ST1511
AI & MACHINE LEARNING

Practical 2
Introduction to Cloud-based Machine Learning Tool

What you will learn / do in this lab

1. *Explore Cloud-based Machine Learning Tool*
2. *Use tool to perform a classification experiment*
3. *Deploy model as a web service*

TABLE OF CONTENTS

1. OVERVIEW 1

Introduction 1

2. AZURE ML 2

Overview 2

Creating a new Experiment 4

Creating a new Saved Dataset 6

3. ML WORKFLOW 8

Formulating the question 8

4. DEPLOYMENT 16

Preparation for deployment 16

1.
OVERVIEW

This practical will introduce to you a typical machine learning workflow for a problem. It will also demonstrate how you can move from the tested data model to deployment as a web service.

INTRODUCTION

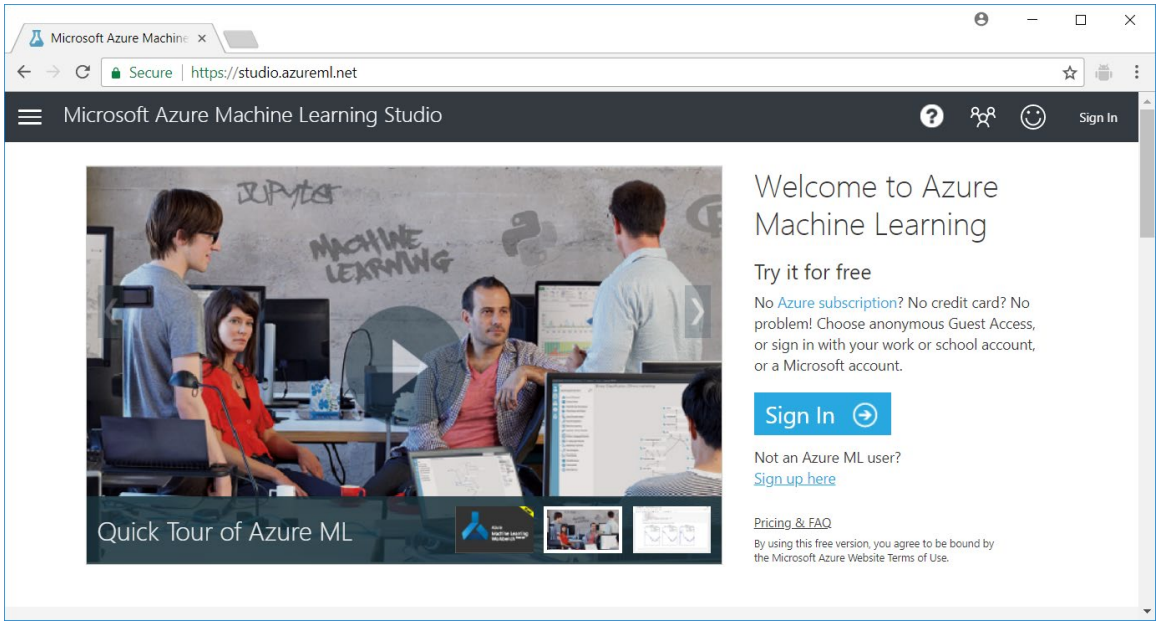
2.
AZURE ML

In this section we will introduce to you the Microsoft Azure Machine Learning Studio.

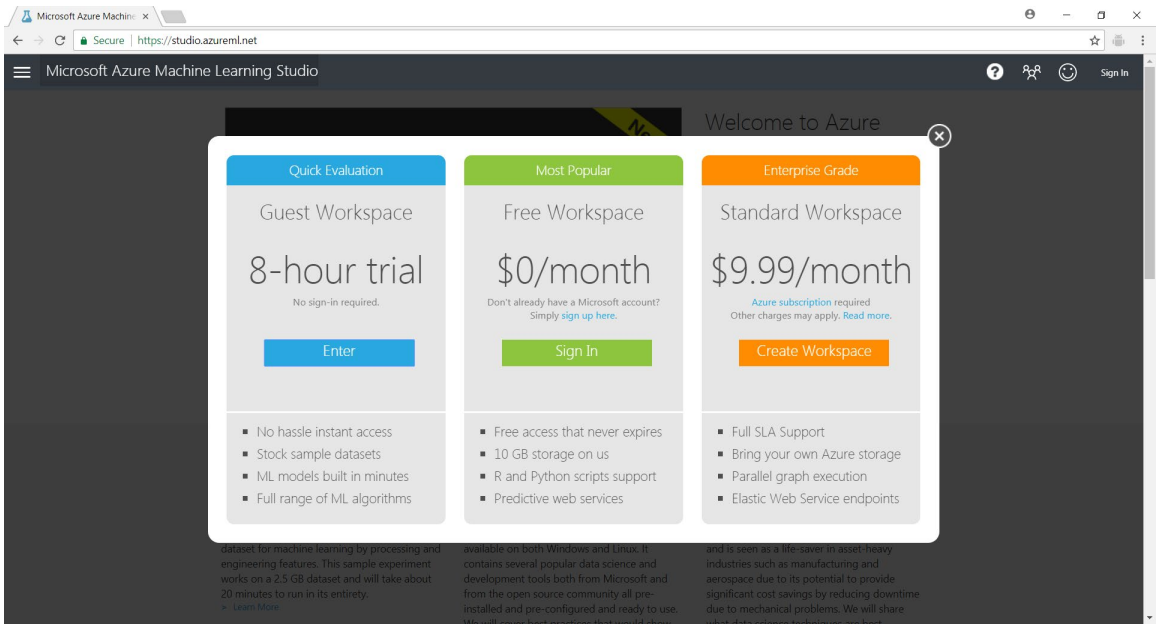
OVERVIEW

Use URL https://studio.azureml.net/

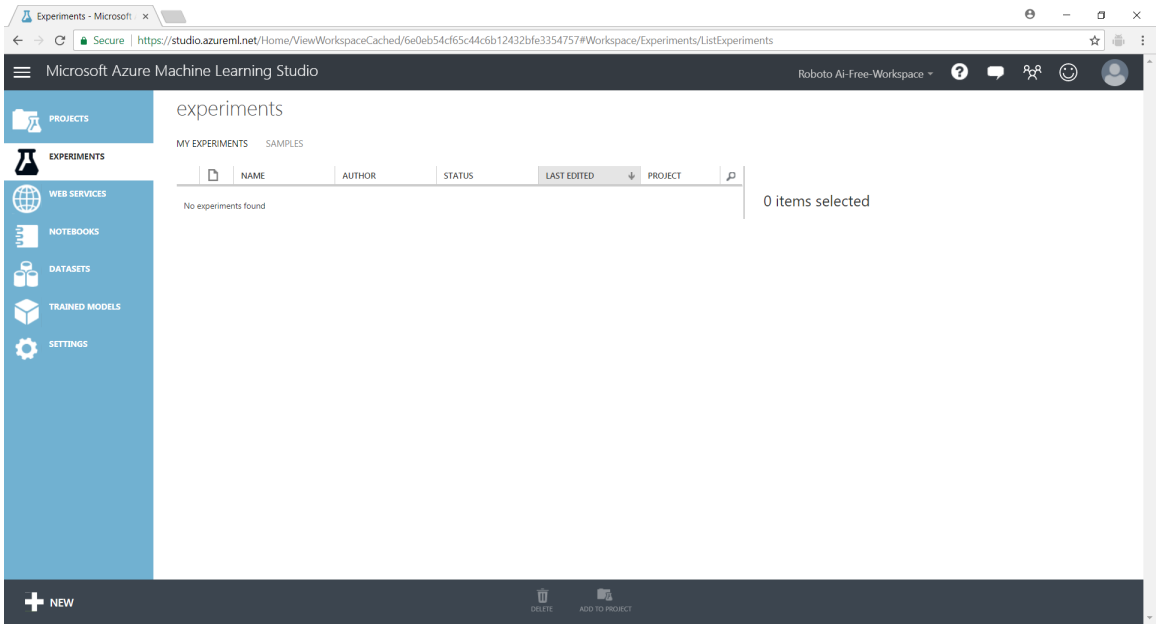
You must login with a Microsoft Account if you wish to complete part 4 (deployment). Otherwise, you can use the anonymous login. There is a free-tier for the Azure Machine Learning (ML) Studio application in the Azure portal.



Click on the link below [Not an Azure ML user?] if you new to Azure ML studio. You should see a screen to select the type of account. Choose the "Free Workspace" option.



If you have login successfully, you will go to the home screen for creating new experiments.



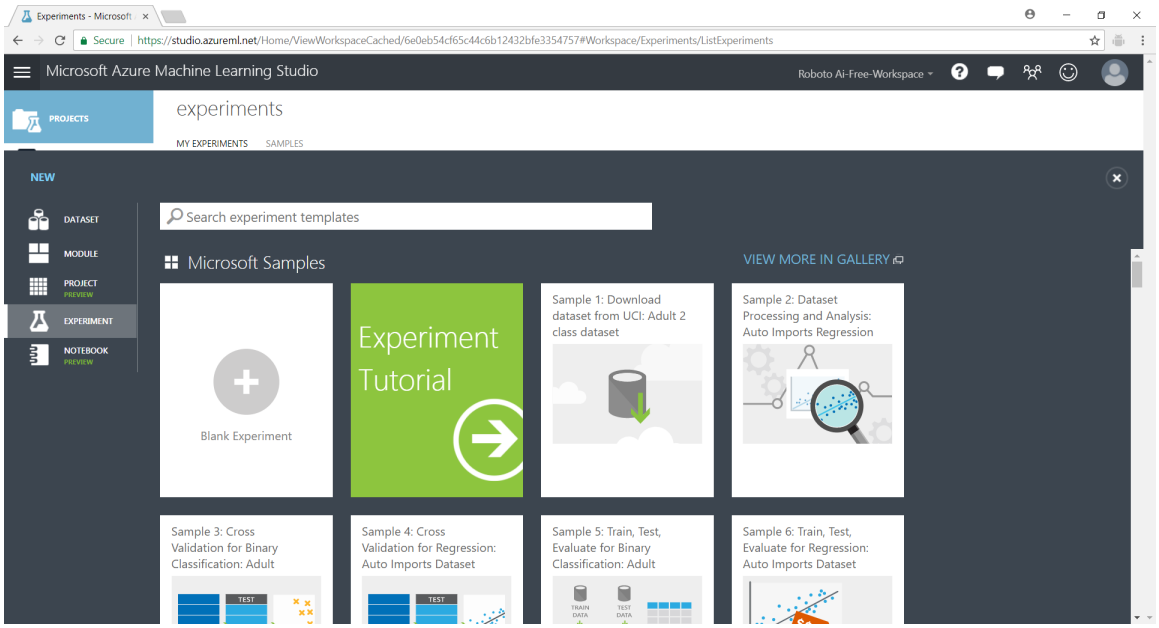
The tab is used to navigate between the different parts of the Azure ML application. Once signed in, you'll see the following tabs on the left:

