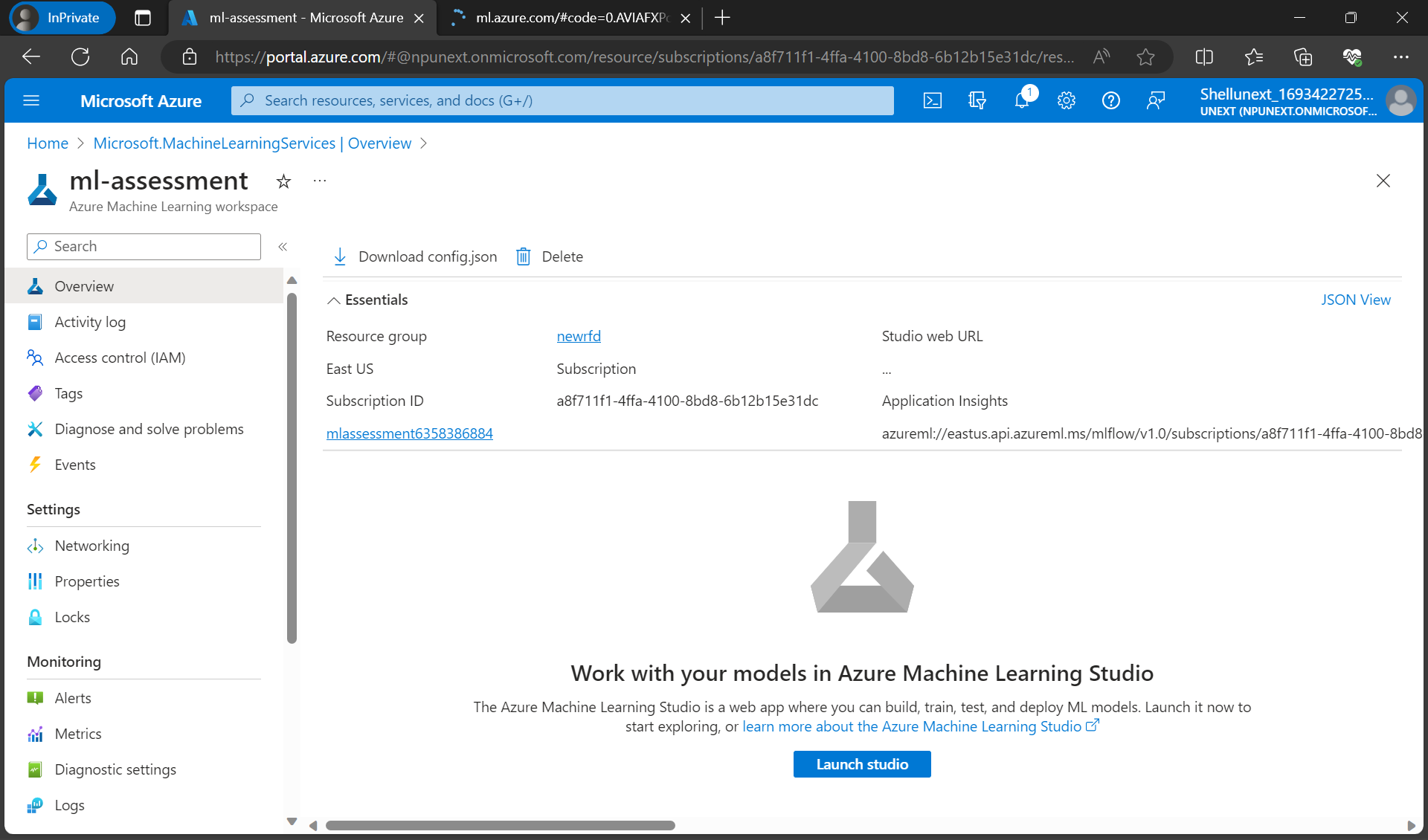
