AI Workshop – Predict employee leave: will they leave, or will they stay?

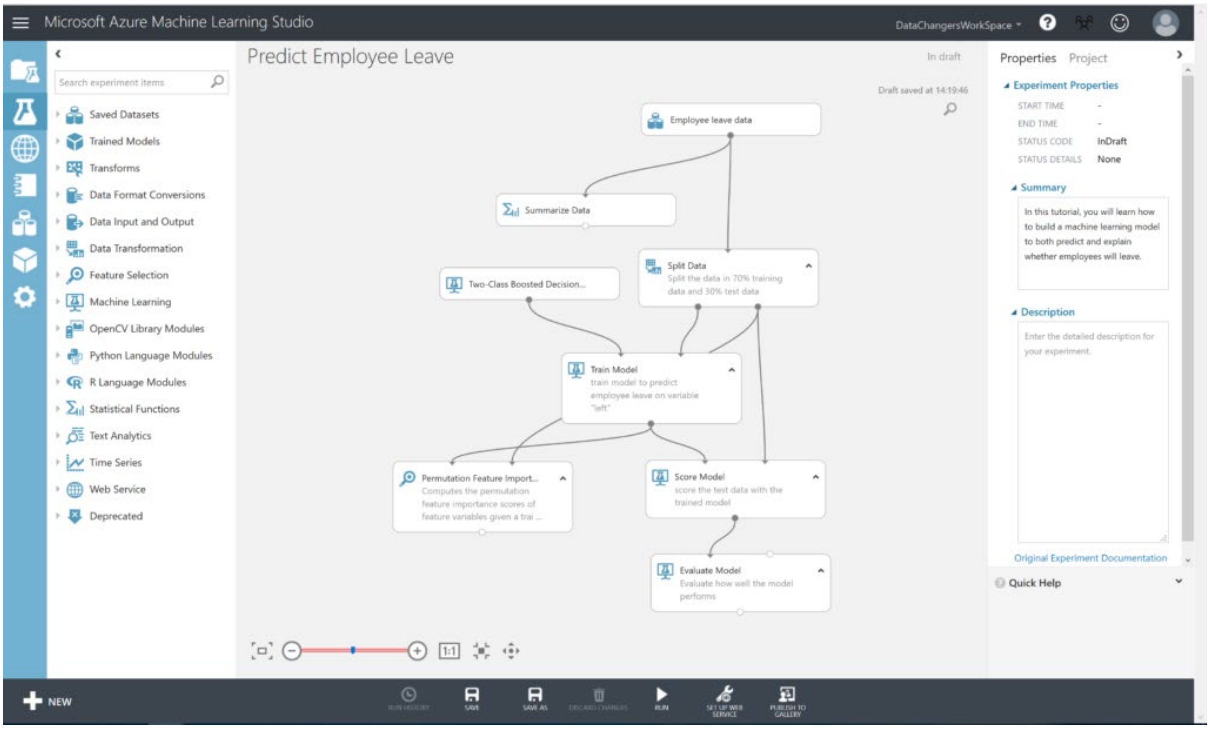
Imagine you are an HR-Manager, and you would like to know which employees are likely to stay, and which might leave your company. Besides you would like to understand which factors contribute to leaving your company. You have gathered data in the past (well, in this case Kaggle simulated a dataset for you, but just imagine), and now you can start with this **Hands On Lab** to build your prediction model to see if that can help you.

In this lab, you will learn how to create a machine learning module with Azure Machine Learning Studio that predicts whether an employee will stay or leave your company. We are aware of the limitations of the dataset but the objective of this hands-on lab is to inspire you to explore the possibilities of using machine learning for your own research, and not to build the next HR-solution.

You will follow several steps to explore the data and build a machine learning model to predict whether an employee will leave or not, and why.

- Step 1: Get the starting experiment
- Step 2: Get a first understanding of the data
- Step 3: Prepare a training and a test set
- Step 4: Train the model
- Step 5: Score the test set
- Step 6: Evaluate the results
- Step 7: Gain insights on the why
- Step 8: Publish your model
- Step 9: Deploy your model as a web service
- Step 10: Test your model

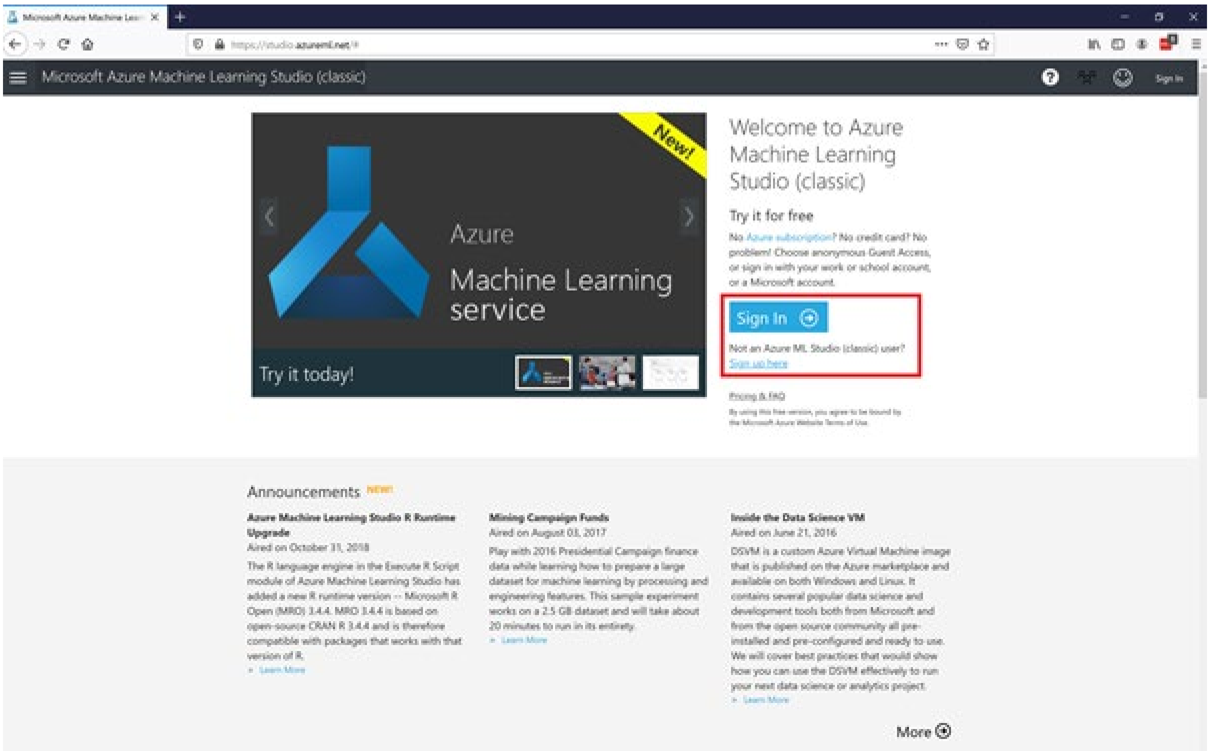
You will build this prediction model with the **Azure Machine Learning Studio**. The complete model will look like this:



In this tutorial, you will learn how to build a machine learning model to both predict and explain whether employees will leave.

Prerequisites: Get Access to Azure Machine Learning Studio

Go to https://studio.azureml.net/ and select **Sign up here** for Azure ML Studio.


Microsoft Azure Machine Lea...
https://studio.azureml.net/#
Microsoft Azure Machine Learning Studio (classic)
Sign In
New!
Azure
Machine Learning service
Try it today!
Sign In
More

Welcome to Azure Machine Learning Studio (classic)

Try it for free

No Azure subscription? No credit card? No problem! Choose anonymous Guest Access, or sign in with your work or school account, or a Microsoft account.

Not an Azure ML Studio (classic) user?
Sign up here

Pricing & FAQ

By using this free version, you agree to be bound by the Microsoft Azure Website Terms of Use.

Announcements NEW!

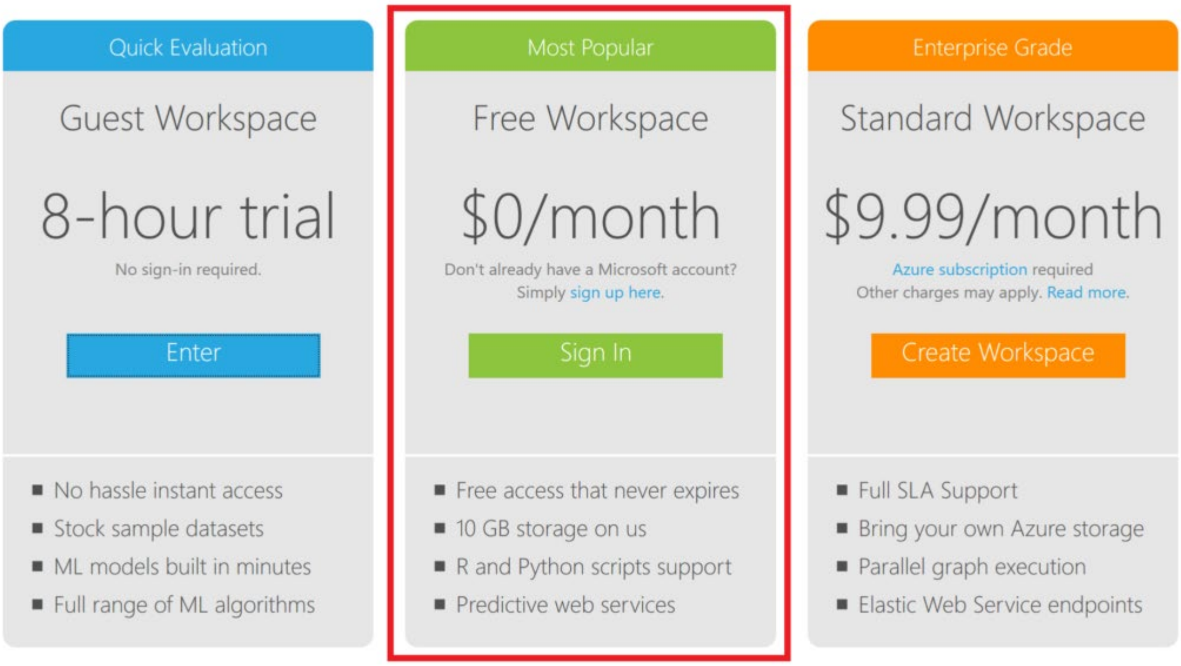
Azure Machine Learning Studio R Runtime Upgrade
Aired on October 31, 2018
The R language engine in the Execute R Script module of Azure Machine Learning Studio has added a new R runtime version -- Microsoft R Open (MRO) 3.4.4. MRO 3.4.4 is based on open-source CRAN R 3.4.4 and is therefore compatible with packages that works with that version of R.
» Learn More

Mining Campaign Funds
Aired on August 03, 2017
Play with 2016 Presidential Campaign finance data while learning how to prepare a large dataset for machine learning by processing and engineering features. This sample experiment works on a 2.5 GB dataset and will take about 20 minutes to run in its entirety.
» Learn More

Inside the Data Science VM
Aired on June 21, 2016
DSVM is a custom Azure Virtual Machine image that is published on the Azure marketplace and available on both Windows and Linux. It contains several popular data science and development tools both from Microsoft and from the open source community all pre-installed and pre-configured and ready to use. We will cover best practices that would show how you can use the DSVM effectively to run your next data science or analytics project.
» Learn More

You will need a Windows LiveID to sign in. If you don't have one, you can sign up here: https://signup.live.com/

Hereafter, you can select the Free Workspace option:



Quick Evaluation

Guest Workspace

8-hour trial

No sign-in required.

Enter

- No hassle instant access
- Stock sample datasets
- ML models built in minutes
- Full range of ML algorithms

Most Popular

Free Workspace

$0/month

Don't already have a Microsoft account? Simply sign up here.

Sign In

- Free access that never expires
- 10 GB storage on us
- R and Python scripts support
- Predictive web services

Enterprise Grade

Standard Workspace

$9.99/month

Azure subscription required
Other charges may apply. Read more.

Create Workspace

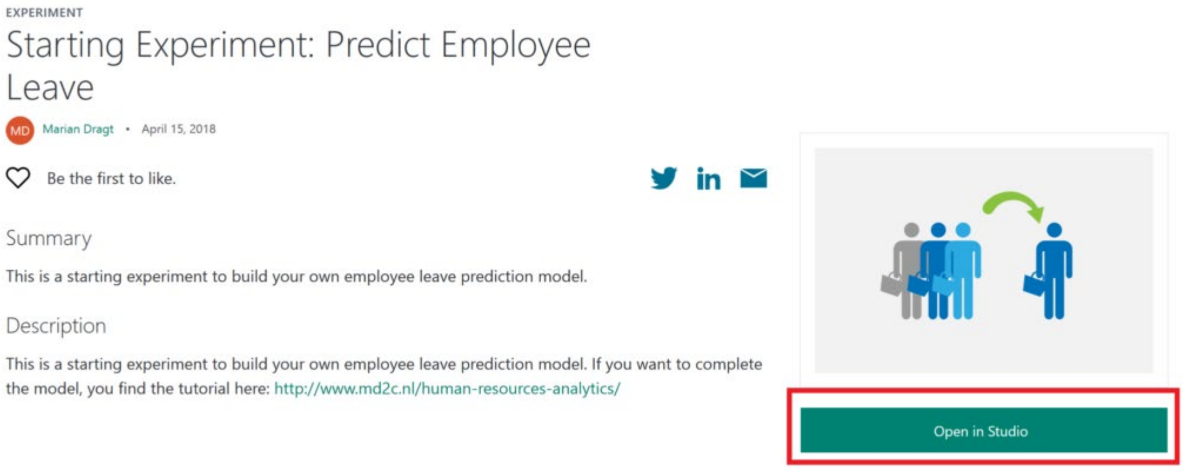
- Full SLA Support
- Bring your own Azure storage
- Parallel graph execution
- Elastic Web Service endpoints

Step 1: Get the starting experiment

We created a starting experiment for you on the **Azure AI Gallery** to give you a smooth start.