- **PROJECTS** - Collections of experiments, datasets, notebooks, and other resources representing a single project
- **EXPERIMENTS** - Experiments that you have created and run or saved as drafts. (DEFAULT HOME PAGE)

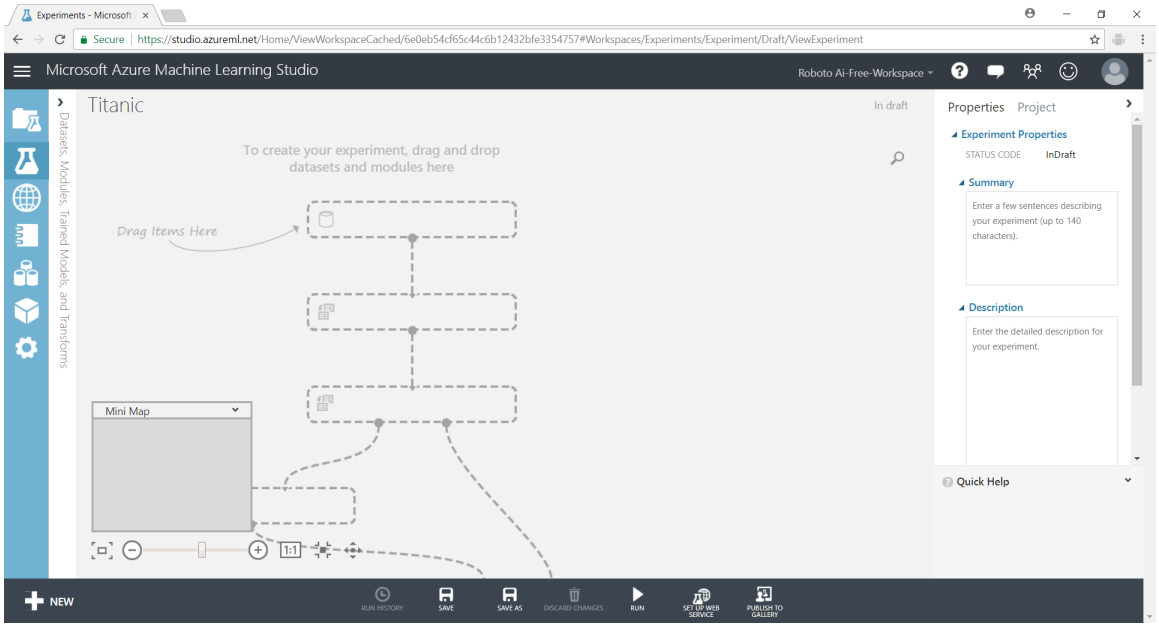
- **WEB SERVICES** - Web services that you have deployed from your experiments
- **NOTEBOOKS** - Jupyter notebooks that you have created
- **DATASETS** - Datasets that you have uploaded into Studio
- **TRAINED MODELS** - Models that you have trained in experiments and saved in Studio
- **SETTINGS** - A collection of settings that you can use to configure your account and resources.

CREATING A NEW EXPERIMENT

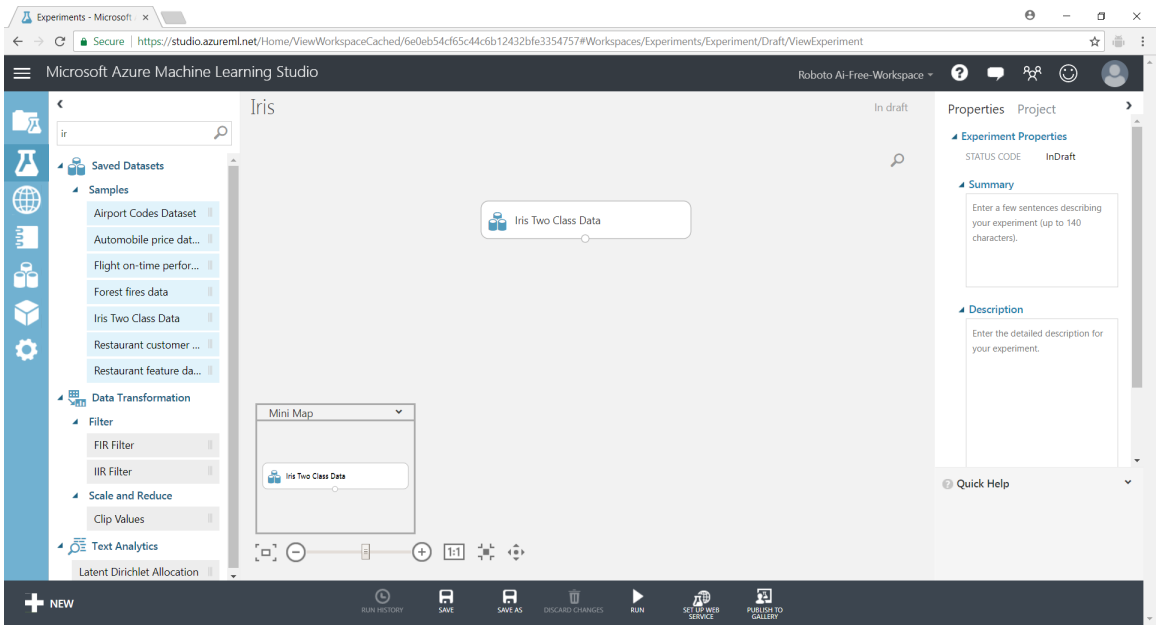
1. Create a new experiment by clicking **+NEW** at the bottom of the Machine Learning Studio window, select **EXPERIMENT**, and then select **Blank Experiment**.



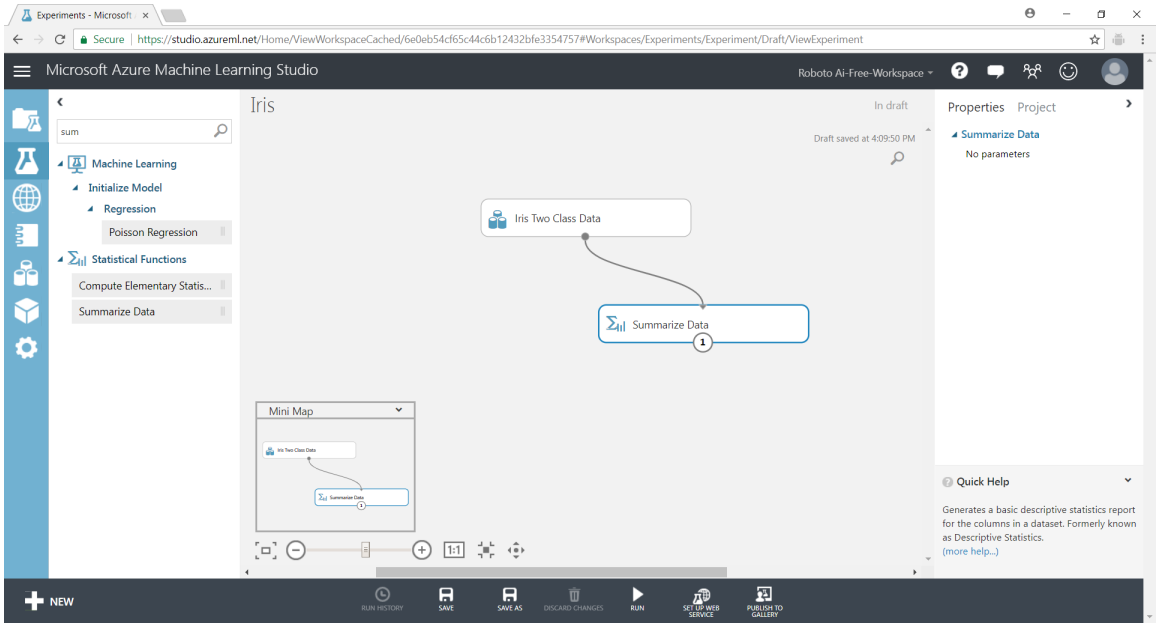
2. The experiment is given a default name that you can see at the top of the canvas. Select this text and rename it to something meaningful, for example, **Iris**. The name doesn't need to be unique.



Select the [Iris Two Class Data] saved data set from modules list on the left and drag it onto the canvas. Note that you can type a few letters into the search bar and the list will automatically narrow. For example, I typed "ir" to filter the list to the iris data set.



You cannot save the experiment unless you have an additional module besides the data source. Add the [Summarize Data] module.

