**Lab Overview**

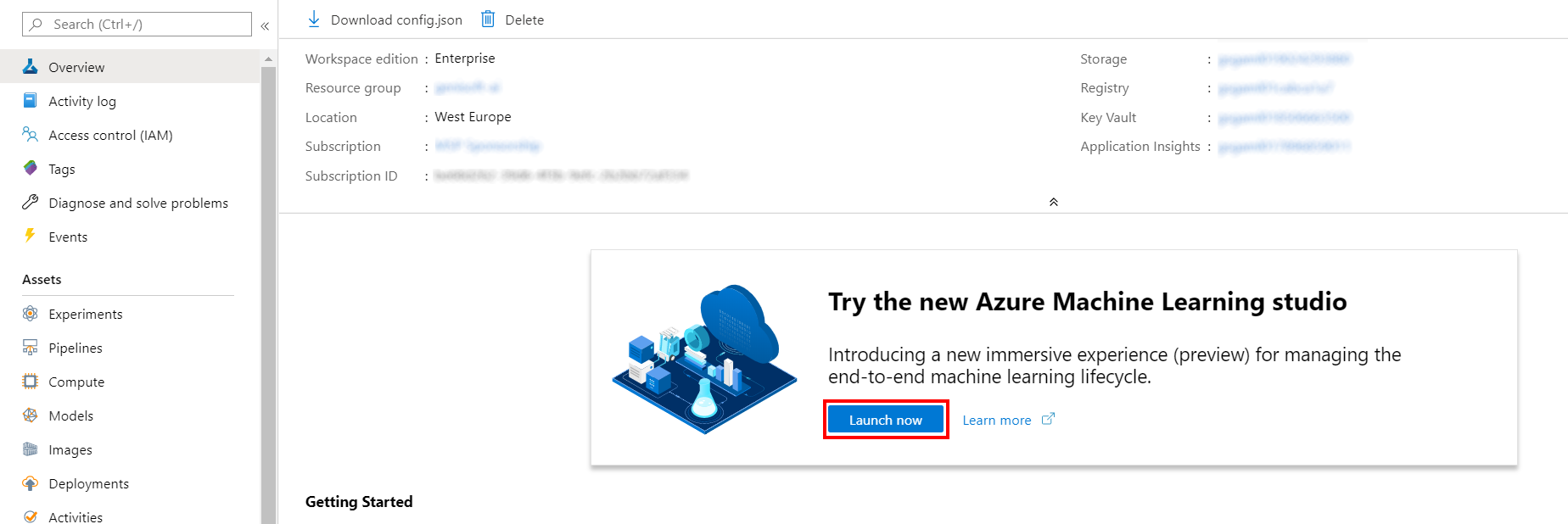
[Azure Machine Learning designer](https://docs.microsoft.com/en-us/azure/machine-learning/service/concept-designer) (preview) gives you a cloud-based interactive, visual workspace that you can use to easily and quickly prep data, train and deploy machine learning models. It supports Azure Machine Learning compute, GPU or CPU. Machine Learning designer also supports publishing models as web services on Azure Kubernetes Service that can easily be consumed by other applications.

In this lab, we will be using the Flight Delays data set that is enhanced with the weather data. Based on the enriched dataset, we will learn to use the Azure Machine Learning Graphical Interface to process data, build, train, score, and evaluate a classification model to predict if a particular flight will be delayed by 15 minutes or more. The classification algorithm used in this lab will be the ensemble algorithm: **Two-Class Boosted Decision Tree**. To train the model, we will use Azure Machine Learning Compute resource. We will do all of this from the Azure Machine Learning designer without writing a single line of code.

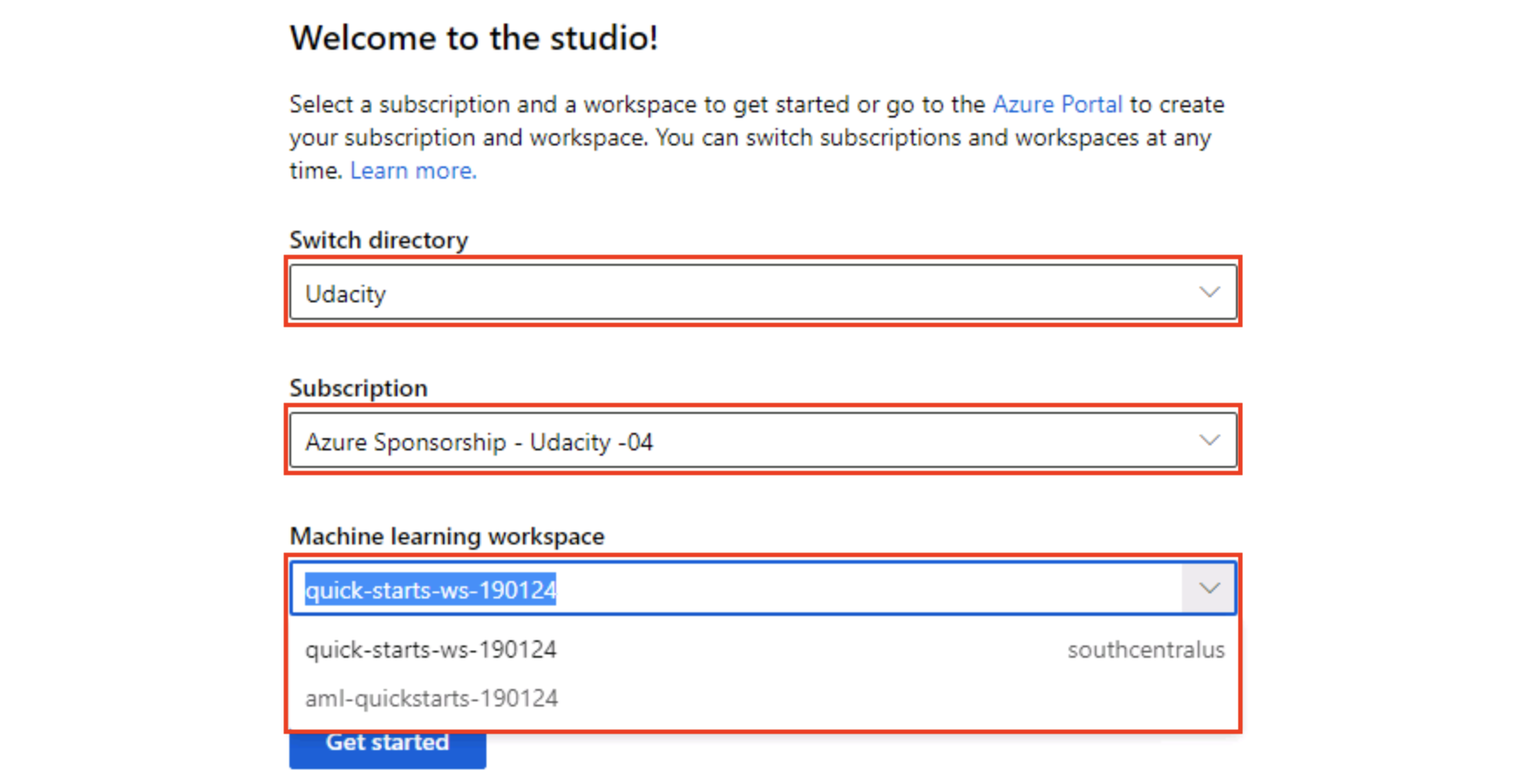
**Exercise 1: Register Dataset with Azure Machine Learning studio**

Task 1: Upload Dataset

1. In [Azure portal](https://portal.azure.com/), open the available machine learning workspace.
2. Select **Launch now** under the **Try the new Azure Machine Learning studio** message.

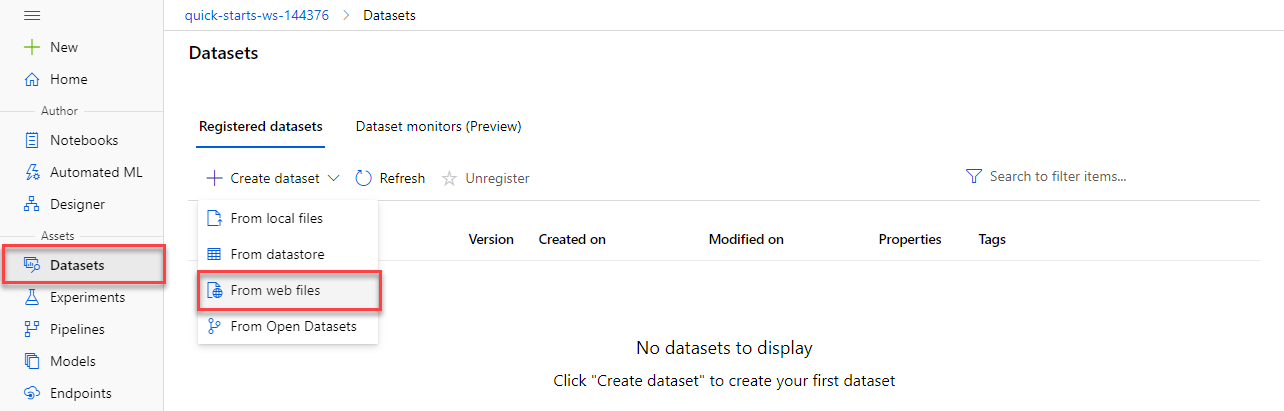


1. When you first launch the studio, you may need to set the directory and subscription. If so, you will see this screen:



For the directory, select **Udacity** and for the subscription, select **Azure Sponsorship**. For the machine learning workspace, you may see multiple options listed. **Select any of these** (it doesn’t matter which) and then click **Get started**.

