Arsh Vohra

Azure ML assignment

Steps:

Launching ML workspace



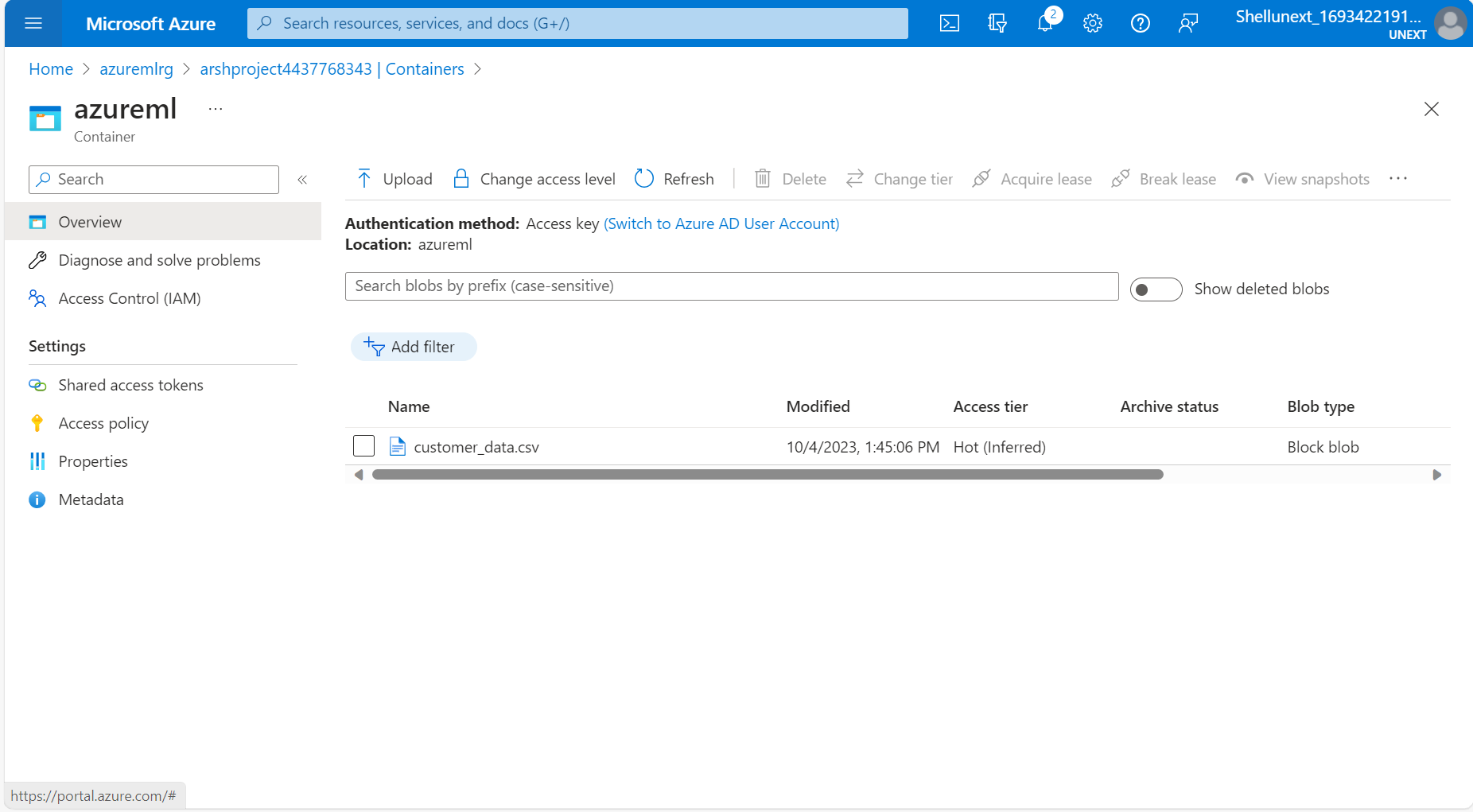
A screenshot of a computer

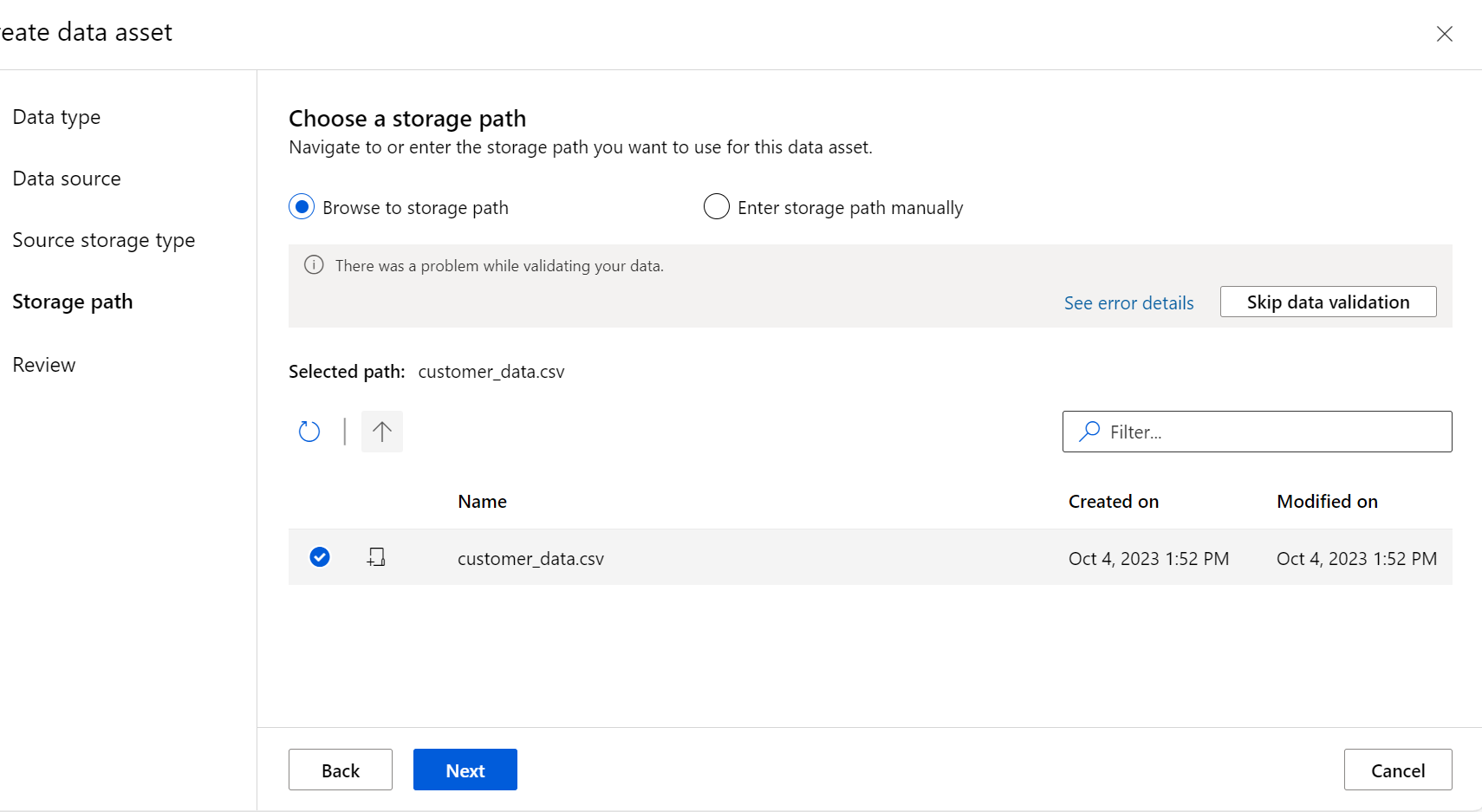
