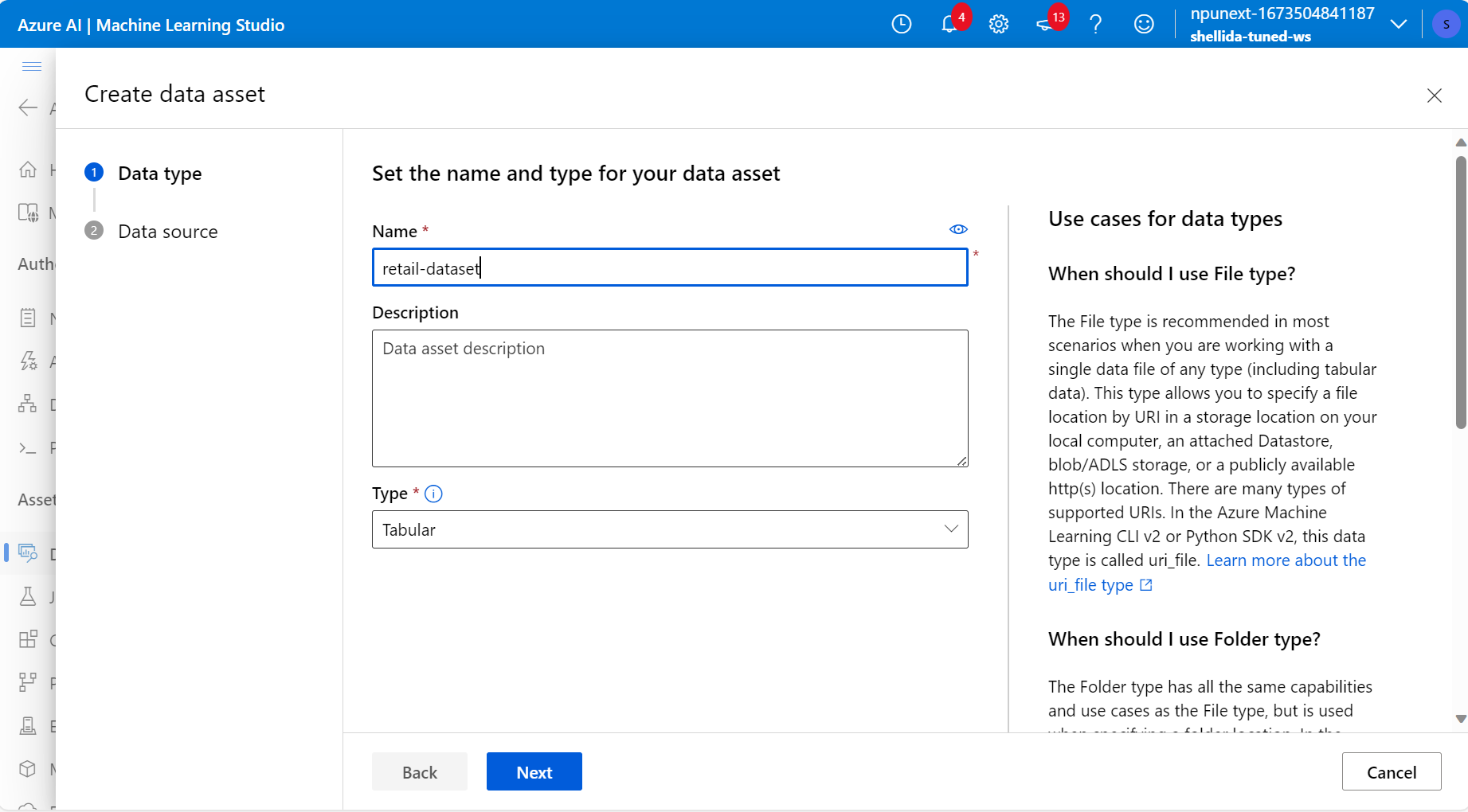
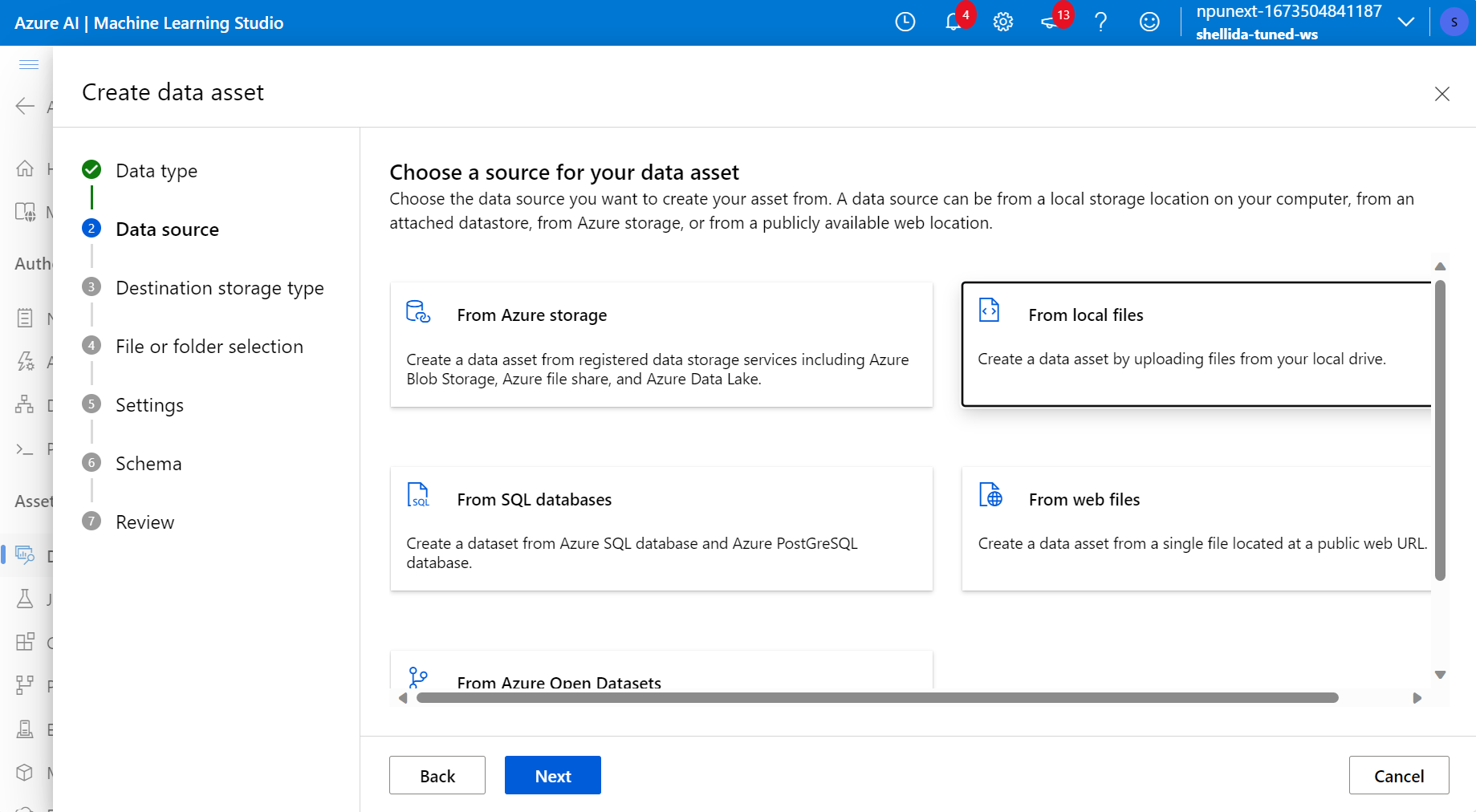
Azure ML Studio - Hands-On Assessment