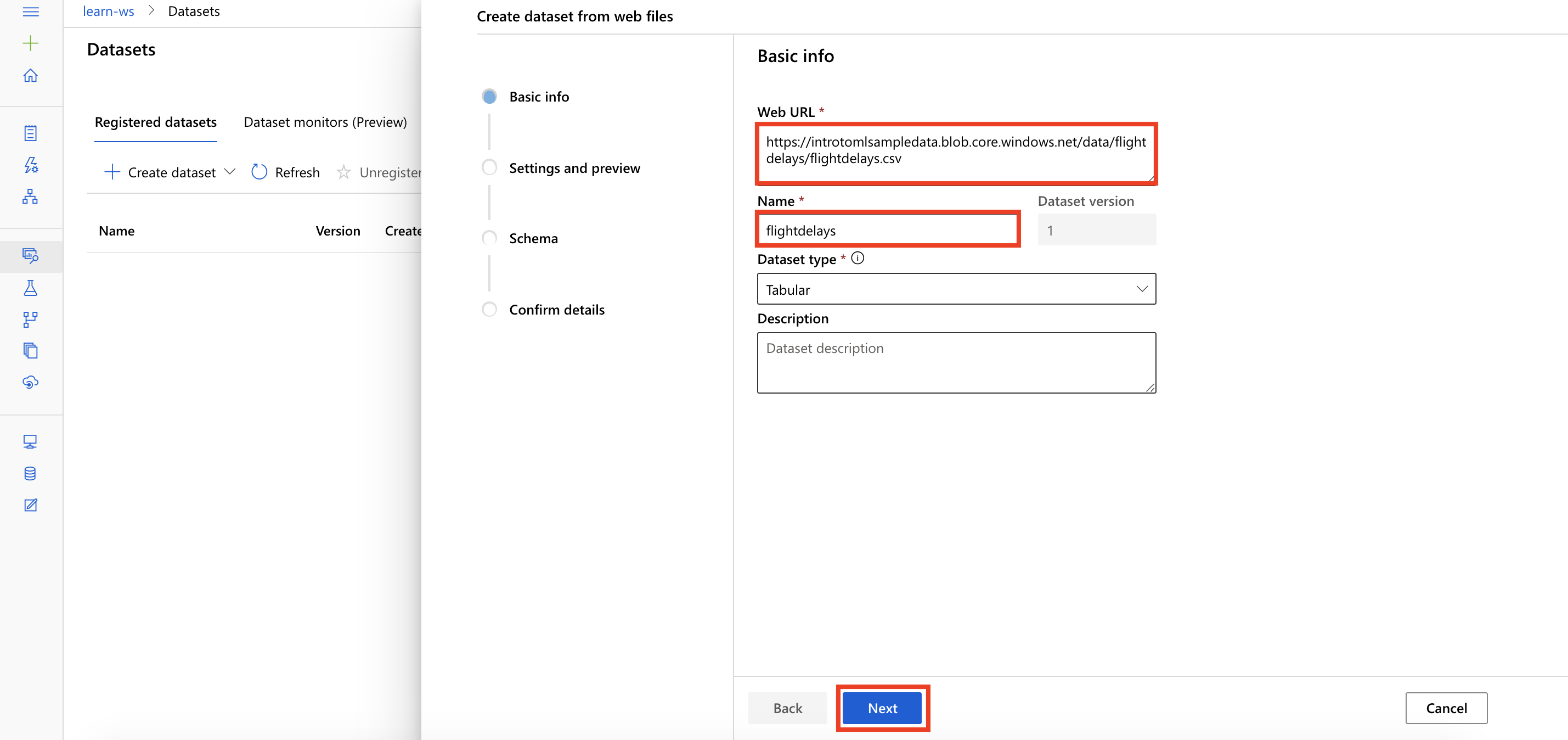
1. From the studio, select **Datasets, + Create dataset, From web files**. This will open the Create dataset from web files dialog on the right.



1. In the Web URL field provide the following URL for the training data file:

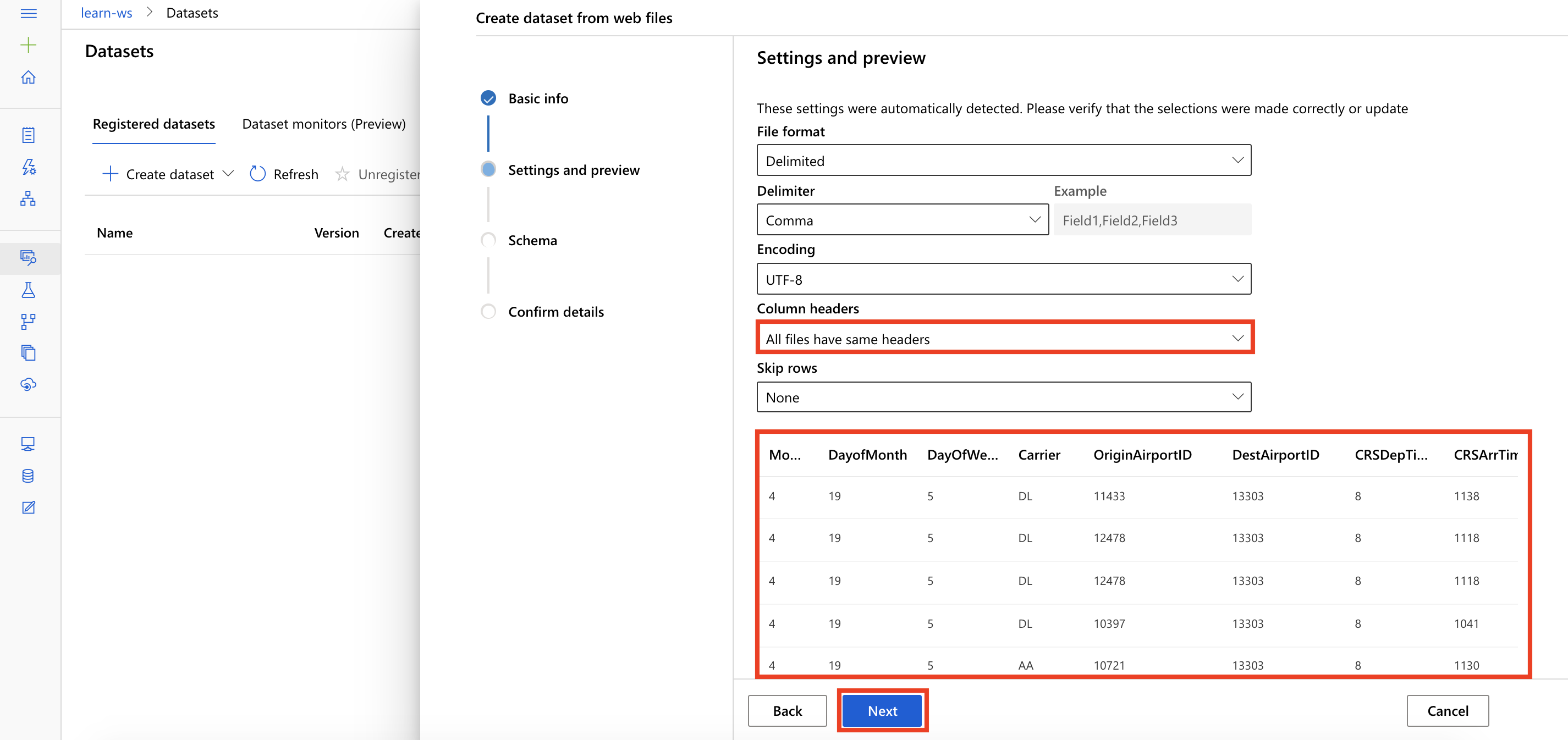
https://introtomlsampledata.blob.core.windows.net/data/flightdelays/flightdelays.csv

1. Provide flightdelays as the Name, leave the remaining values at their defaults and select **Next**.



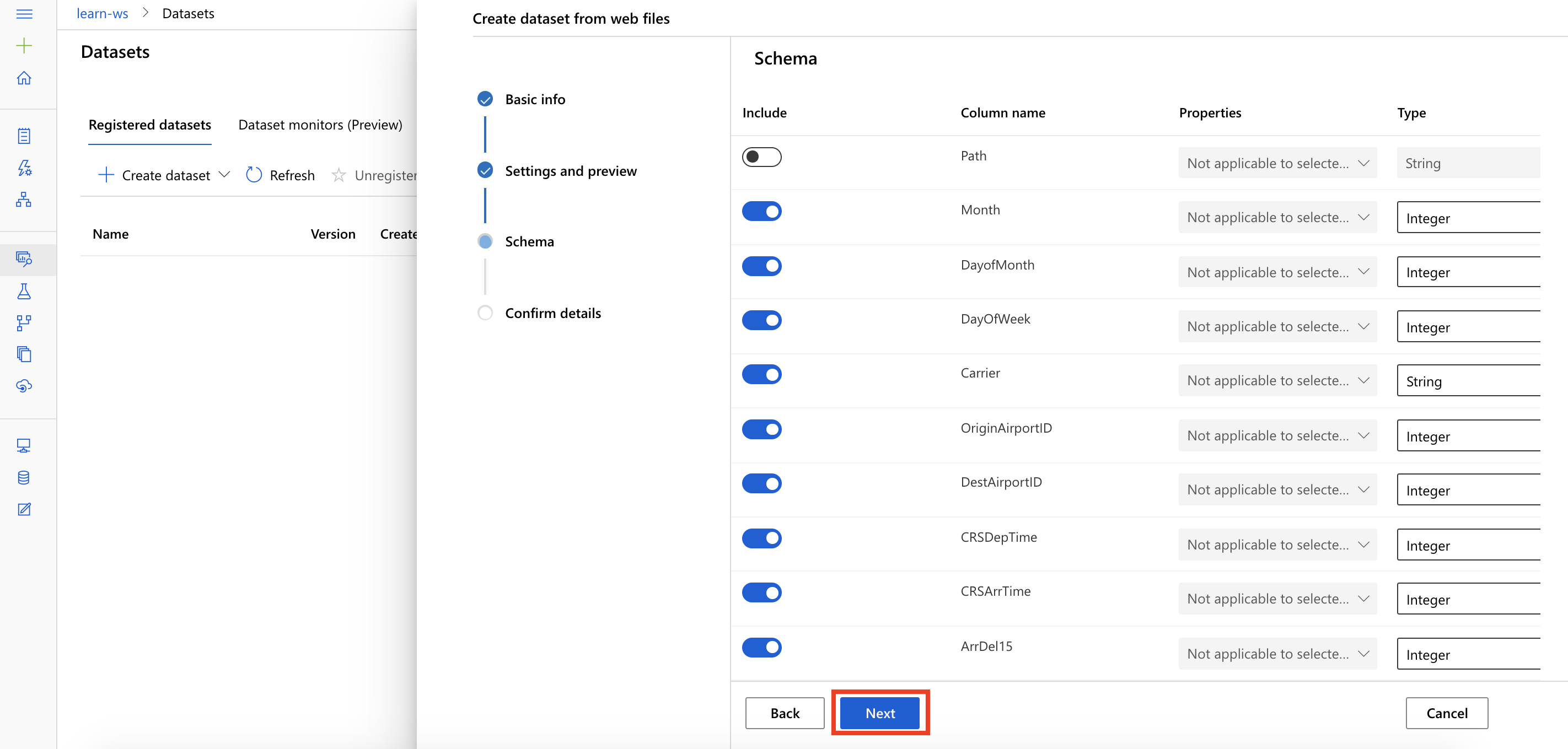
Task 2: Preview Dataset

1. On the Settings and preview panel, set the column headers drop down to All files have same headers.
2. Review the dataset and then select **Next**



