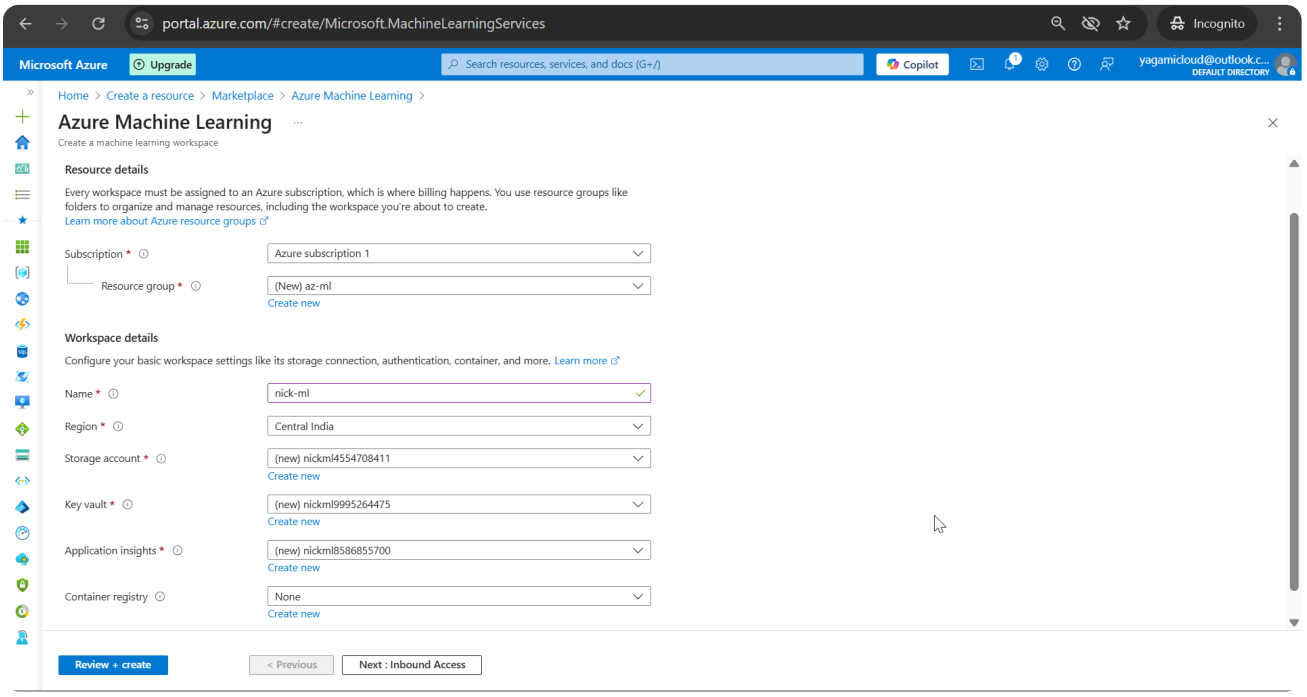
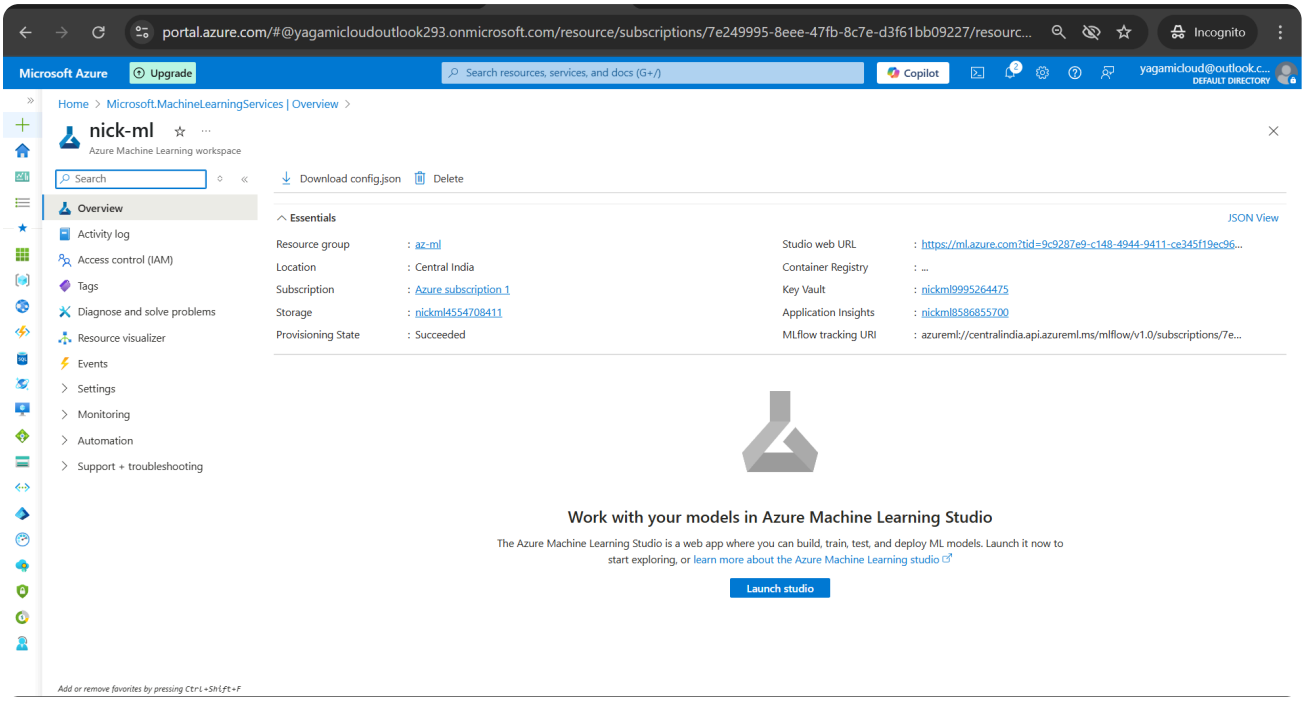
01 binary-classification Azure ML Designer