Description automatically generated

**Data Preparation**

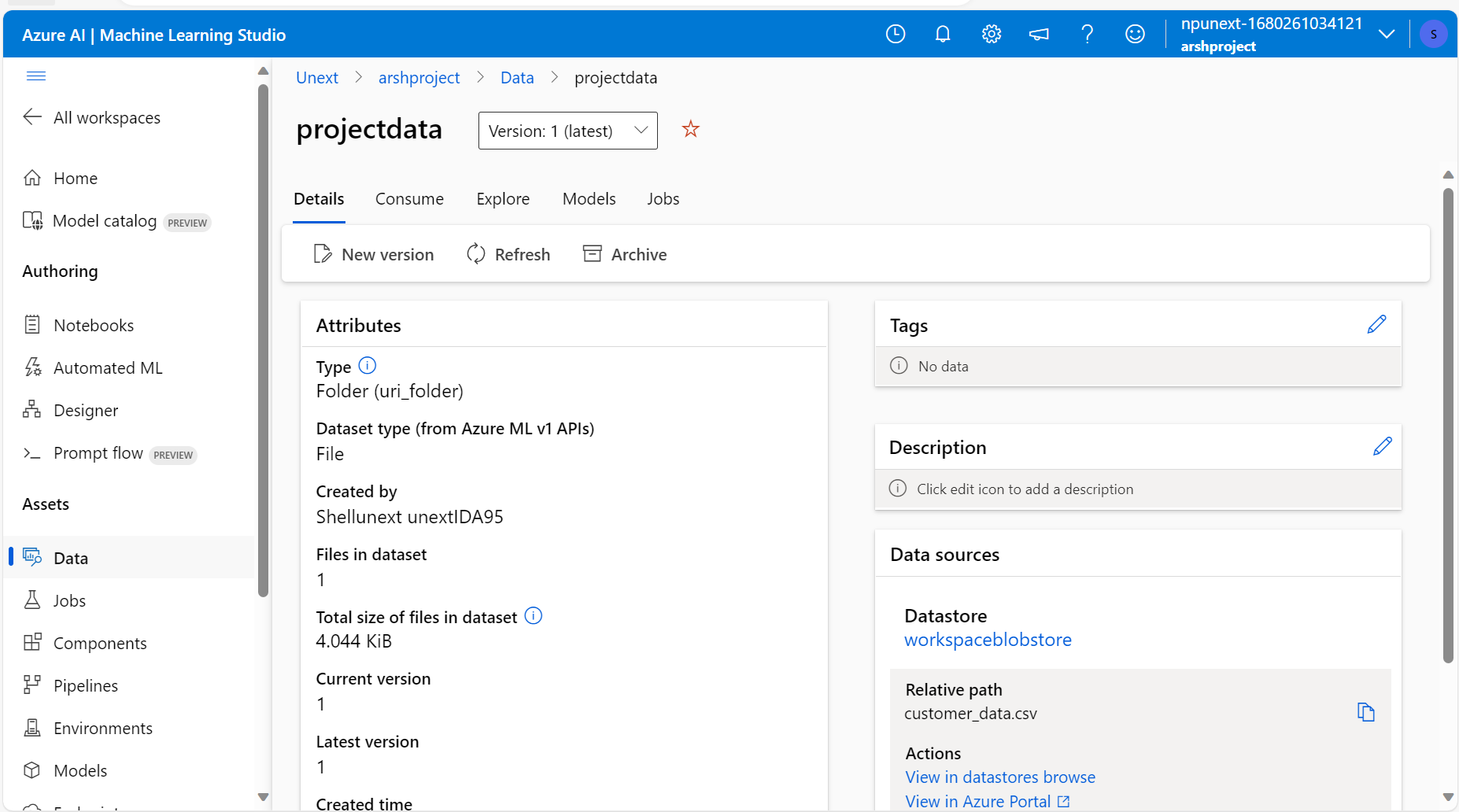
1. Uploading csv file in blob storage account