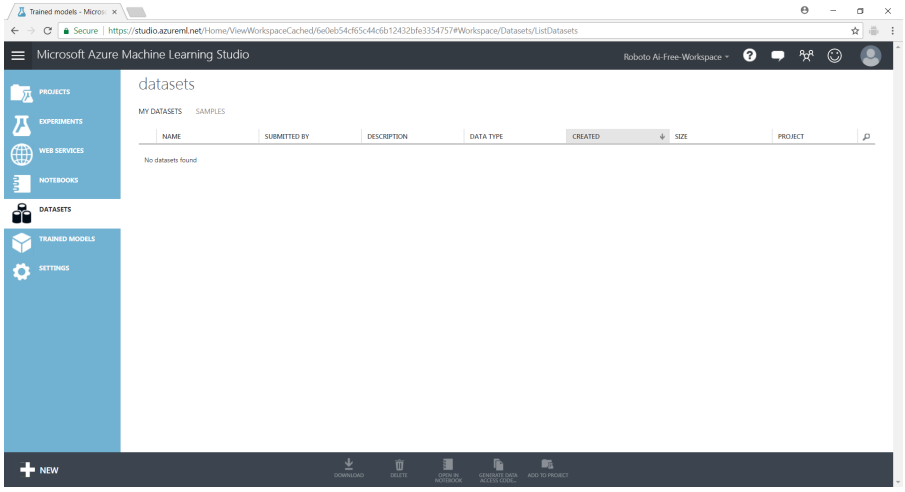


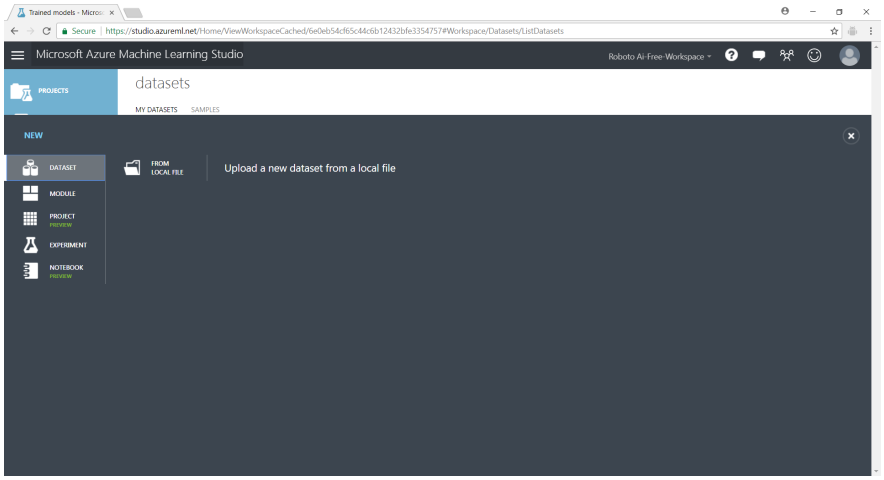
Connect the [Iris Two Class Data] data to the [Summarize Data] by clicking the bottom output port (o-shape) of [Iris] and dragging the connector to the top input port (o-shape) of [Summarize].

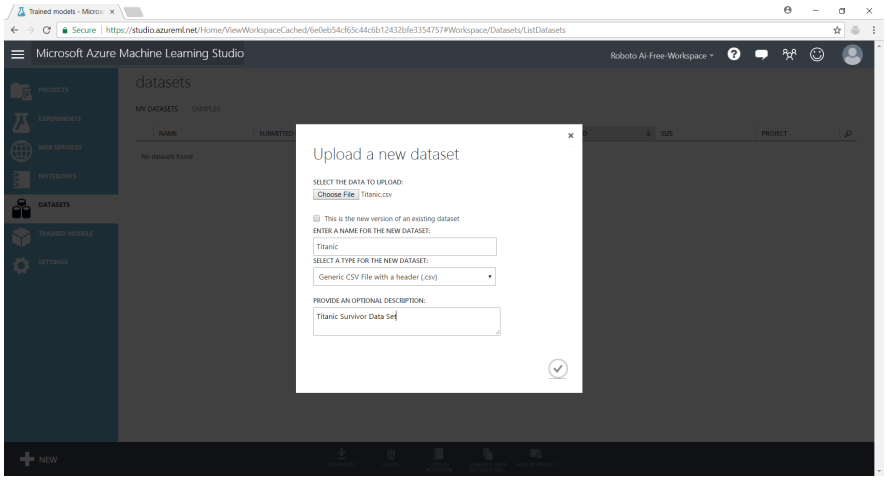
Click on **Save** at the bottom action bar to save the experiment.

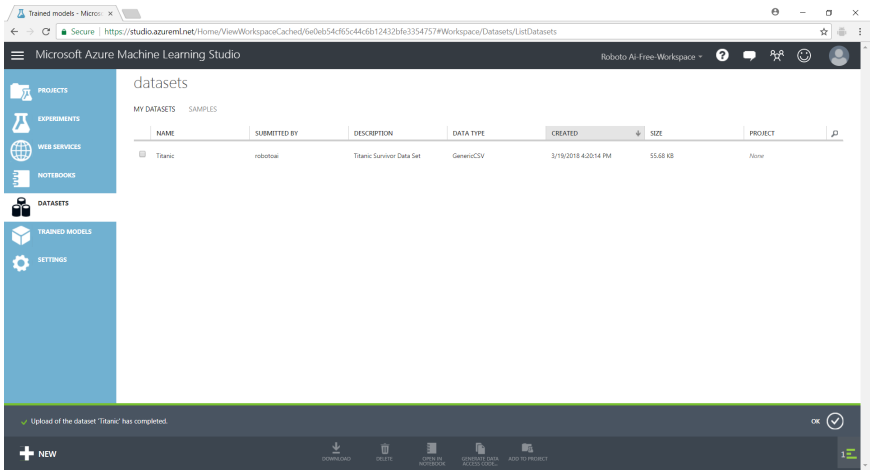
CREATING A NEW SAVED DATASET

Click on the Data Set tab. Create a new Data Set by uploading the titanic.csv file.



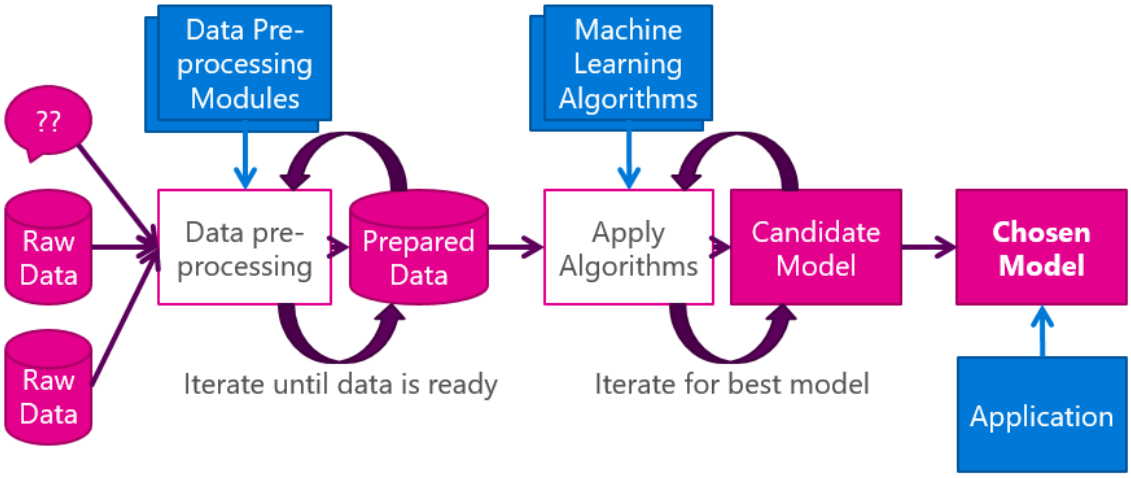

Microsoft Azure Machine Learning Studio
Roboto Ai-Free-Workspace
datasets
MY DATASETS
SAMPLES
NEW
DATASET
MODULE
PROJECT
EXPERIMENT
NOTEBOOK
FROM LOCAL FILE
Upload a new dataset from a local file


Microsoft Azure Machine Learning Studio
Roboto Ai-Free-Workspace
datasets
Upload a new dataset
SELECT THE DATA TO UPLOAD:
Choose File
Titanic.csv
This is the new version of an existing dataset
ENTER A NAME FOR THE NEW DATASET:
Titanic
SELECT A TYPE FOR THE NEW DATASET:
Generic CSV File with a header (.csv)
PROVIDE AN OPTIONAL DESCRIPTION:
Titanic Survivor Data Set
NEW


Microsoft Azure Machine Learning Studio
Roboto Ai-Free-Workspace
datasets
MY DATASETS
SAMPLES
NAME
SUBMITTED BY
DESCRIPTION
DATA TYPE
CREATED
SIZE
PROJECT
Titanic
robotoai
Titanic Survivor Data Set
GenericCSV
3/19/2018 4:20:14 PM
55.68 KB
None
PROJECTS
EXPERIMENTS
WEB SERVICES
NOTEBOOKS
DATASETS
TRAINED MODELS
SETTINGS
Upload of the dataset 'Titanic' has completed.
OK
NEW

3.
ML WORKFLOW

In this section, we will use Azure ML Studio to develop a data model for a machine learning problem.