Click **HERE** to open the starting experiment.

This experiment uses a **simulated dataset** from **Kaggle**. You have to open the experiment in your studio by clicking on the green button "Open in Studio". This will open the Azure Machine Learning Studio in a browser, and you can copy the experiment to your free workspace.



EXPERIMENT

Starting Experiment: Predict Employee Leave

MD Marian Dragt • April 15, 2018

Be the first to like.

Summary

This is a starting experiment to build your own employee leave prediction model.

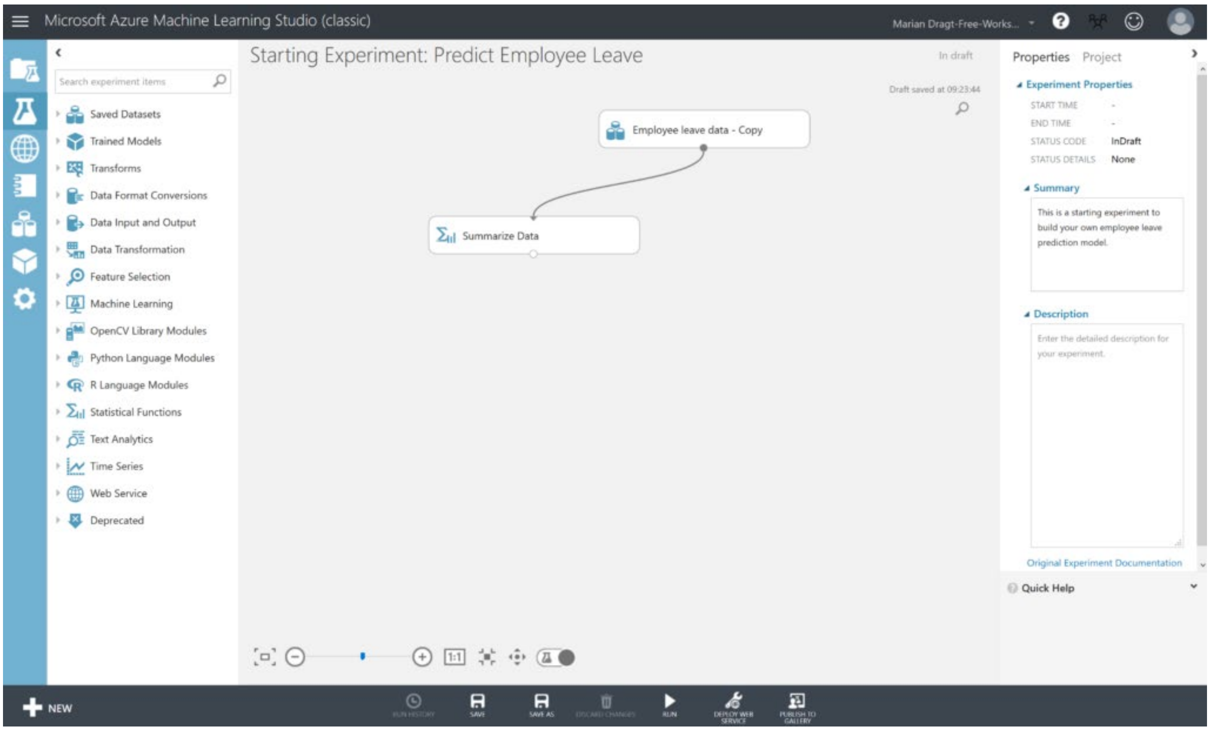
Description

This is a starting experiment to build your own employee leave prediction model. If you want to complete the model, you find the tutorial here: http://www.md2c.nl/human-resources-analytics/

Open in Studio

Step 2: Inspecting the data

In the **Starting experiment: Predict Employee Leave** experiment, you will find the **Employee Leave data** on the canvas, together with a **Summarize Data** module.



Starting Experiment: Predict Employee Leave

In draft

Draft saved at 09:23:44

Experiment Properties

START TIME	-
END TIME	-
STATUS CODE	InDraft
STATUS DETAILS	None

Summary

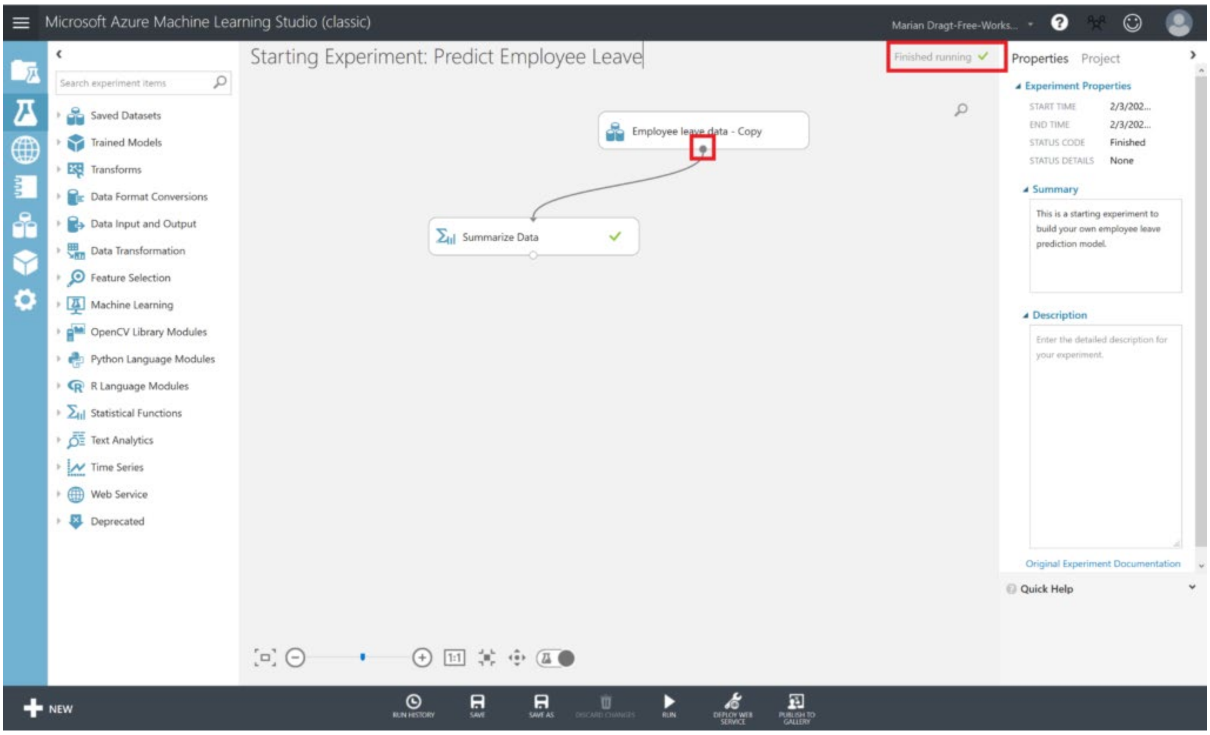
This is a starting experiment to build your own employee leave prediction model.

Description

Enter the detailed description for your experiment.

Original Experiment Documentation

If you look at the top corner right, you can see that the experiment is “in draft”. This means that it hasn’t been saved, nor that it has been executed before. Therefore, we start with running the experiment, by clicking on the **RUN** button in the menu at the bottom. After running the experiment, the top corner message will change into “Finished running”, and we can start inspecting our data.



Experiment Properties

START TIME	2/3/202...
END TIME	2/3/202...
STATUS CODE	Finished
STATUS DETAILS	None

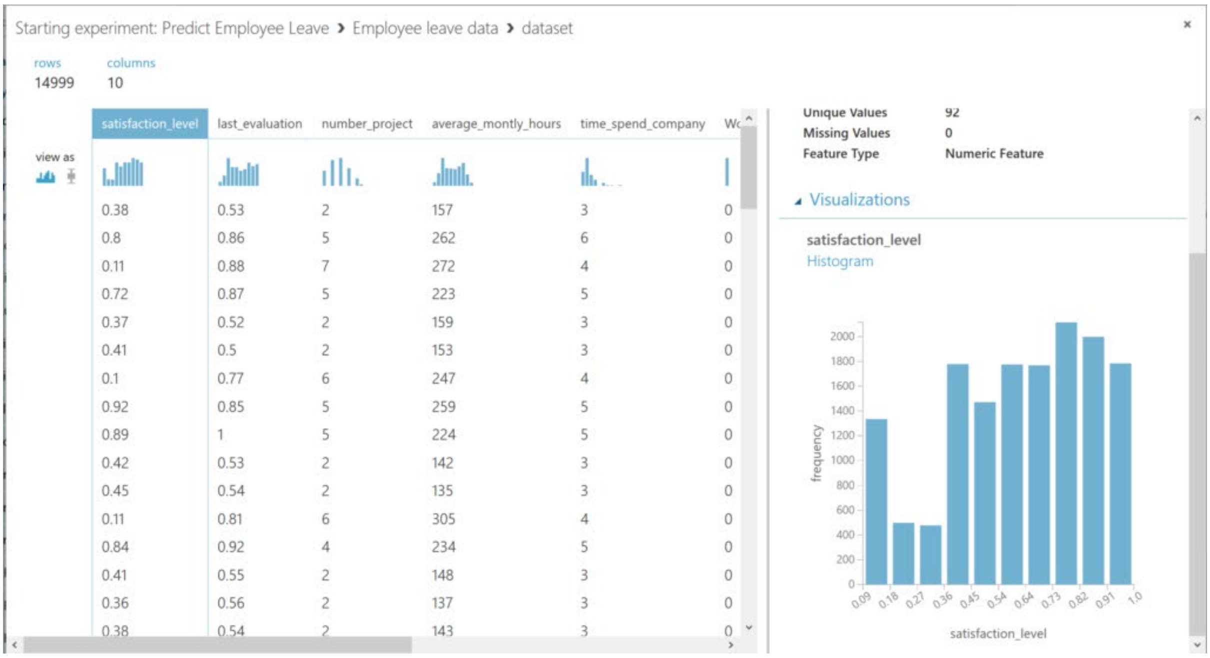
Summary

This is a starting experiment to build your own employee leave prediction model.

Description

Enter the detailed description for your experiment.

To get a first impression of the data, you can right-click the output port of the dataset and select “Visualize” from the menu to **visualize** the data. The output port is the little circle under every module on the canvas.


Starting experiment: Predict Employee Leave › Employee leave data › dataset

rows	columns
14999	10

satisfaction_level	last_evaluation	number_project	average_montly_hours	time_spend_company	Wc
0.38	0.53	2	157	3	0
0.8	0.86	5	262	6	0
0.11	0.88	7	272	4	0
0.72	0.87	5	223	5	0
0.37	0.52	2	159	3	0
0.41	0.5	2	153	3	0
0.1	0.77	6	247	4	0
0.92	0.85	5	259	5	0
0.89	1	5	224	5	0
0.42	0.53	2	142	3	0
0.45	0.54	2	135	3	0
0.11	0.81	6	305	4	0
0.84	0.92	4	234	5	0
0.41	0.55	2	148	3	0
0.36	0.56	2	137	3	0
0.38	0.54	2	143	3	0

Unique Values	92
Missing Values	0
Feature Type	Numeric Feature

You can scroll through the different columns, and by selecting them, you get an overview in the panel on the right.

We can continue inspecting the dataset. The data comprise a wide range of topics which allow to explain employees’ leave behaviour in relation with A) organizational factors (department); B) employment relational factors (i.e. tenure, the number of projects participated in; the average working hours per month; objective career development; salary); and C) job-related factors (performance evaluation; involvement in workplace accidents).