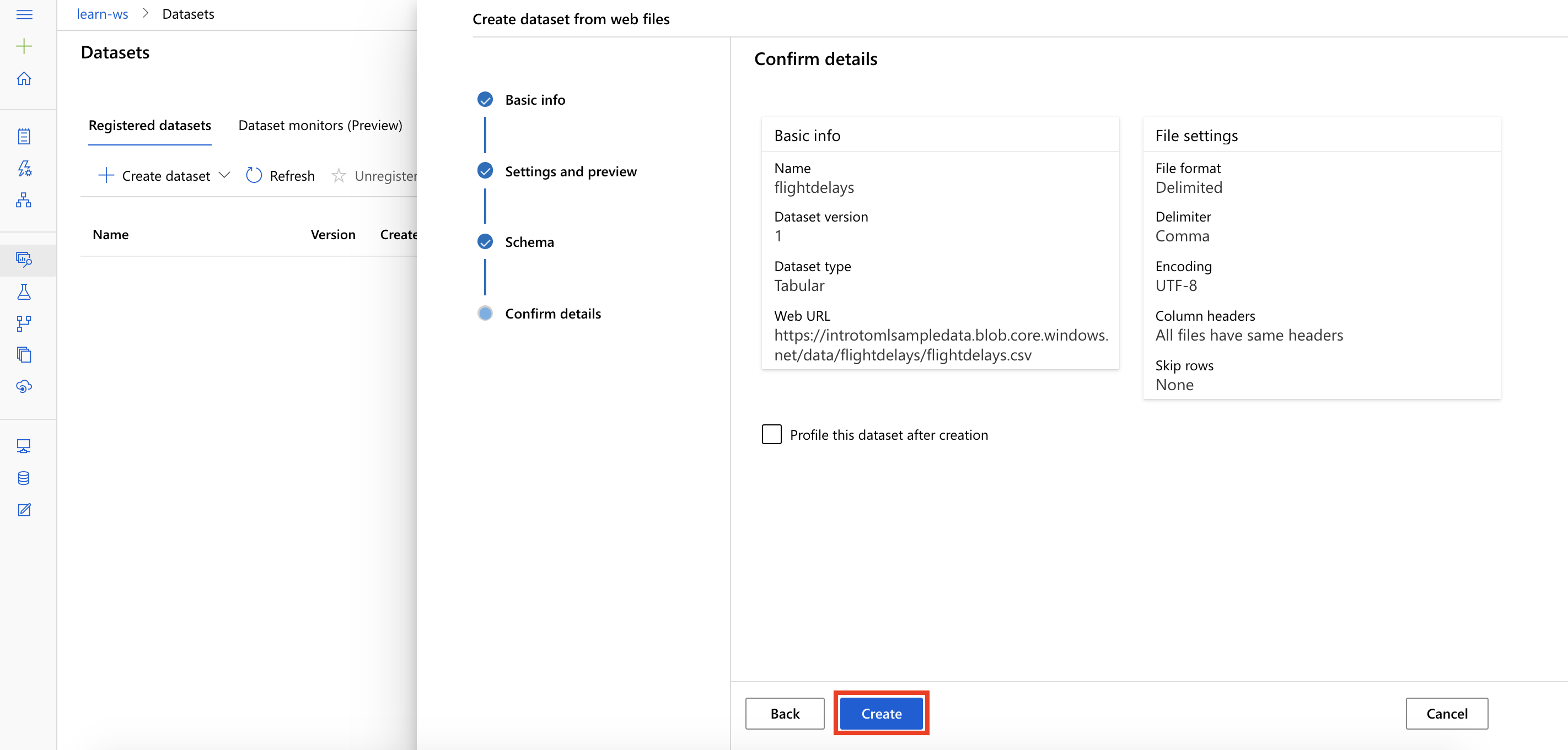
Task 3: Select Columns

1. Select columns from the dataset to include as part of your training data. Leave the default selections and select **Next**



Task 4: Create Dataset

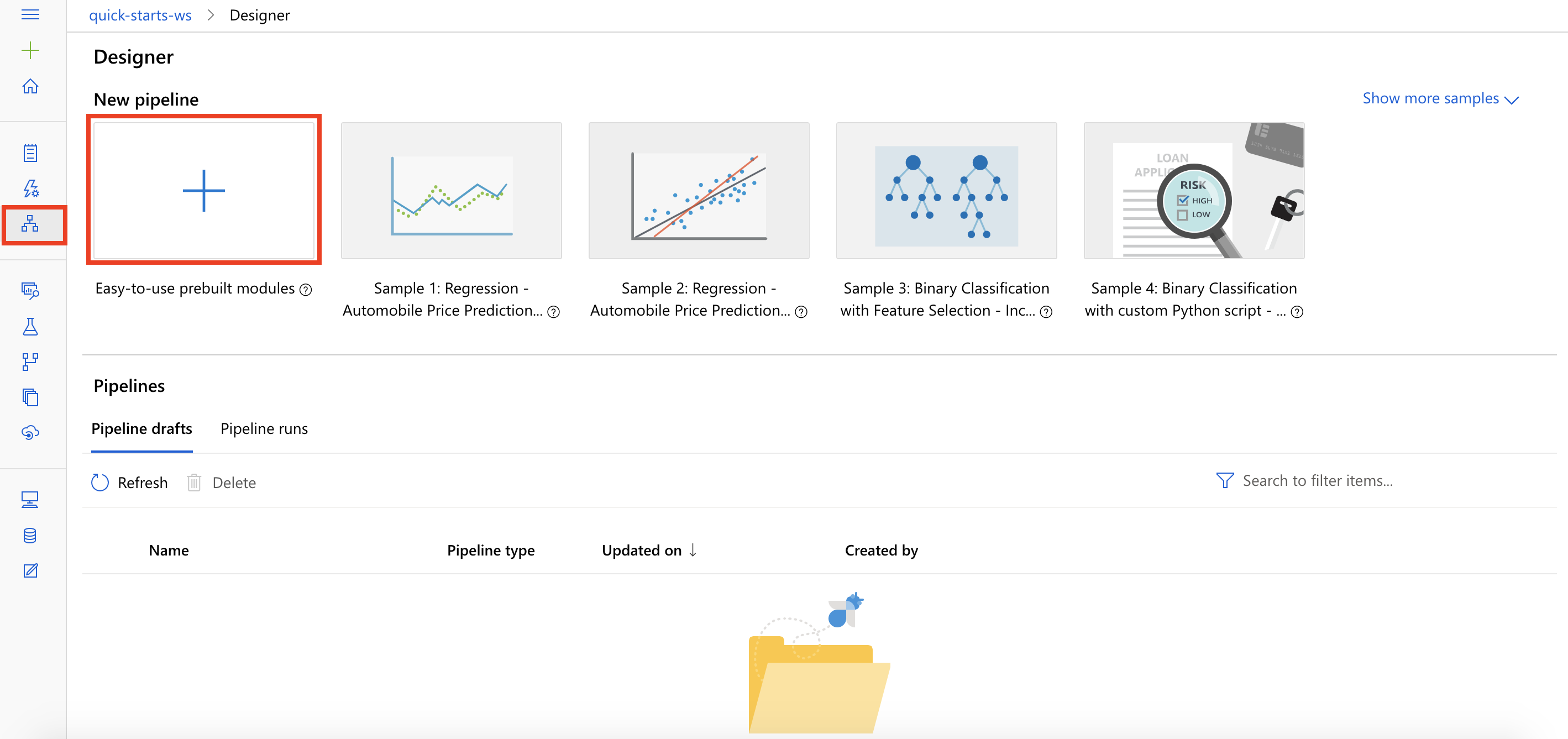
1. Confirm the dataset details and select **Create**



