Raport Lab: Azure ML & DevOps & Mlflow

Mouad GUEDAD

Year : 2023/2024

Major : QFM

Problem:

You work as a data scientist in a small team with three other data scientists for a shipping company based in the port of Turku in Finland. 90% of goods imported into Finland pass through cargo ships in the country's ports. For the transport of goods, weather conditions and logistics can sometimes be difficult at ports. Rainy conditions can distort operations and logistics at ports, which can affect supply chain operations. Predicting rainy conditions in advance helps optimize resources such as human, logistics and transportation for efficient supply chain operations at ports. From a business perspective, forecasting rainy conditions in advance allows ports to reduce operating costs by up to approximately 20% by enabling efficient planning and scheduling of human resources, logistics and shipping resources. transportation for supply chain operations.

Task:

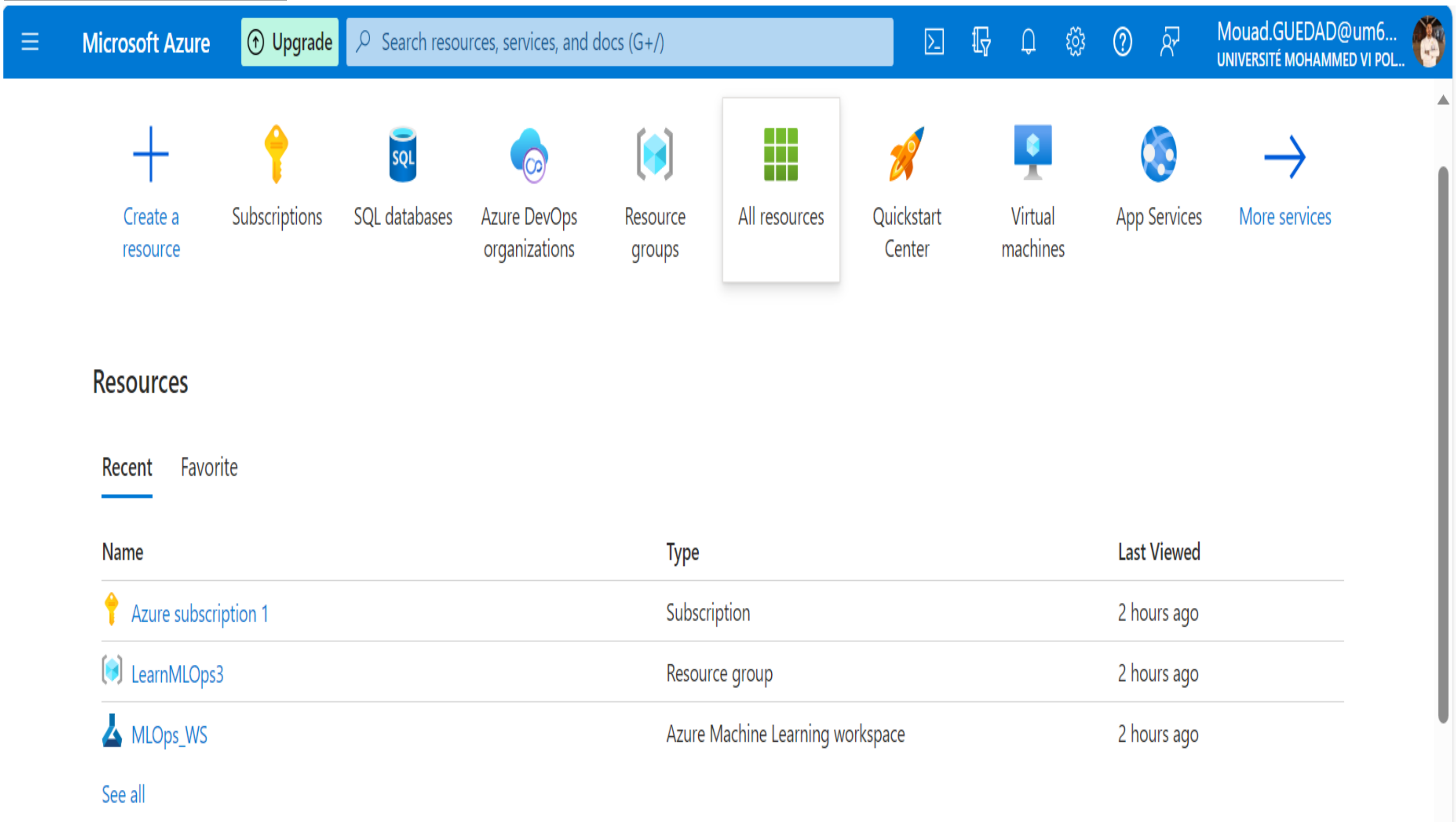
As a data scientist, you are responsible for developing an ML-based solution to predict weather conditions 4 hours in advance at the port of Turku in Finland. This will allow the port to optimize its resources, resulting in savings of up to 20%. To start, you have a set of historical weather data covering a period of 10 years from the port of Turku. Your task is to create an ML solution focused on continuous learning to optimize operations at the Port of Turku