FORMULATING THE QUESTION

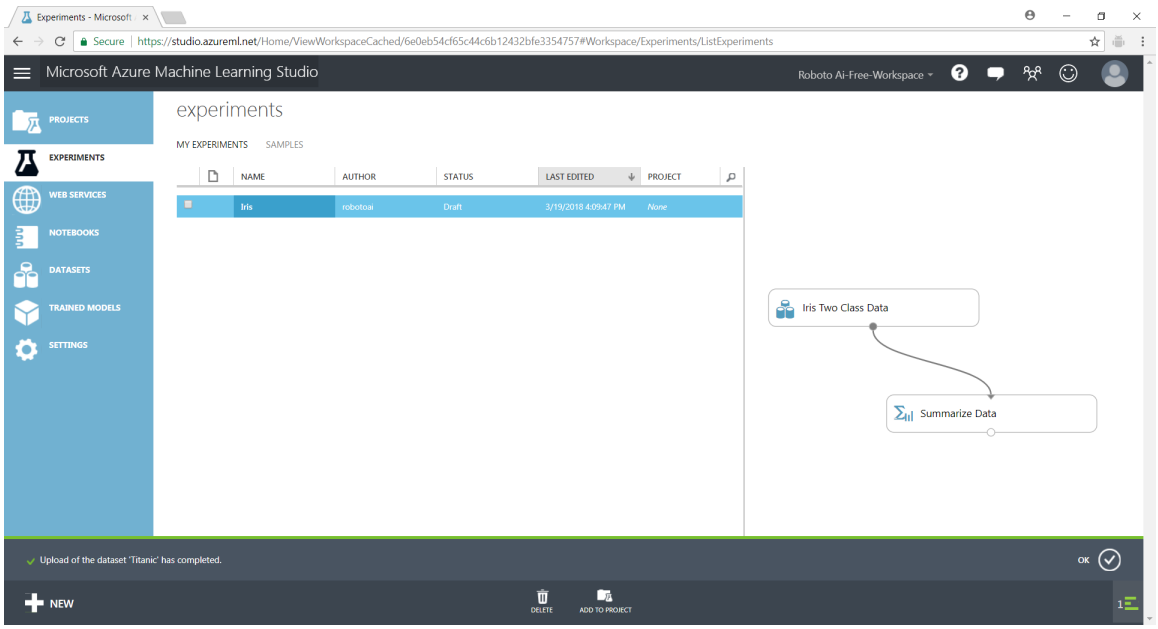
The ML workflow process begins with the formulation of the question (no shown in diagram); what is it that we desire as the answer from the machine learning algorithm? Do we want to label a new data with certain attributes in a particular known grouping (classification problem)? Do we want to know which of the attributes in the data help us to label the data correctly (classification problem)? Do want to predict a numeric value (for example, the housing price) for a new property (regression problem)?

It is important to consider the question first, because if we don't know what the question is, it is impossible to determine if the data is sufficient or if the algorithm is producing the output that we want.

DATA PREPARATION

The next step often consumes the most time. Given the question that we want to have answers to; do we have the right data. If the data is available, the data be cleaned of errors and duplicates. Missing data values in the data must be handled.

In the earlier section, we had already created an experiment using a clean [Iris Two Class Data] dataset. Let's continue using the previous experiment. You can click on the [Experiments] tab on the left side.

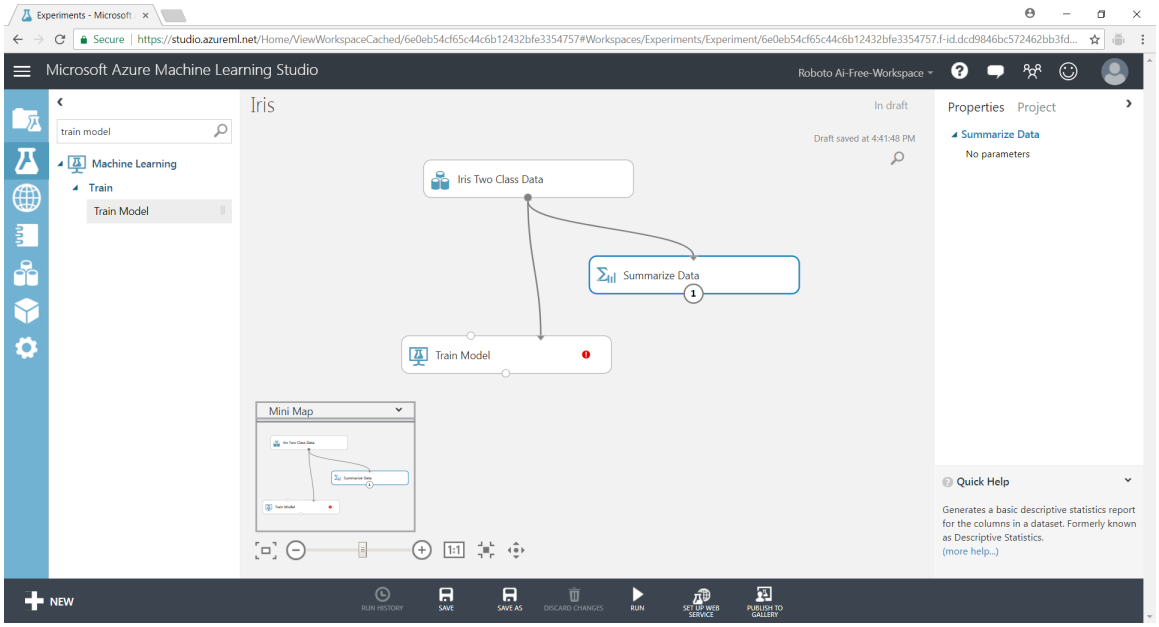


MODELING

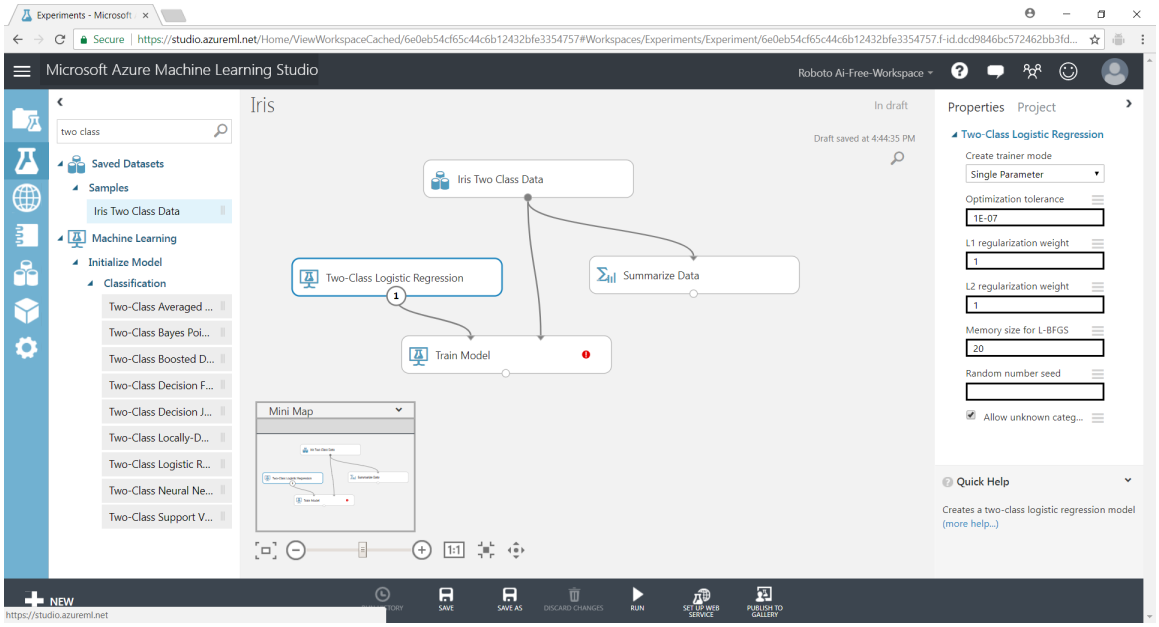
As seen in the diagram above we would apply a suitable learning algorithm to create the model.

Re-open the Iris experiment

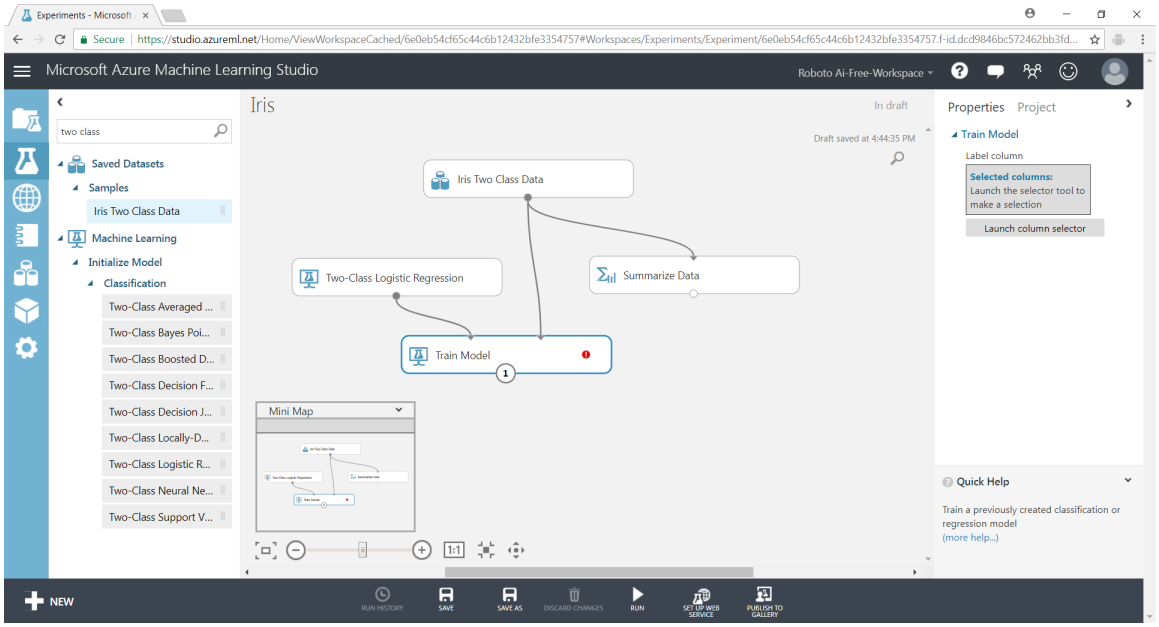
Drag in the [Train Model] module and connect the [Iris] data to the right port.