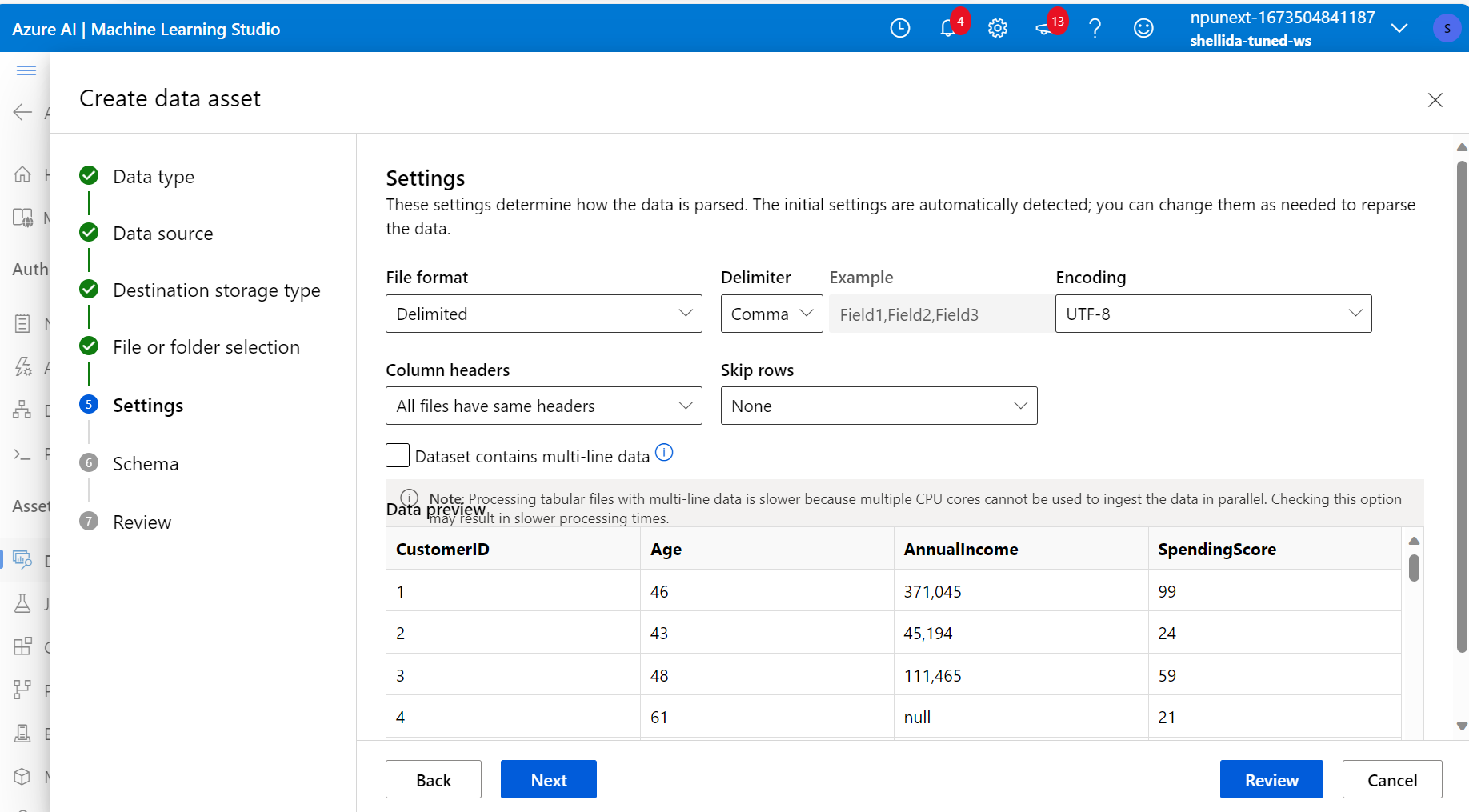
Name: Meghna Sabharwal

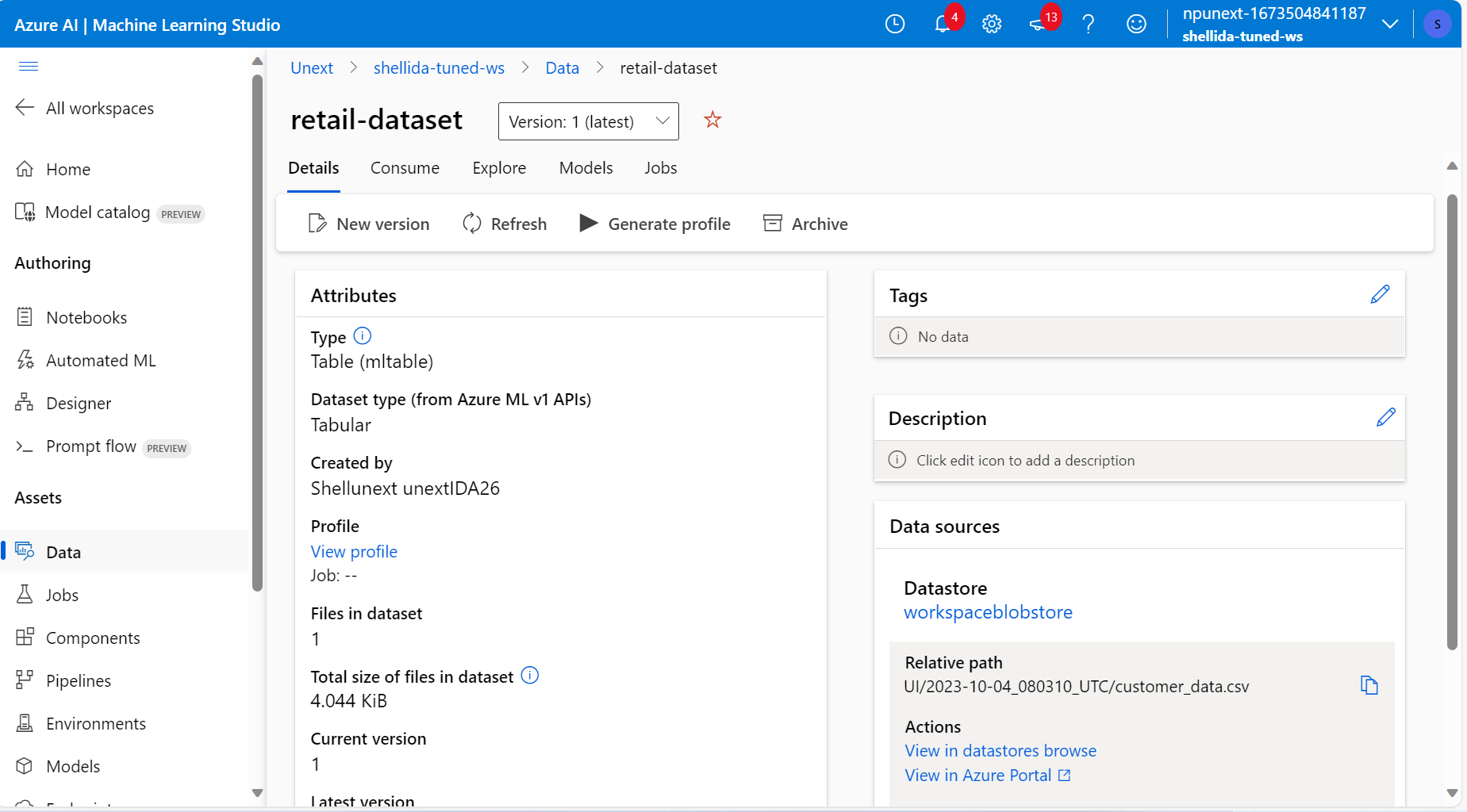
Employee id: 654971

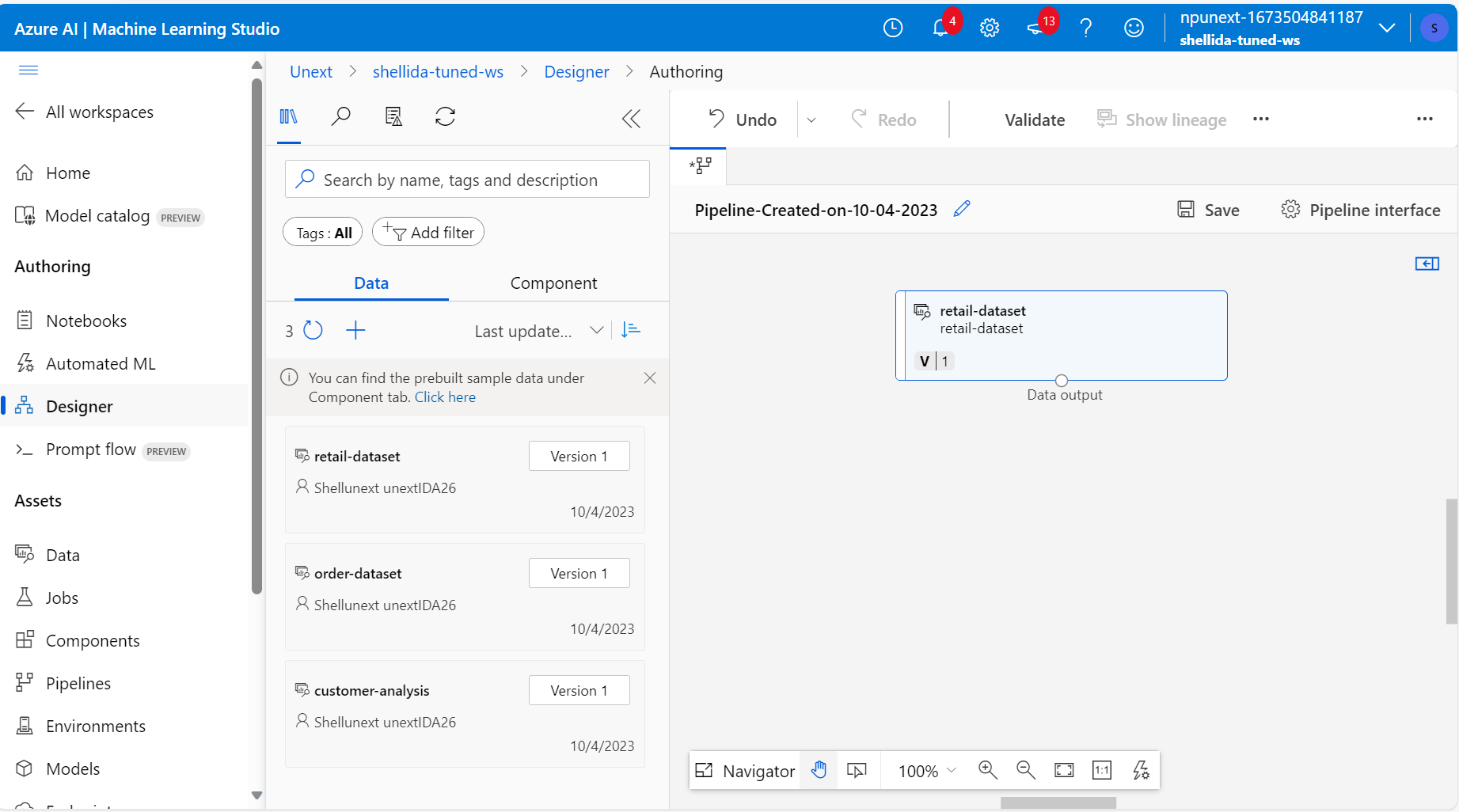
Data uploading



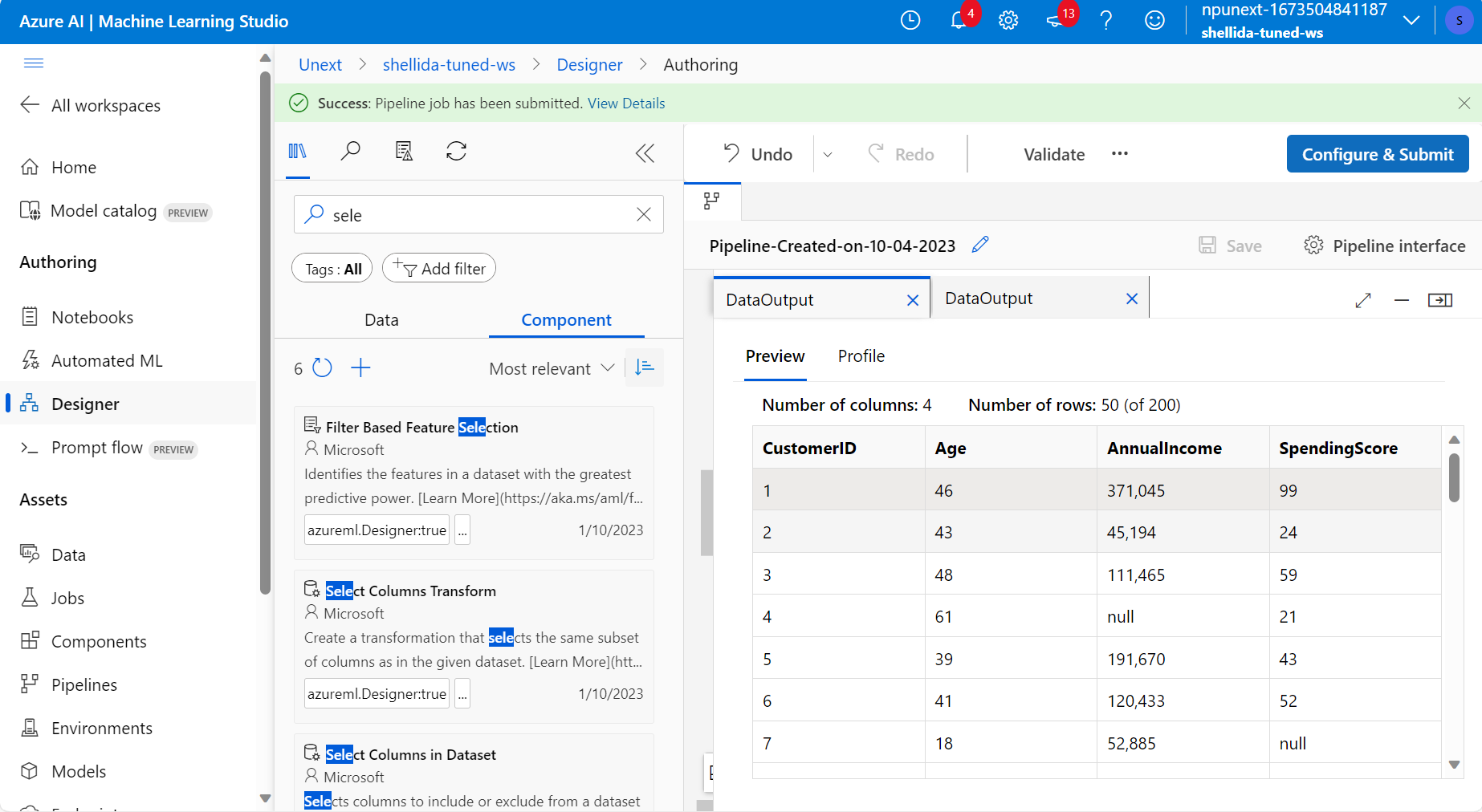


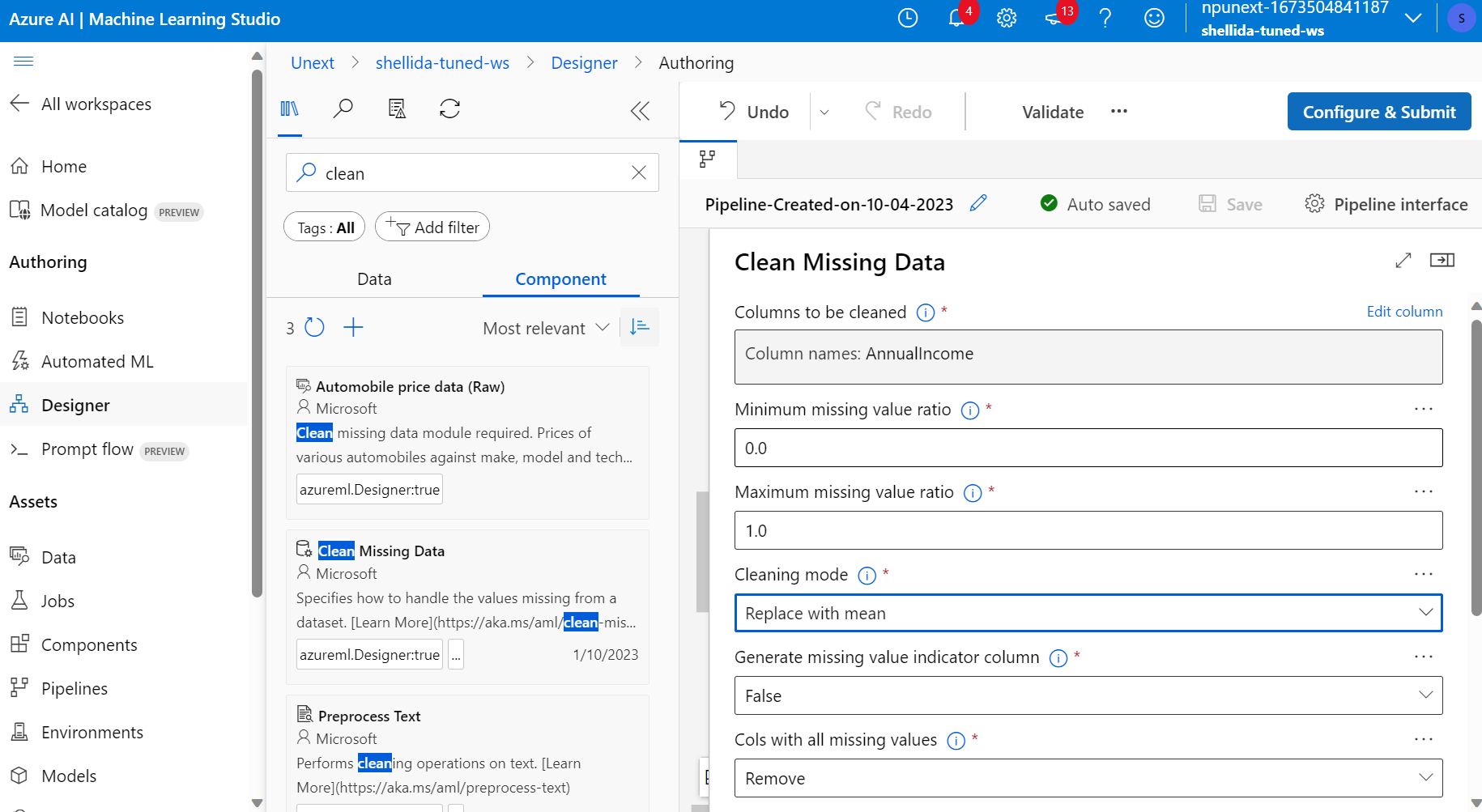


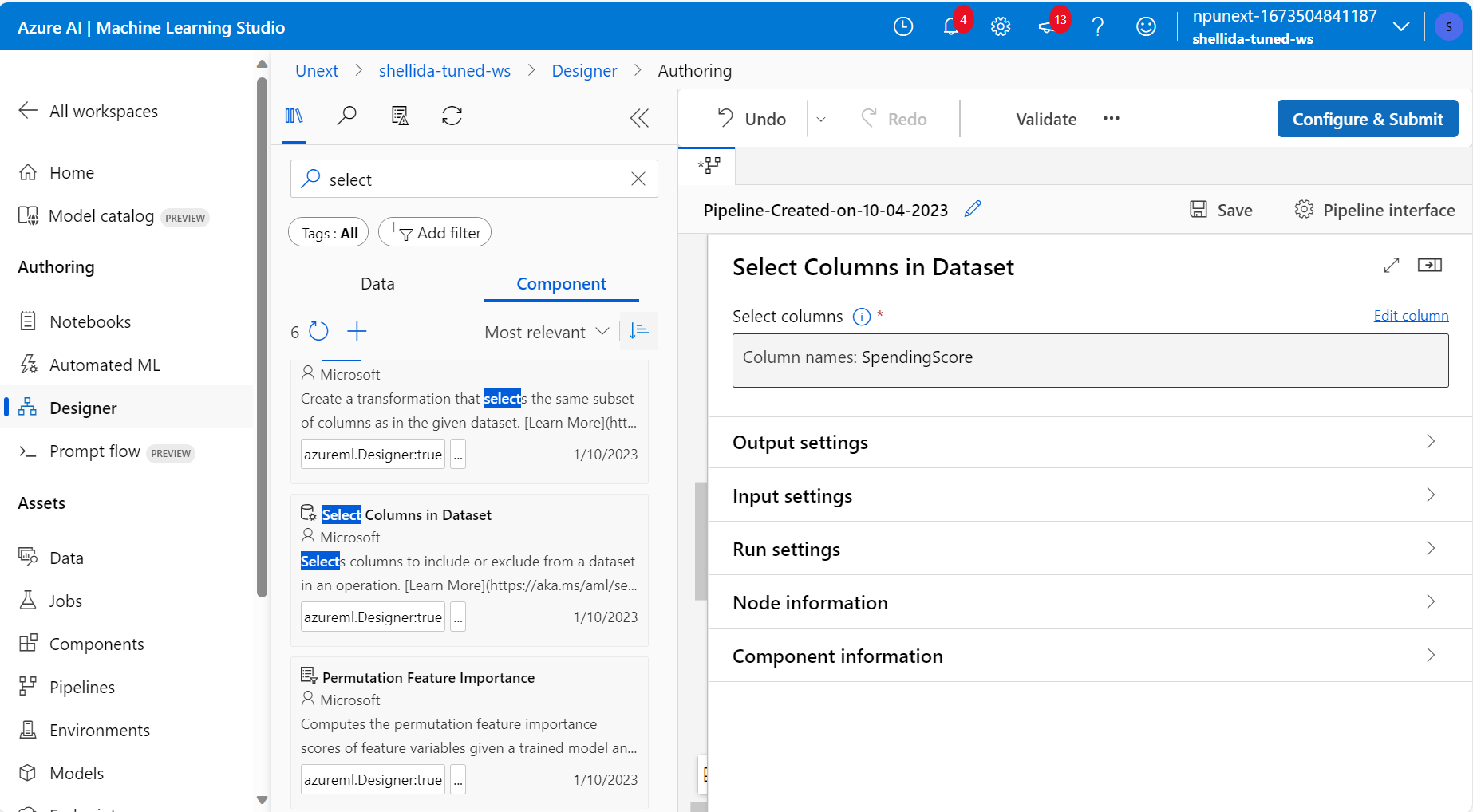


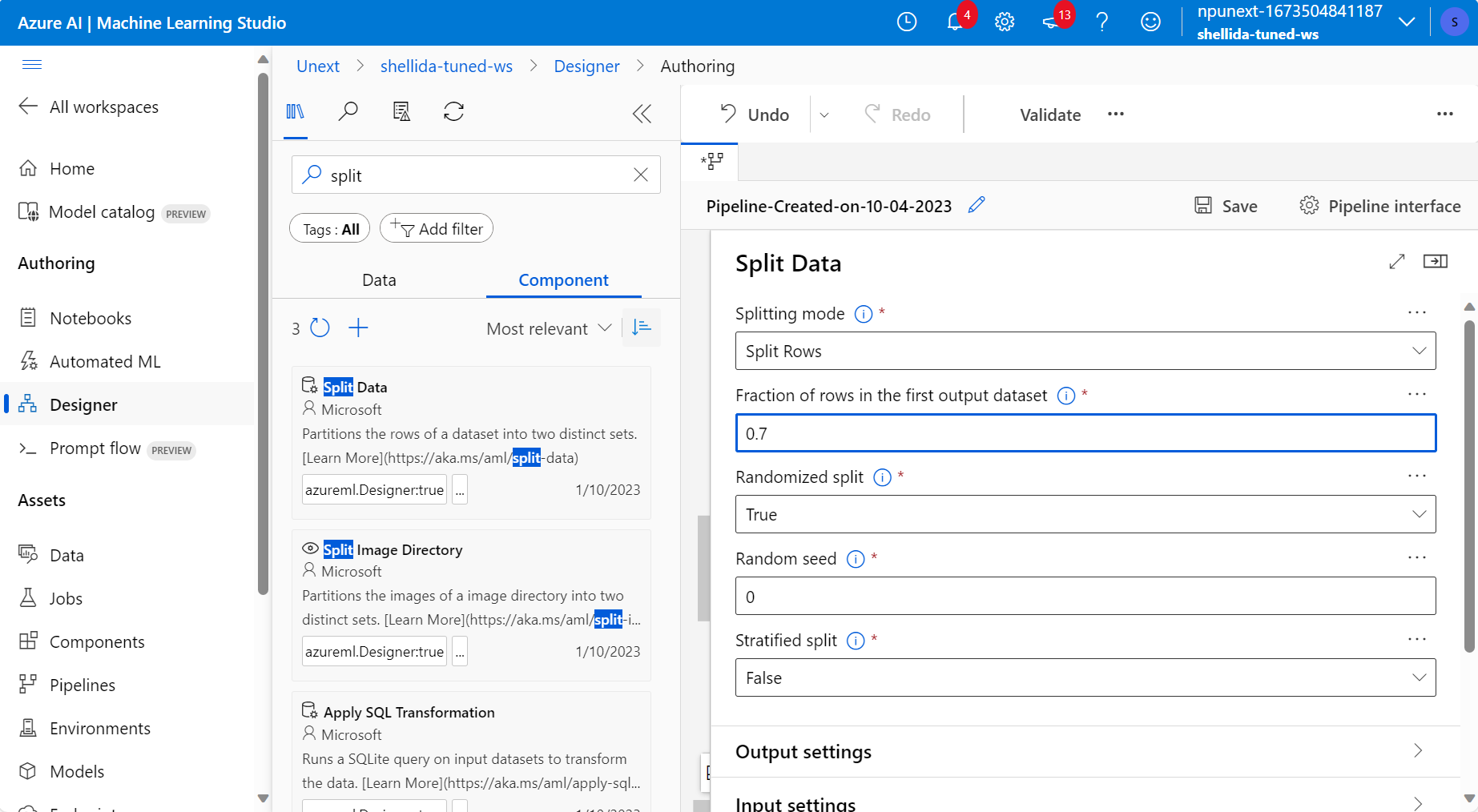


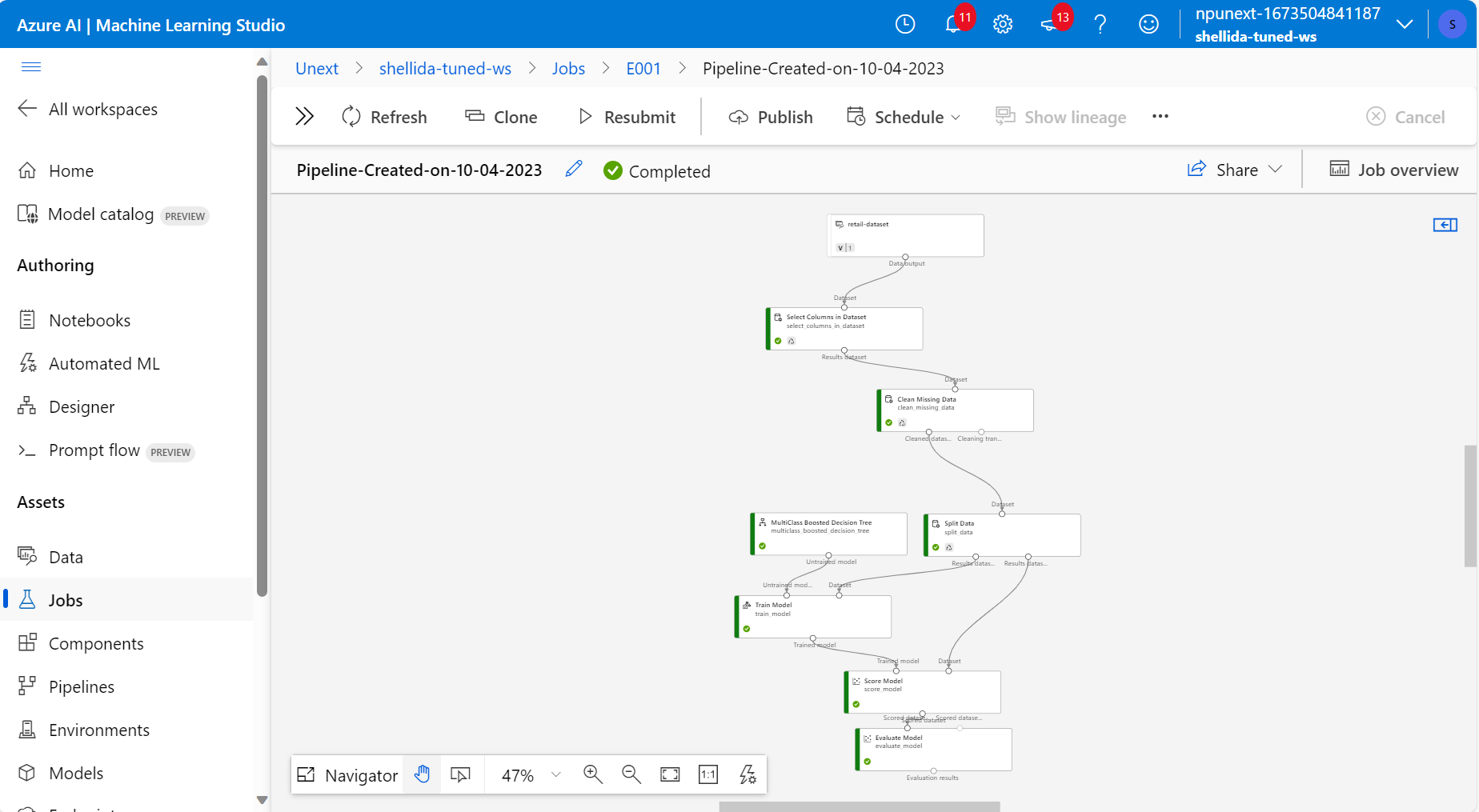
Data preprocessing



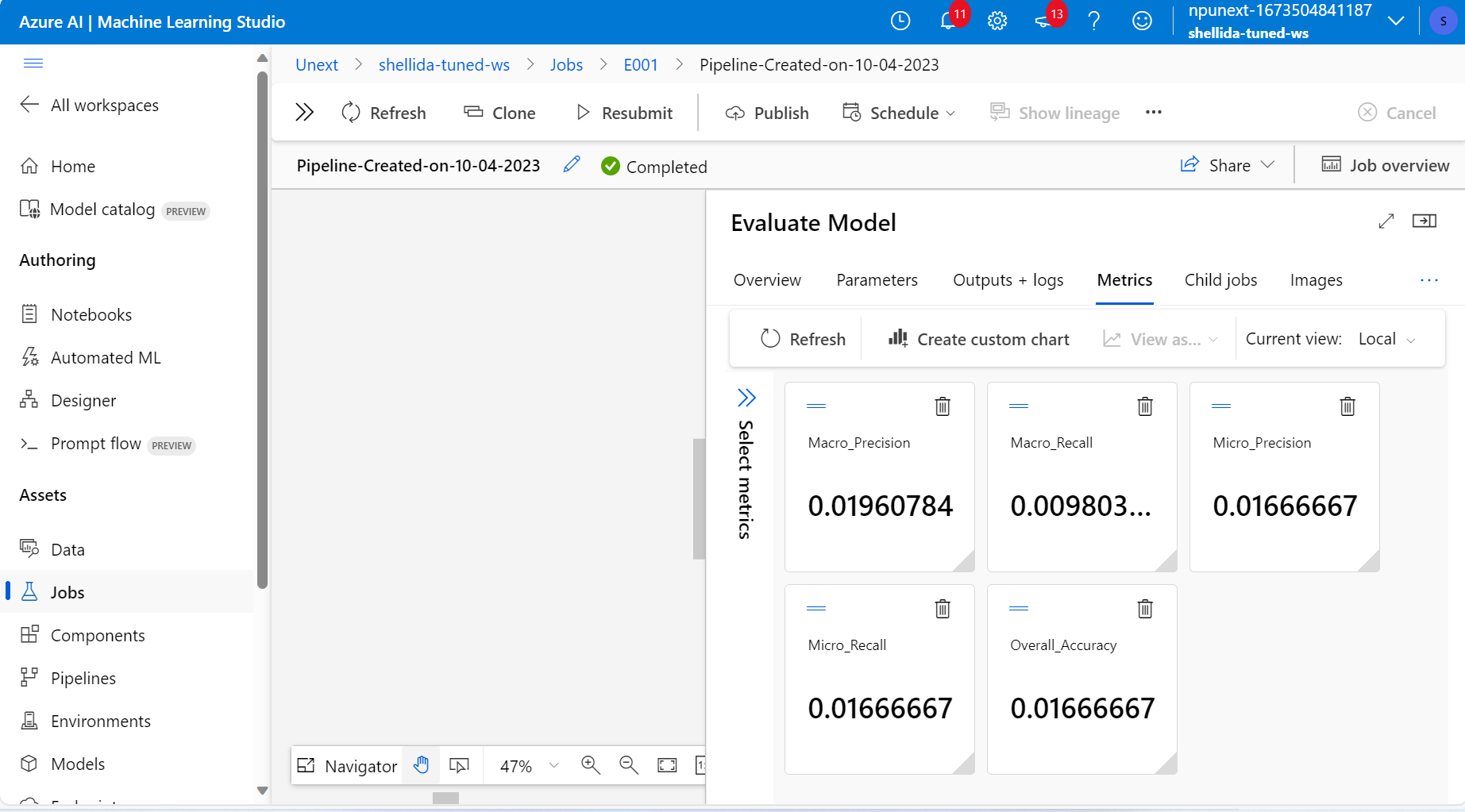




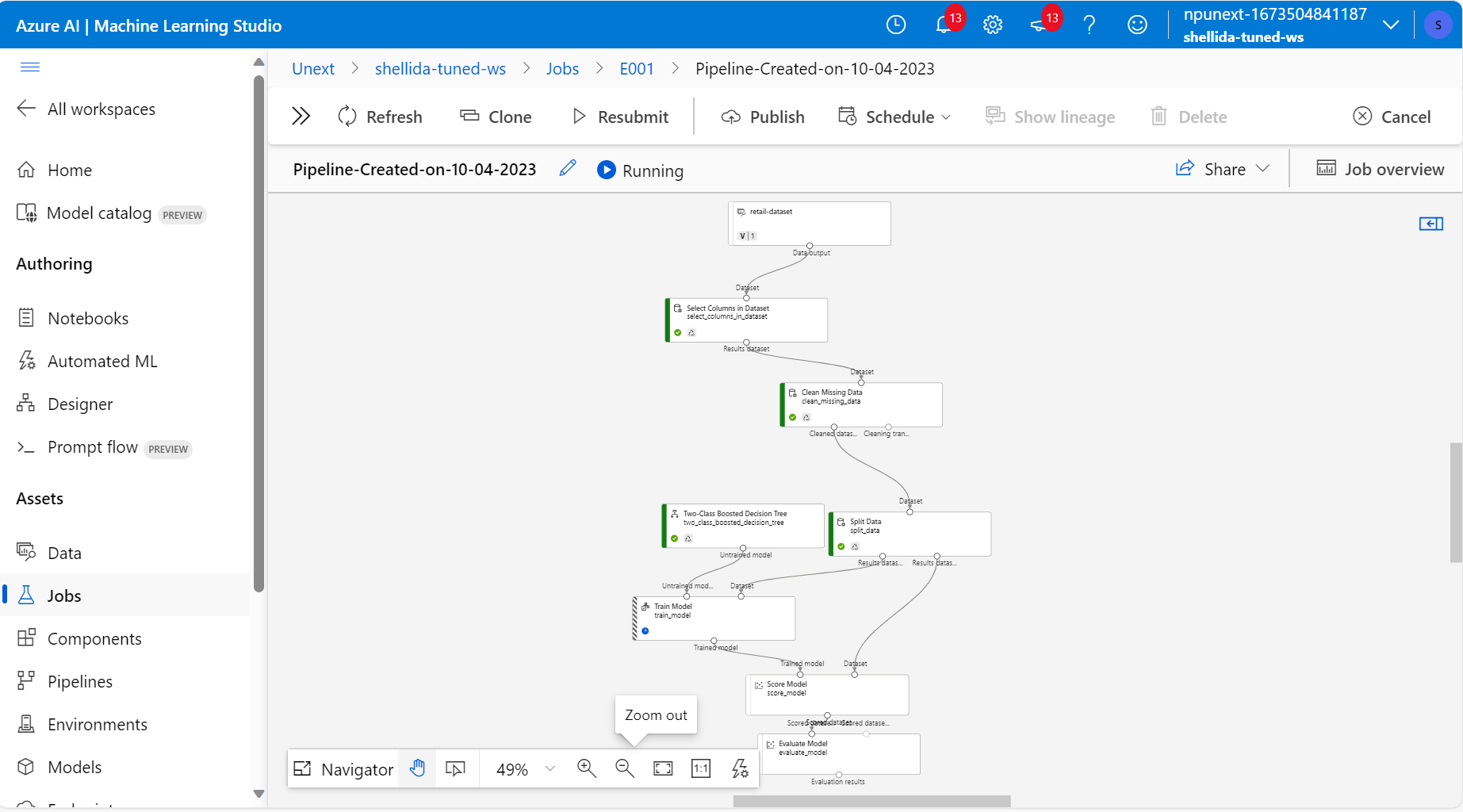




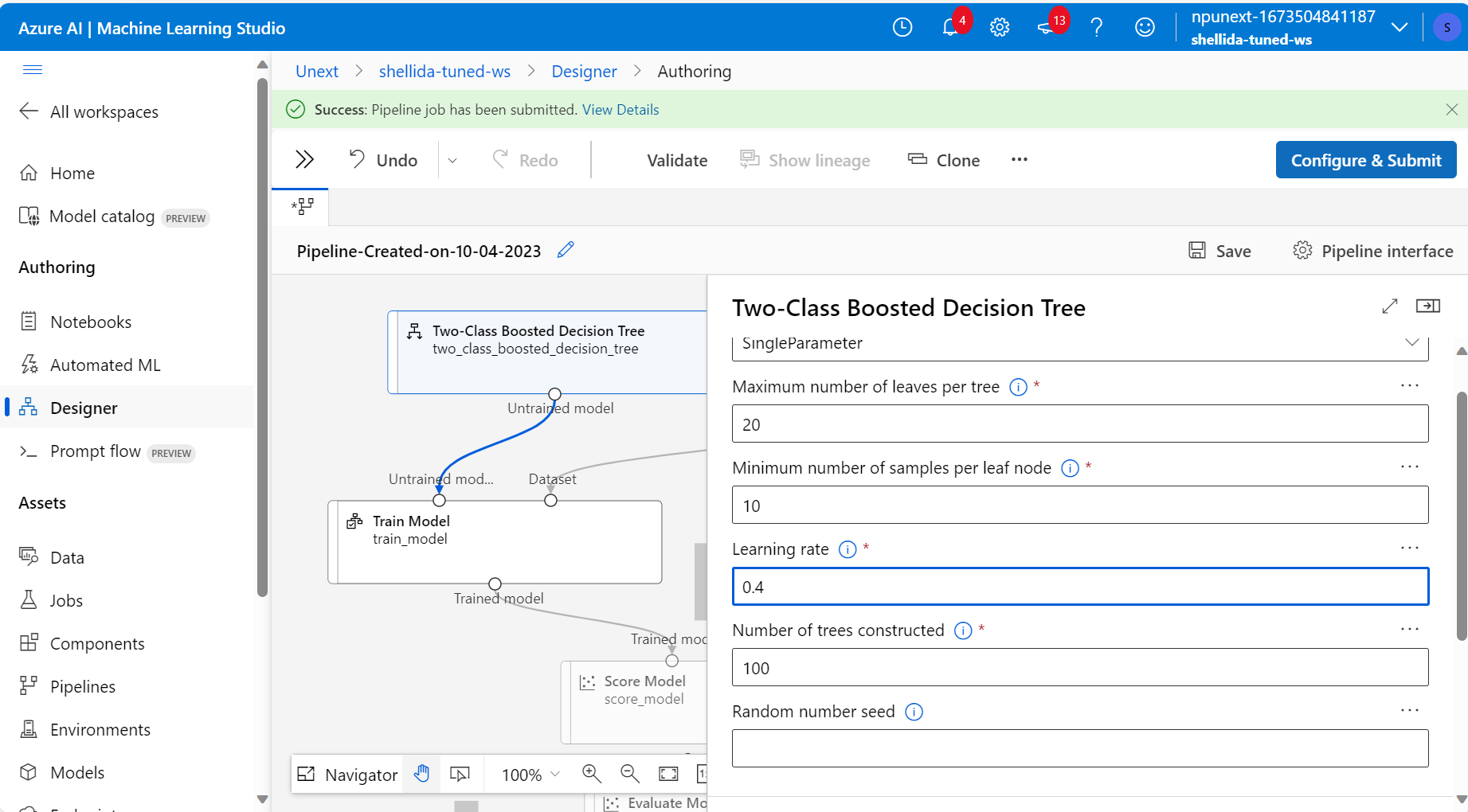
Evaluation metrices



Decision tree



Hyperparameter tuning



1. **What are the key steps involved in preparing the dataset for training a machine learning model using Azure Machine Learning? Briefly explain each step**