1. Setting up resources and tools

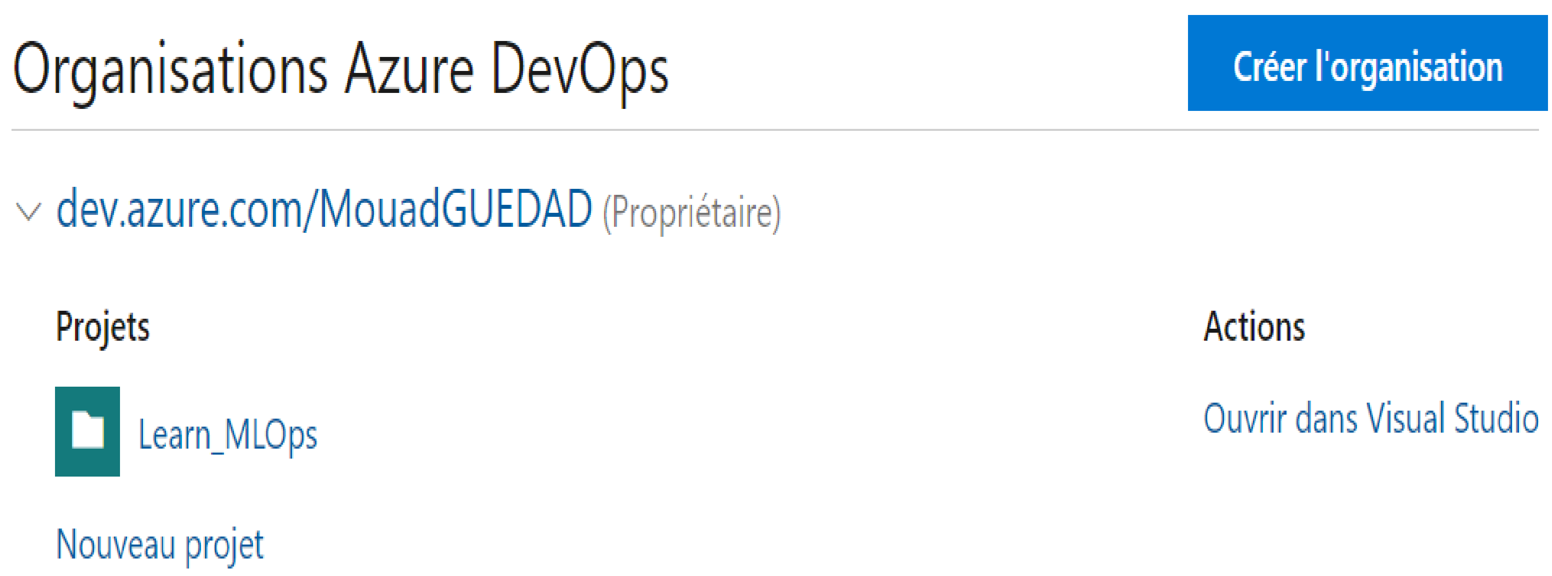
1.1 MLflow

After installing MLflow in my local and I checked I found it is already in my local .