We have the following available variables in the dataset:

Organizational factors

- Department

Employment relational factors

- Time spent at the company
- Number of projects
- Average monthly hours
- Salary
- Whether they have had a promotion in the last 5 years

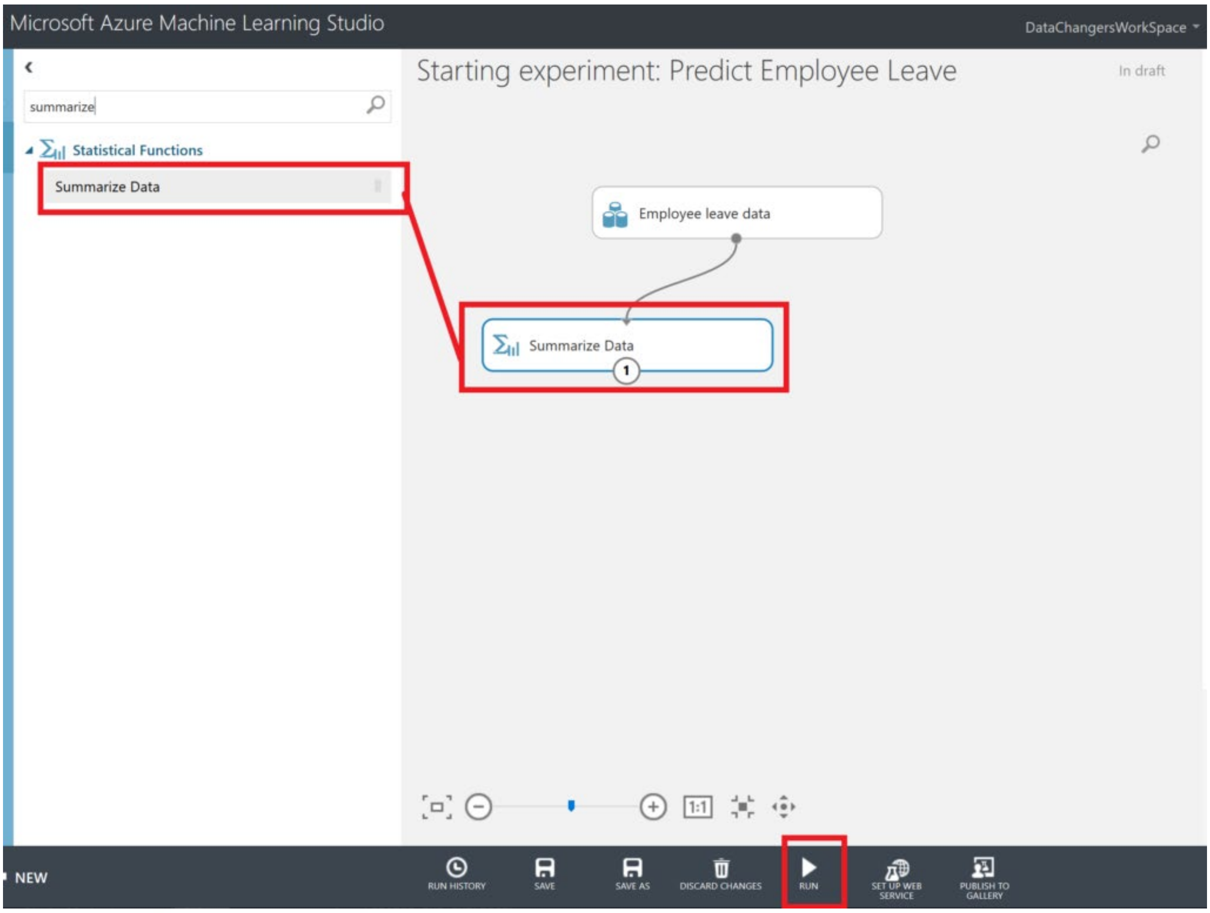
Job-related factors

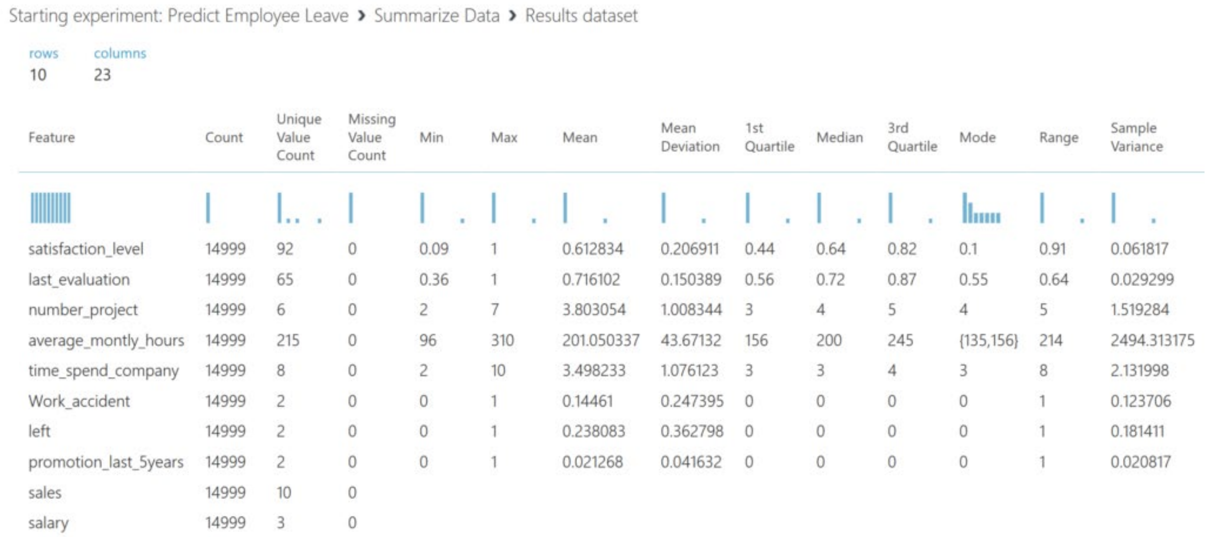
- Last evaluation
- Whether they have had a work accident

Dependent variable

- Whether the employee has left

Another way to get a first impression of the data. Therefore, we use the **Summarize Data module**, which gives us insights about the data.





Starting experiment: Predict Employee Leave › Summarize Data › Results dataset

rows 10 columns 23

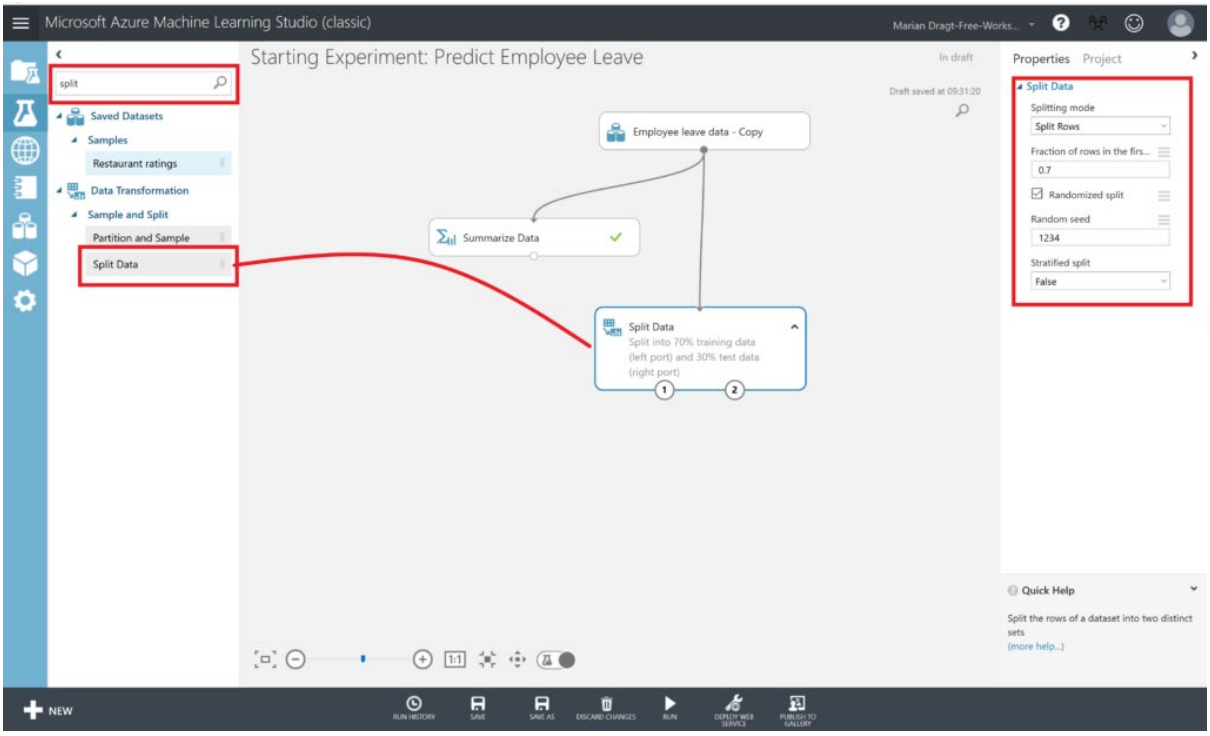
Feature	Count	Unique Value Count	Missing Value Count	Min	Max	Mean	Mean Deviation	1st Quartile	Median	3rd Quartile	Mode	Range	Sample Variance
satisfaction_level	14999	92	0	0.09	1	0.612834	0.206911	0.44	0.64	0.82	0.1	0.91	0.061817
last_evaluation	14999	65	0	0.36	1	0.716102	0.150389	0.56	0.72	0.87	0.55	0.64	0.029299
number_project	14999	6	0	2	7	3.803054	1.008344	3	4	5	4	5	1.519284
average_montly_hours	14999	215	0	96	310	201.050337	43.67132	156	200	245	{135,156}	214	2494.313175
time_spend_company	14999	8	0	2	10	3.498233	1.076123	3	3	4	3	8	2.131998
Work_accident	14999	2	0	0	1	0.14461	0.247395	0	0	0	0	1	0.123706
left	14999	2	0	0	1	0.238083	0.362798	0	0	0	0	1	0.181411
promotion_last_5years	14999	2	0	0	1	0.021268	0.041632	0	0	0	0	1	0.020817
sales	14999	10	0										
salary	14999	3	0										

After you ran the model, you can **right-click** on the output port of the **Summarize Data** module and select **Visualize**. We see that we have 14999 observations, and that we don't miss any data. We also get an idea about the variance and distribution of the data.

Step 3: Prepare a training and a test set

We split the dataset into a training and a test set, using 70% of the data to train the model with, and 30% of the data to test the model later on. Therefore, we drag the **Split Data**

module on the canvas. You can find this module in the menu left, next to the canvas. You can either click through the various options or use the search function.



Split Data

Splitting mode: Split Rows

Fraction of rows in the firs...: 0.7

☑ Randomized split

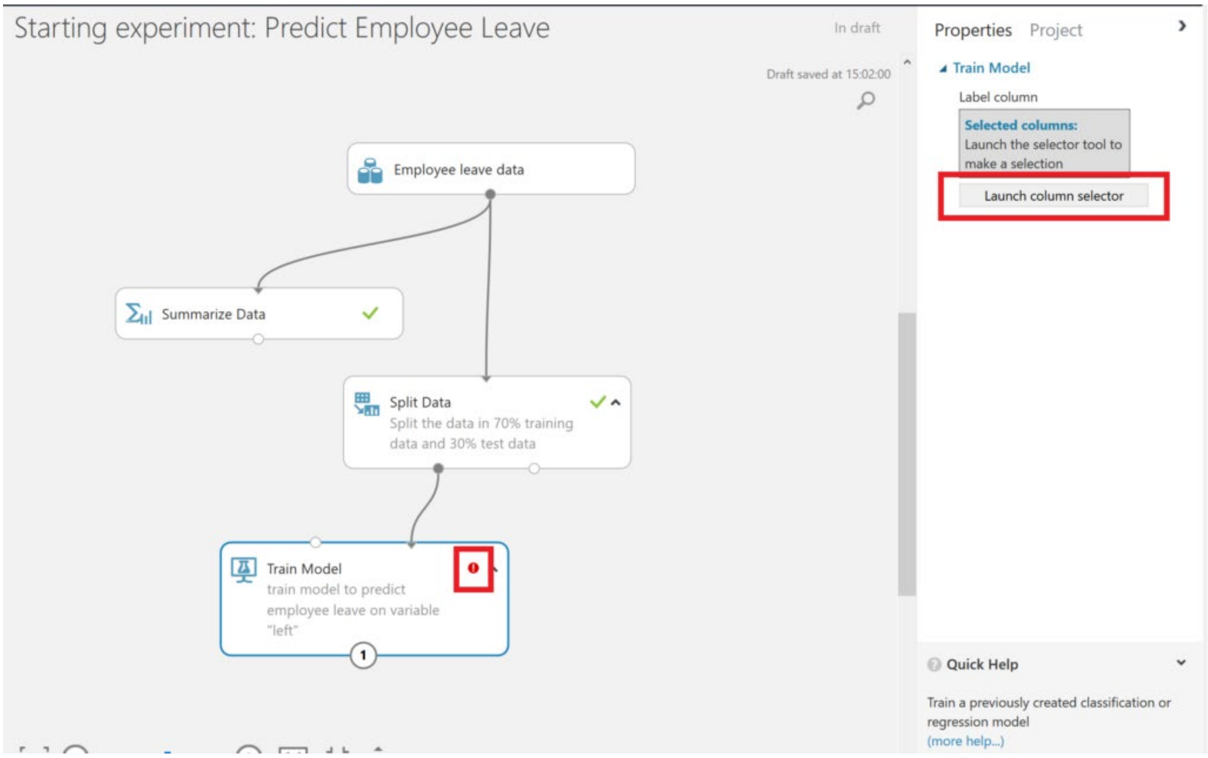
Random seed: 1234

Stratified split: False

When you have found the Split Data module, you can drag it on the canvas, and connect the output port of the dataset to the input port of the Split Data module. You can connect the modules by left clicking on the output port, and keep you mouse button down while draging it to the module you want to connect it to. We set a seed, so we can repeat this experiment. Make sure you **RUN the model** after every step.

Step 4: Train the model

Since we have split the data, we can continue to work with the training data set. We first select the **Train model** module and drag it on the canvas. But when we do so you will a little red exclamation mark. This is because we haven’t selected the variable that we want to predict, and we haven’t defined the algorithm that we want to use to train the model with. First, we will select the dependent variable. Therefore, we have to click on the **Launch column selector**.