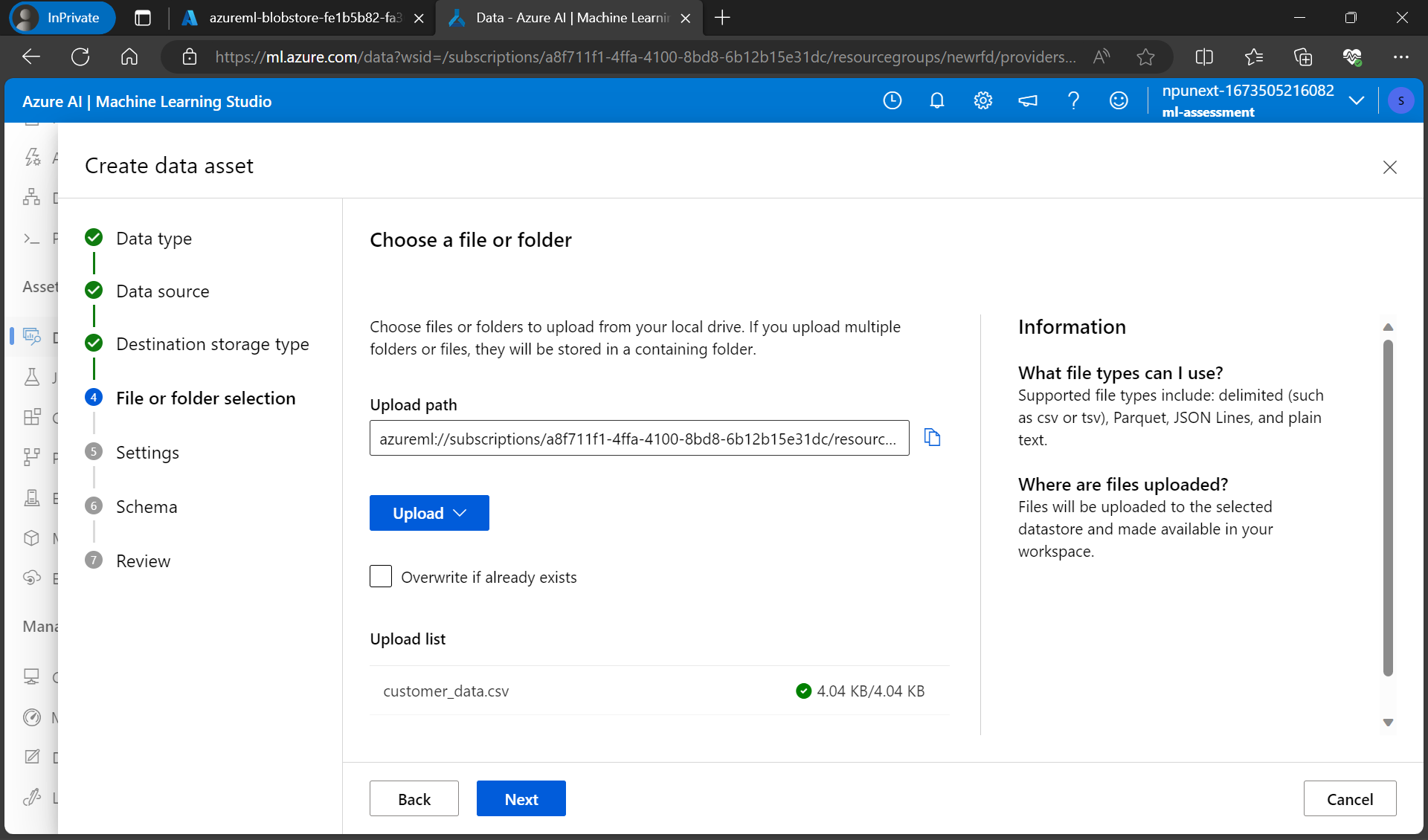
Azure ML Assessment