1.2 Azure Machine learning

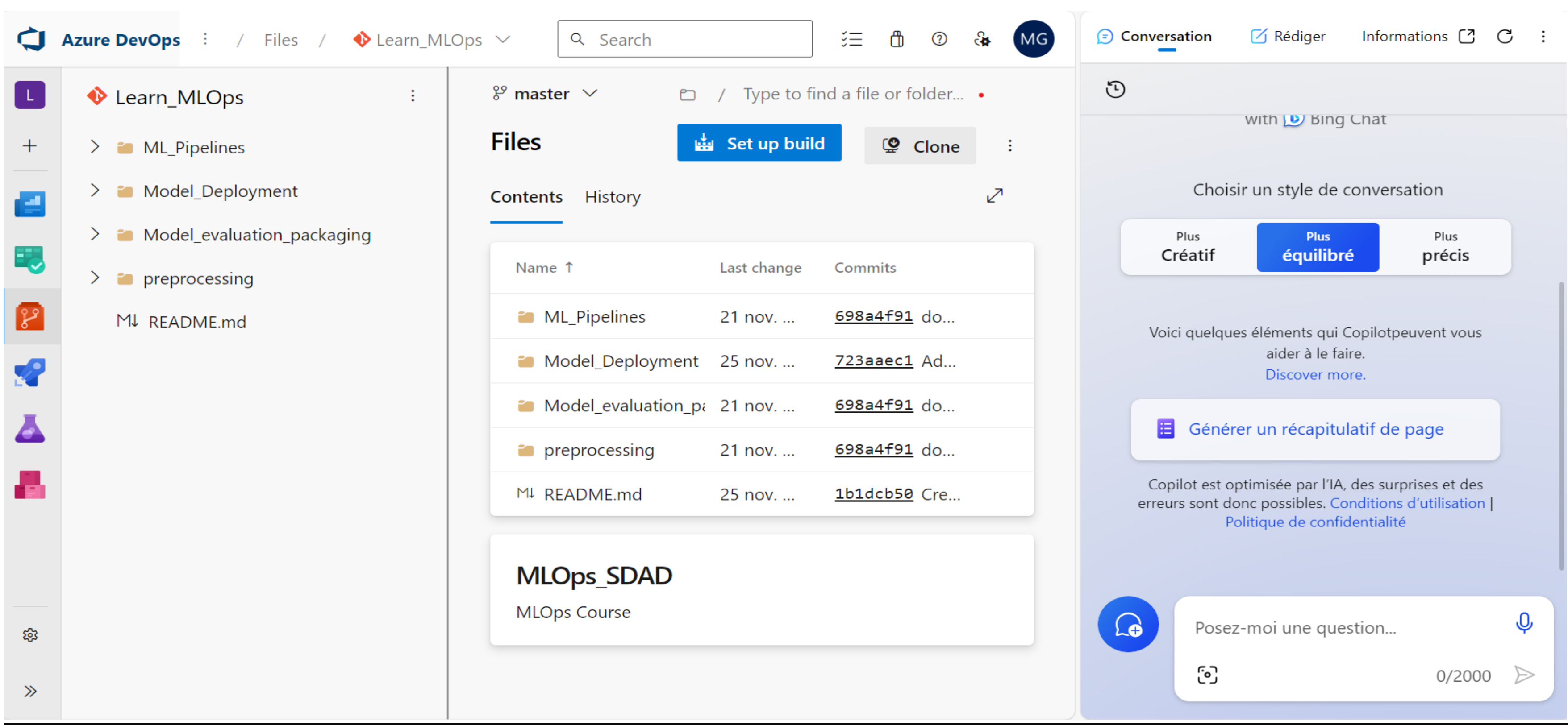


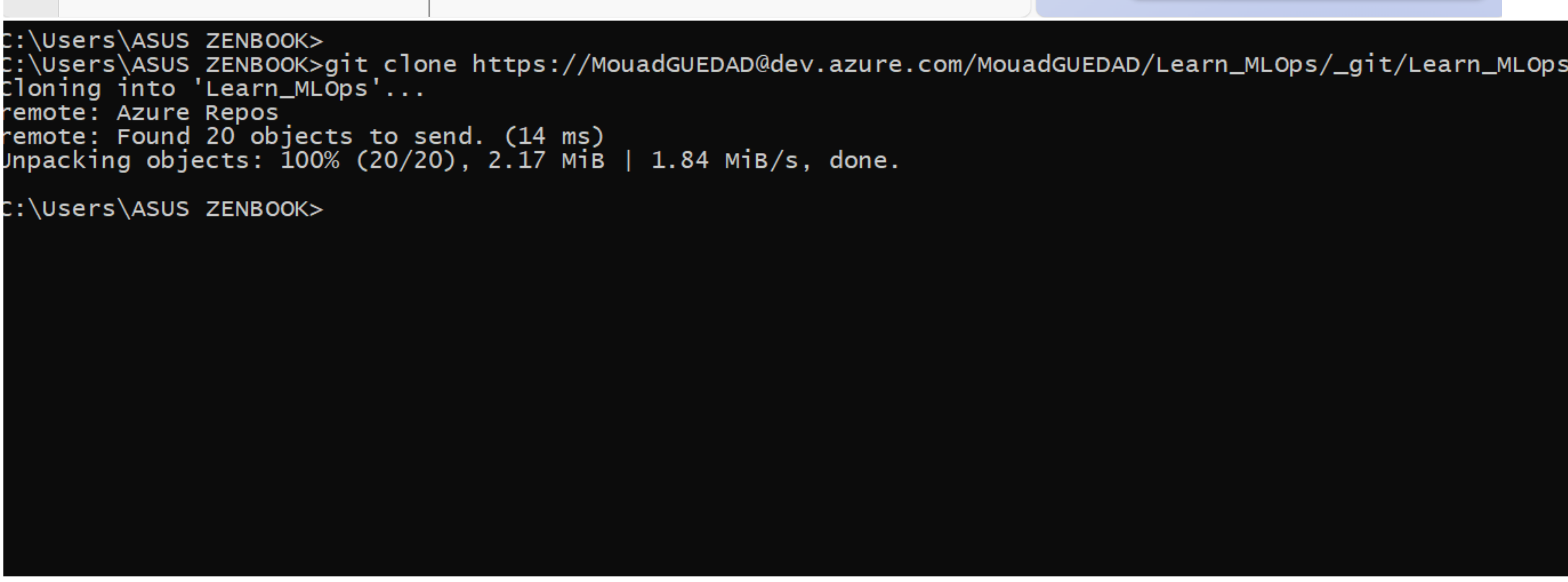
1.3 Azure DevOps