**Exercise 2: Create New Training Pipeline**

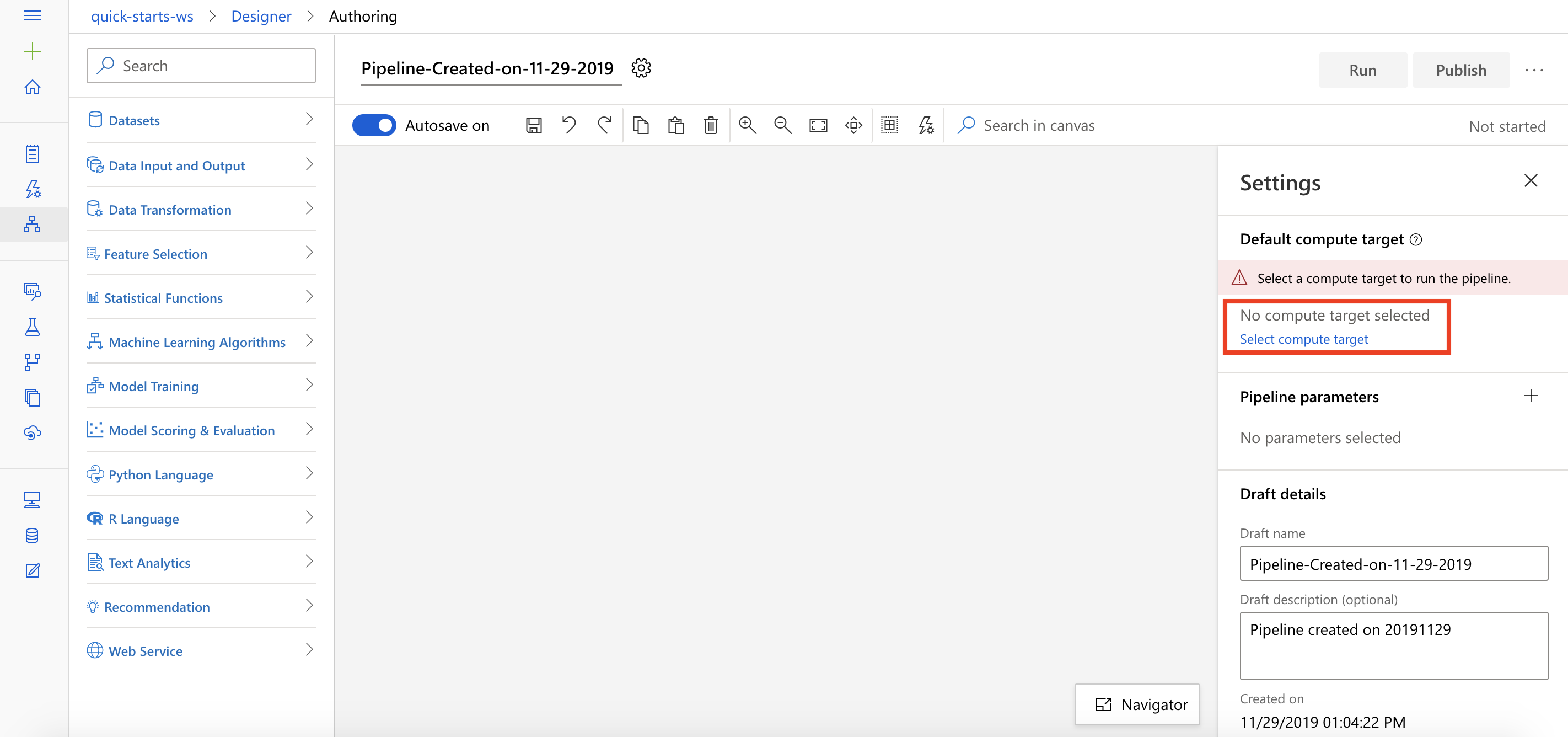
Task 1: Open Pipeline Authoring Editor

1. From the studio, select **Designer, +**. This will open a visual pipeline authoring editor.



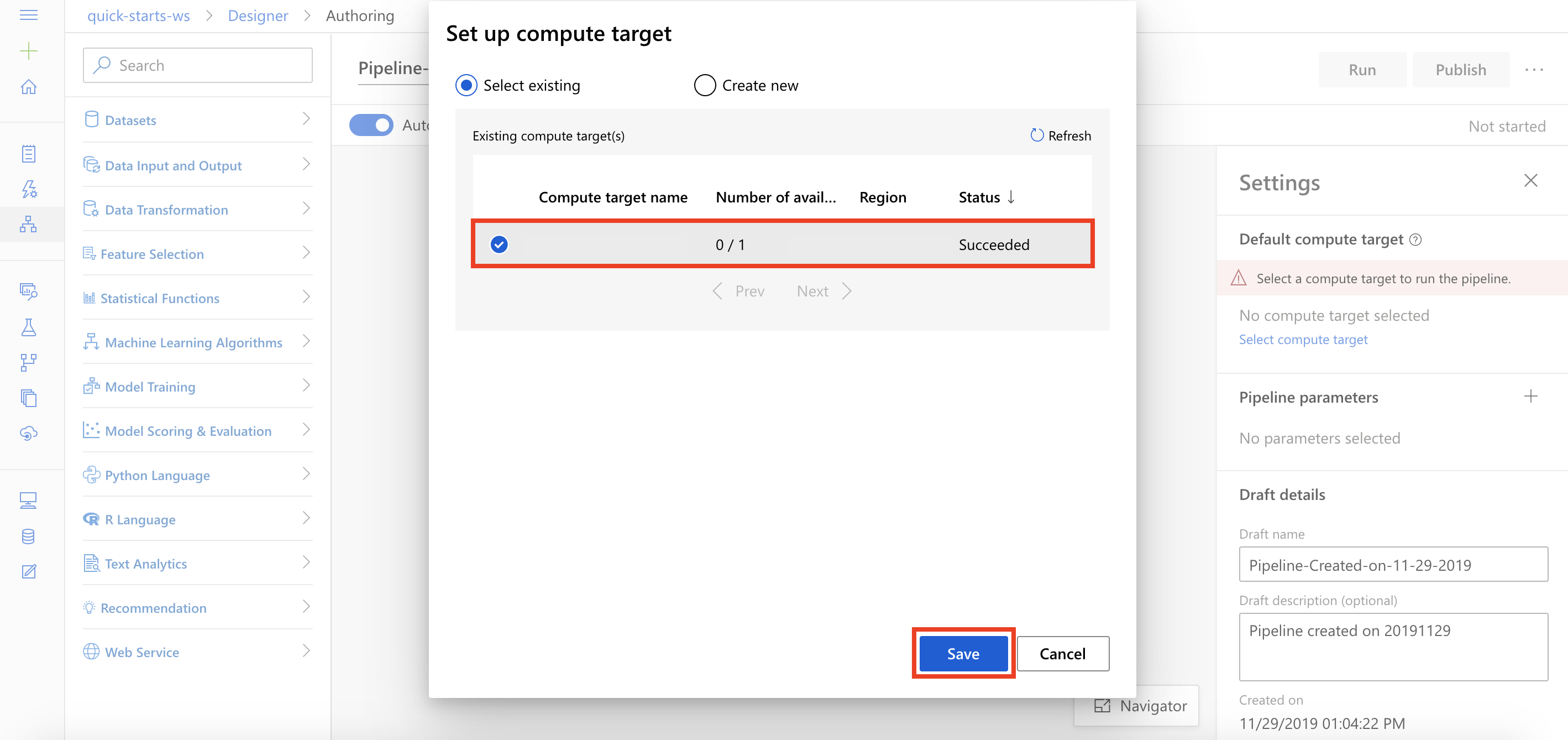
Task 2: Setup Compute Target

1. In the settings panel on the right, select **Select compute target**.



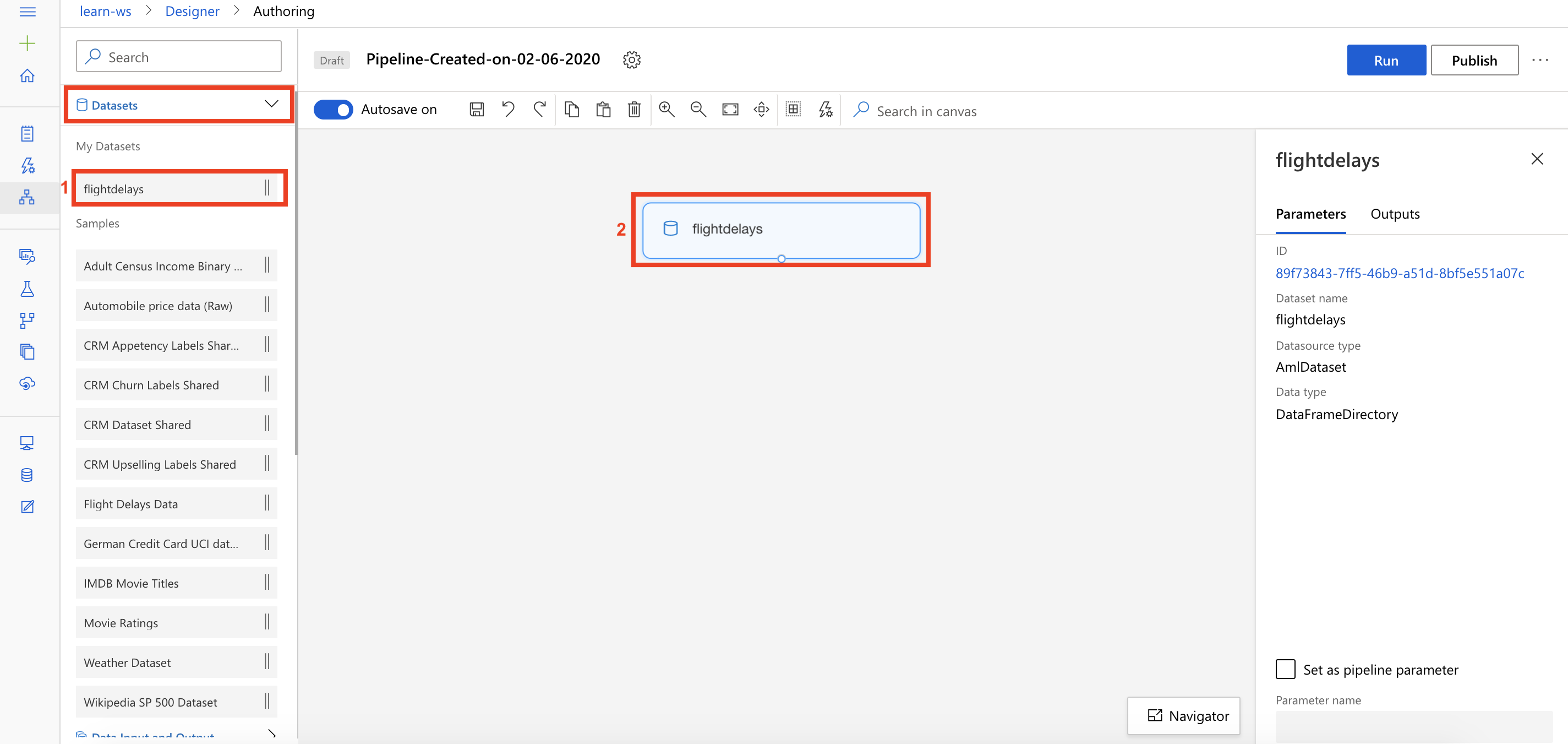
1. In the Set up compute target editor, select the available compute, and then select **Save**.

Note: If you are facing difficulties in accessing pop-up windows or buttons in the user interface, please refer to the Help section in the lab environment.



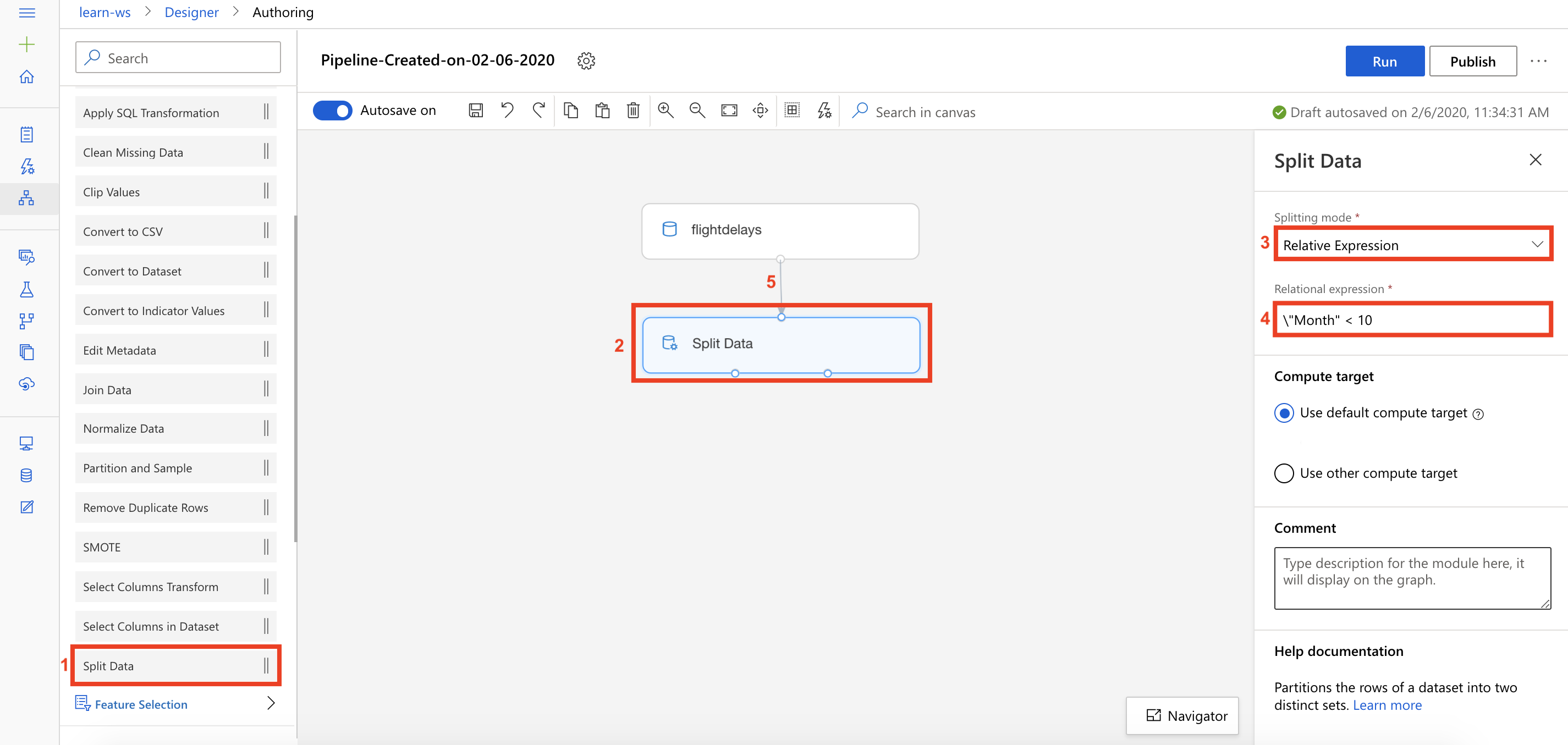
Task 3: Add Dataset

1. Select **Datasets** section in the left navigation. Next, select **My Datasets, flightdelays** and drag and drop the selected dataset on to the canvas.