Data asset created



1. Preparing the data

Cleaning missing data

A screenshot of a computer

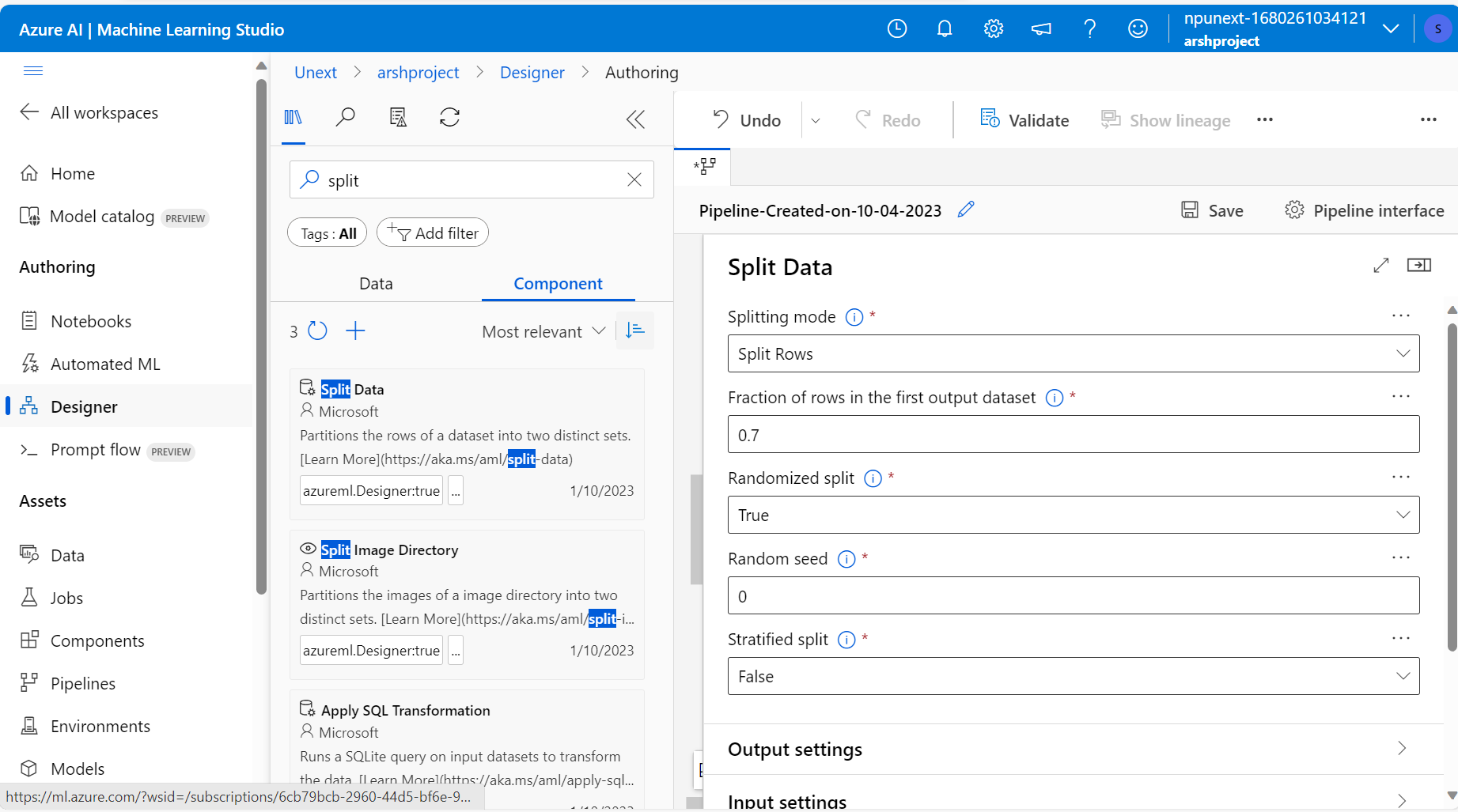
Description automatically generated

A screenshot of a computer

Description automatically generated

**Model Development**

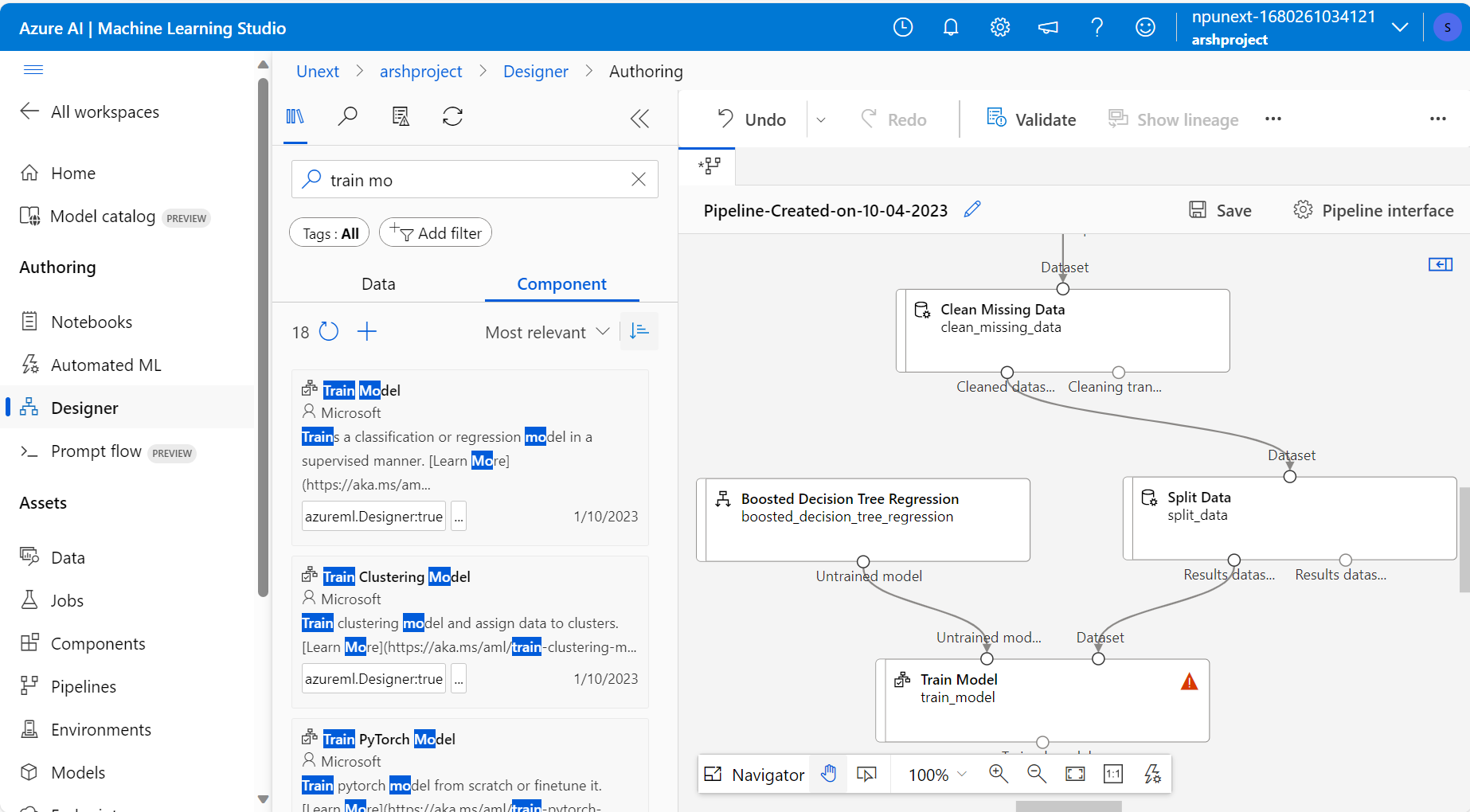
Splitting data into train and test (70/30)