Harsh Gaurav Batch: Shell7-SK-23-B4 Room no. 07

ML Workspace created



Dataset Creation

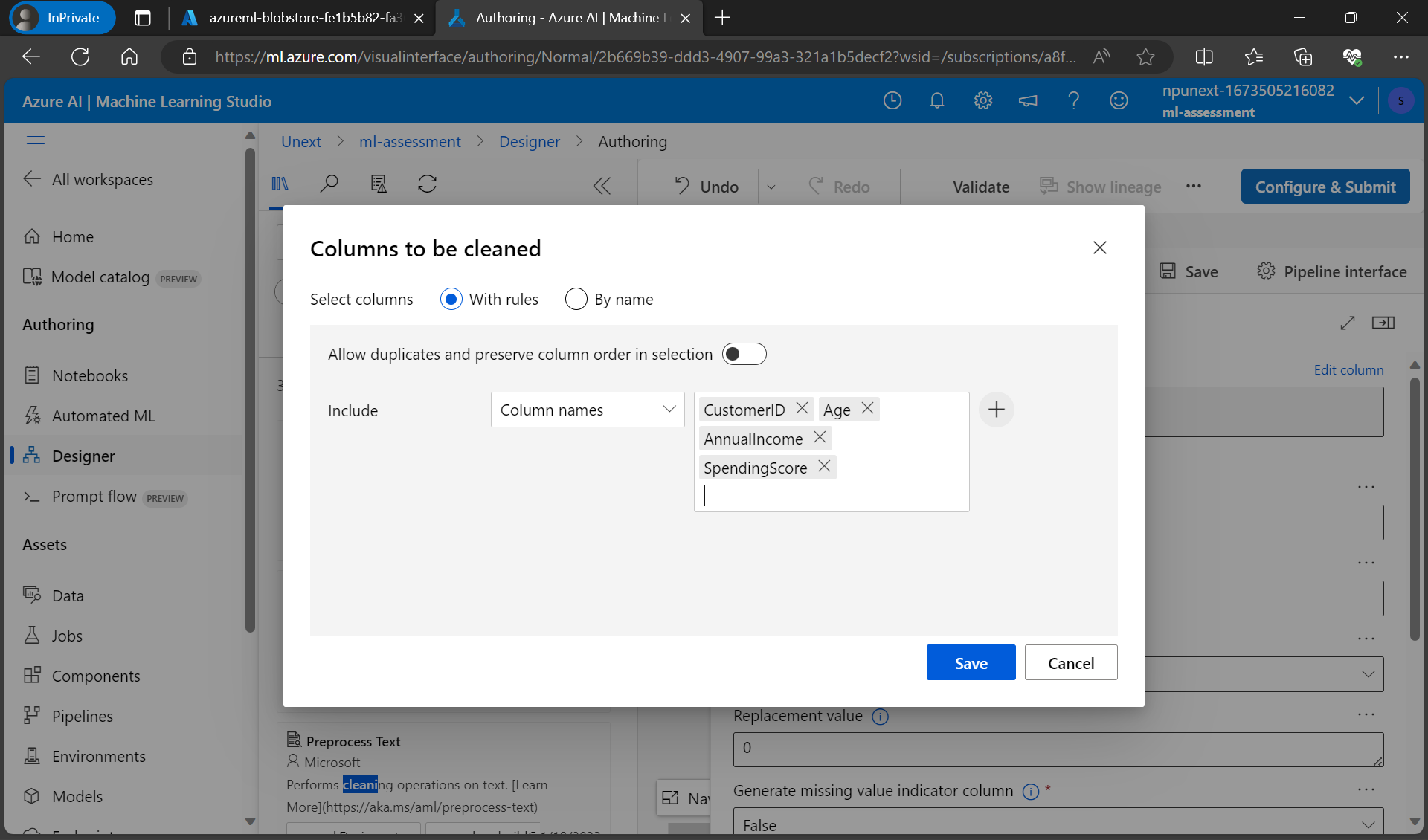