Task 4: Split Dataset

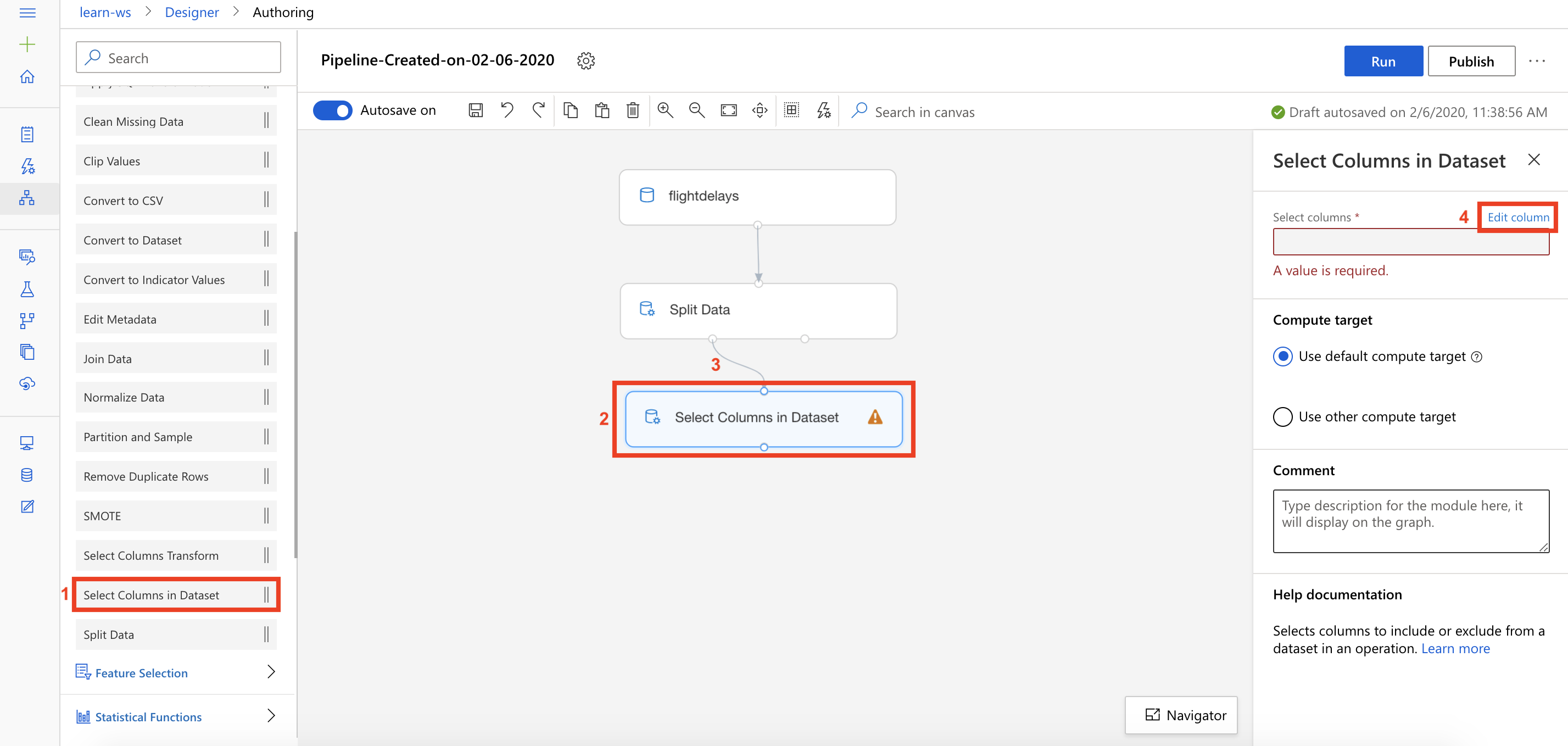
1. We will split the dataset such that months prior to October will be used for model training and months October to December will be used for model testing.
2. Select **Data Transformation** section in the left navigation. Follow the steps outlined below:
   1. Select the **Split Data** prebuilt module
   2. Drag and drop the selected module on to the canvas
   3. Splitting mode: **Relative Expression**
   4. Relational expression: **\”Month” < 10**
   5. Connect the Dataset to the Split Data module



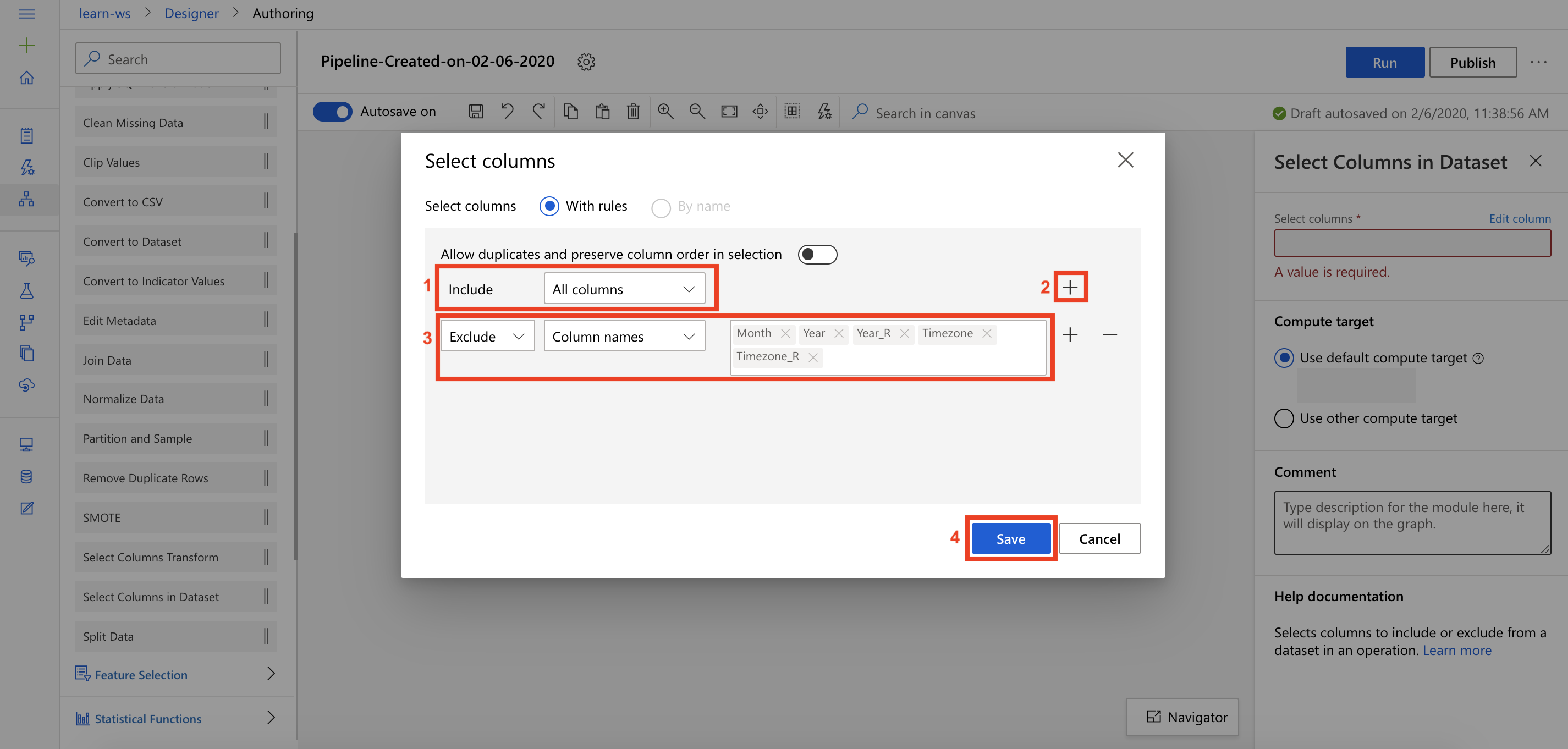
*Note that you can submit the pipeline at any point to peek at the outputs and activities. Running pipeline also generates metadata that is available for downstream activities such selecting column names from a list in selection dialogs.*

Task 5: Select Columns in Dataset

1. Select **Data Transformation** section in the left navigation. Follow the steps outlined below:
   1. Select the **Select Columns in Dataset** prebuilt module
   2. Drag and drop the selected module on to the canvas
   3. Connect the first output of the Split Data module to the Select Columns in Dataset module
   4. Select **Edit column** link to open the Select columns` editor

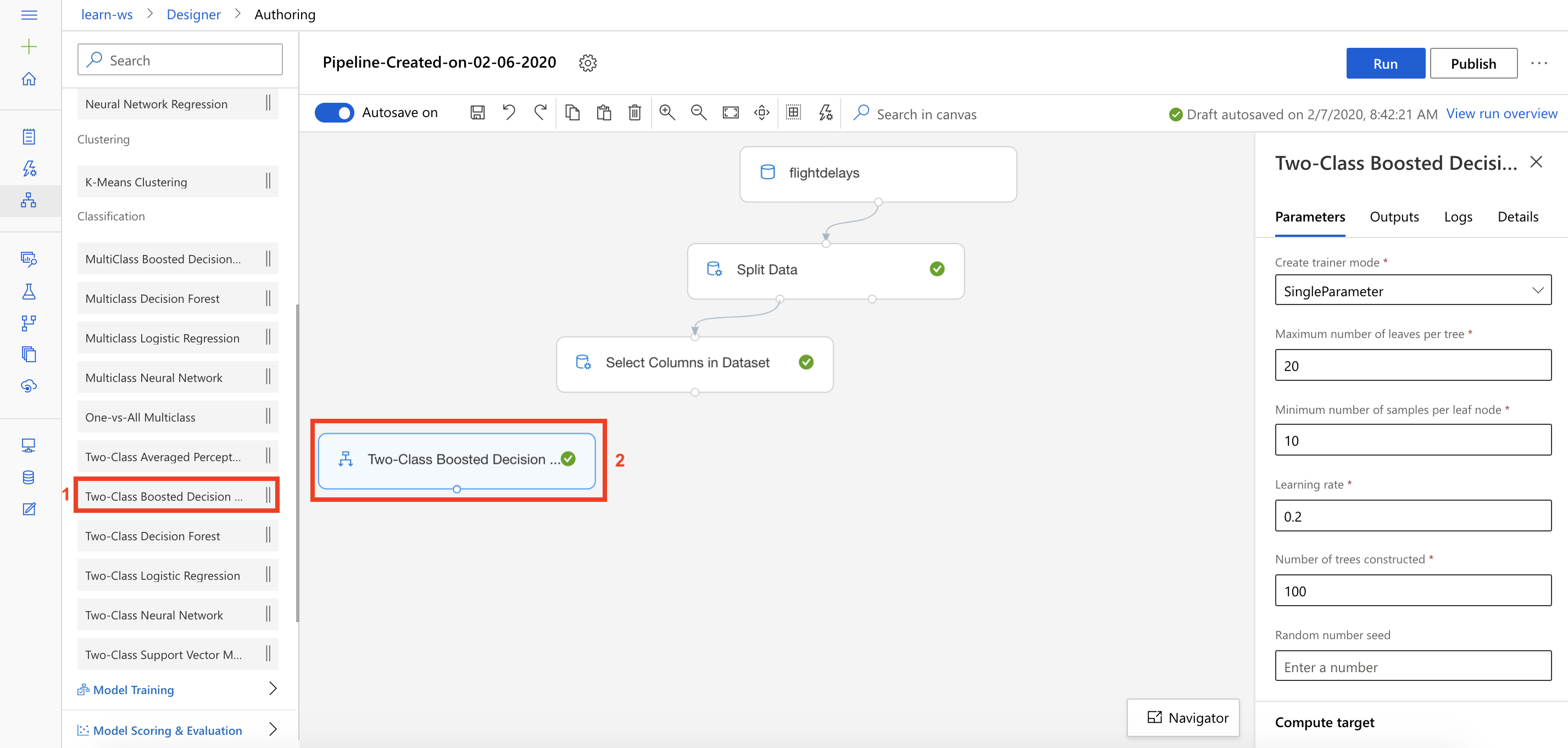


1. In the Select columns editor, follow the steps outlined below:
   1. Include: **All columns**
   2. Select **+**
   3. Exclude: **Column names**, provide the following column names to exclude: **Month, Year, Year\_R, Timezone, Timezone\_R**
   4. Select **Save**