1. go to marketplace/azure machine learning



2.Review and create... and then go to the resource



3. Click on launch studio

4. Explore the UI

5...

Building a Classification Machine Learning Pipeline

The classification model is used to predict values that fall into simple categories.

We are going to use Azure Machine Learning Studio to build a Machine Learning model.

We will use a pre-built data set that is already available in ML studio - adult_census_income_binary_classification

Rows 32,561 Columns 15

age	workclass	fnlwgt	education	education-num	marital-status	occupation	relationship	race	sex	capital-gain	capital-loss	hours-per-week	native-country	income
39	State-gov	77516	Bachelors	13	Never-married	Adm-clerical	Not-in-family	White	Male	2174	0	40	United-States	<=50K
50	Self-emp-not-inc	83311	Bachelors	13	Married-civ-spouse	Exec-managerial	Husband	White	Male	0	0	13	United-States	<=50K
38	Private	215646	HS-grad	9	Divorced	Handlers-cleaners	Not-in-family	White	Male	0	0	40	United-States	<=50K
53	Private	234721	11th	7	Married-civ-spouse	Handlers-cleaners	Husband	Black	Male	0	0	40	United-States	<=50K
28	Private	338409	Bachelors	13	Married-civ-spouse	Prof-specialty	Wife	Black	Female	0	0	40	Cuba	<=50K
37	Private	284582	Masters	14	Married-civ-spouse	Exec-managerial	Wife	White	Female	0	0	40	United-States	<=50K
49	Private	160187	9th	5	Married-spouse-	Other-service	Not-in-	Black	Female	0	0	16	Jamaica	<=50K