Customers data asset created

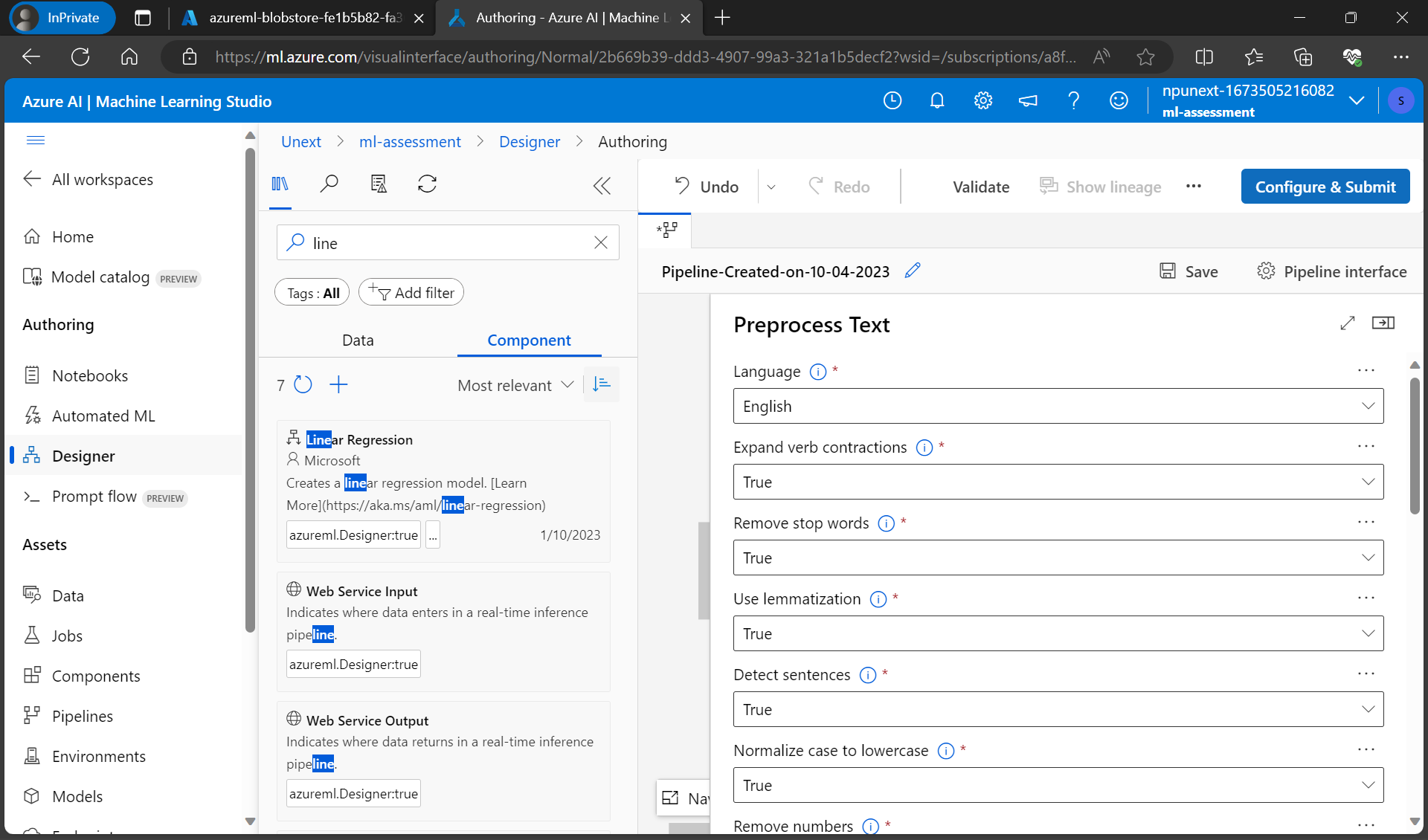
A screenshot of a computer