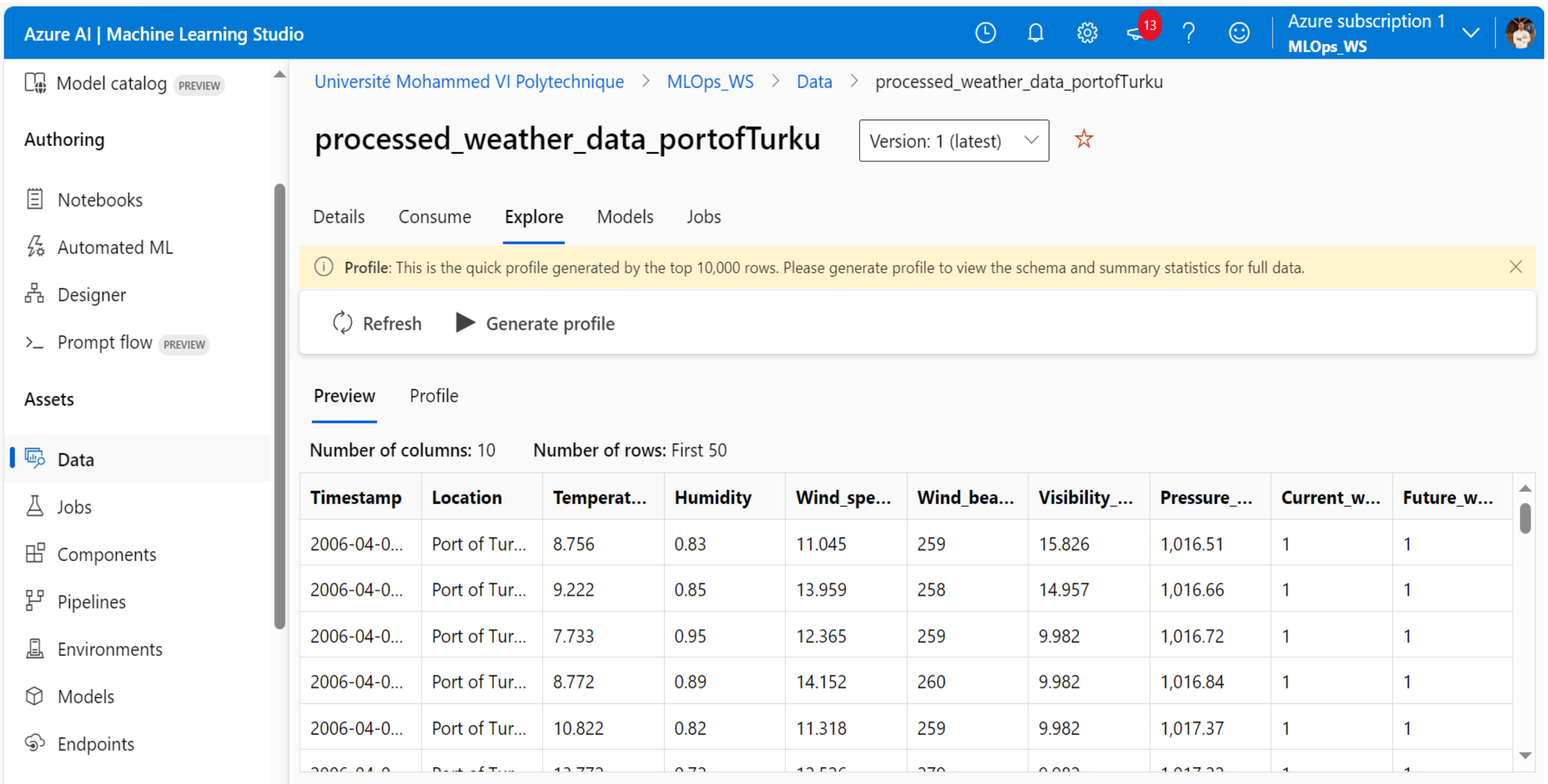
2. Data preprocessing

After I follow the instruction you gave us in the second steep (data preprocessing) I get :