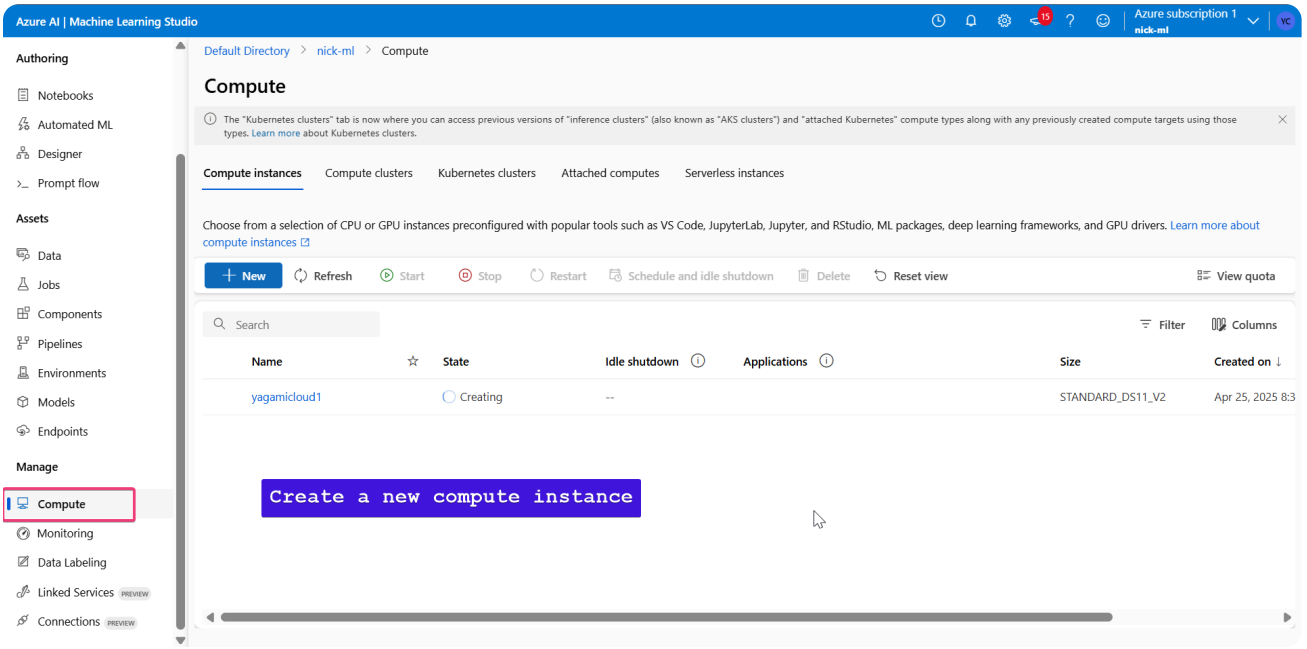
The model will take in a set of features and the label will be the income.

We are going to create a pipeline to create the Machine Learning model.

An Azure Machine Learning pipeline is a workflow that is used to execute a machine learning task.

We can develop the pipeline in the designer of the Azure Machine Learning studio. We can drag data sets and components onto the canvas for our pipeline.

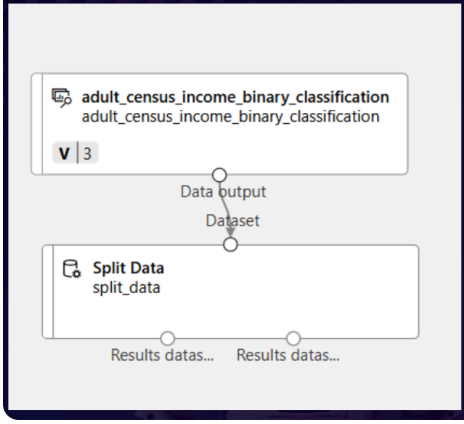
6. Compute instance



6. Run the pipeline and check the splits

7.

Split data



- Our aim is to build a machine learning model. This model will be used to classify the income of individuals into 2 classes - <=50K or > 50K
- The first step was to get the data set and split the set into a training and test set. The training set would have 70% of the data and the test set would have 30% of the data.

8.

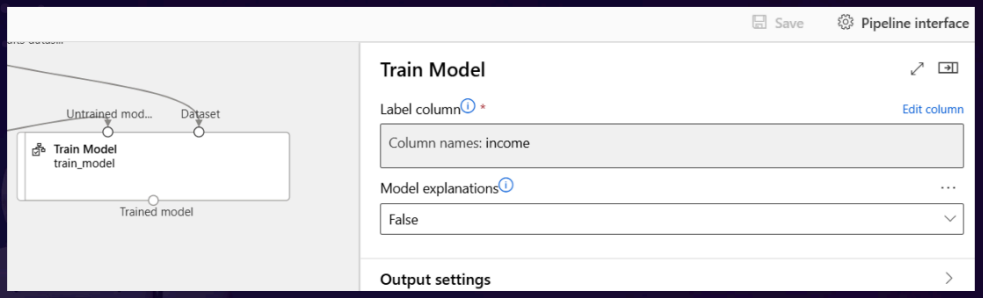
Train Model

- Next we need to add the component to train the model. We will add as the input the 70% of the split data phase.
- We will add the algorithm which will be used to train the machine learning model.

9.