Starting experiment: Predict Employee Leave

In draft

Draft saved at 15:02:00

Employee leave data

Summarize Data

Split Data
Split the data in 70% training data and 30% test data

Train Model
train model to predict employee leave on variable "left"

Properties Project

Train Model

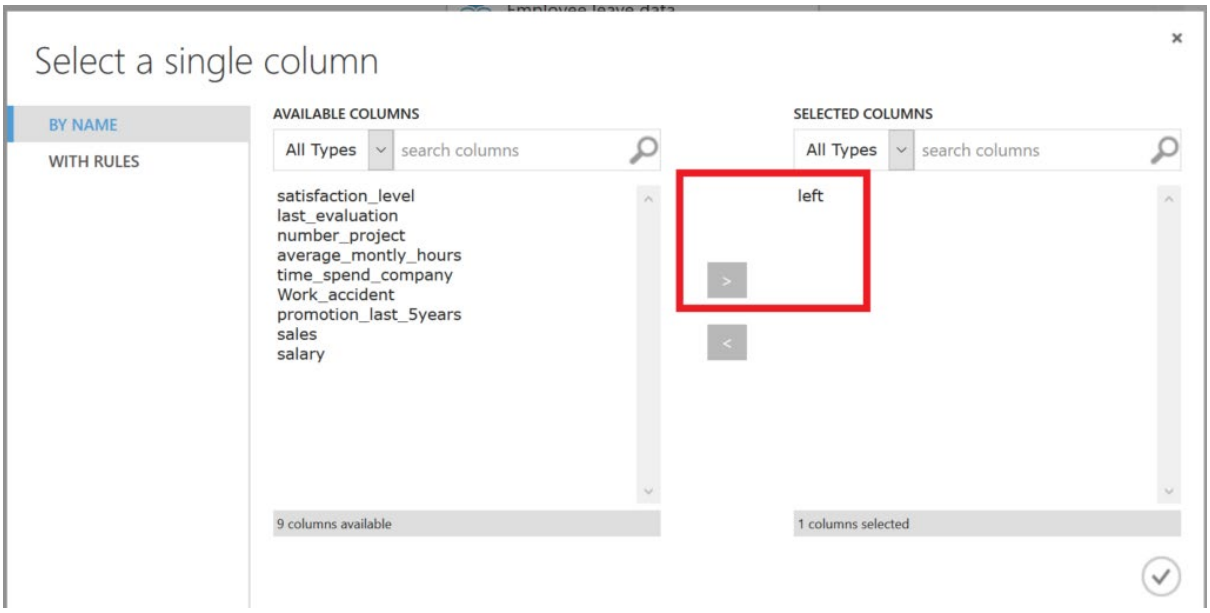
Label column

Selected columns:
Launch the selector tool to make a selection

Quick Help

Train a previously created classification or regression model
(more help...)

In order to set the dependent variable, we select the variable “left” (indicating whether an employee has left or not) from AVAILABLE COLUMNS and use the arrow button to get it to the right side, under “SELECTED COLUMNS”.


Select a single column

AVAILABLE COLUMNS

satisfaction_level
last_evaluation
number_project
average_montly_hours
time_spend_company
Work_accident
promotion_last_5years
sales
salary

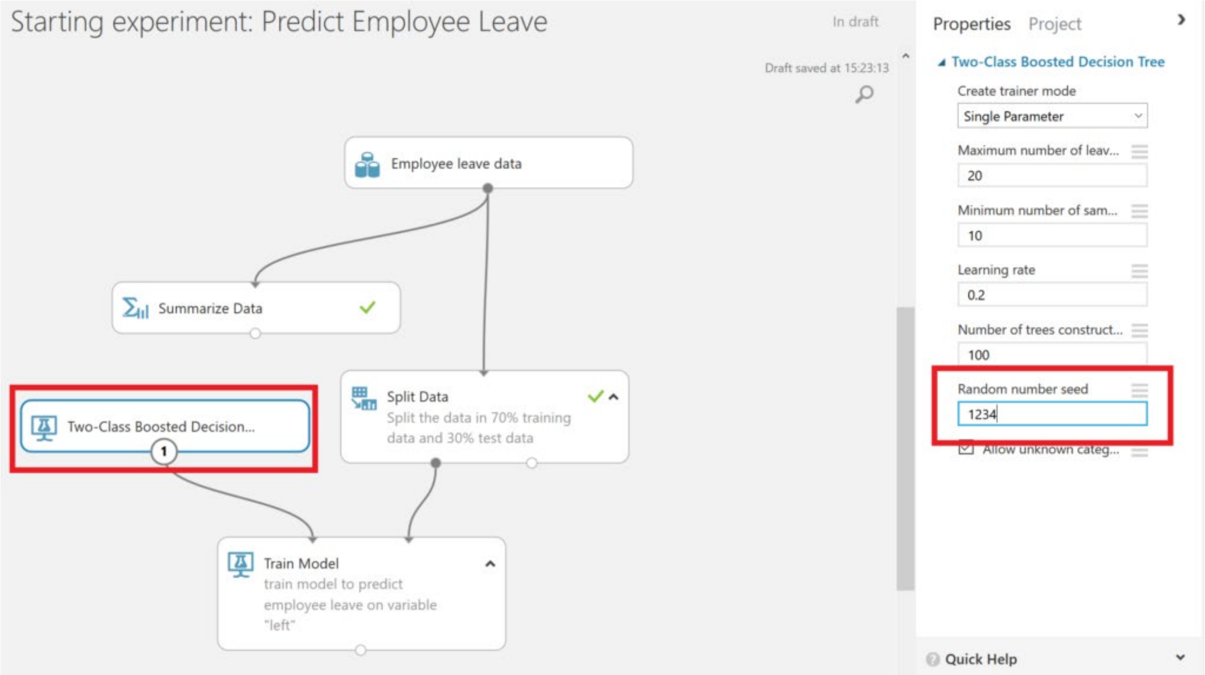
9 columns available

SELECTED COLUMNS

left

1 columns selected

Furthermore, we have to select the algorithm to train the model with. In this experiment we use the **Two-Class Boosted Decision Tree algorithm** with the standard parametrization. We do add a seed to make this experiment replicable.


Starting experiment: Predict Employee Leave
In draft
Draft saved at 15:23:13
Employee leave data
Summarize Data
Two-Class Boosted Decision...
1
Split Data
Split the data in 70% training data and 30% test data
Train Model
train model to predict employee leave on variable "left"
Properties
Project
Quick Help

Two-Class Boosted Decision Tree

Create trainer mode: Single Parameter

Maximum number of leav...: 20

Minimum number of sam...: 10

Learning rate: 0.2

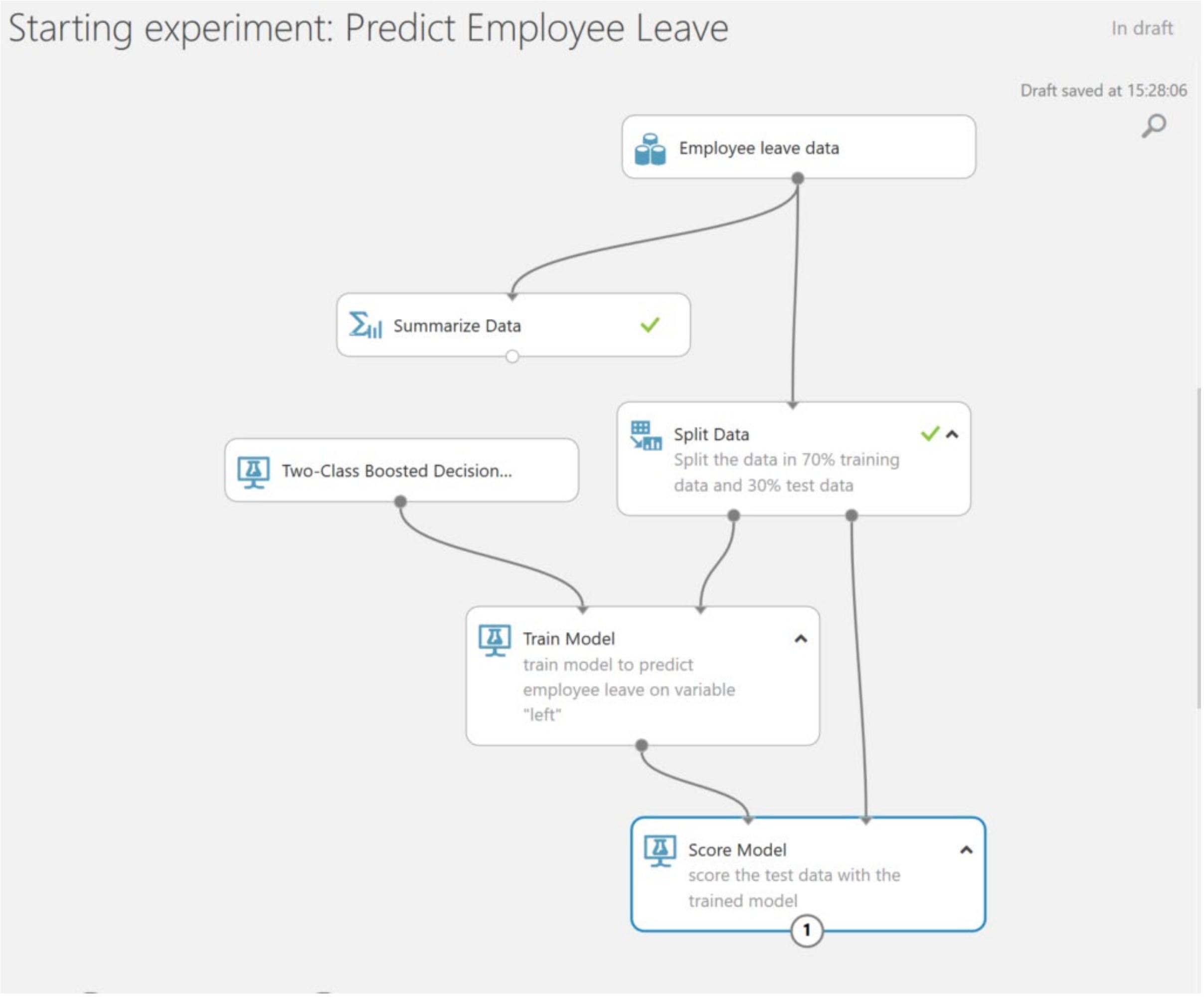
Number of trees construct...: 100

Random number seed: 1234

☑ Allow unknown categ...

Step 5: Score the test set

After this, we are prepared to score the test set and see how our model performs. Therefore, we use the **Score Model module** and we connect both the output port of the Train model module, which contains the trained model, as the outcome of the Split Data set, containing the test data.



Starting experiment: Predict Employee Leave

In draft

Draft saved at 15:28:06

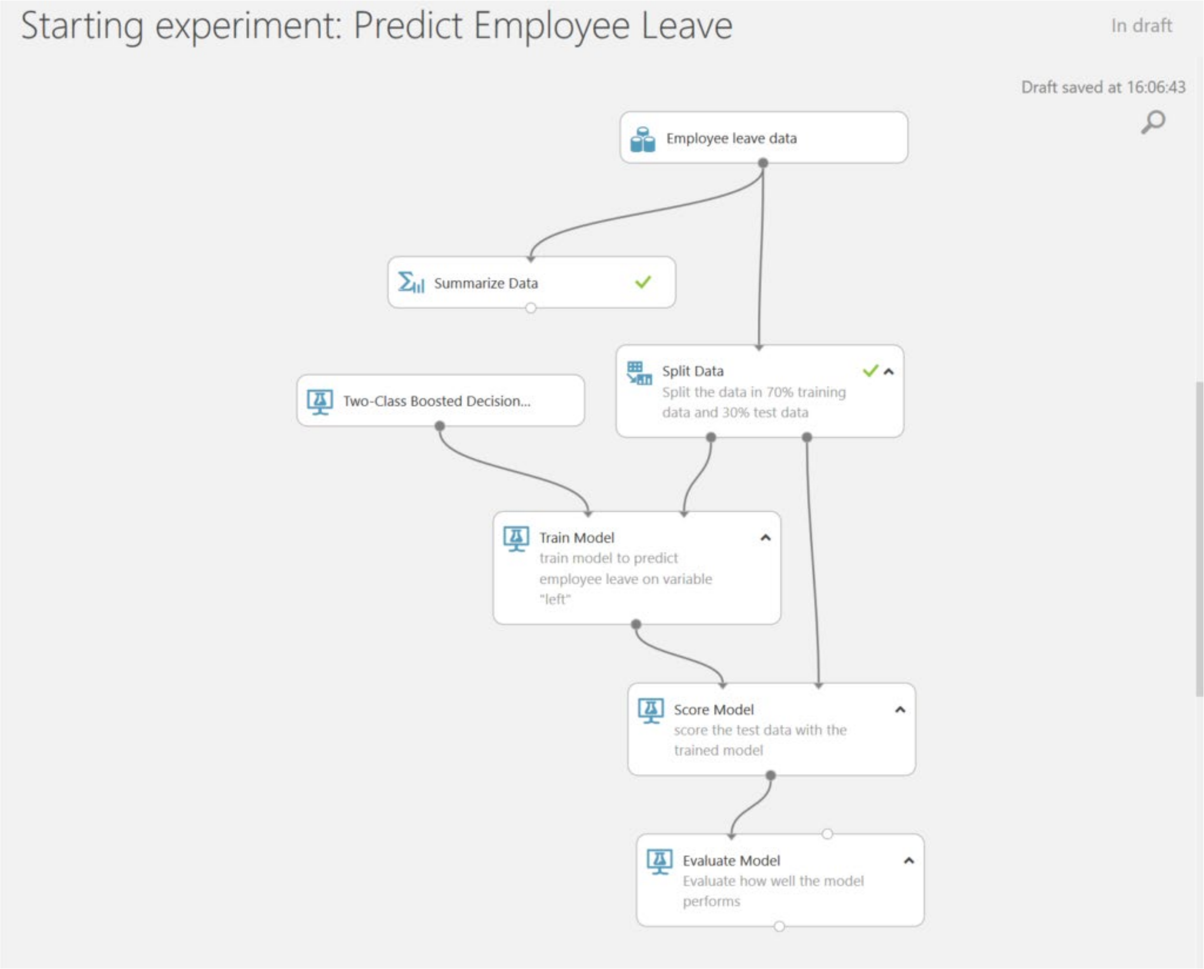
Split the data in 70% training data and 30% test data

train model to predict employee leave on variable "left"