Description automatically generated

Columns Cleaned



Preprocess Text

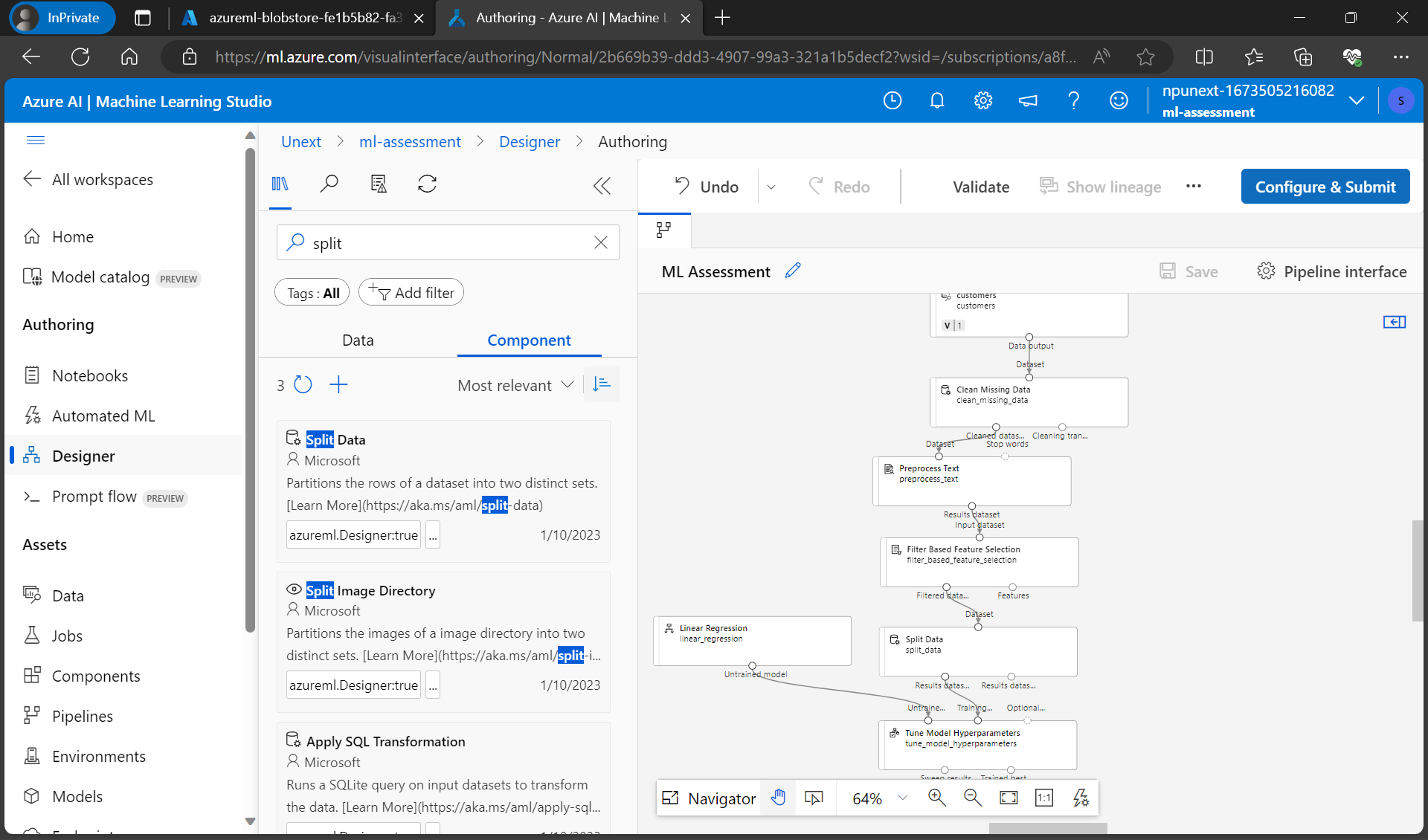


Feature Selection