Train Model

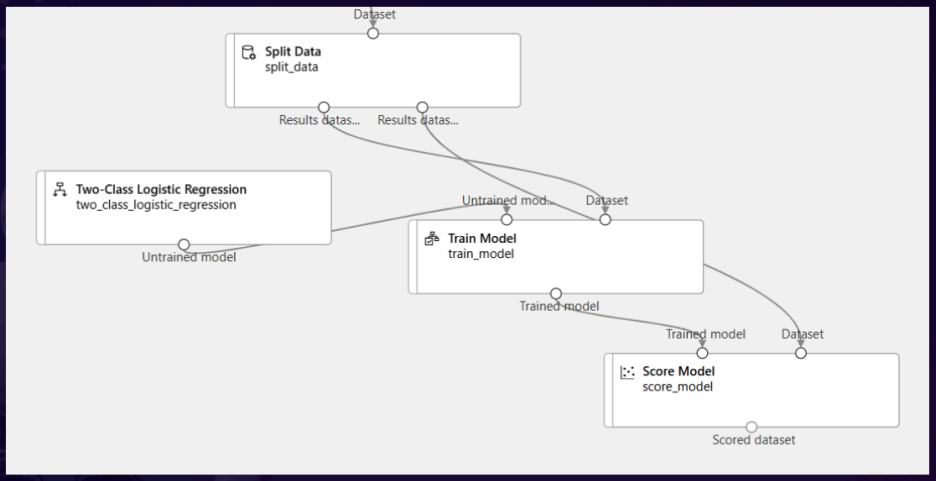
- We will also ensure to train the model based on the income column. This is our label.



10

Score Model

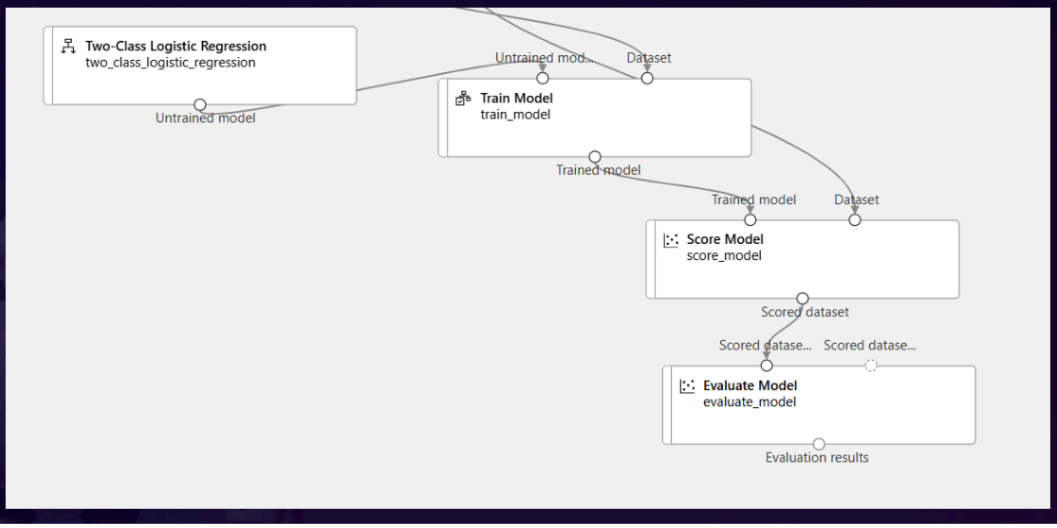
- We can now score the model based on the 30% test data. This will help us understand how efficient the model is when it comes to predicting the income value.



11.

Evaluate Model

- Based on the scoring and the test data, we can finally evaluate the model.

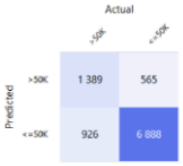


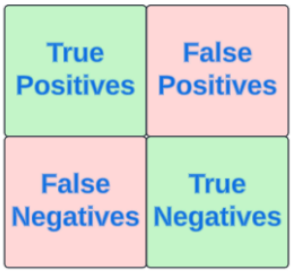
12.

Two-class Logistic regression

Confusion Matrix

This table gives us a summary of the results when it comes to what the model was supposed to predict and the actual prediction.





Accuracy	0.847
Precision	0.711
Recall	0.6
F1 Score	0.651
AUC	0.901

Accuracy - This tells how often the classifier is right in predicting results.

$$\frac{\text{True Positives + True Negatives}}{\text{True Positives + True Negatives + False Positives + False Negatives}}$$

Precision - This tells to what extend does the model accurately predict results.