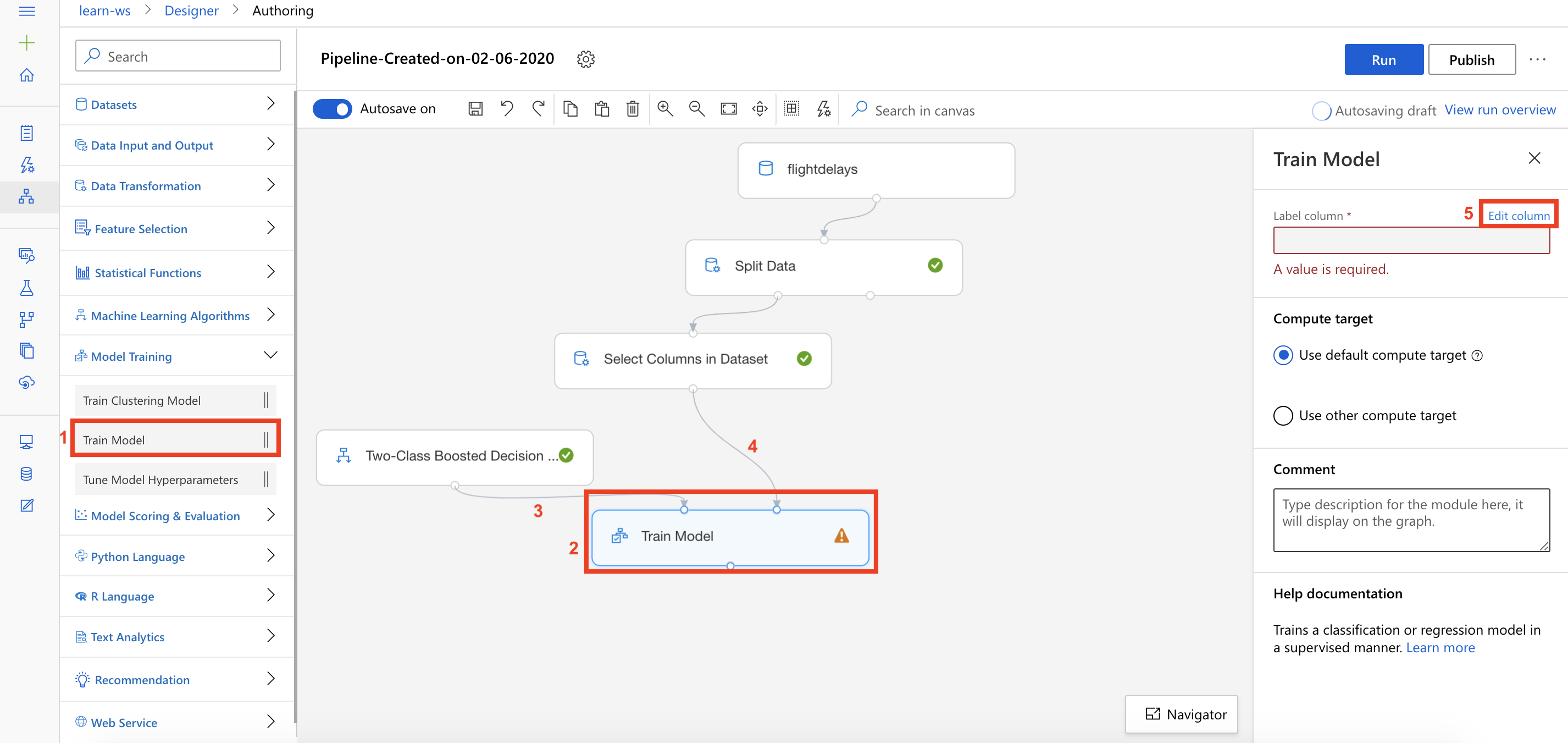
Task 6: Initialize Classification Model

1. Select **Machine Learning Algorithms** section in the left navigation. Follow the steps outlined below:
   1. Select the **Two-Class Boosted Decision Tree** prebuilt module
   2. Drag and drop the selected module on to the canvas

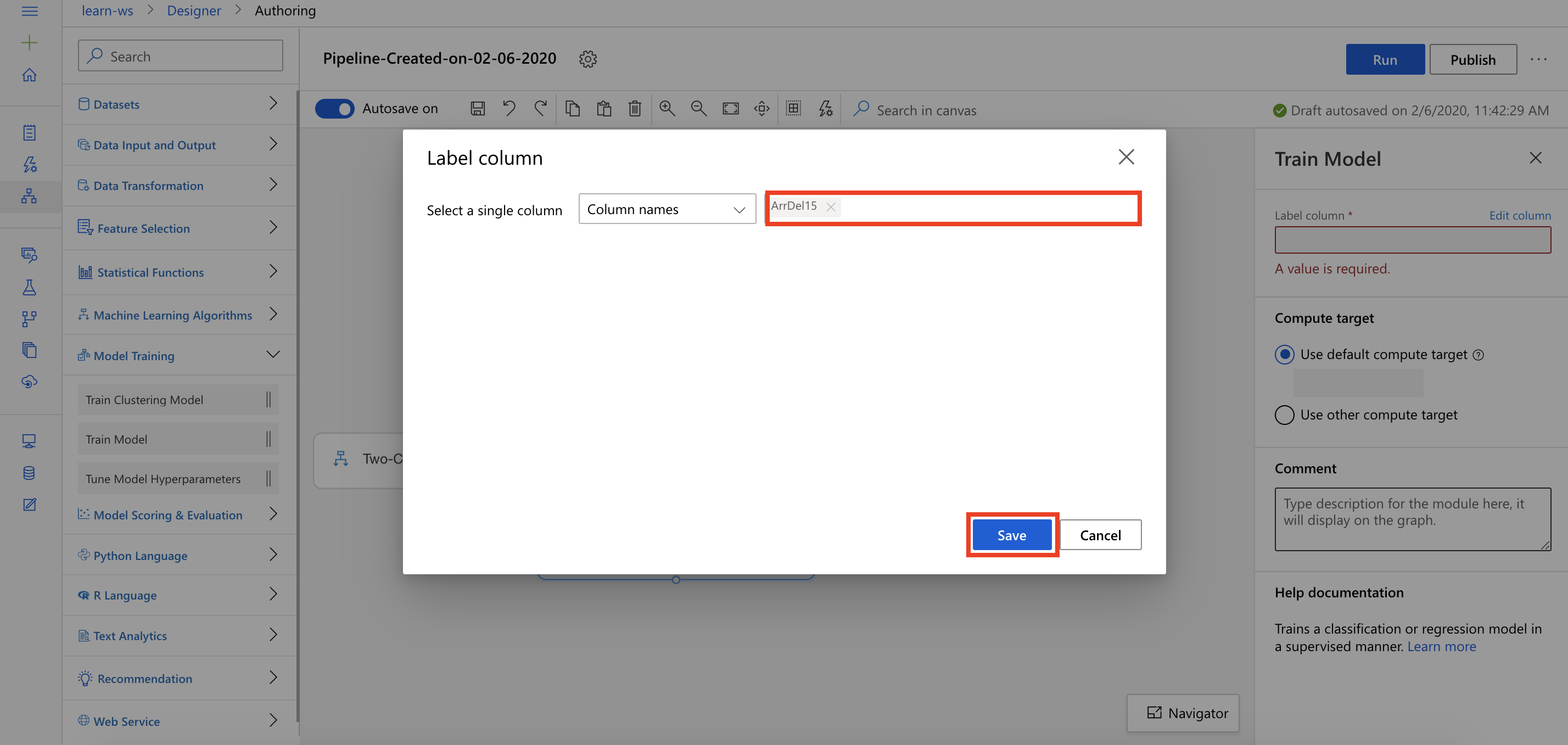


Task 7: Setup Train Model Module

1. Select **Model Training** section in the left navigation. Follow the steps outlined below:
   1. Select the **Train Model** prebuilt module
   2. Drag and drop the selected module on to the canvas
   3. Connect the Two-Class Boosted Decision Tree module to the first input of the Train Model module
   4. Connect the Select Columns in Dataset module to the second input of the Train Model module
   5. Select the **Edit column** link to open the Label column editor

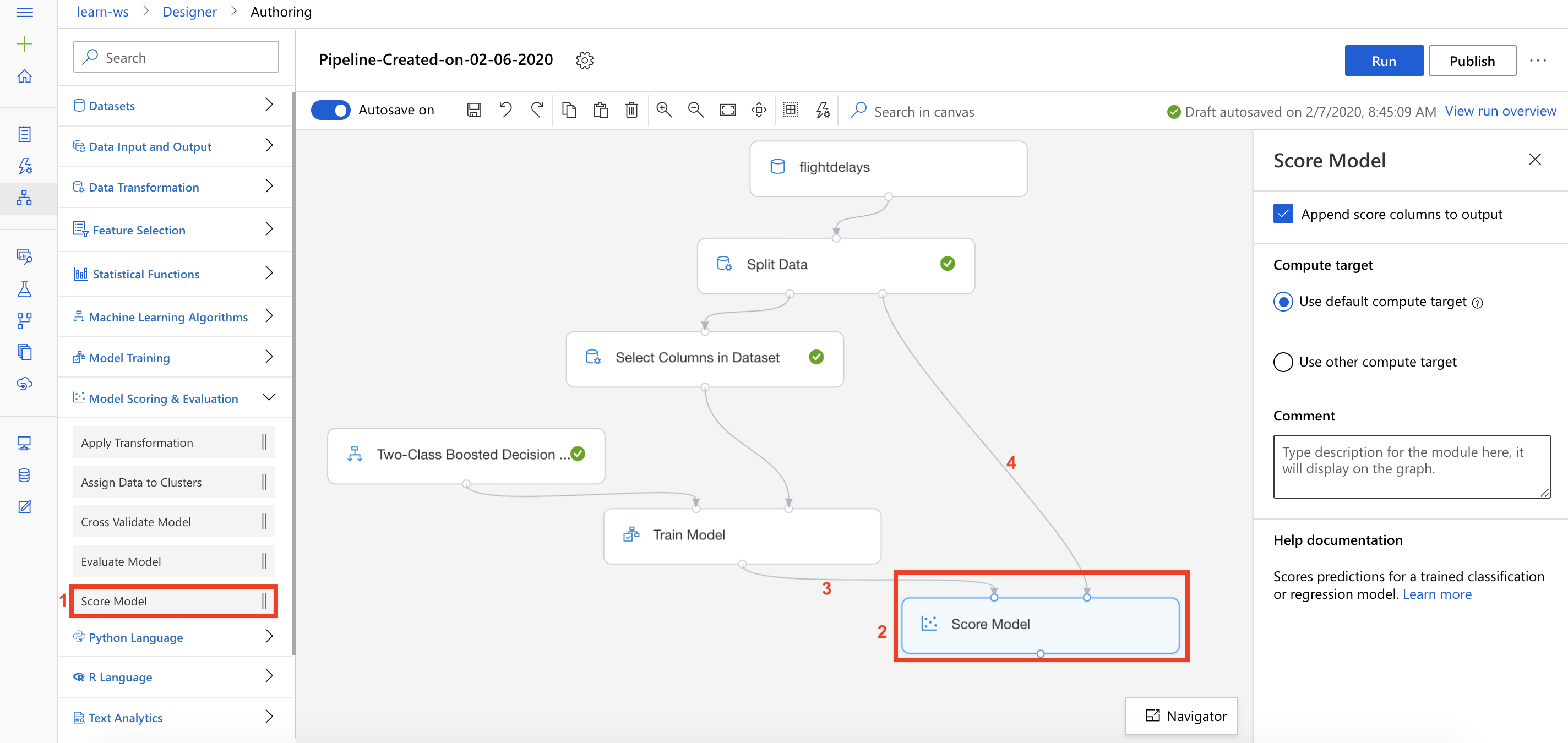


1. The Label column editor allows you to specify your Label or Target column. Type in the label column name **ArrDel15** and then select **Save**.