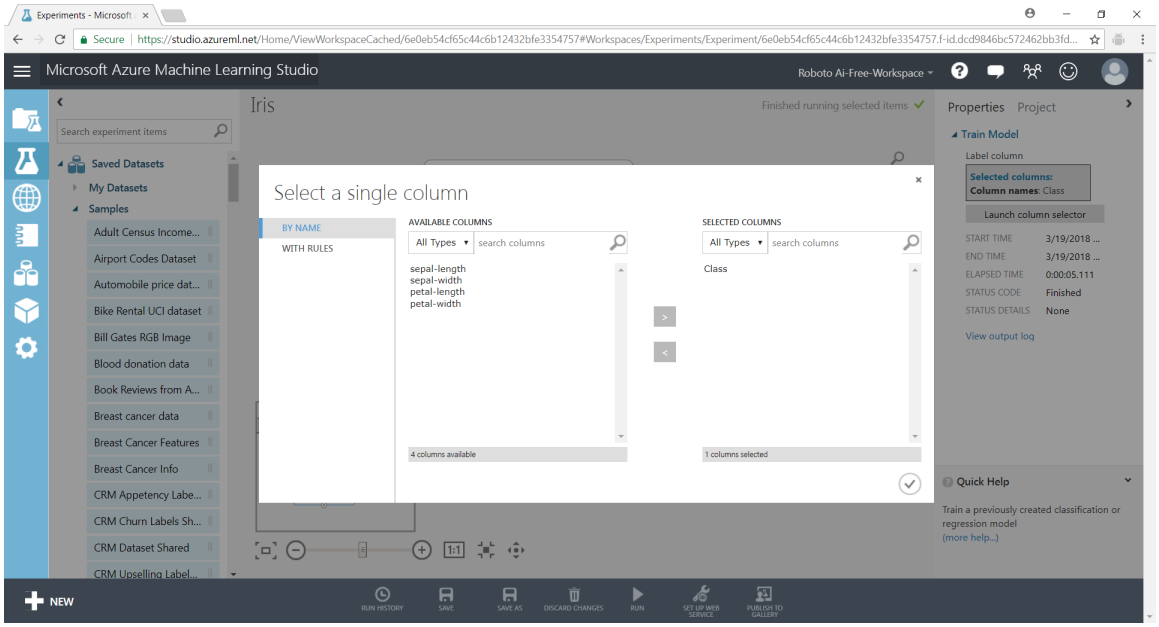


Select the learning algorithm you want to use to apply to the [Train Model]. Use the [Two Class Logistic Regression] model.

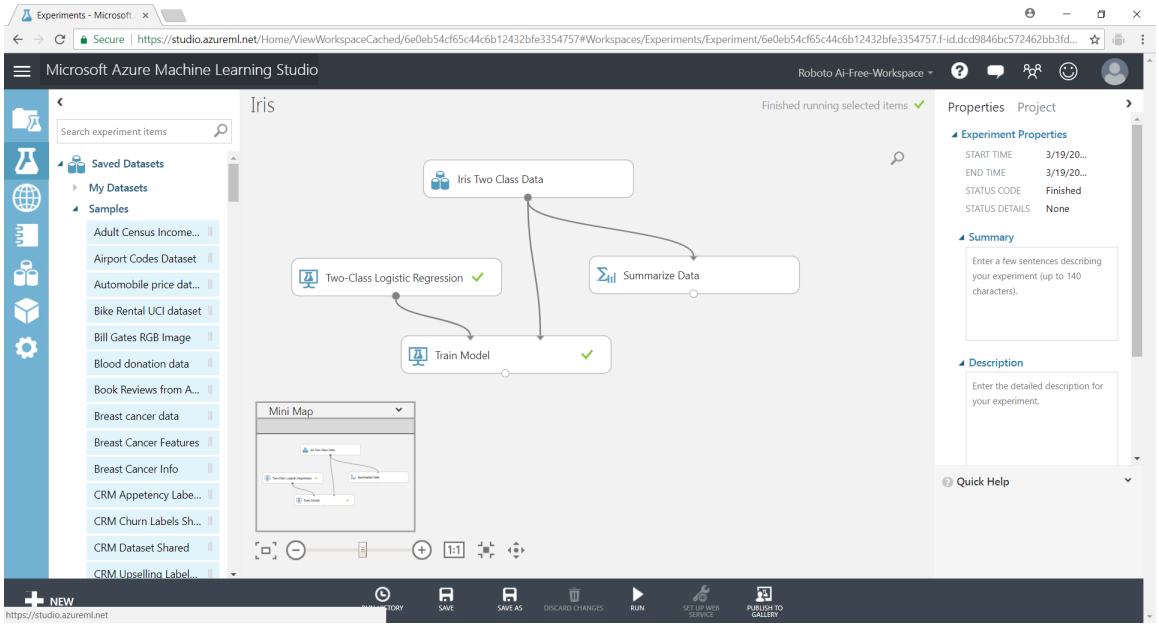


The red exclamation mark denotes an error that needs to be correct. Click on the [Train Model]. The properties tool for the module appears on the right-hand side. Use the column selector to choose the output or target class.



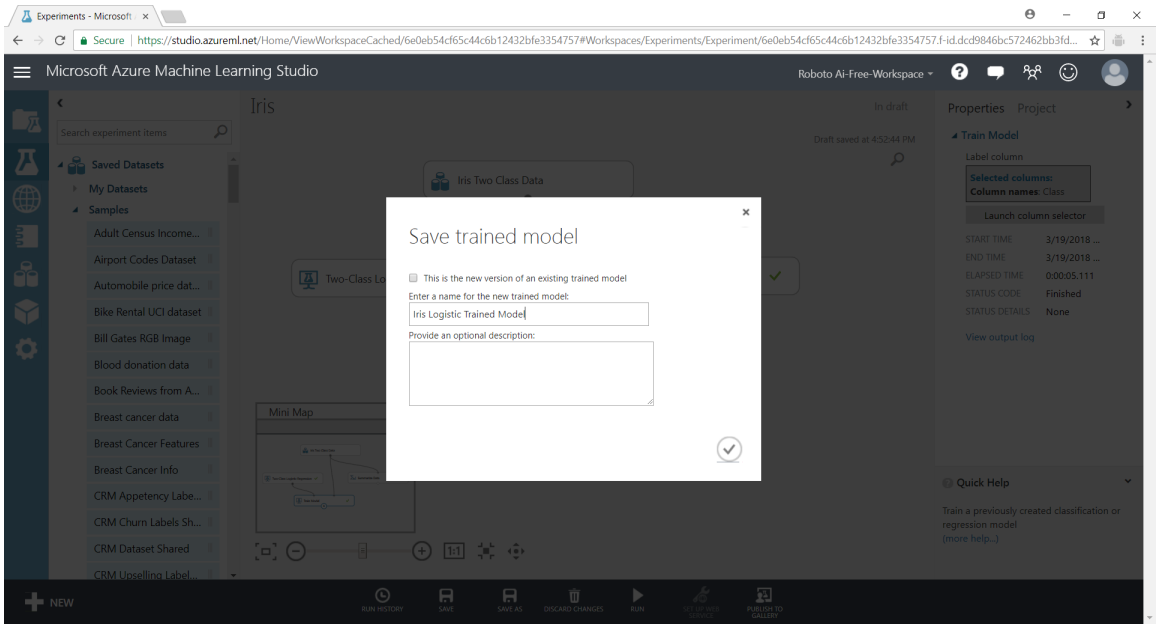


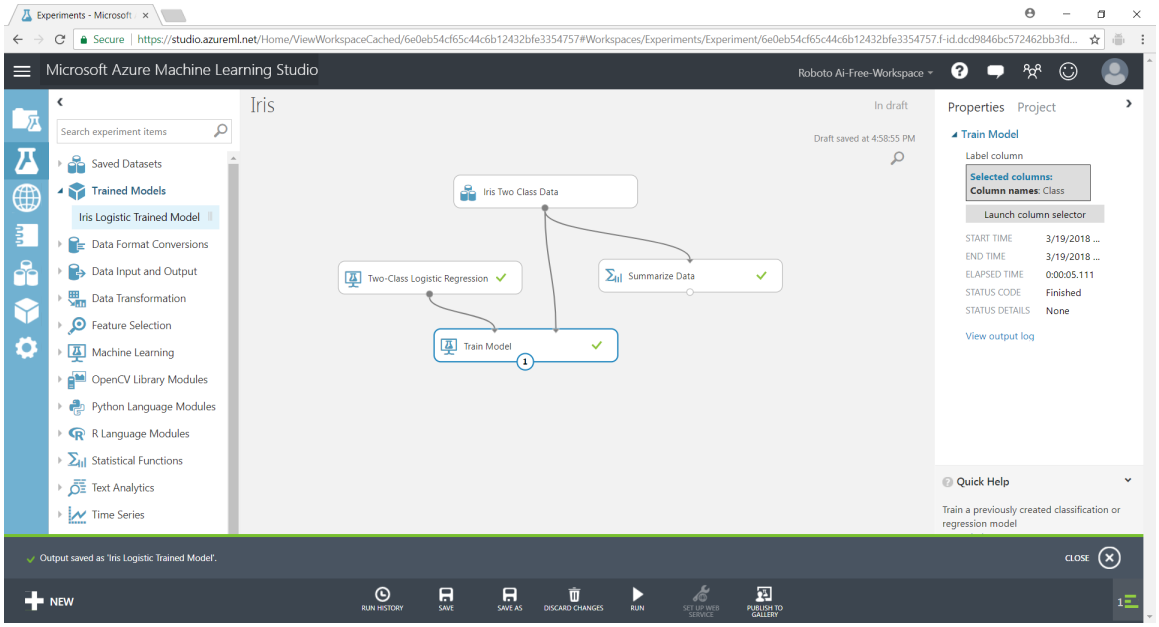
Run the model. Right-click on the [Train Model] and select **Run Selected**. Right on the [Train Model] and select Trained Model > **Visualize**.



Remember to click **Save** at the bottom action bar.

After training the model successfully, we can save the trained model for later use. Right-click on the [Train Model] module and select Trained model > **Saved as Trained Model**.