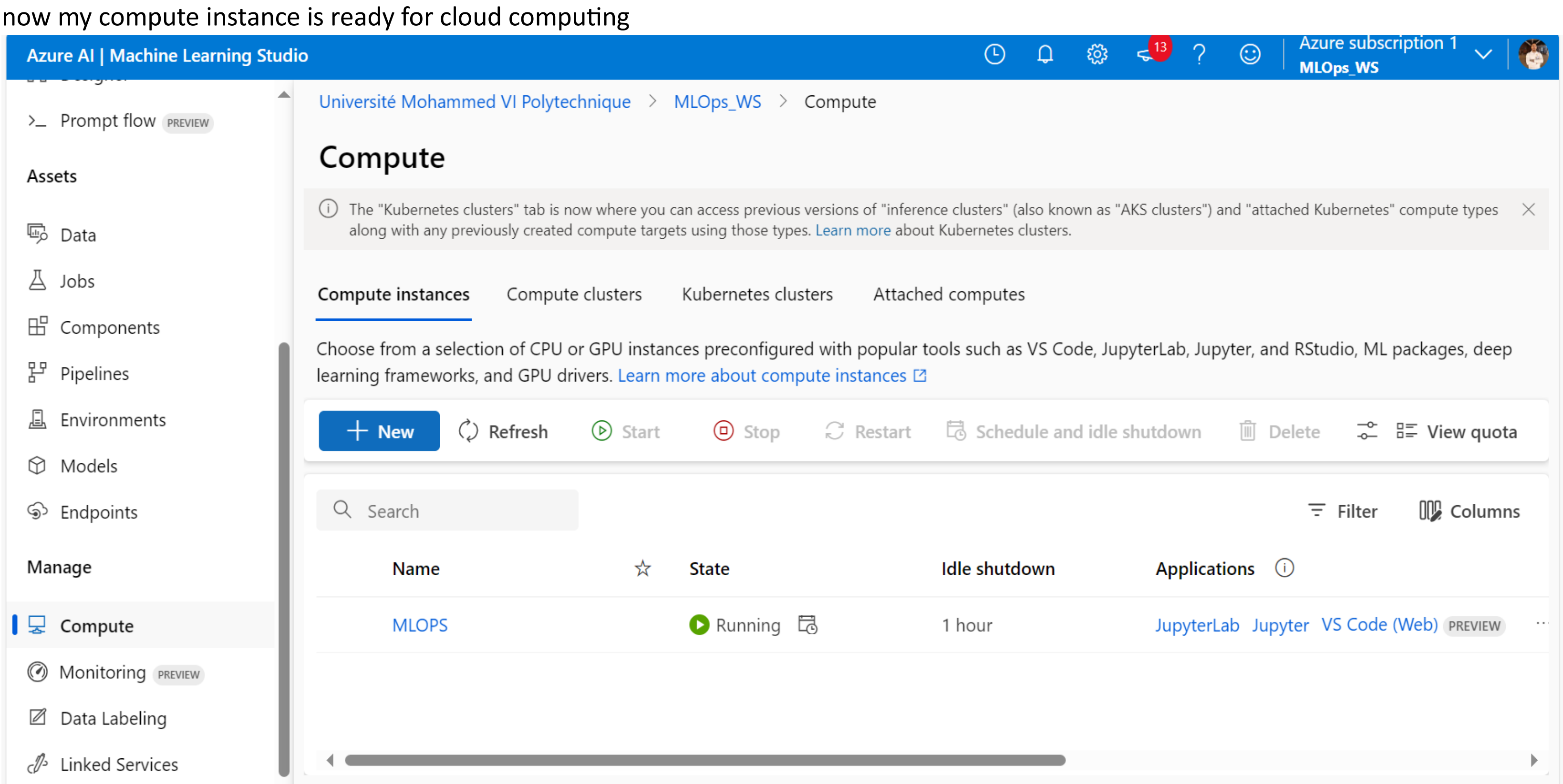


Then after the clonage the the dataset saved on the workspace is :