Task 8: Setup Score Model Module

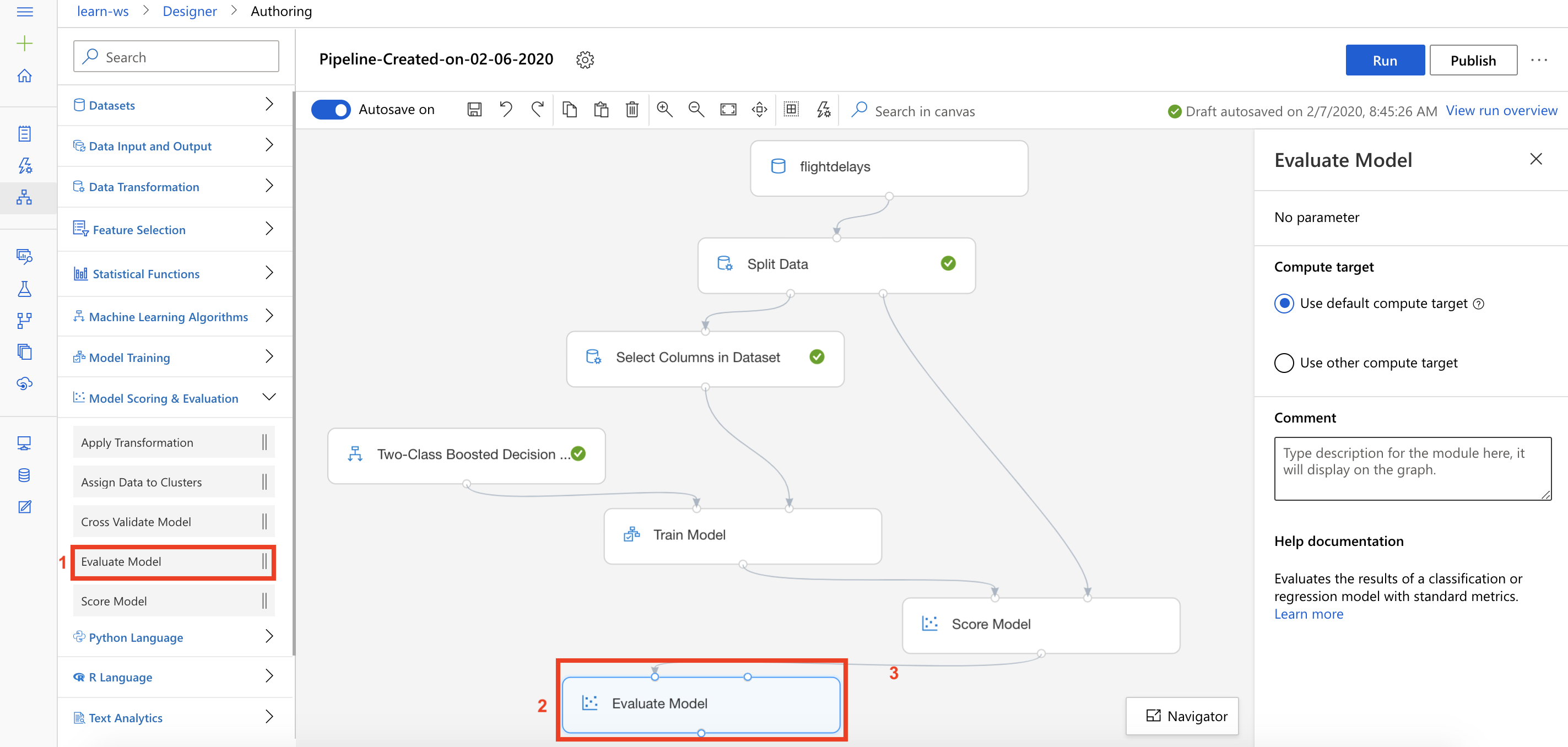
1. Select **Model Scoring & Evaluation** section in the left navigation. Follow the steps outlined below:
   1. Select the **Score Model** prebuilt module
   2. Drag and drop the selected module on to the canvas
   3. Connect the Train Model module to the first input of the Score Model module
   4. Connect the second output of the Split Data module to the second input of the Score Model module



*Note that Split Data module will feed data for both model training and model scoring.*

Task 9: Setup Evaluate Model Module

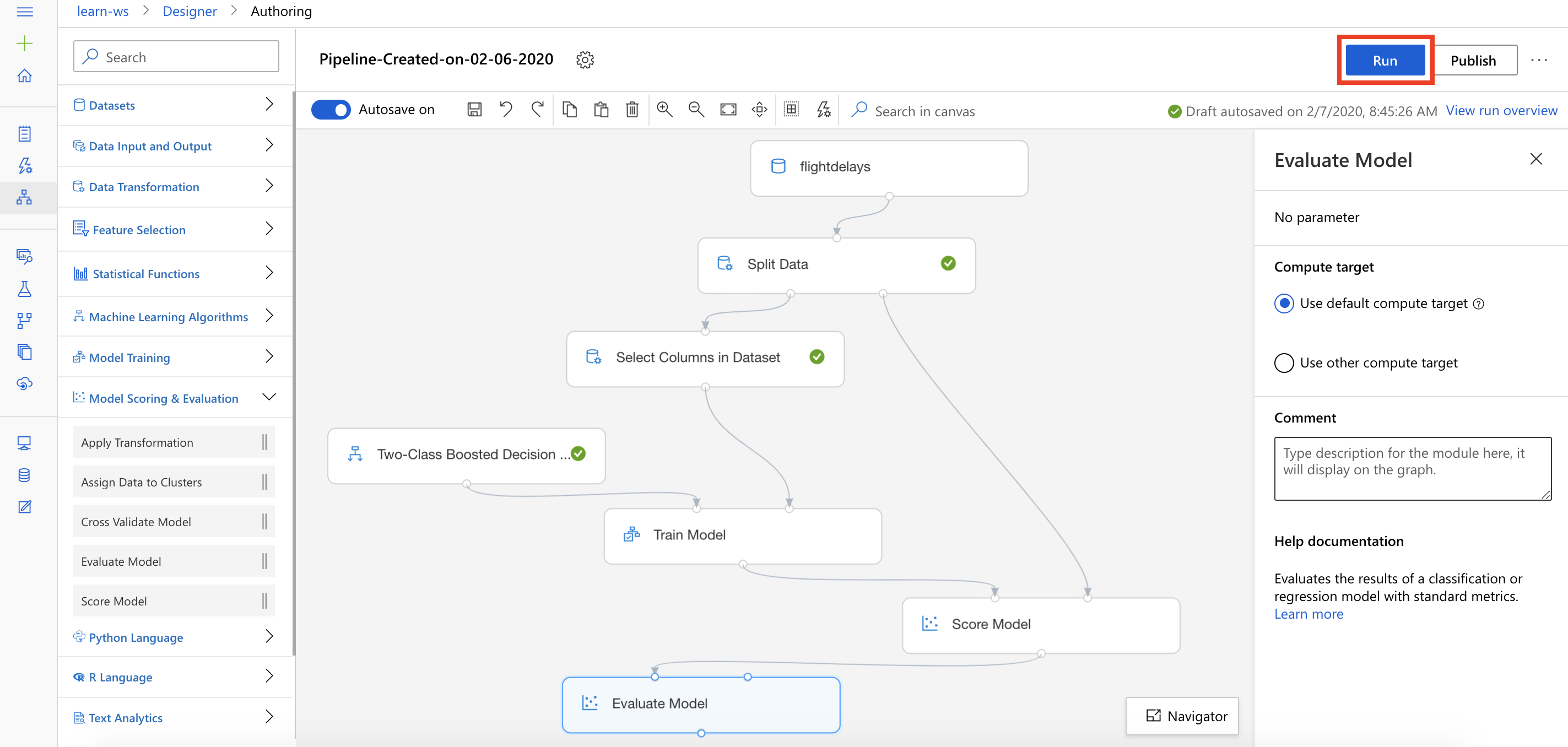
1. Select **Model Scoring & Evaluation** section in the left navigation. Follow the steps outlined below:
   1. Select the **Evaluate Model** prebuilt module
   2. Drag and drop the selected module on to the canvas
   3. Connect the Score Model module to the first input of the Evaluate Model module