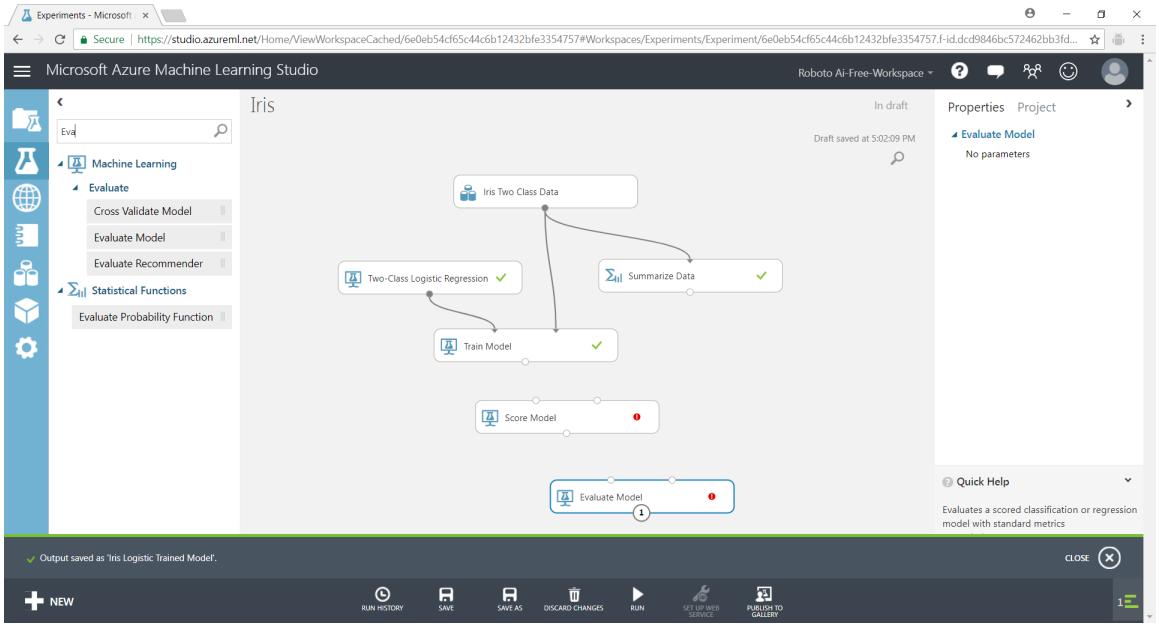




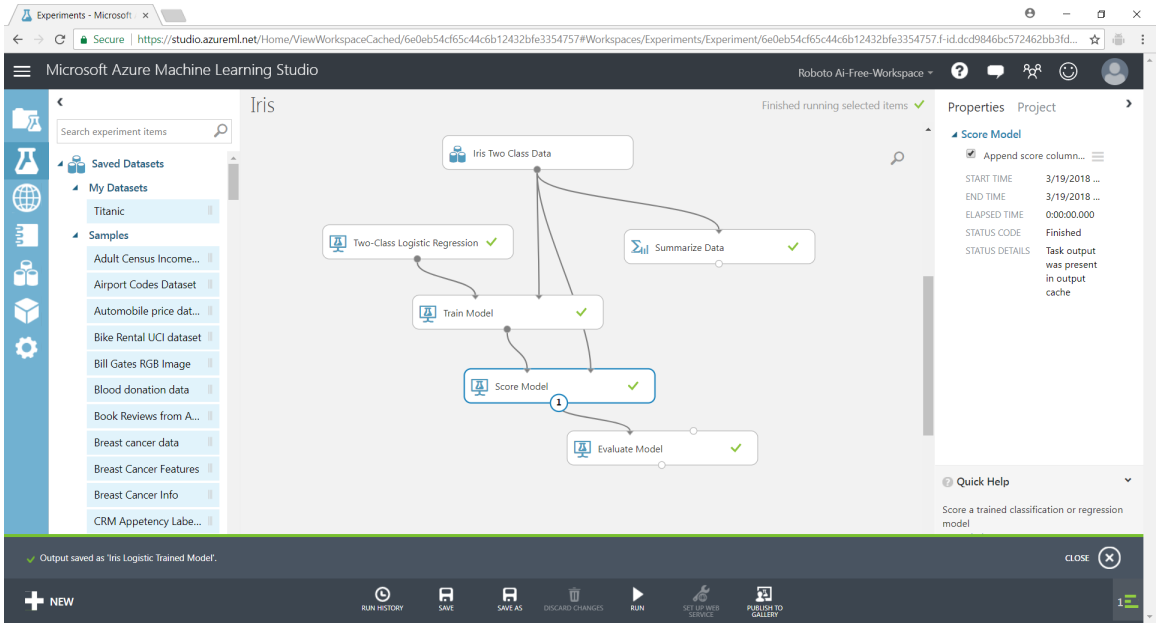
EVALUATING THE MODEL

To evaluate the model is to check that the model is giving us the output that we want.

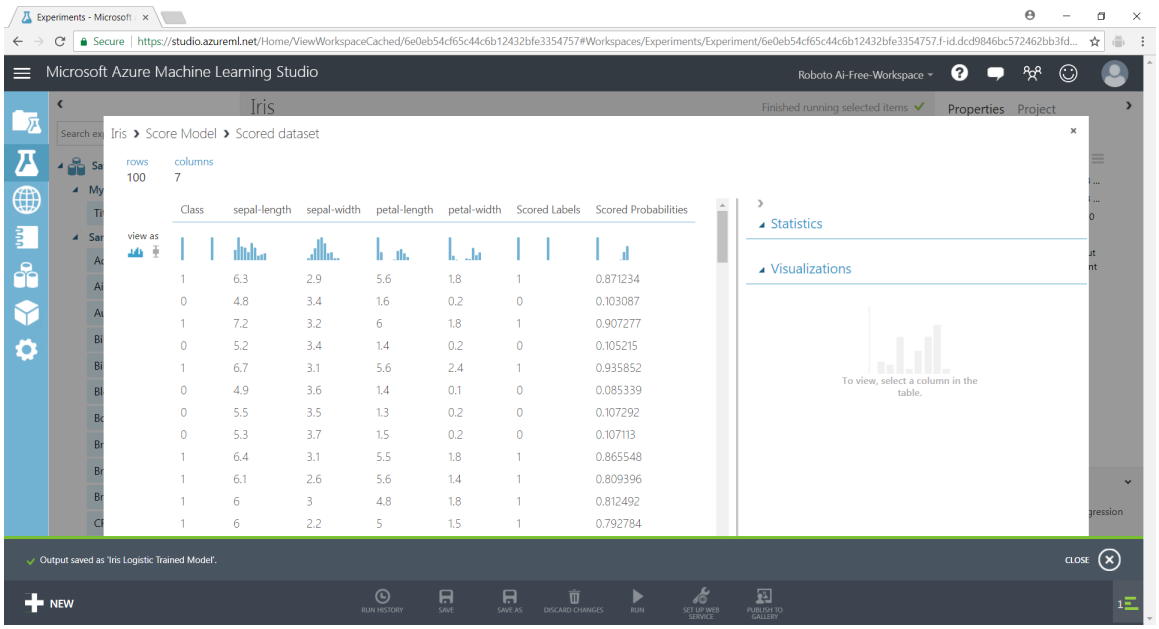
Select and add the [Score Model] and [Evaluate Model] modules to the experiment. In this simple experiment, I have not partitioned the data into a separate test set. I am assuming that all the data is for training and also testing.



Connect the ports correctly as shown in the diagram below.

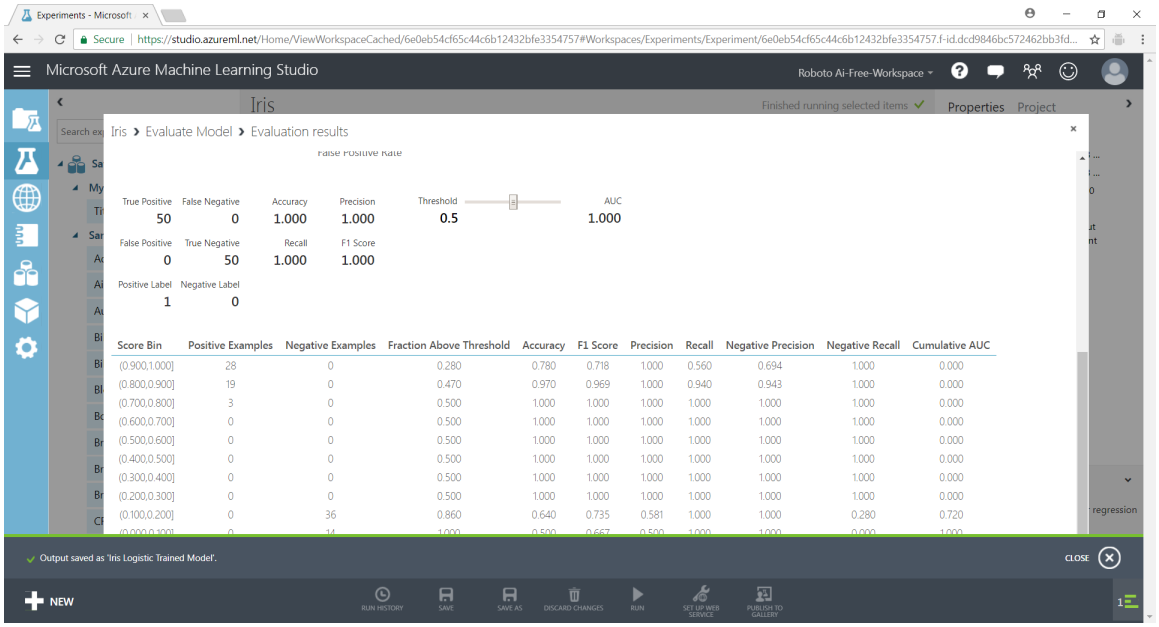


Right-click [Score Model] and select Scored Dataset > **Visualize**.



This shows a table of the actual Class/label the "Scored Labels" or predicted label for the data. It appears to be 100% accurate.

Right-click [Evaluate Model] and select Evaluation results > **Visualize**.

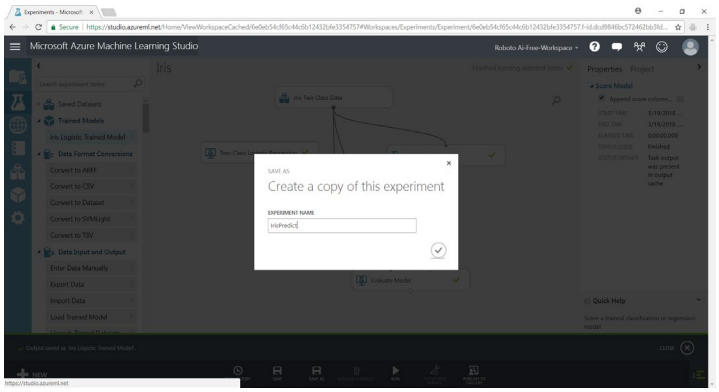


The accuracy is reported to be 100% which confirms what we saw in the table from the [Score model]

DEPLOYING THE MODEL

We need to use the Saved Trained Model [Iris Logistic Trained Model] from earlier to create another experiment.