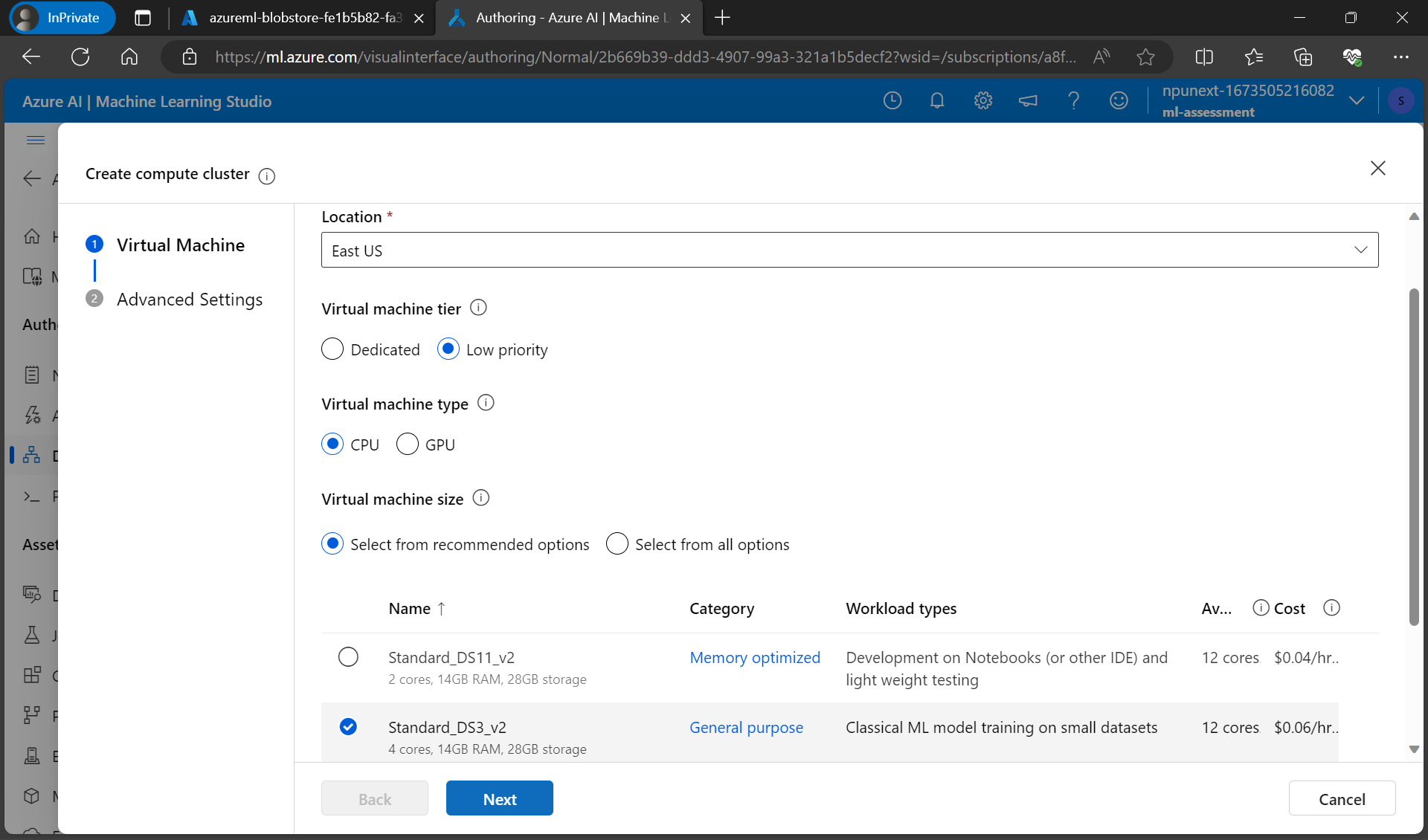
A screenshot of a computer

Description automatically generated