**Exercise 3: Submit Training Pipeline**

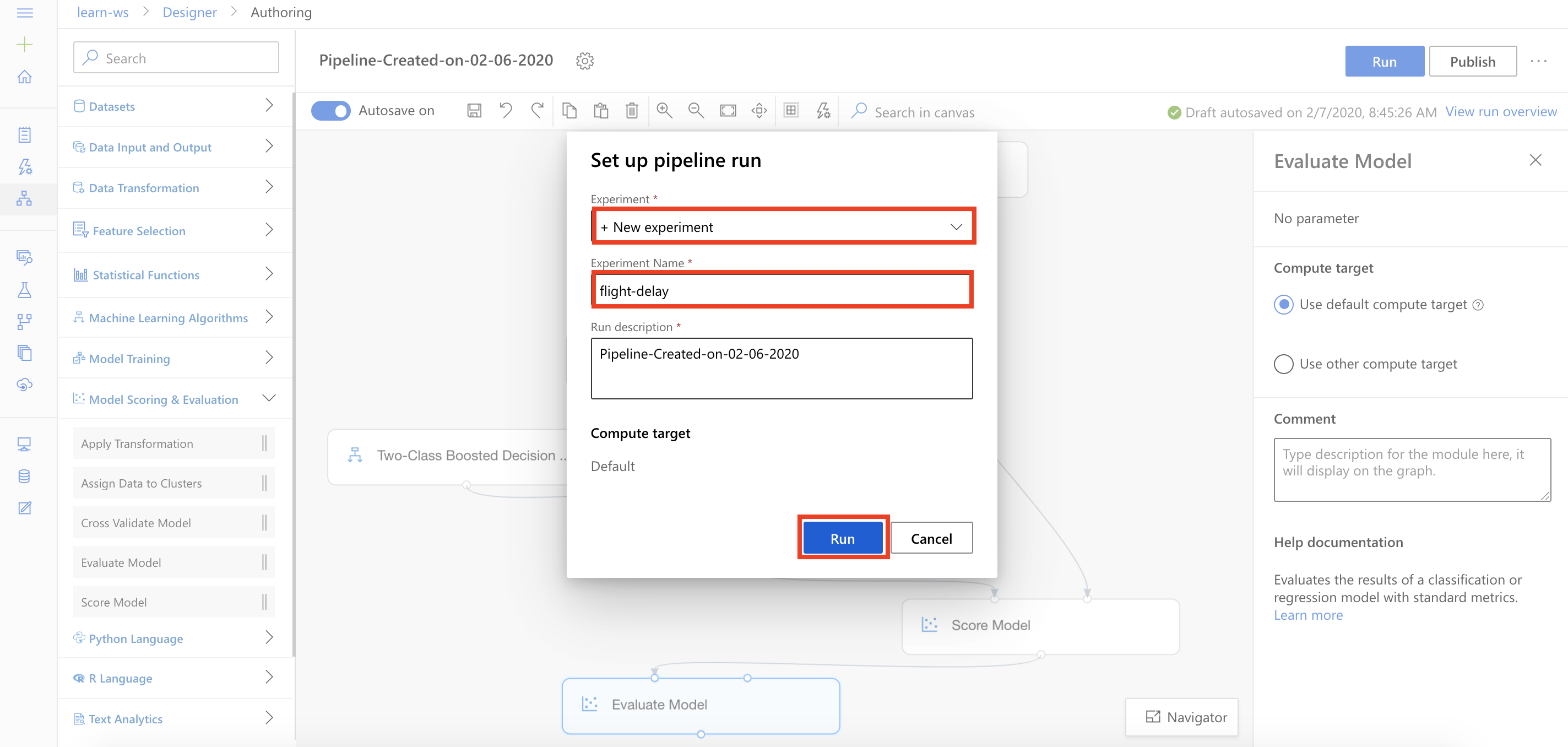
Task 1: Create Experiment and Submit Pipeline

1. Select **Submit** to open the Setup pipeline run editor.



Please note that the button name in the UI is changed from **Run** to **Submit**.

1. In the Setup pipeline run editor, select **Experiment, Create new** and provide New experiment name: **flight-delay**, and then select **Submit**.

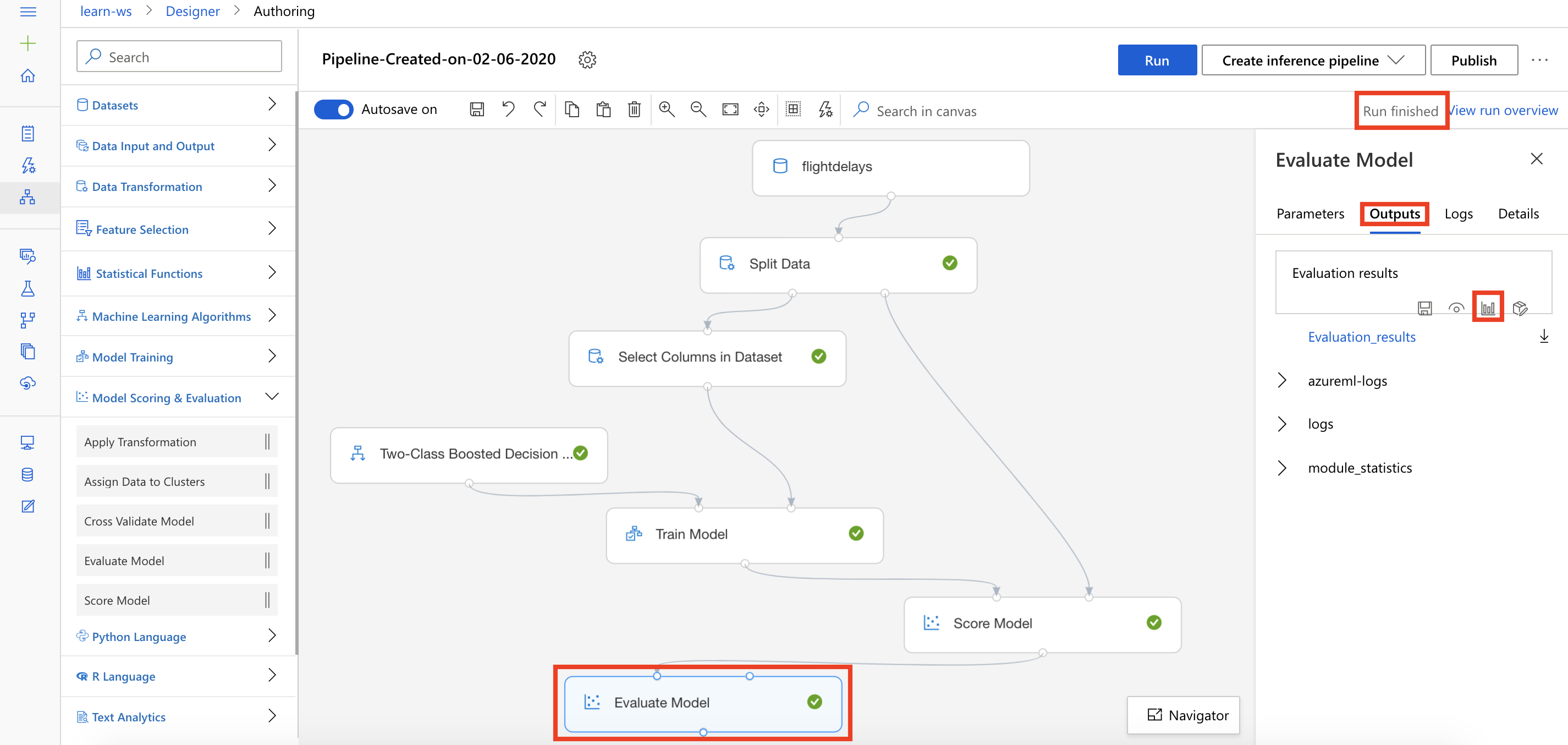


1. Wait for pipeline run to complete. It will take around **10 minutes** to complete the run.
2. While you wait for the model training to complete, you can learn more about the classification algorithm used in this lab by selecting [Two-Class Boosted Decision Tree](https://docs.microsoft.com/en-us/azure/machine-learning/algorithm-module-reference/two-class-boosted-decision-tree).

**Exercise 4: Visualize the Evaluation Results**

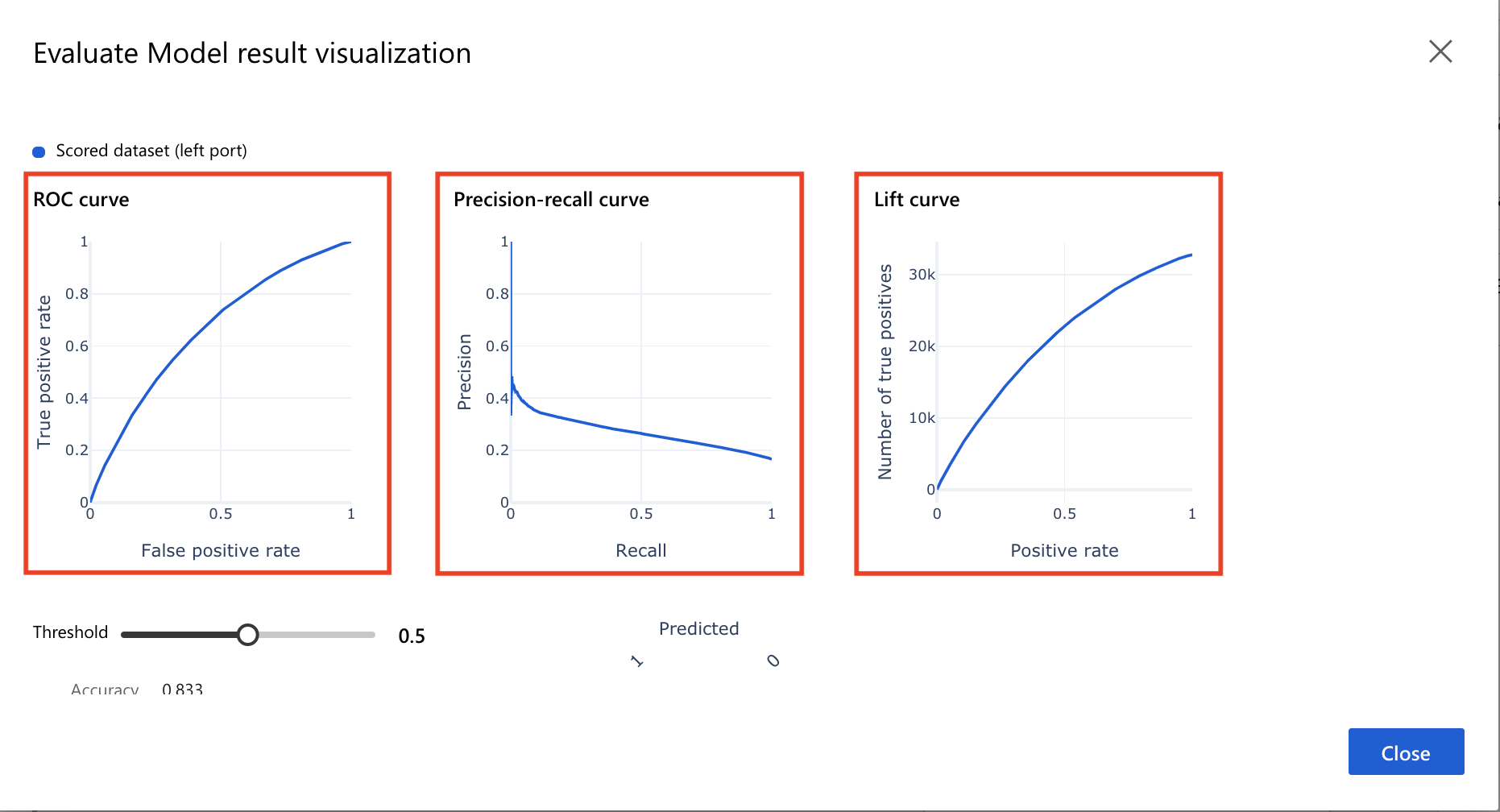
Task 1: Open the Result Visualization Dialog

1. Select **Evaluate Model, Outputs, Visualize** to open the Evaluate Model result visualization dialog.



Task 2: Evaluate Model Performance

1. Evaluate the model performance by reviewing the various evaluation curves, such as **ROC curve**, **Precision-recall curve**, and **Lift curve**.



1. Scroll down to review the following:
   1. Review the key metrics for classifiers: **Accuracy, Precision, Recall, F1 Score, and AUC**
   2. Review the binary classifier’s **Confusion Matrix**



**Next Steps**

Congratulations! You have trained and evaluated your first ensemble machine learning model. You can continue to experiment in the environment but are free to close the lab environment tab and return to the Udacity portal to continue with the lesson.