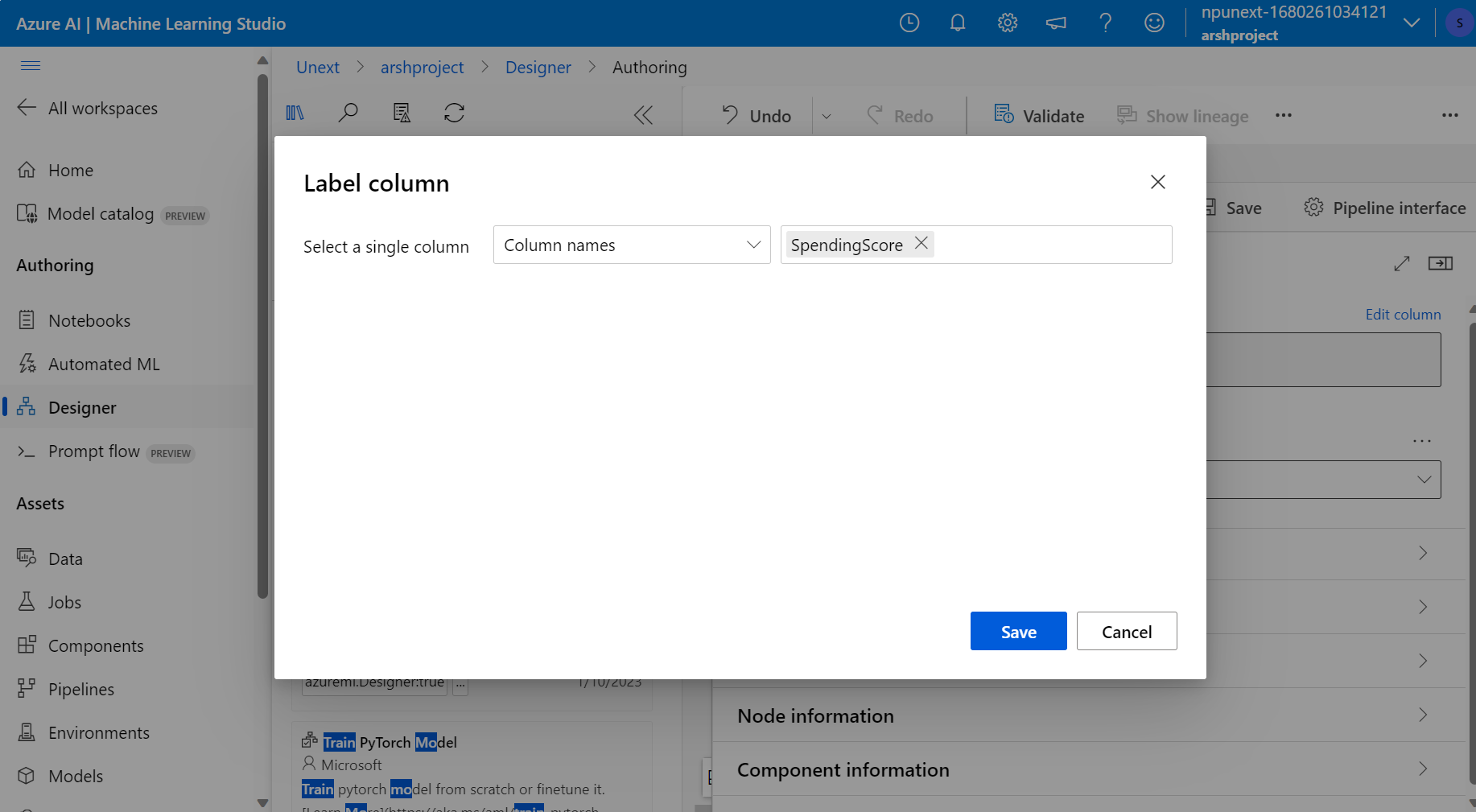
A screenshot of a computer

Description automatically generated

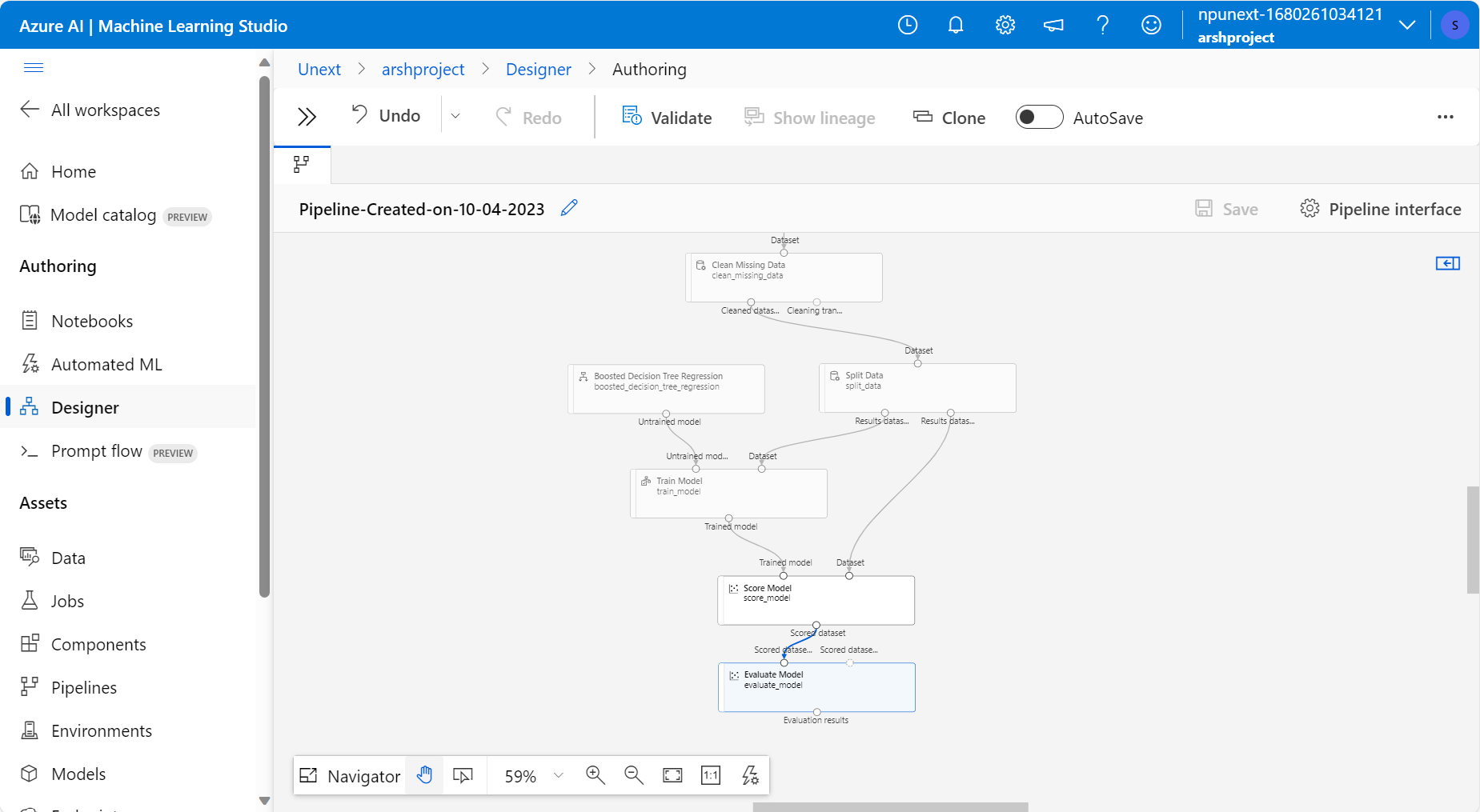
Using regression model



Training the model on selected SpendingScore feature from test data



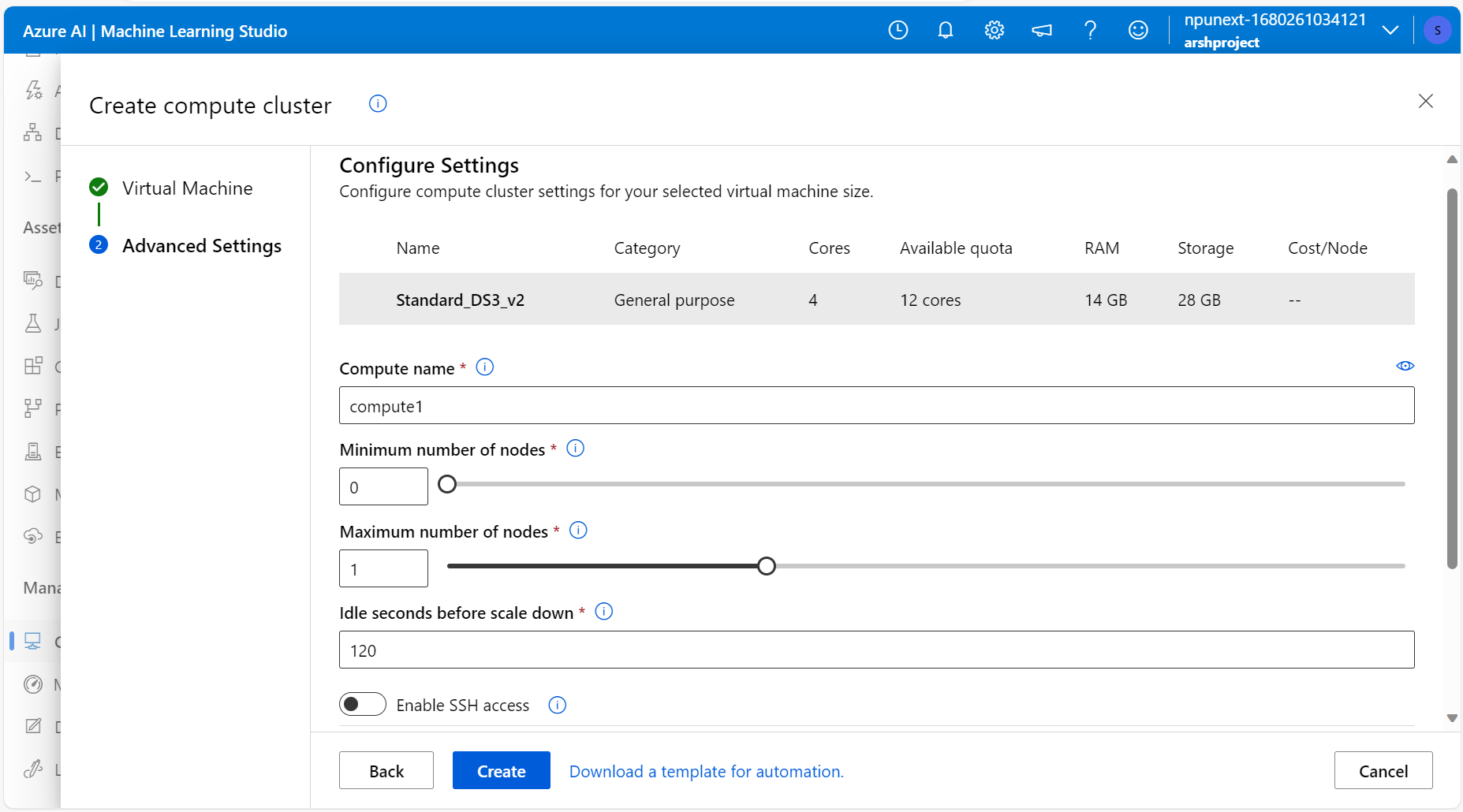
Final pipeline



Creating compute cluster

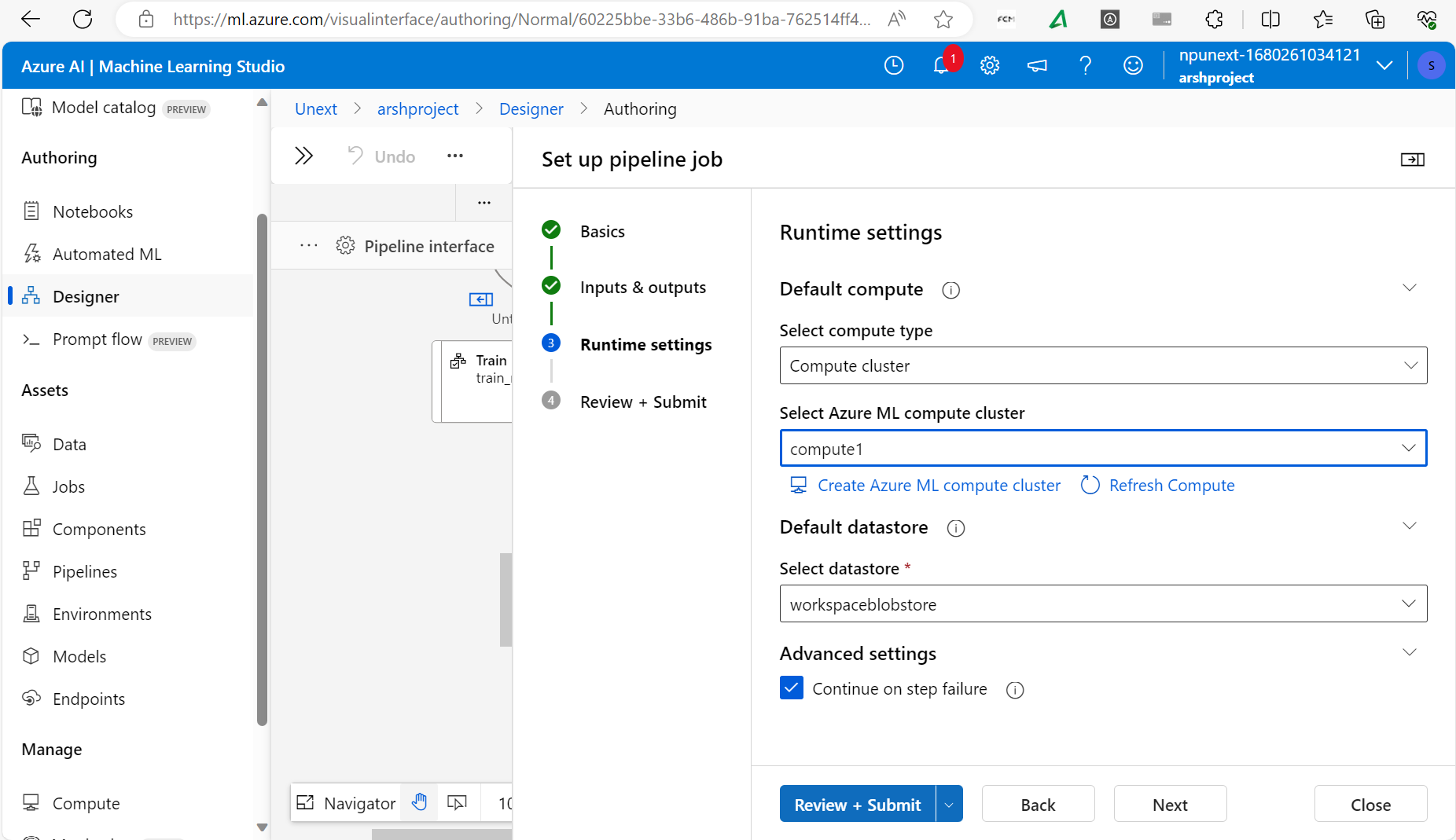
A computer screen shot of a computer

Description automatically generated