Save the current experiment as a copy using the **Save As** at the bottom to IrisPredict.

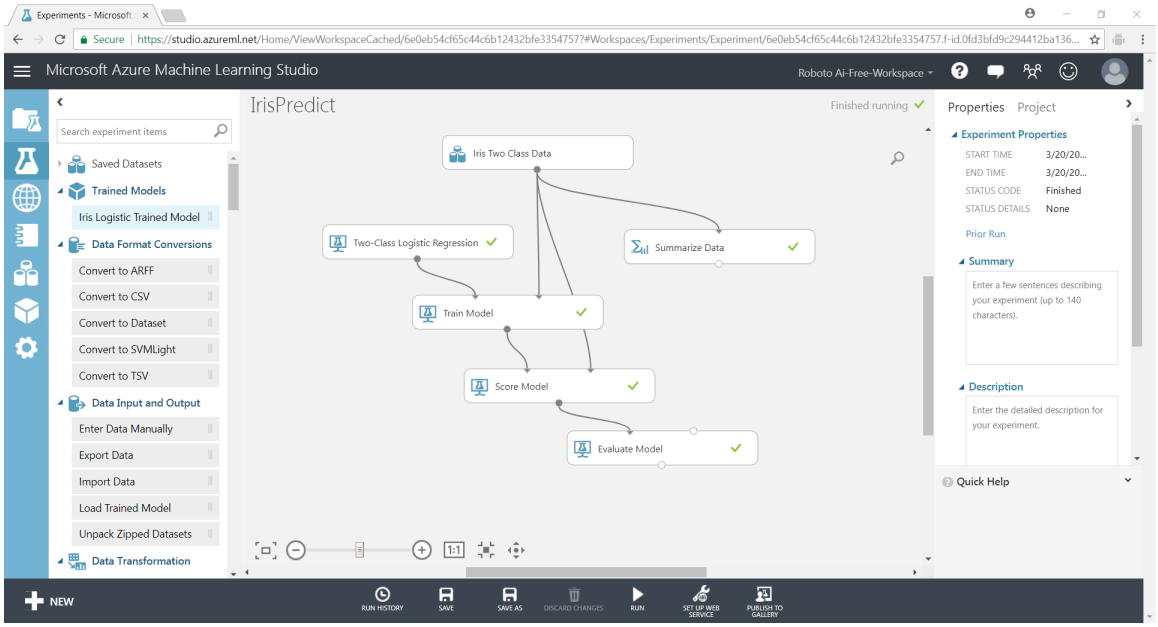


4.
DEPLOYMENT

In this section, we will deploy the data model we had developed earlier into a web service

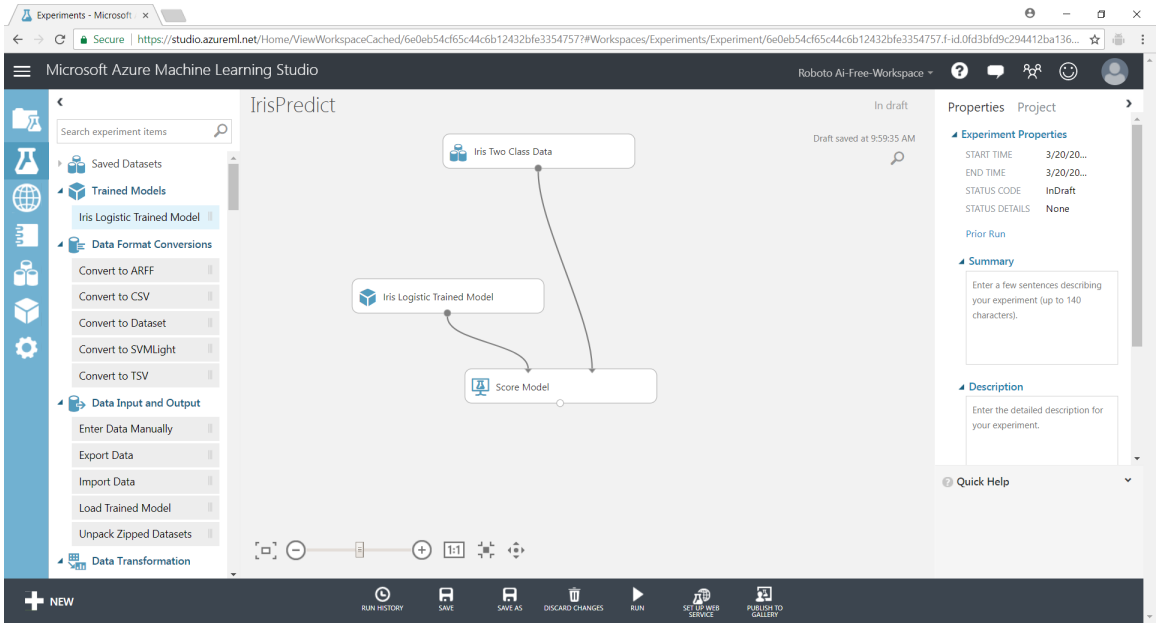
PREPARATION FOR DEPLOYMENT

Open the [IrisPredict] experiment which we saved earlier. And run it.

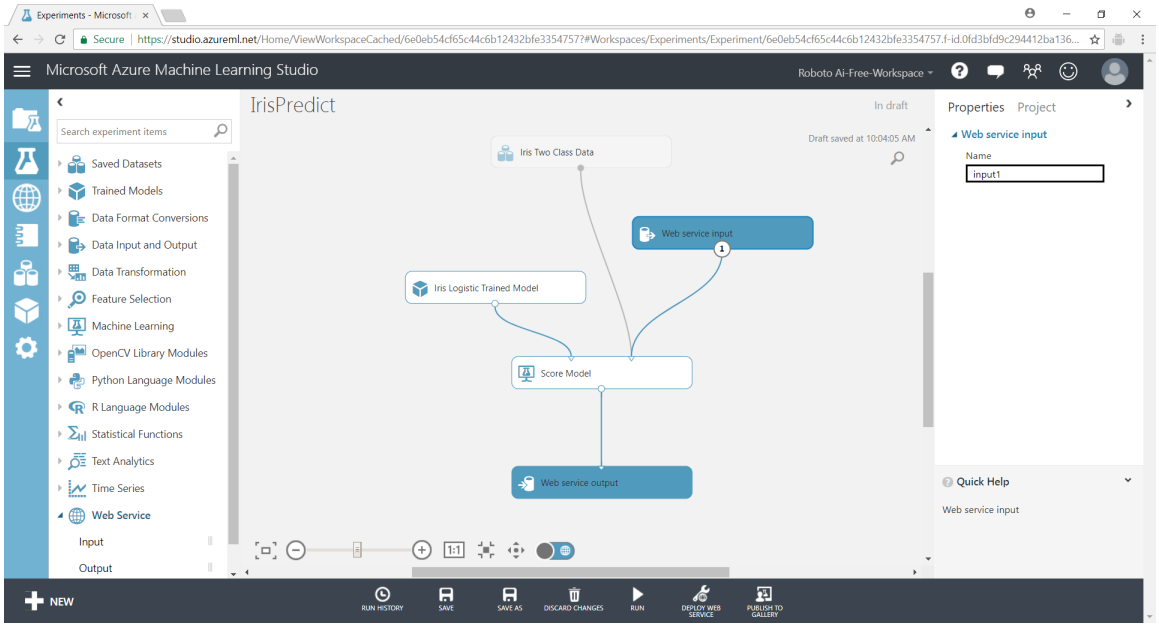


Remove [Evaluate Model], [Summarize Data]
Replace the [Train Model] and [Two-Class Logistic Regression] with [Iris Logistic Trained Model] from the Saved Trained Models

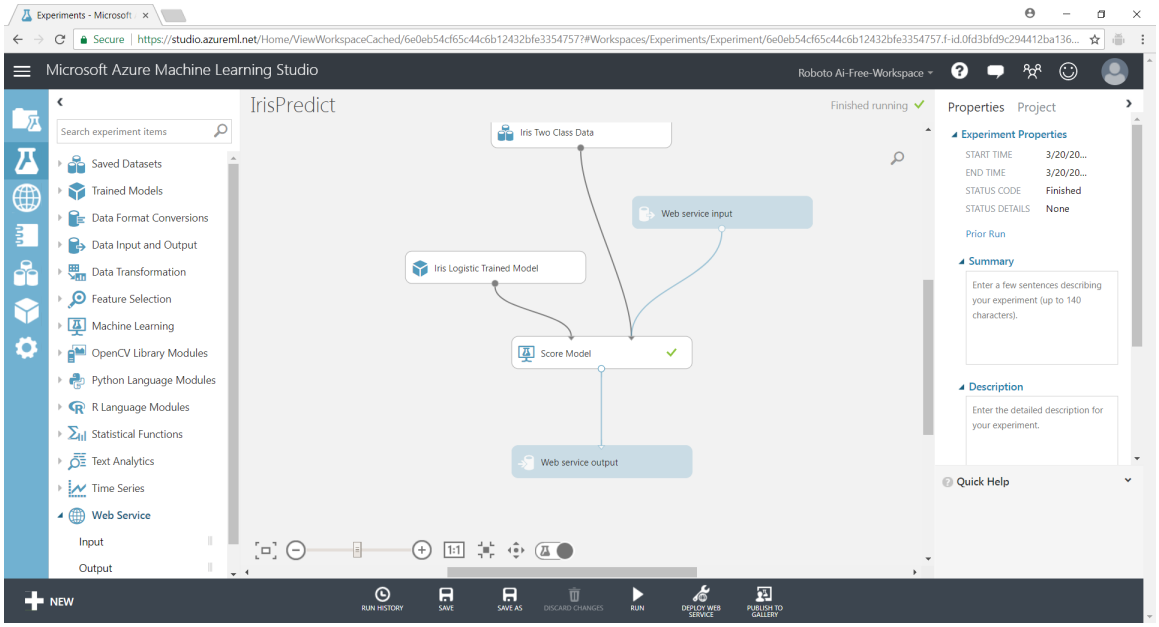
Run the model by clicking **Run** on the bottom action bar