**Ans:** There are various steps involved in preparing the data for training in machine learning

Data Collection:

Gather relevant data from various sources, such as databases, files, APIs, or external data providers. For example in the assessment we have used the data from the local files, uploaded it in the azure storage blob

Data Cleaning:

Clean the dataset to remove inconsistencies, missing values, duplicates, and outliers. Cleaning is an essential part of data preprocessing as it helps to maintain the consistency of the data

Data Transformation:

Preprocess the data by performing feature engineering, scaling, encoding categorical variables, and any other necessary transformations to make it suitable for machine learning algorithms.

Data Splitting:

Divide the dataset into training, validation, and test sets. The training set is used to train the model, the validation set helps tune hyperparameters and avoid overfitting, and the test set is used to evaluate the model's performance.

Data Uploading:

Upload the cleaned and transformed dataset to Azure Machine Learning's data storage, such as Azure Blob Storage or Azure Data Lake Storage, so it can be easily accessed by your machine learning experiments.

Data Integration with Compute:

Configure your machine learning compute environment to access the registered dataset. This ensures that your training script can access and use the dataset during model training.

Model Training:

Develop and train your machine learning model using Azure Machine Learning's compute resources.

Model Evaluation:

Evaluate the model's performance on the validation and test datasets to assess its accuracy, precision, recall, F1-score, or other relevant metrics, depending on the problem type (e.g., classification, regression).

Iteration and Optimization:

Iterate on the model training process, adjusting hyperparameters, and experimenting with different algorithms to improve model performance.

Deployment:

Once satisfied with the model's performance, deploy it as a web service or containerized application using Azure Machine Learning's deployment capabilities.

Monitoring and Maintenance:

Continuously monitor the deployed model's performance and retrain it as needed to keep it up-to-date and accurate with new data.

1. **Why is it important to split the dataset into training and testing sets when developing a machine learning model? How does this help in model evaluation?**

**Ans** It is important to split the dataset as it helps in better learning of the model, more data is feeded into the model the better results it would provide. However, the data must not be over fed as it can work only well for a provided data so the balance must be maintained

How does this help in model evaluation: The primary goal of splitting the dataset is to evaluate how well your machine learning model generalizes to unseen data. By training the model on one subset (the training set) and testing it on another subset (the testing set), you can assess its performance on data it has never encountered during training. The training set allows the model to learn from the data (reducing bias), while the testing set evaluates how well it can generalize to new, unseen examples (assessing variance). This helps you strike a balance and avoid overfitting (high variance) or underfitting (high bias).

1. **Describe a machine learning algorithm suitable for predicting customer purchasing behaviour in the given scenario. Explain why you chose this algorithm**

**Ans:**  Random Forest is a versatile and powerful ensemble learning technique that is well-suited for this scenario, and here's why it's a good choice:

It has high accuracy, helps to avoid and combat overfitting in the model, helps to tolerate outliers

1. **What is hyperparameter tuning, and why is it important in machine learning? Explain a technique used for hyperparameter tuning and its benefits**

**Ans** Hyperparameter tuning, also known as hyperparameter optimization, is the process of systematically searching for the best combination of hyperparameters for a machine learning model. Hyperparameters are configuration settings that are set before training begins and cannot be learned from the data. They control various aspects of the learning process, such as model complexity, convergence speed, and regularization strength

It highs high efficiency and also avoids overfitting striking a good balance between variance and bias.