Running and checking the score





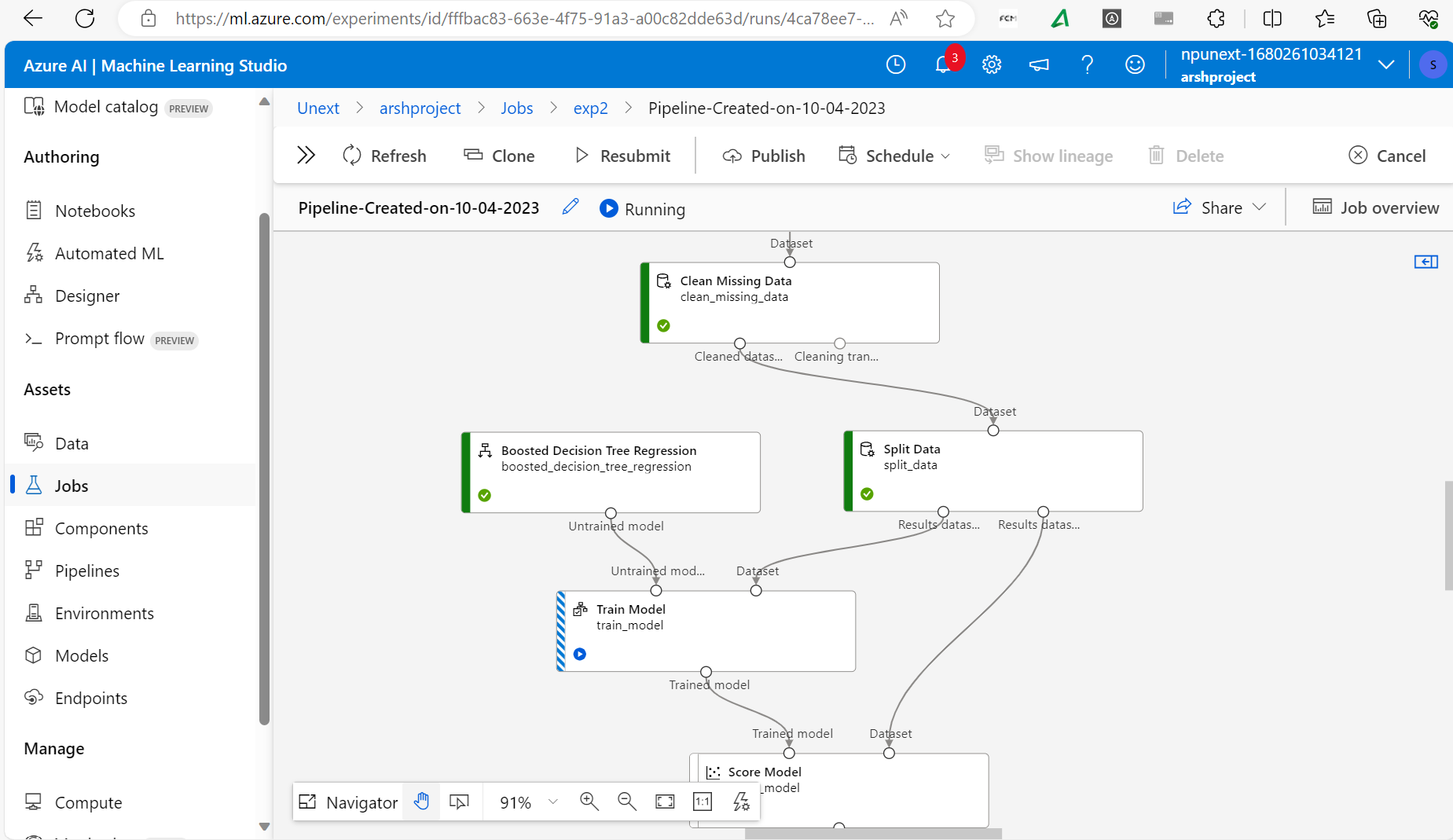
A screenshot of a computer

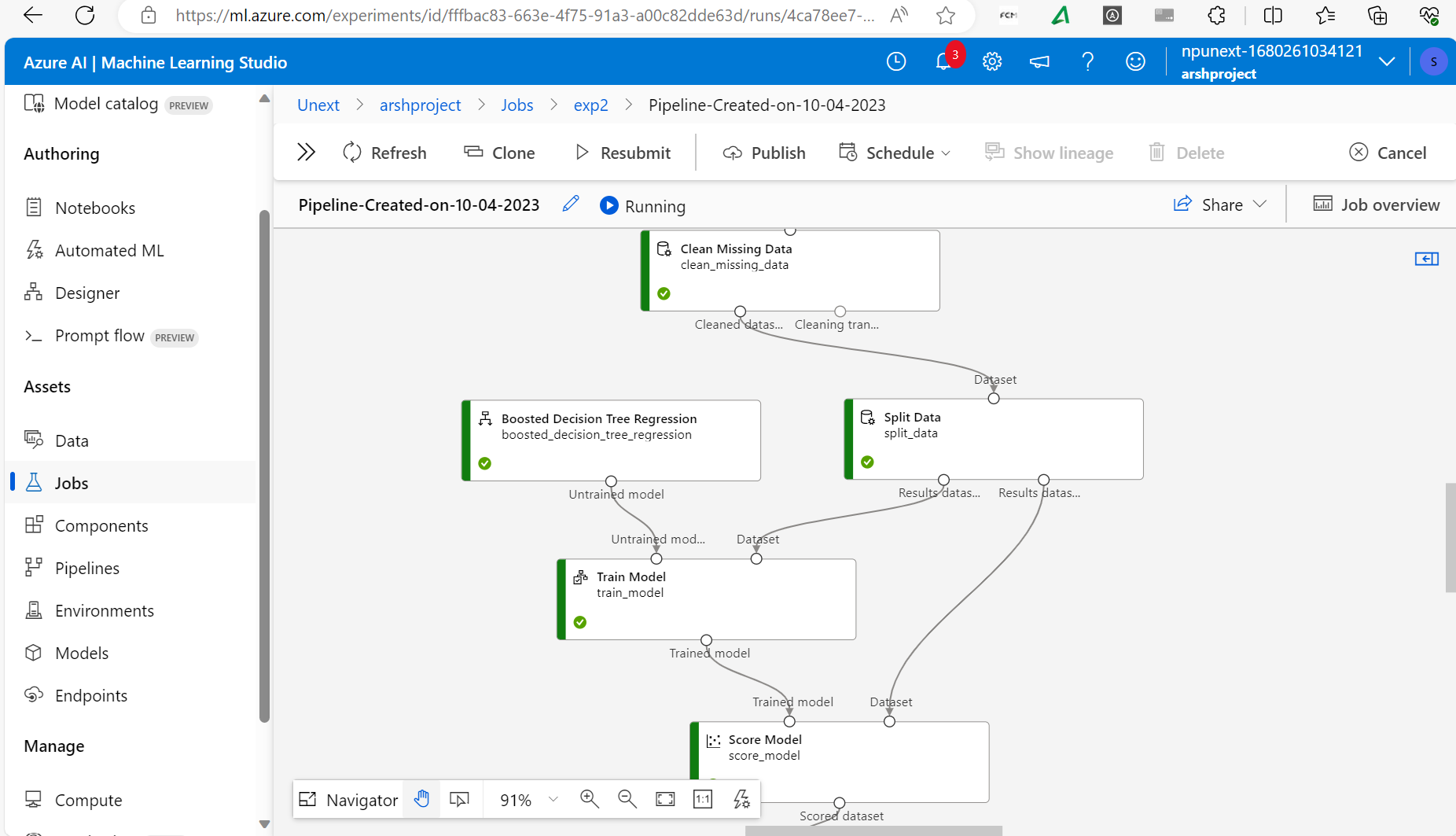
Description automatically generated

A screenshot of a computer

Description automatically generatedA screenshot of a computer

Description automatically generated



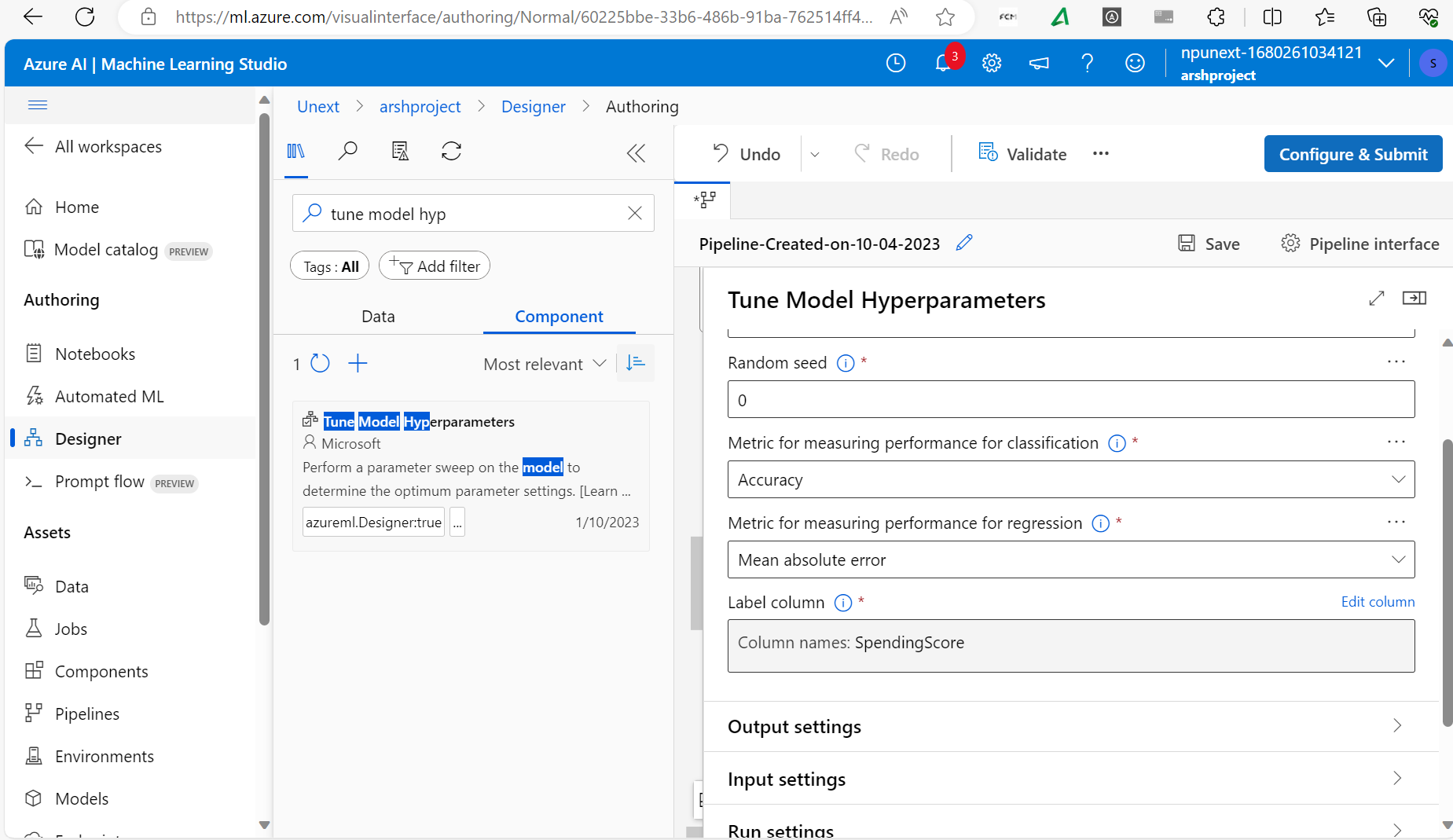


Score

A screenshot of a computer

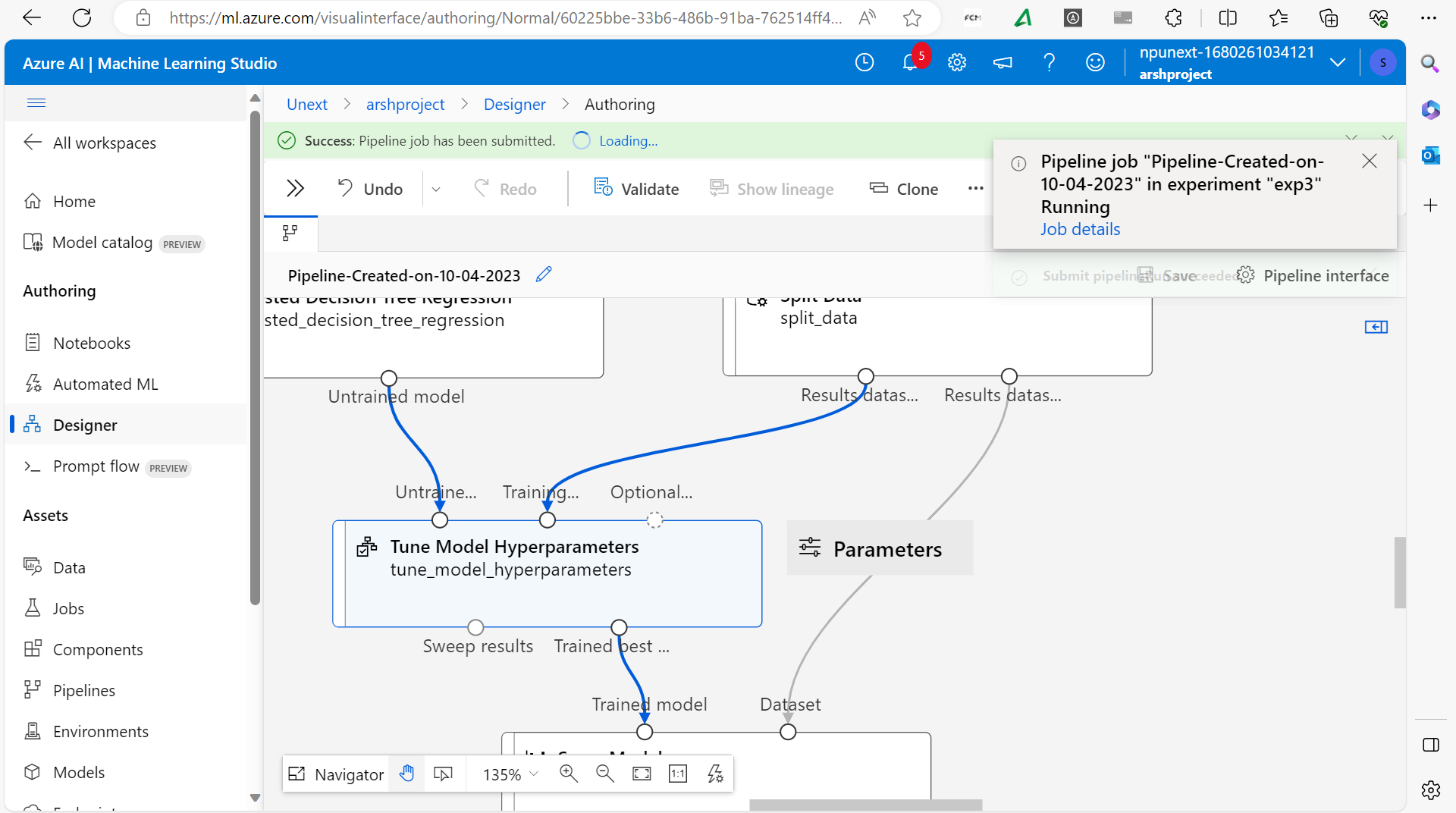
Description automatically generated