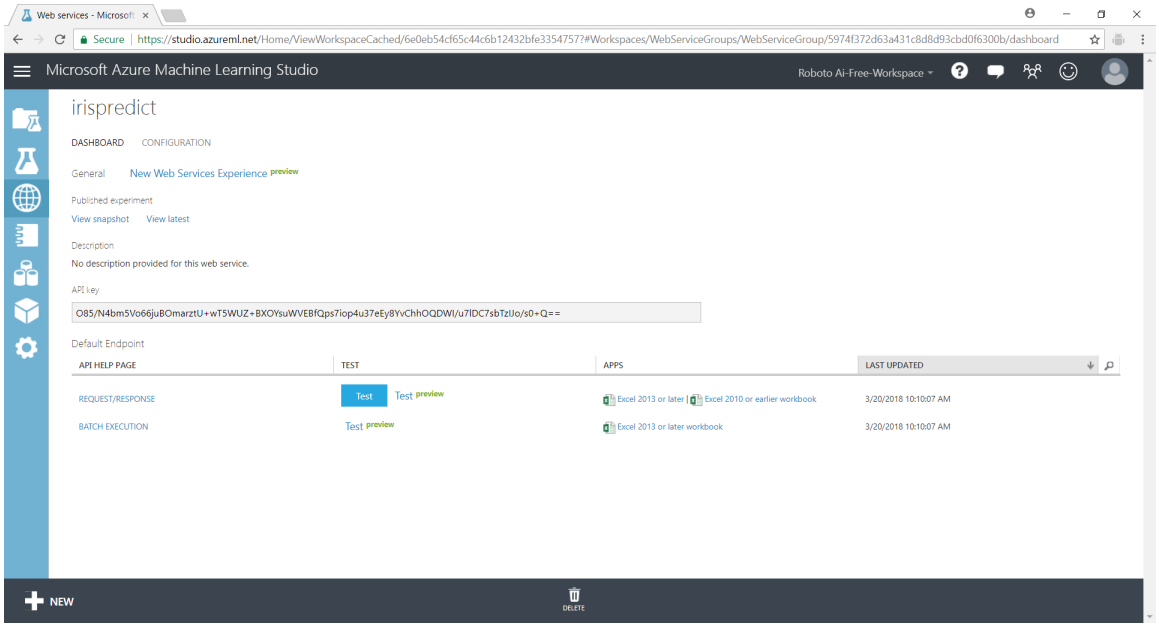
Click on the [SET UP WEB SERVICE] button on the bottom action bar.



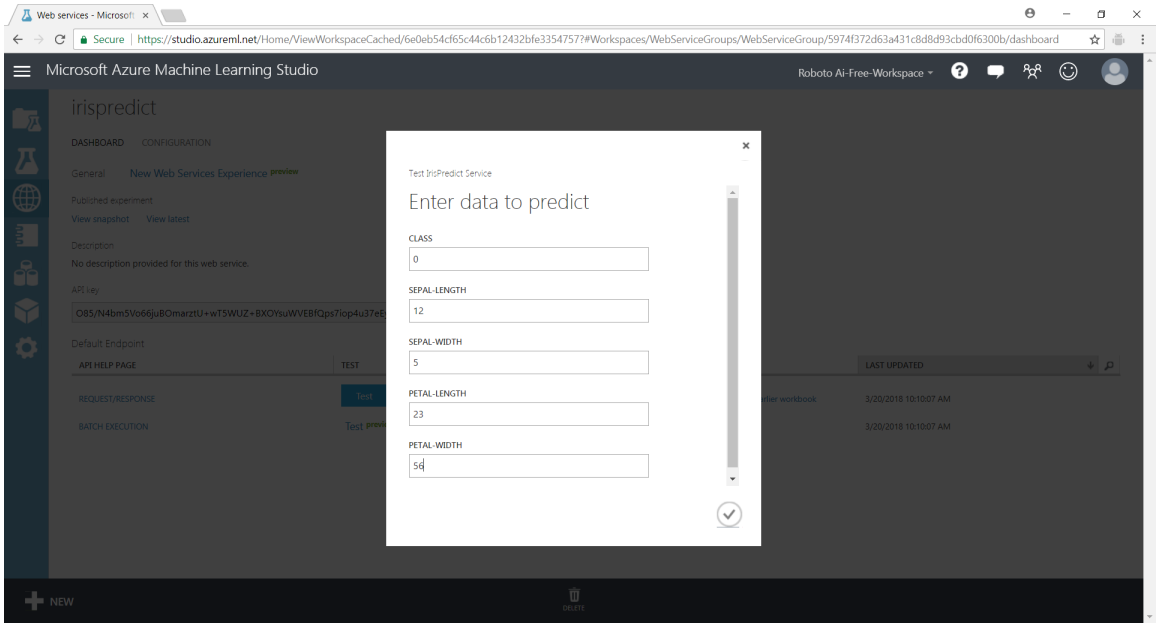
A new “predictive experiment” is created. Click on the **Run**.



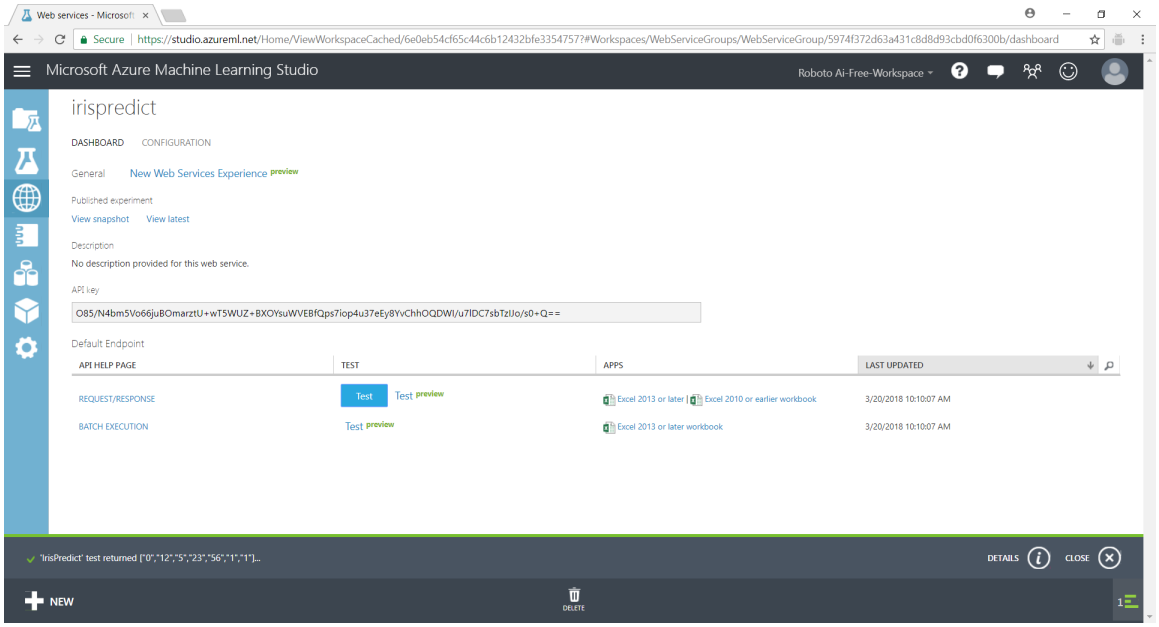
Click on the [**DEPLOY AS WEB SERVICE**] button on the bottom action bar.



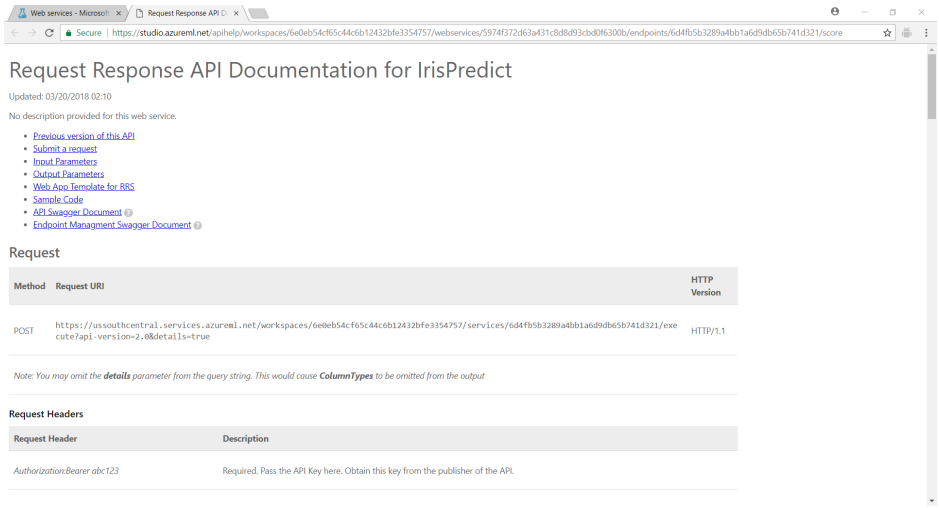
To test the web service click on the [Test] button and enter some values.



Click on the [Details] to see the details of the results returned.



Click on the REQUEST/RESPONSE link under API HELP PAGE column. You will be brought to a page containing details about the web service deployed and how to use it.



Clicking on the [Sample Code] brings you to some code samples in different programming languages including Python.

ADDITIONAL EXERCISE

Try this on your own

Activity

- *Create a new experiment to use the Titanic dataset*
- *Find the best algorithm to use for predicting who will survive the Titanic disaster.*