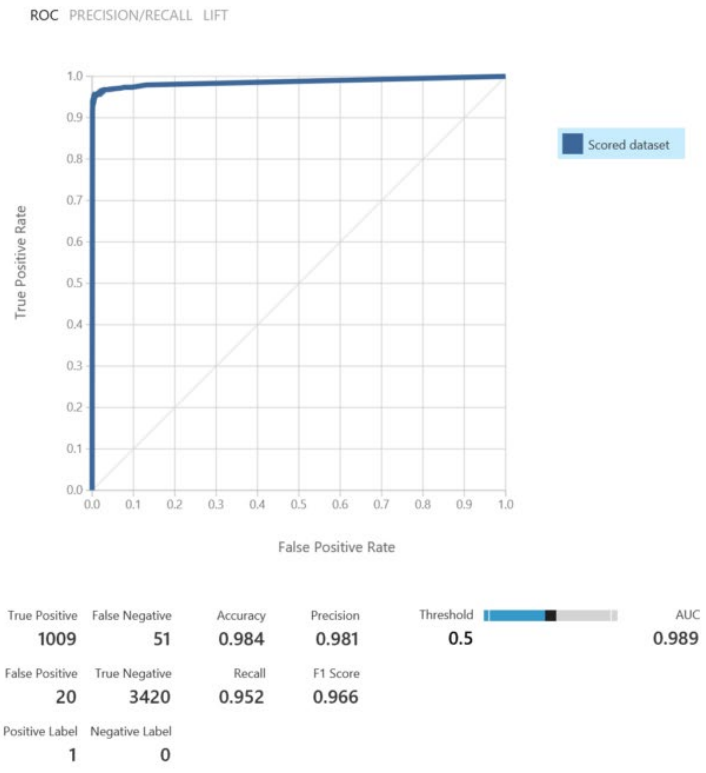
score the test data with the trained model

Step 6: Evaluate the results

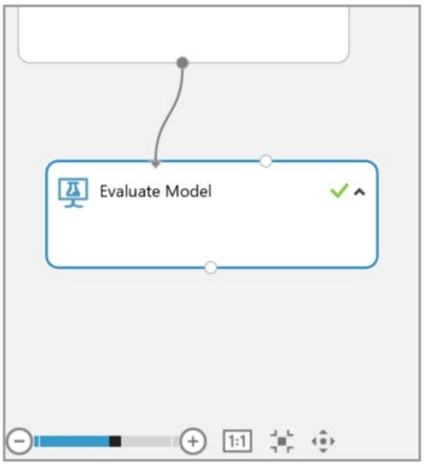
Finally, it's time to evaluate the results of our model. We use the **Evaluate Model module** which we connect to the results of our prior scoring.


Starting experiment: Predict Employee Leave
In draft
Draft saved at 16:06:43
Employee leave data
Summarize Data ✓
Two-Class Boosted Decision...
Split Data ✓
Split the data in 70% training data and 30% test data
Train Model
train model to predict employee leave on variable "left"
Score Model
score the test data with the trained model
Evaluate Model
Evaluate how well the model performs

Let's run the model, and then right click on the Evaluate Model module to visualize the results. We can predict with 98% accuracy and 98% precision.

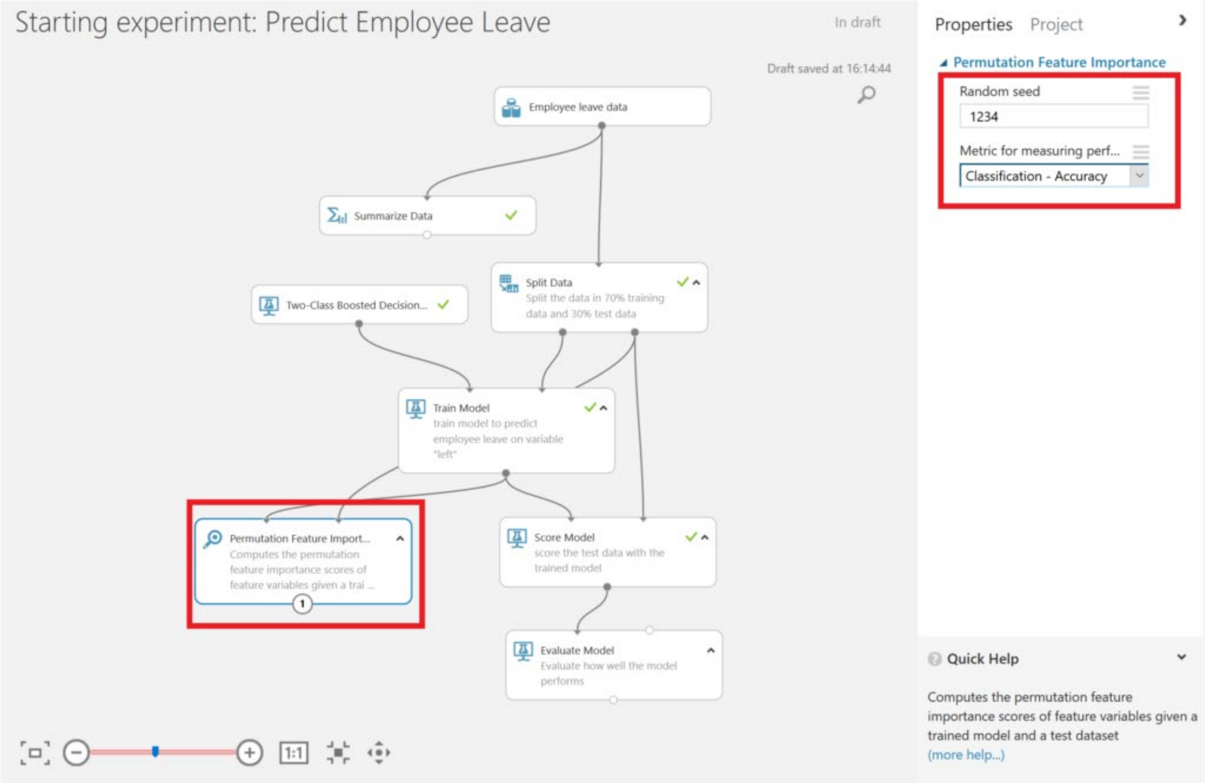


True Positive	False Negative	Accuracy	Precision	Threshold	AUC
1009	51	0.984	0.981	0.5	0.989
False Positive	True Negative	Recall	F1 Score		
20	3420	0.952	0.966		
Positive Label	Negative Label				
1	0				



Step 7: Gain insights on the why

Our final question was why employees were leaving. Therefore, we could add the **Permutation Feature Importance module**. We connect the output port of the Train Model module and the output port of the Split Data module. Now we can compute the permutation feature importance scores of feature variables given this trained model and the test dataset. We set a seed to make the experiment replicable, and we focus on accuracy, meaning that we are both interested in selected correctly the people that leave, and the people that will not leave.



Starting experiment: Predict Employee Leave

Properties Project

Permutation Feature Importance

Random seed: 1234

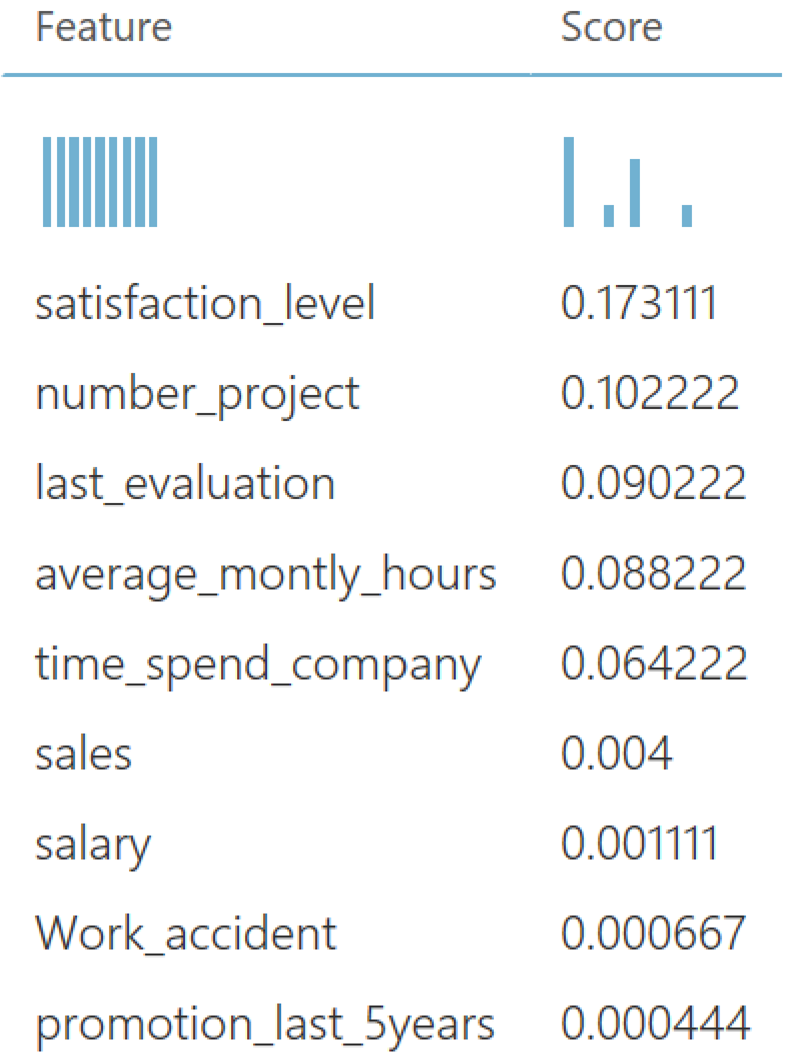
Metric for measuring perf...: Classification - Accuracy

Quick Help

Computes the permutation feature importance scores of feature variables given a trained model and a test dataset (more help...)

In draft
Draft saved at 16:14:44
Employee leave data
Summarize Data
Two-Class Boosted Decision...
Split Data
Split the data in 70% training data and 30% test data
Train Model
train model to predict employee leave on variable "left"
Permutation Feature Import...
Computes the permutation feature importance scores of feature variables given a trai ...
Score Model
score the test data with the trained model
Evaluate Model
Evaluate how well the model performs
1:1

If we run the model, and right-click on the output port of the **Permutation Feature Importance** module, we find that satisfaction was one of the main factors when leaving, according to this dataset. Next to that, the number of projects an employee got was important.

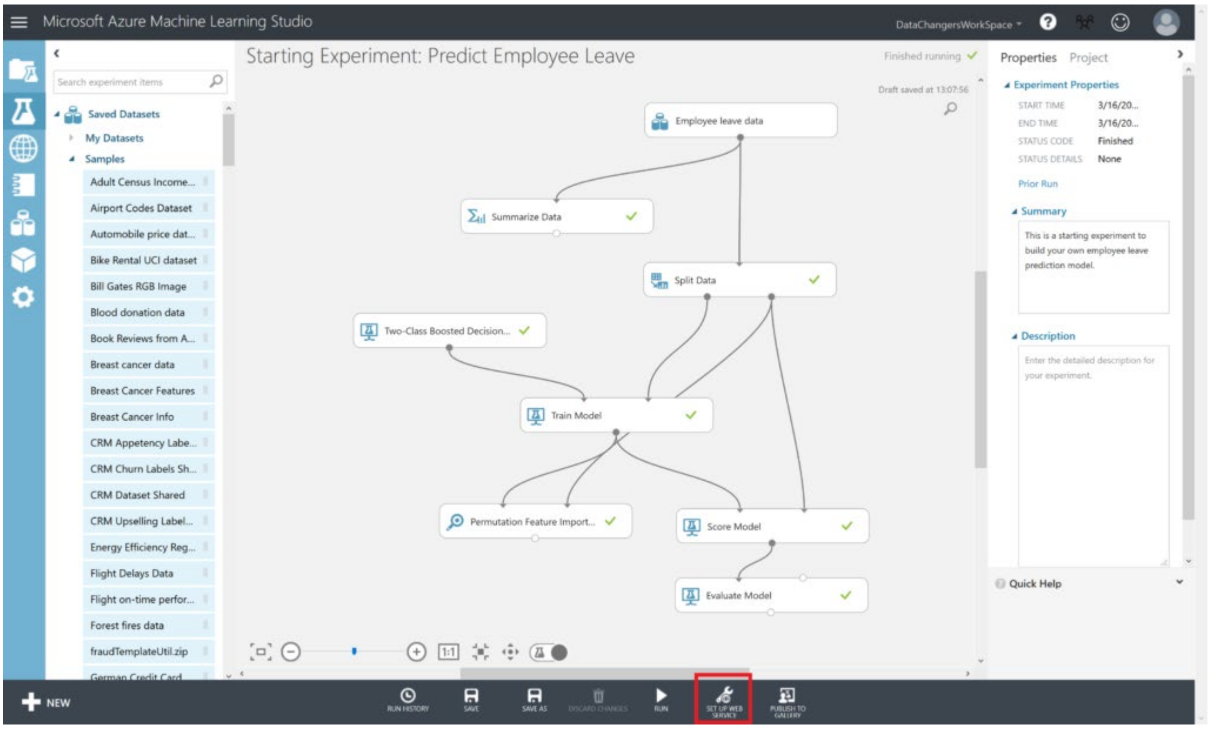


Feature	Score
satisfaction_level	0.173111
number_project	0.102222
last_evaluation	0.090222
average_montly_hours	0.088222
time_spend_company	0.064222
sales	0.004
salary	0.001111
Work_accident	0.000667
promotion_last_5years	0.000444

Step 8: Publish your Model

Publish the Model as a Web Service

Make sure you have saved and ran the experiment. With the **Starting experiment: Predict Employee Leave** experiment open, click the **SET UP WEB SERVICE** icon at the bottom of the Azure ML Studio page and click **Predictive Web Service [Recommended]**.