Hyperparameter



A screenshot of a computer

Description automatically generated

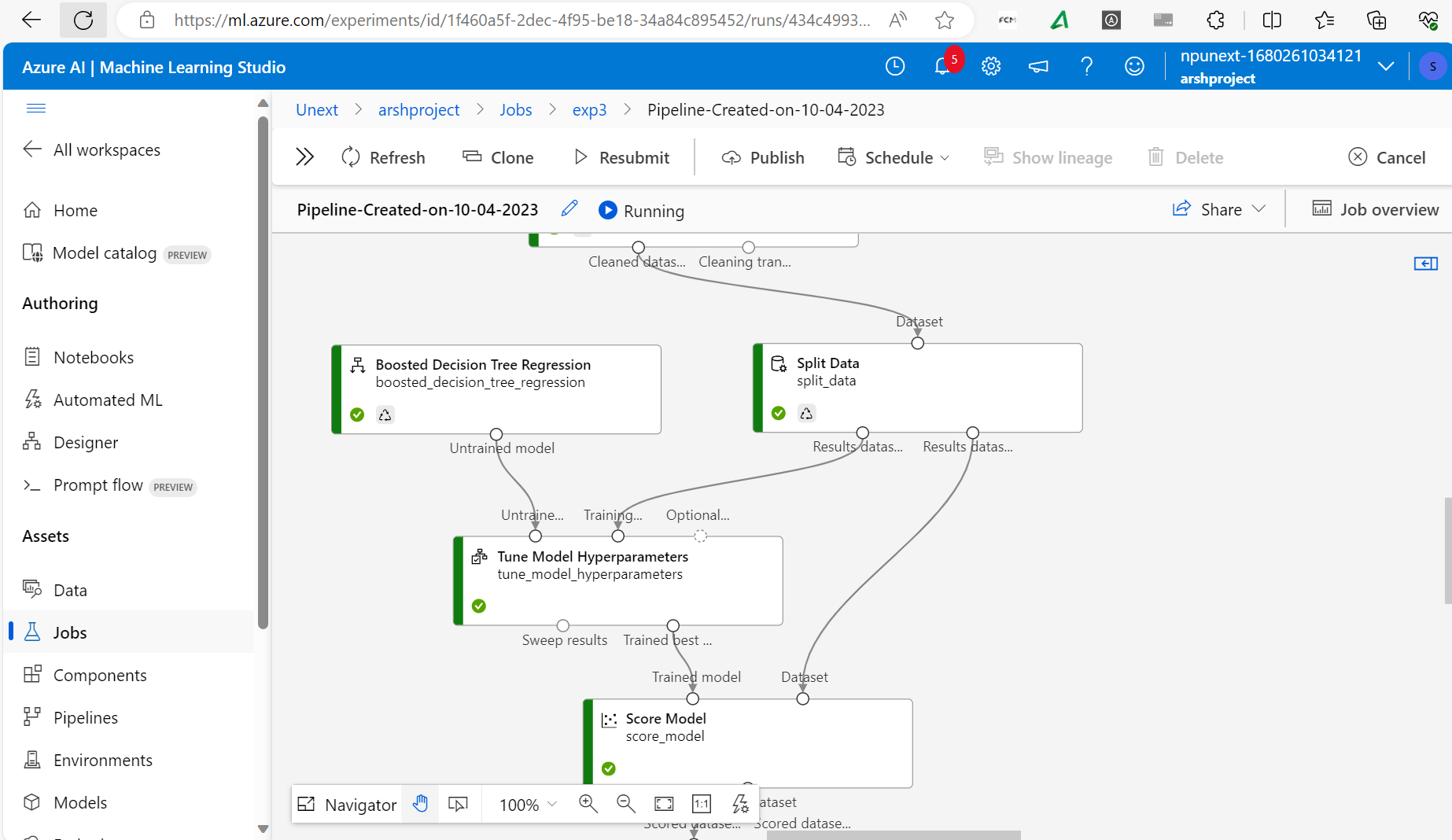


A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated



A screenshot of a computer

Description automatically generated

**Assessment answers**

1. The key steps involved in preparing the dataset for training a machine learning model are:
2. Creating a Machine Learning workspace. Blob storage is made by default. Adding a csv file to the default ml container in blob storage account.
3. Launch the ML studio. Create a data asset from azure file.
4. Design a pipeline and clean the data asset using various actions like clean data
5. Split the data after cleaning into train and test sets
6. Pass it to the training model.
7. Splitting into train and test datasets is necessary in order to evaluate the performance of our training model. Our model will be trained on 70% of the dataset and on the rest 30%, since we know how or model should have predicted the answer, we can test the accuracy of our model.
8. We could’ve gone with any regression model since our output is going to be binary. Either the customer purchases(1) or no (0). So if our model predicts percentage of a customer somewhere close to 1 that means he/she buys otherwise not.
9. Hyperparameter tuning is necessary for efficiently training the model and improves its accuracy score