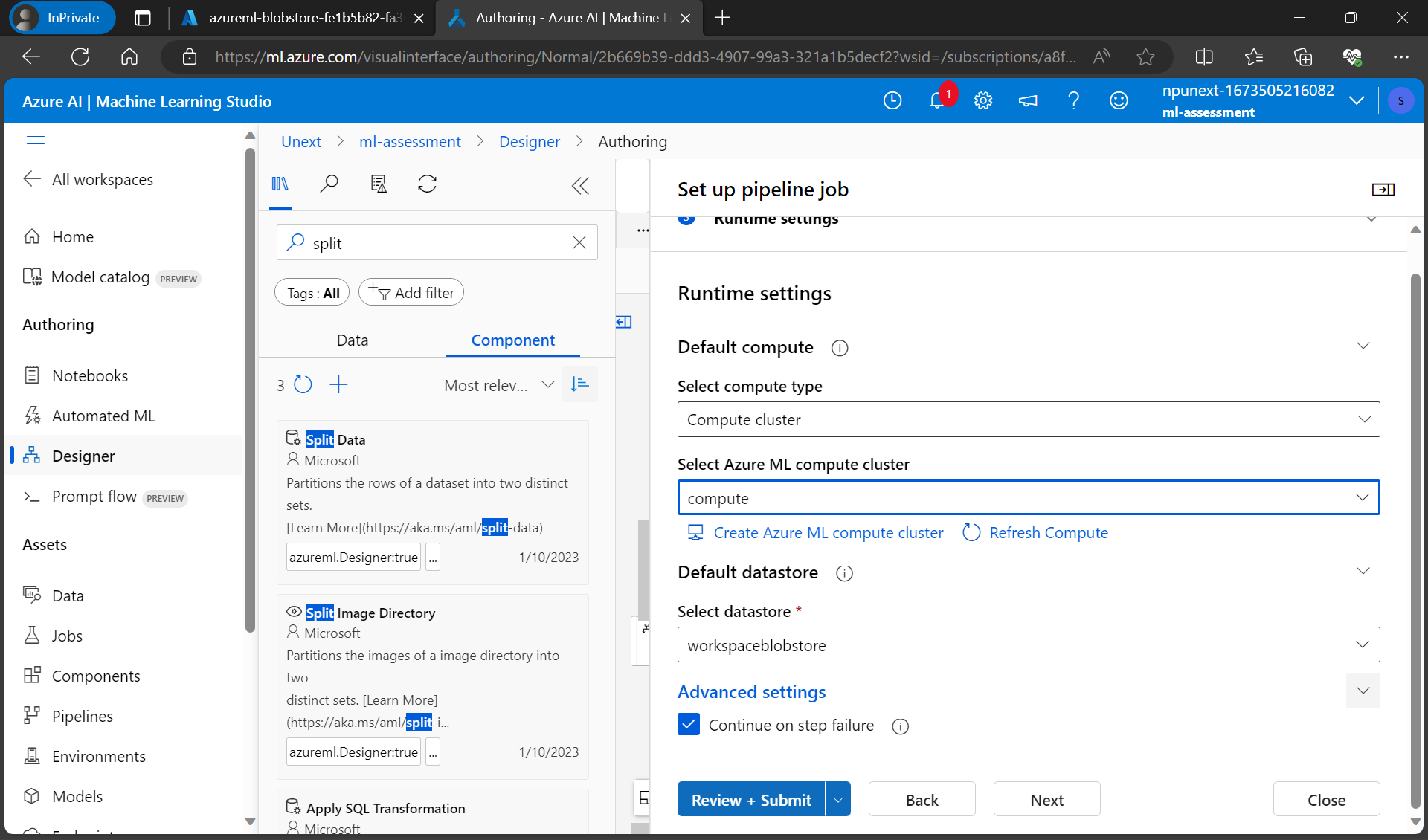
Pipeline Created with all the steps and Split the data into training and testing datasets