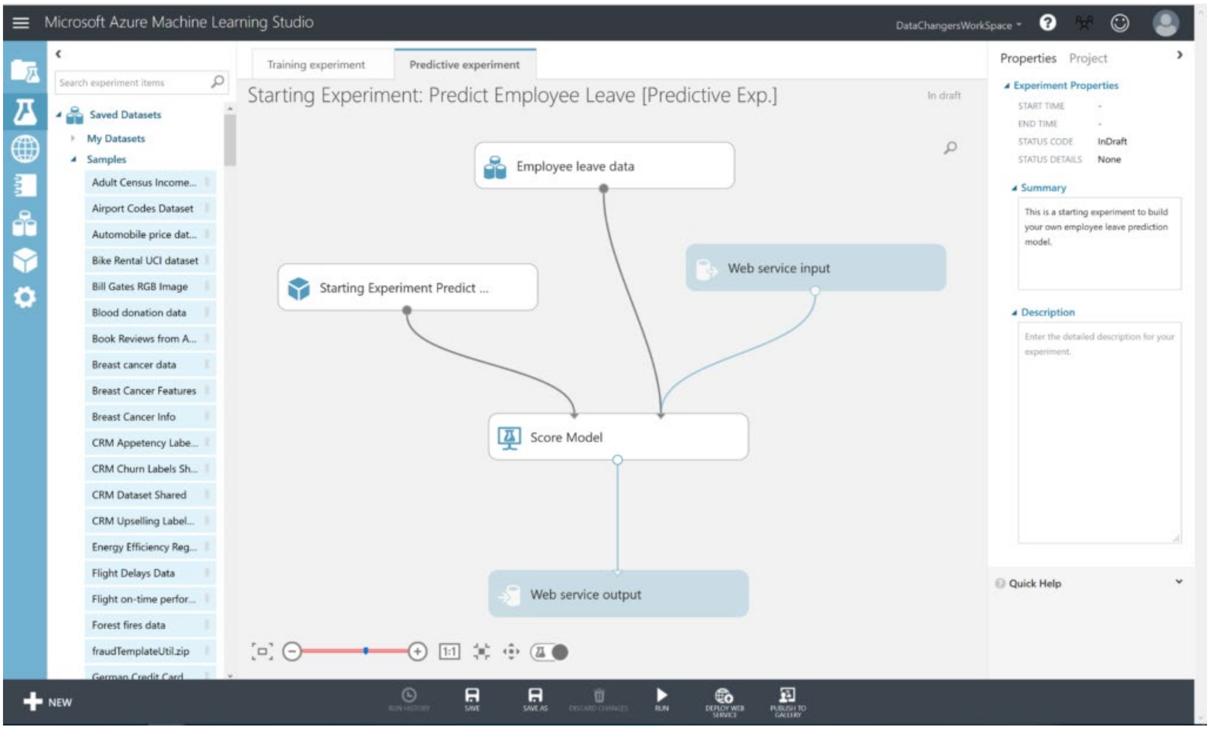


START TIME 3/16/20...
END TIME 3/16/20...
STATUS CODE Finished
STATUS DETAILS None

This is a starting experiment to build your own employee leave prediction model.

Enter the detailed description for your experiment.

A new **Predictive Experiment** tab will be automatically created. Verify that, with a bit of rearranging, the Predictive Experiment resembles this figure:



Experiment Properties
START TIME -
END TIME -
STATUS CODE InDraft
STATUS DETAILS None

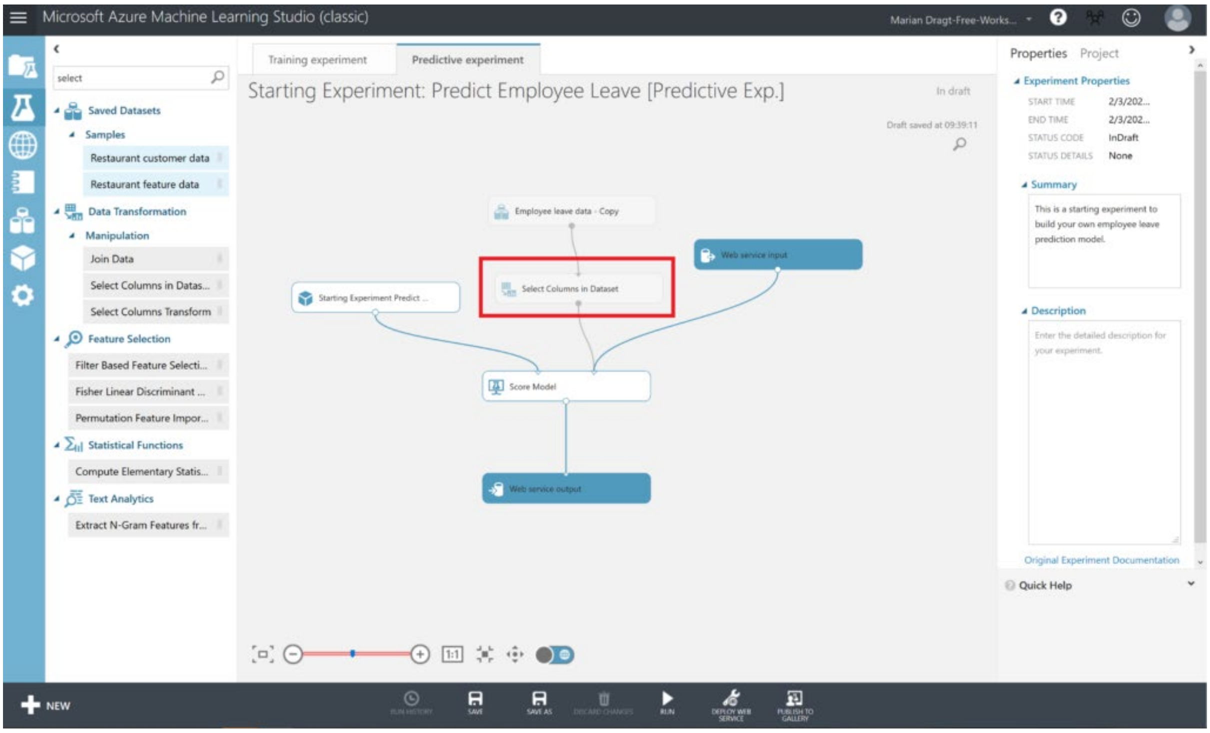
Summary

This is a starting experiment to build your own employee leave prediction model.

Description

Enter the detailed description for your experiment.

An important step to carry out now, is eliminating the dependent variable “left”. Because this is exactly what you want to predict and you can’t leave it as input variable, although it will be ignored by the model itself. In order to do so, select **Select Columns in Dataset** and put it between the dataset and the **Score Model** module.


Microsoft Azure Machine Learning Studio (classic)
Marian Dragt-Free-Works...
Training experiment
Predictive experiment
Starting Experiment: Predict Employee Leave [Predictive Exp.]
In draft
Draft saved at 09:39:11
Employee leave data - Copy
Web service input
Starting Experiment Predict ...
Select Columns in Dataset
Score Model
Web service output
select
Saved Datasets
Samples
Restaurant customer data
Restaurant feature data
Data Transformation
Manipulation
Join Data
Select Columns in Datas...
Select Columns Transform
Feature Selection
Filter Based Feature Selecti...
Fisher Linear Discriminant ...
Permutation Feature Impor...
Statistical Functions
Compute Elementary Statis...
Text Analytics
Extract N-Gram Features fr...
Properties
Project
Original Experiment Documentation
Quick Help
NEW
RUN HISTORY
SAVE
SAVE AS
DISCARD CHANGES
RUN
DEPLOY WEB SERVICE
PUBLISH TO GALLERY

Experiment Properties
START TIME 2/3/202...
END TIME 2/3/202...
STATUS CODE InDraft
STATUS DETAILS None

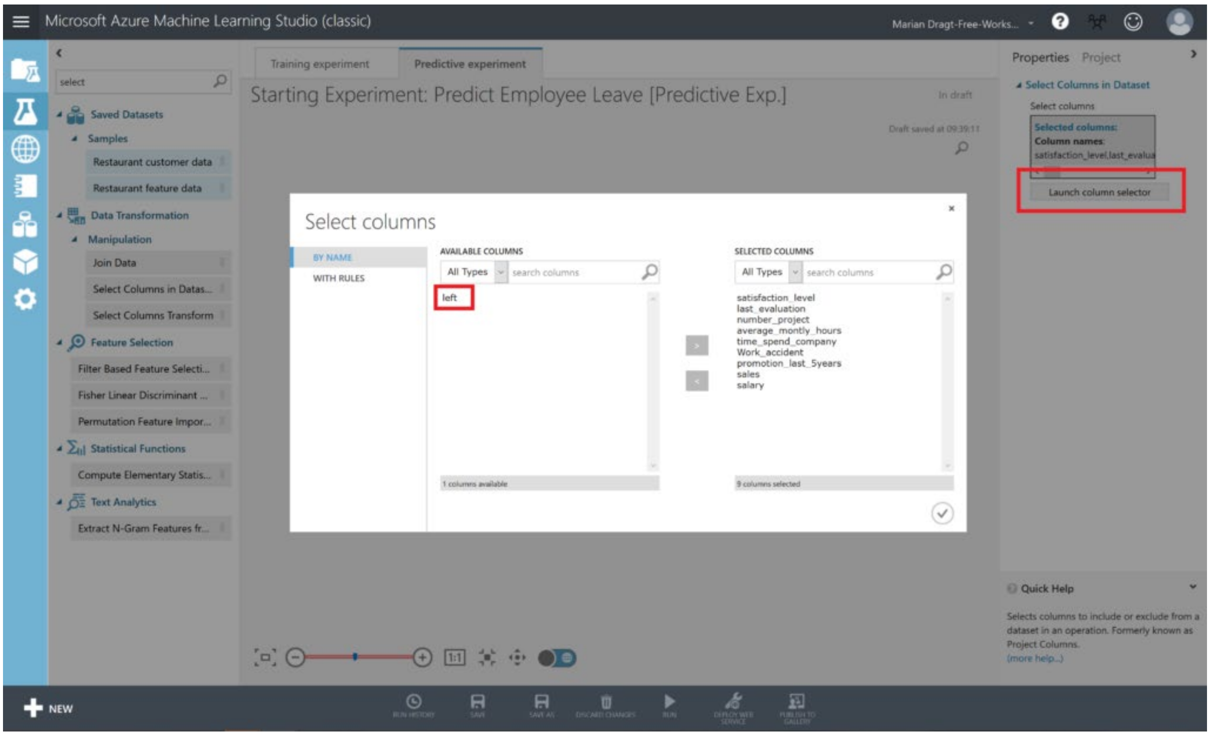
Summary

This is a starting experiment to build your own employee leave prediction model.

Description

Enter the detailed description for your experiment.

Select the **Select Columns in Dataset** module, open the **Launch column selector**, and remove the variable “left”. Now RUN the predictive experiment.



Microsoft Azure Machine Learning Studio (classic)
Marian Dragt-Free-Works...
Training experiment
Predictive experiment
Starting Experiment: Predict Employee Leave [Predictive Exp.]
In draft
Draft saved at 09:39:11
Properties
Project
Select Columns in Dataset
Select columns
Launch column selector
Quick Help
NEW

Selected columns:
Column names: satisfaction_level,last_evalua

Select columns

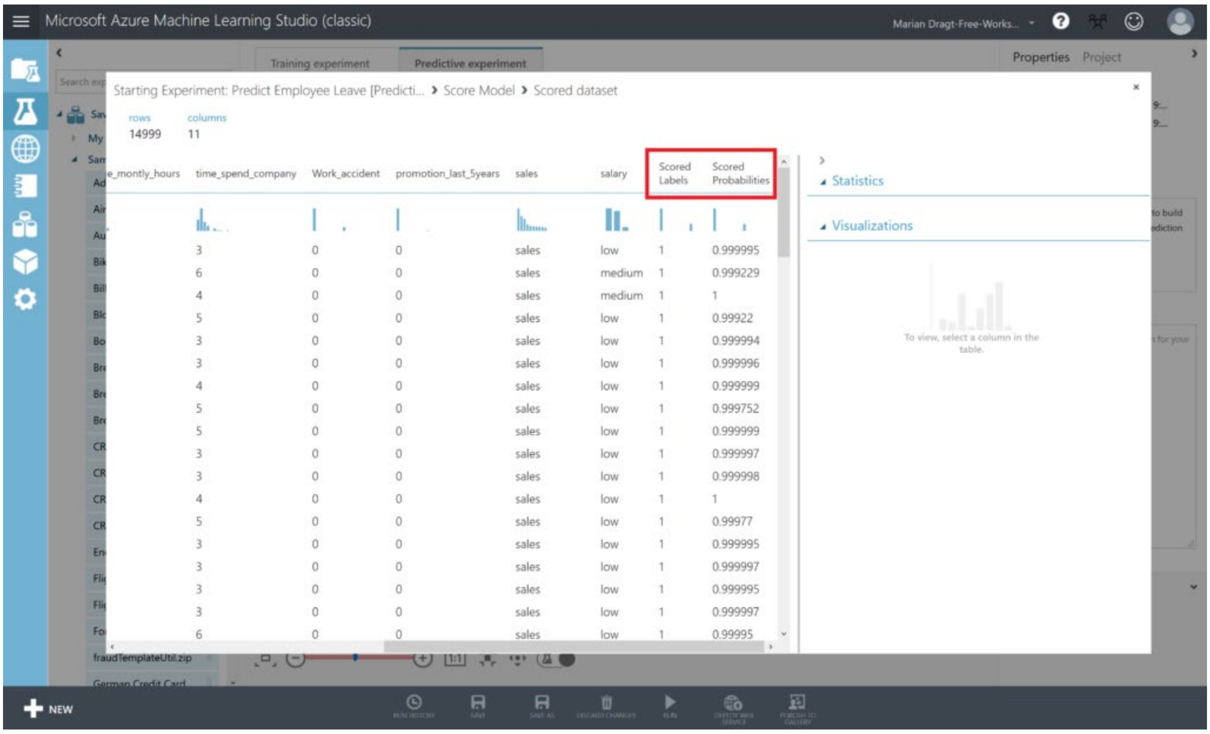
BY NAME
WITH RULES

AVAILABLE COLUMNS
All Types | search columns
left
1 columns available

SELECTED COLUMNS
All Types | search columns
satisfaction_level
last_evaluation
number_project
average_montly_hours
time_spend_company
Work_accident
promotion_last_5years
sales
salary
9 columns selected

Selects columns to include or exclude from a dataset in an operation. Formerly known as Project Columns.
(more help...)

