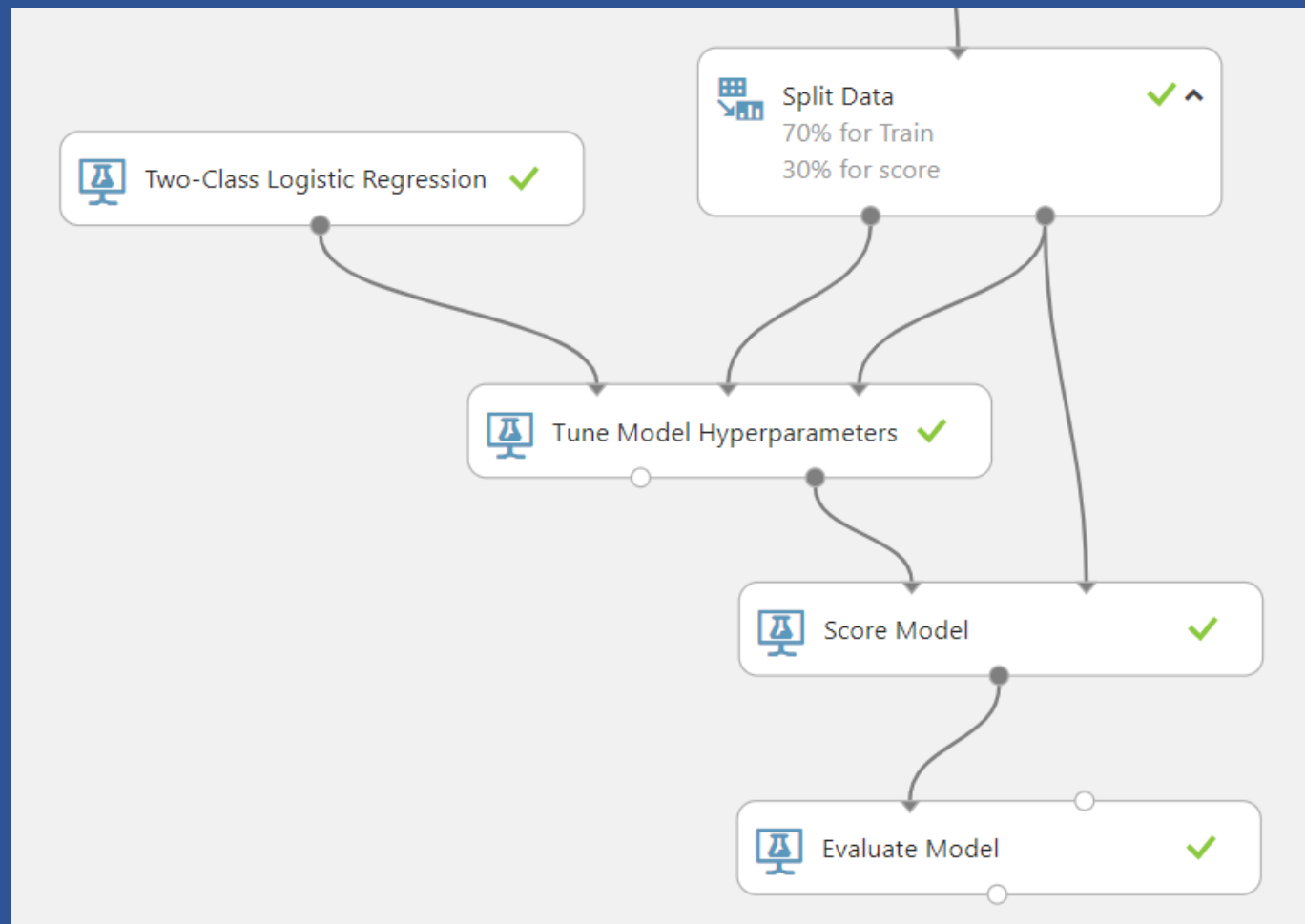
Fraction of rows in the first...
0.7

☒ Randomized split

Random seed
0

Stratified split
False ▼

Add Tune Model Hyperparameters, Two-class Logistic Regression, Score Model, and Evaluate Model



Run and View evaluation Results

True Positive	False Negative	Accuracy	Precision
106	46	0.690	0.693
False Positive	True Negative	Recall	F1 Score
47	101	0.697	0.695
Positive Label	Negative Label		
1	0		

Next Step

Sentiment AutoML