Azure ML Compute



Submitted Pipeline job

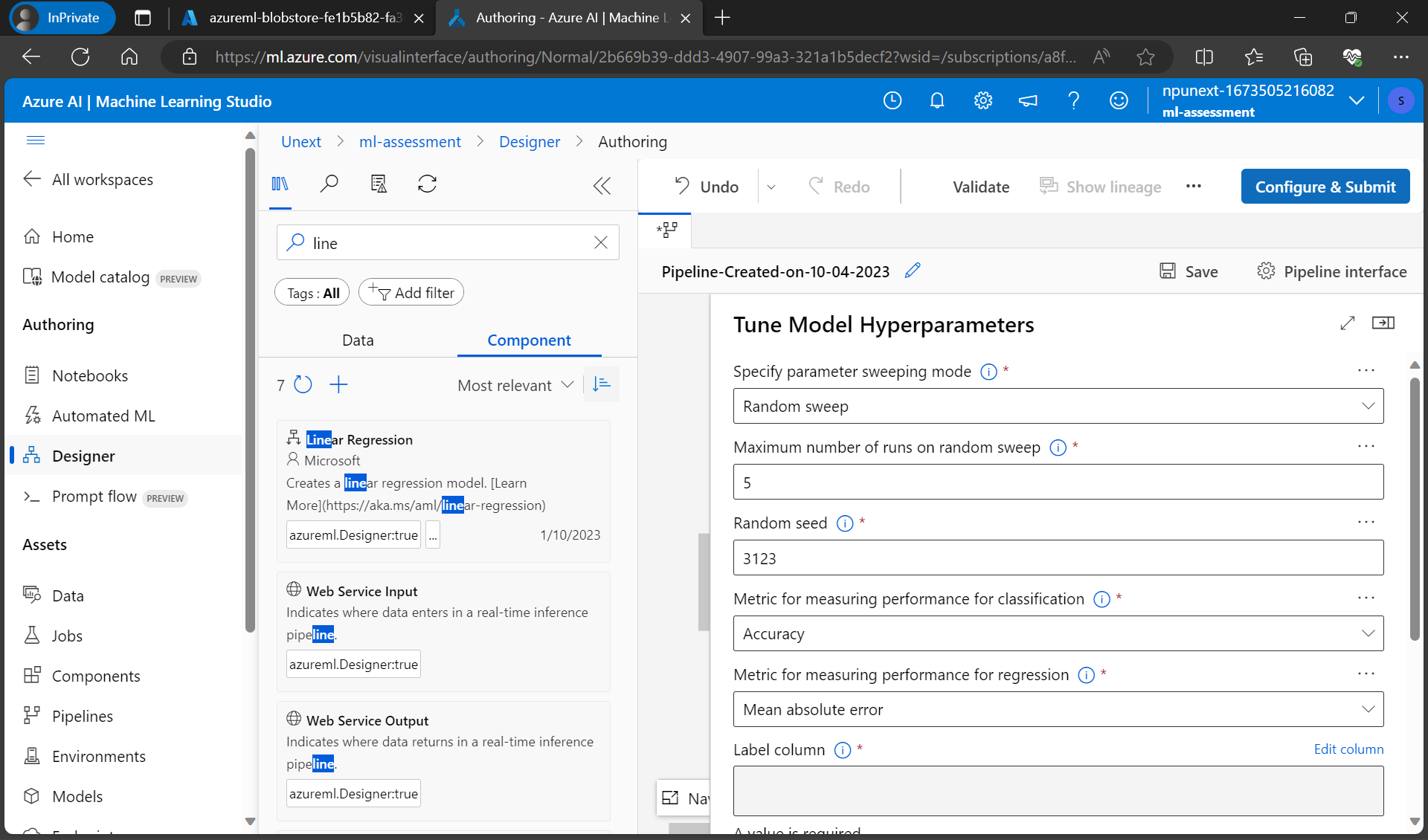


Pipeline running

**A screenshot of a computer

Description automatically generated**

Hyperparameters Model Tuning



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

There are various processes involved in getting a dataset ready for Azure Machine Learning's machine learning model training.

• Data collection: Gather pertinent information from a variety of sources, such as databases or files.

• Data cleaning: Correct issues like missing data or inaccuracies to verify the accuracy of the data.

• Data transformation: Restructure data or add new features for machine learning.

• Data Splitting: Split data into training and testing sets to assess the performance of the model.

• Uploading of Data: Save prepared data in Azure for model creation and training.

These six steps are crucial for getting your data ready for machine learning in Azure.

* 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?

It is crucial to divide the dataset into training and testing sets so that we can evaluate how effectively our machine learning model works. The training set aids the algorithm in discovering patterns from previously collected data, whilst the testing set serves as fresh obstacles. We may evaluate the model's generalizability and ability to make precise predictions on unobserved data by comparing how well it predicts the testing set. This division makes sure that our model is discovering relevant patterns rather than merely memorising the training data.

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

The Random Forest algorithm is a useful method for forecasting client purchasing behaviour in scenarios requiring various factors (such demographics, prior purchases, and website interactions). Because it can handle complicated, nonlinear interactions within the data and performs well with both category and numerical characteristics, Random Forest is an effective option. It creates a collection of decision trees, each trained on a different set of data. This ensemble method lessens overfitting and boosts precision.

Additionally, Random Forest offers feature significance scores to assist you comprehend the elements that have the most effect on consumer behaviour. It also doesn't need substantial feature engineering and is quite resilient to outliers.

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

Finding the ideal collection of parameters for a machine learning model is called hyperparameter tweaking. Hyperparameters are configuration options that are chosen before training rather than ones that are learnt from the data. The effectiveness and generalizability of the model are substantially influenced by these characteristics. Grid Search is a popular method for hyperparameter tweaking. For each hyperparameter, a range of values is specified, and all potential combinations are then carefully tested using grid search. Cross-validation is used to train and assess each combination's model, Gauging its effectiveness.

Benefits of Grid Search:

* Comprehensive Search: Grid Search exhaustively explores various hyperparameter combinations, ensuring that you don't miss the optimal configuration.
* Automation: It automates the hyperparameter tuning process, saving time and effort compared to manual tuning.
* Optimization: Grid Search helps improve a model's performance by finding the best hyperparameters, leading to more accurate predictions.