When you inspect the output of the Score Model module, by right-clicking on the output port and selecting **Visualize**, you will see that there are 2 extra columns in your dataset, named **Scored Labels** and **Scored Probabilities**. Scored Labels contains the prediction whether an employee will leave or not, and is based on the Scored Probabilities value, where 0.5 is the cut-off: from 0.5 an employee will leave.



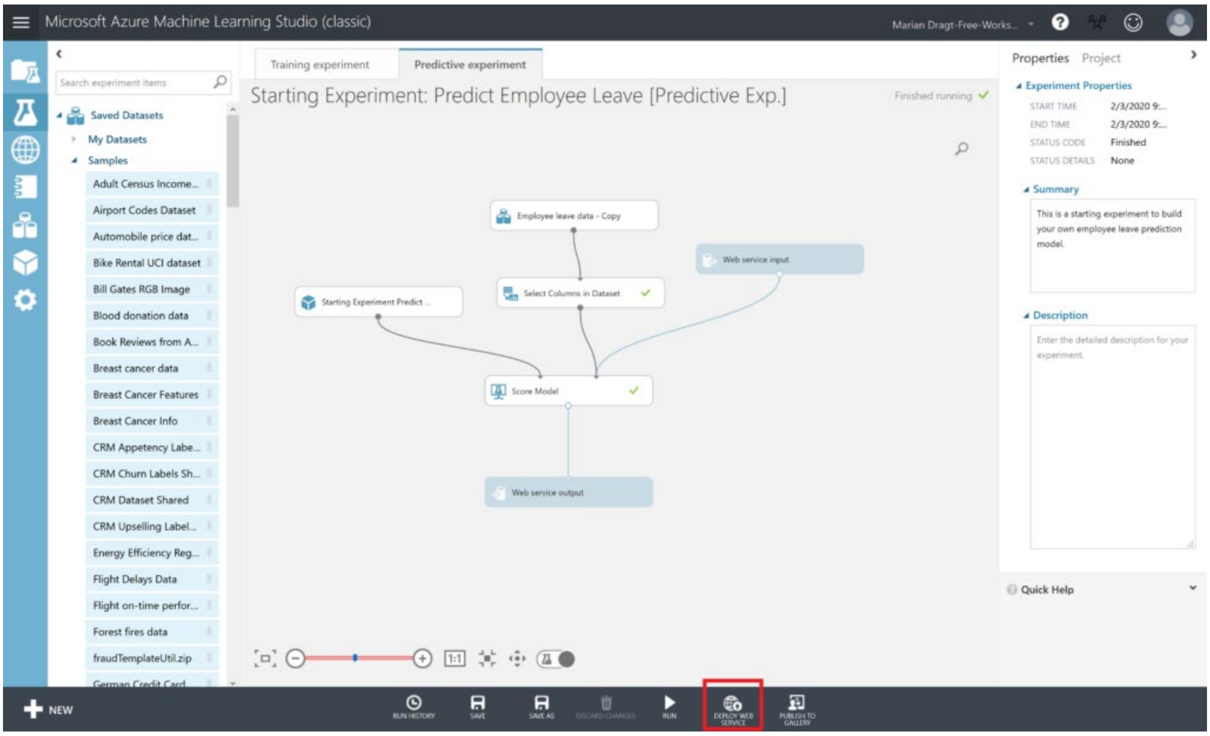
Starting Experiment: Predict Employee Leave [Predicti... › Score Model › Scored dataset

rows 14999
columns 11

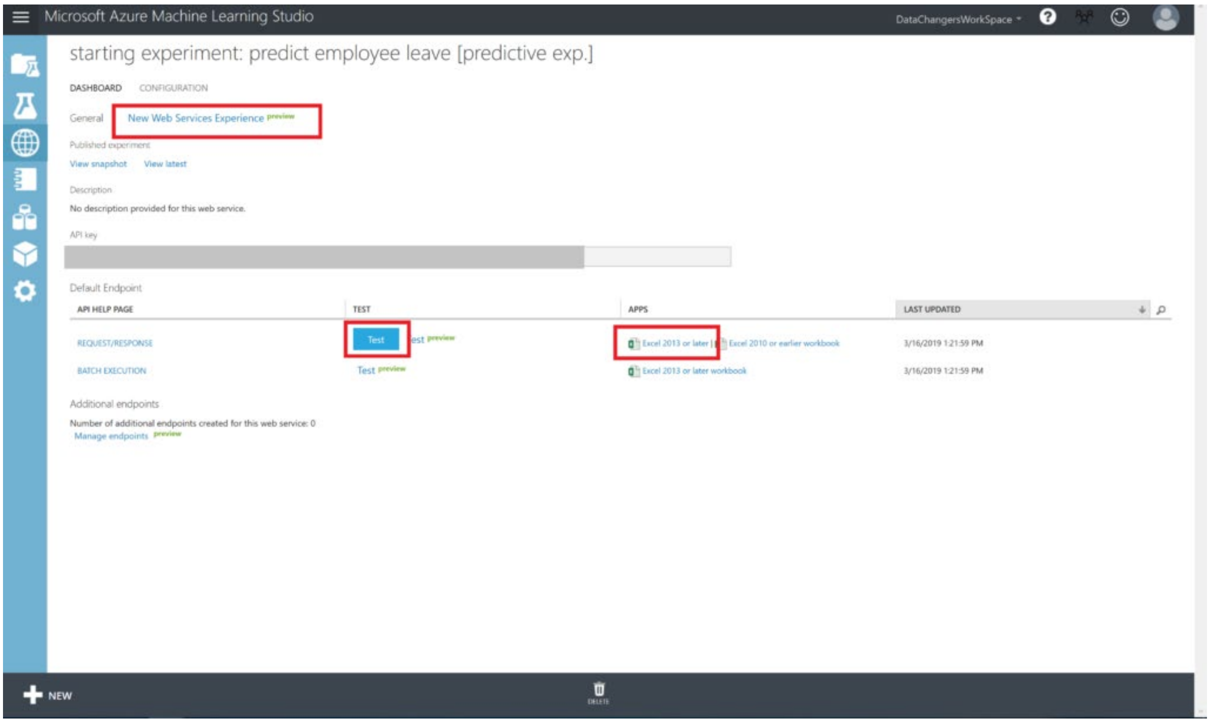
e_montly_hours	time_spend_company	Work_accident	promotion_last_5years	sales	salary	Scored Labels	Scored Probabilities
	3	0	0	sales	low	1	0.999995
	6	0	0	sales	medium	1	0.999229
	4	0	0	sales	medium	1	1
	5	0	0	sales	low	1	0.99922
	3	0	0	sales	low	1	0.999994
	3	0	0	sales	low	1	0.999996
	4	0	0	sales	low	1	0.999999
	5	0	0	sales	low	1	0.999752
	5	0	0	sales	low	1	0.999999
	3	0	0	sales	low	1	0.999997
	3	0	0	sales	low	1	0.999998
	4	0	0	sales	low	1	1
	5	0	0	sales	low	1	0.99977
	3	0	0	sales	low	1	0.999995
	3	0	0	sales	low	1	0.999997
	3	0	0	sales	low	1	0.999995
	3	0	0	sales	low	1	0.999997
	6	0	0	sales	low	1	0.99995

9: Deploy and Use the Web Service

In the **Starting Experiment: Predict Employee Leave [Predictive Exp.]** experiment, click the **Deploy Web Service** button at the bottom of the window.



Wait a few seconds for the dashboard page to appear. You have several options to connect to the webservice.



starting experiment: predict employee leave [predictive exp.]

DASHBOARD CONFIGURATION

General New Web Services Experience preview

Published experiment

View snapshot View latest

Description

No description provided for this web service.

API key

Default Endpoint

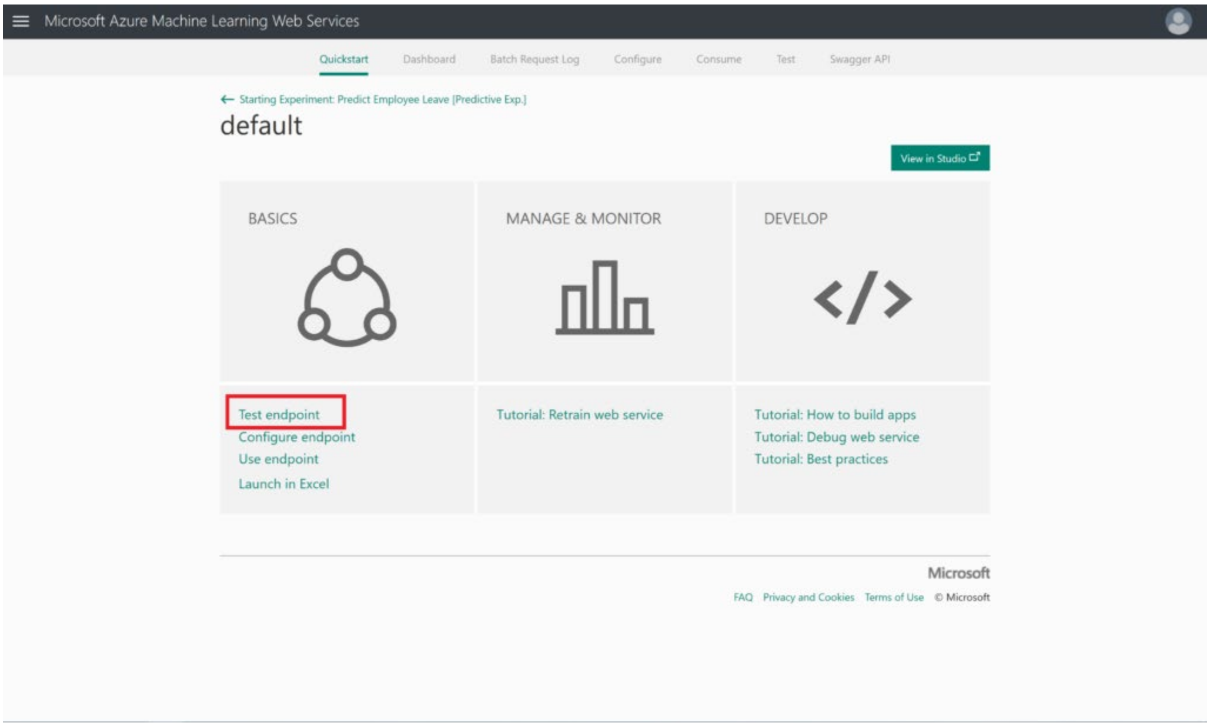
API HELP PAGE	TEST	APPS	LAST UPDATED
REQUEST/RESPONSE	Test Test preview	Excel 2013 or later | Excel 2010 or earlier workbook	3/16/2019 1:21:59 PM
BATCH EXECUTION	Test preview	Excel 2013 or later workbook	3/16/2019 1:21:59 PM

Additional endpoints

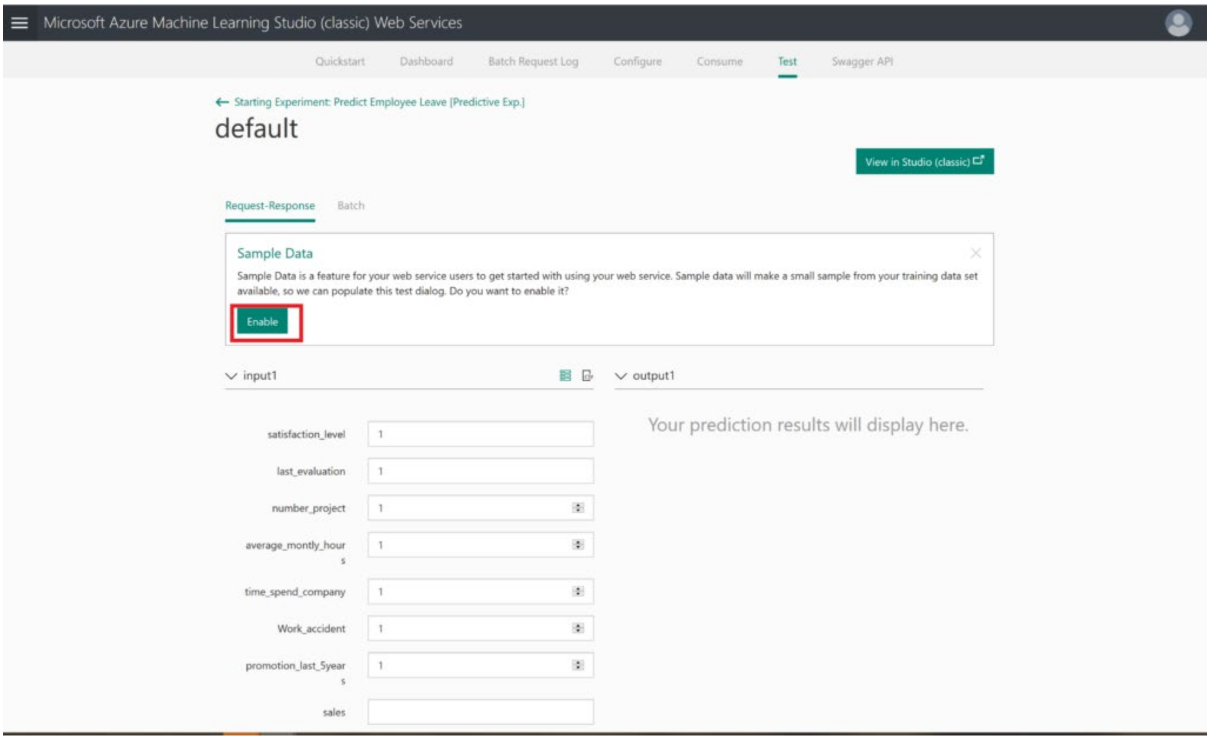
Number of additional endpoints created for this web service: 0
Manage endpoints preview

Step 10: Test your model

To test this webservice, you can i.e. click on **New Web Services Experience (preview)**. This will open a new browser where you have the option to test your model (**Test endpoint** option under BASICS)



When clicking on **Test endpoint**, you have the option to enable the usage of sample data, which will generate a sample record to test your model with.



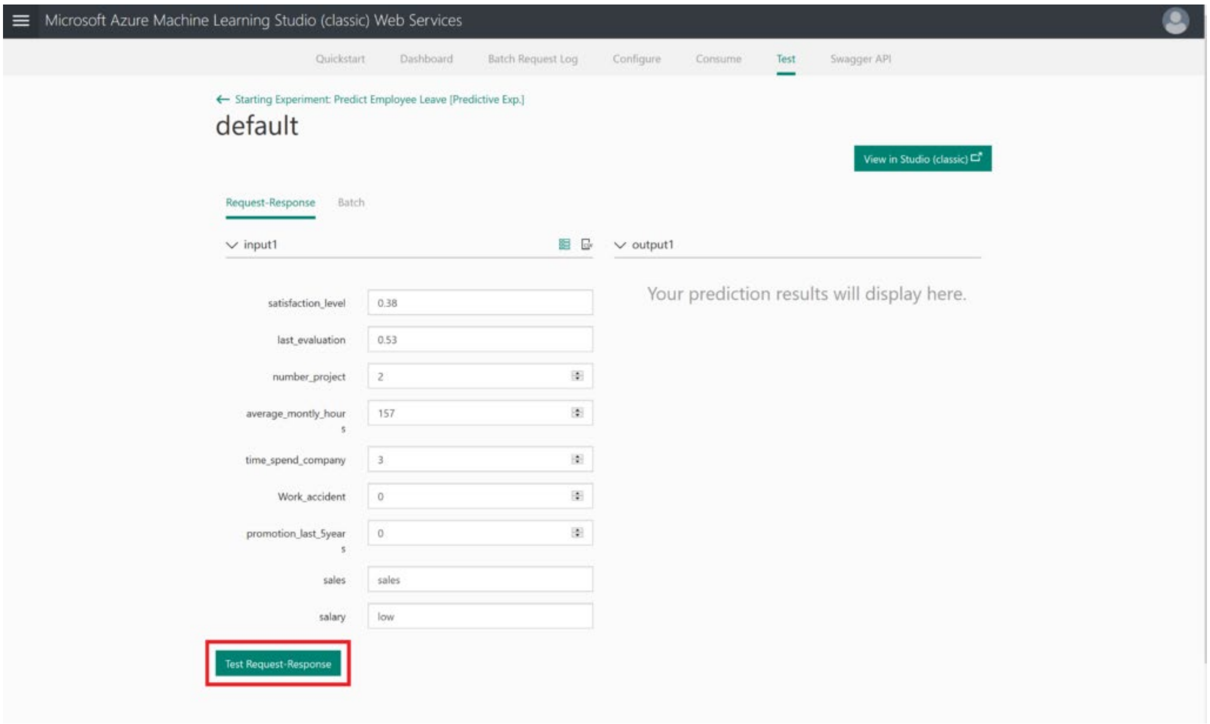
Sample Data

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?

satisfaction_level	1
last_evaluation	1
number_project	1
average_montly_hours	1
time_spend_company	1
Work_accident	1
promotion_last_5years	1
sales	

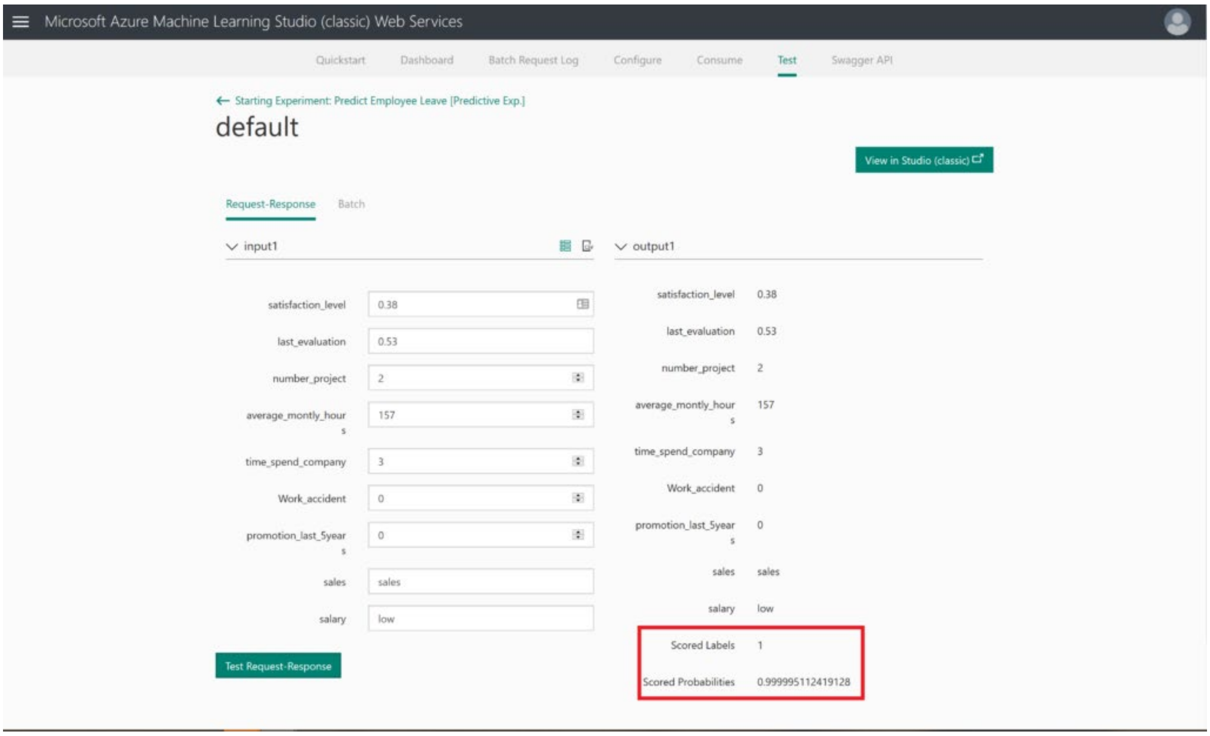
Your prediction results will display here.

After enabling this sample data, you will see the generated sample data.


Microsoft Azure Machine Learning Studio (classic) Web Services
Quickstart
Dashboard
Batch Request Log
Configure
Consume
Test
Swagger API
← Starting Experiment: Predict Employee Leave [Predictive Exp.]
default
View in Studio (classic)
Request-Response
Batch
input1
output1
Your prediction results will display here.
Test Request-Response

input1	
satisfaction_level	0.38
last_evaluation	0.53
number_project	2
average_montly_hours	157
time_spend_company	3
Work_accident	0
promotion_last_5years	0
sales	sales
salary	low

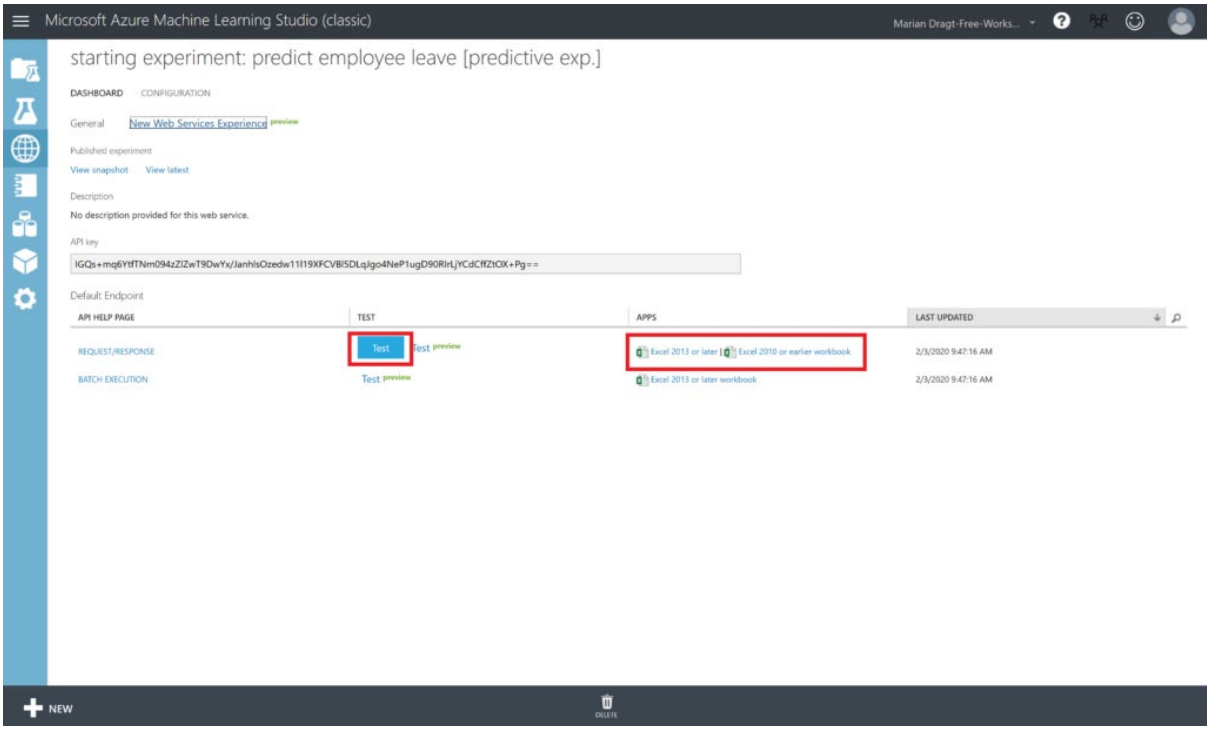
The final step would be pressing the **Test Request-Response** button: will this person leave the company?



input1	
satisfaction_level	0.38
last_evaluation	0.53
number_project	2
average_montly_hours	157
time_spend_company	3
Work_accident	0
promotion_last_5years	0
sales	sales
salary	low

output1	
satisfaction_level	0.38
last_evaluation	0.53
number_project	2
average_montly_hours	157
time_spend_company	3
Work_accident	0
promotion_last_5years	0
sales	sales
salary	low
Scored Labels	1
Scored Probabilities	0.999995112419128

If the webservice does not work, you can also use the option to click on the blue TEST button, or to launch an Excel file. It is up to you to explore these options now.



Microsoft Azure Machine Learning Studio (classic)
Marian Dragt-Free-Works...
NEW
DELETE

starting experiment: predict employee leave [predictive exp.]

DASHBOARD CONFIGURATION

General New Web Services Experience preview

Published experiment

View snapshot View latest

Description

No description provided for this web service.

API key

IGQs+mq6YtfTNm094zZlZwT9DwYx/JanhlsOzedw11I19XFCVBI5DLqJgo4NeP1ugD90RIrLjYCdCffZtOX+Pg==

Default Endpoint

API HELP PAGE	TEST	APPS	LAST UPDATED
REQUEST/RESPONSE	Test Test preview	Excel 2013 or later | Excel 2010 or earlier workbook	2/3/2020 9:47:16 AM
BATCH EXECUTION	Test preview	Excel 2013 or later workbook	2/3/2020 9:47:16 AM

Summary

By completing this lab, you have prepared your environment and data, and built and deployed your own Azure Machine Learning model. We hope you enjoyed this introductory hands-on lab and that you will build many more machine learning solutions!

Limitations

Of course, there is much information missing. We don’t know anything about the dates of the obtained data, nor do we know anything between the data gathering and the moment that the employee left.

Inspiration

As mentioned before, this hands-on lab is created to inspire you. If for whatever reason you were struggling to get the model built, you can also download the complete model from the Azure AI Gallery.

We hope you have enjoyed this workshop and hopefully it inspired you to build your own models. If you want to take your models into production, then please use another environment: https://ml.azure.com/

Here you can find a similar interface, called Designer, but with this interface, you can also deploy and manage your models.