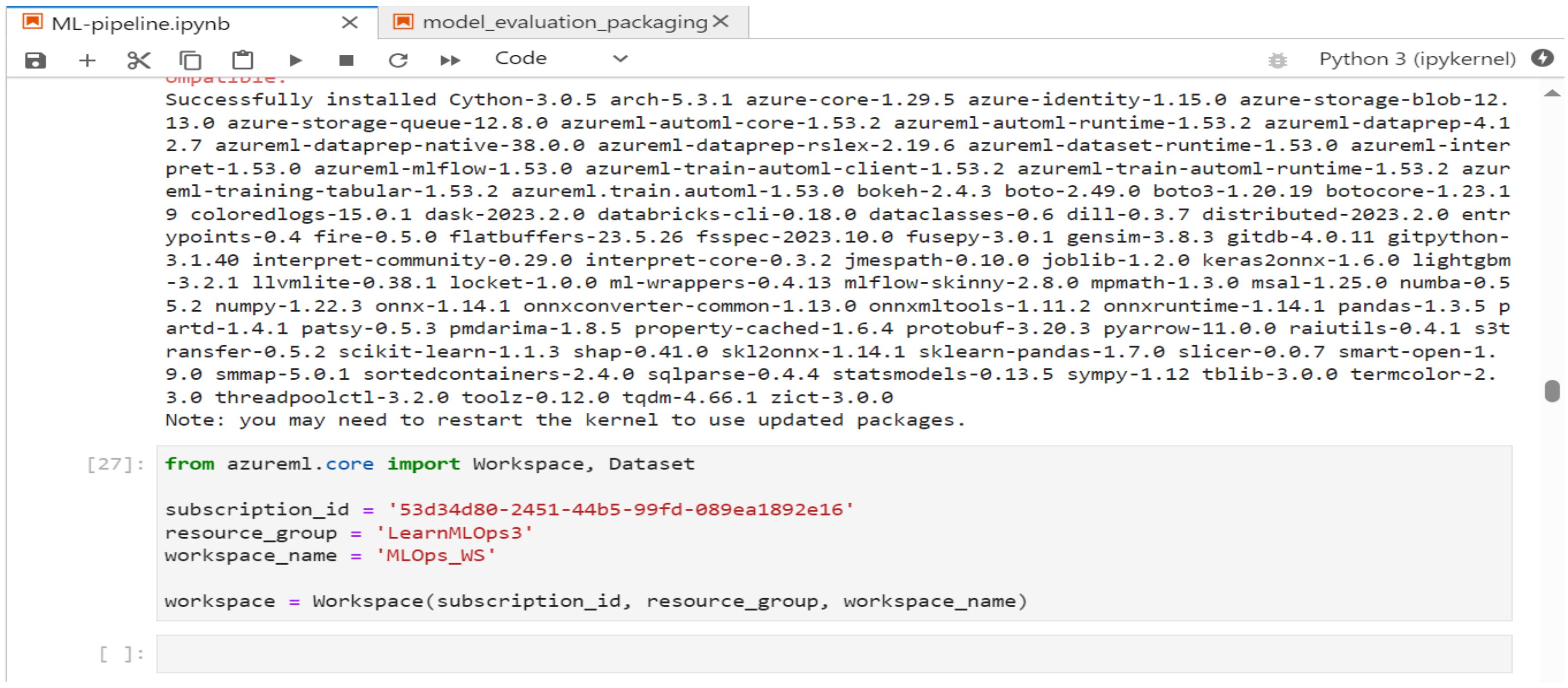
3. Pipeline construction

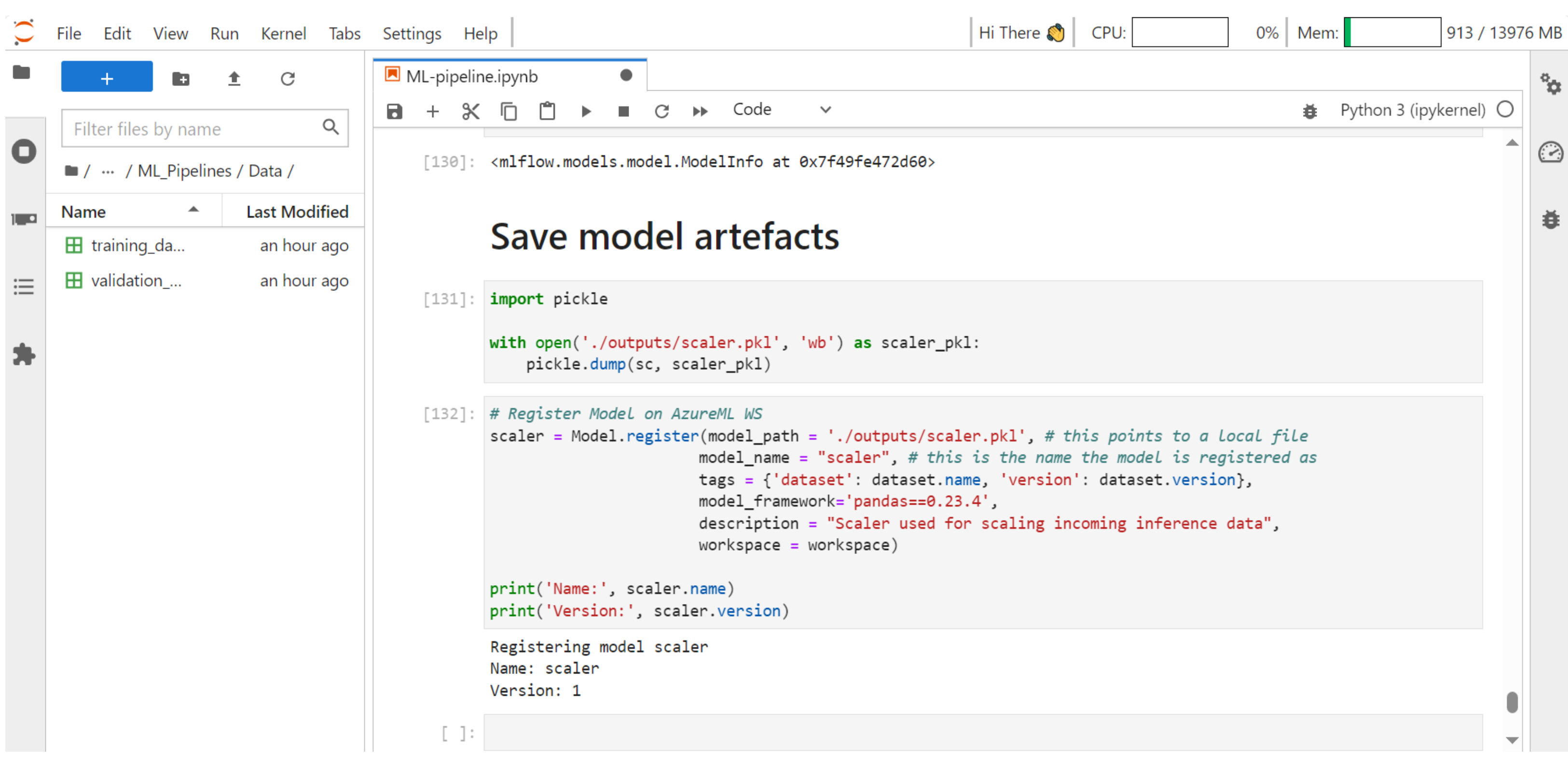
I processed the data on my local computer. For training ML models and implementing the pipeline, I use computing resources provisioned in the cloud (Microsoft Azure). In the following, I will create an Azure compute instance.
now my compute instance is ready for cloud computing



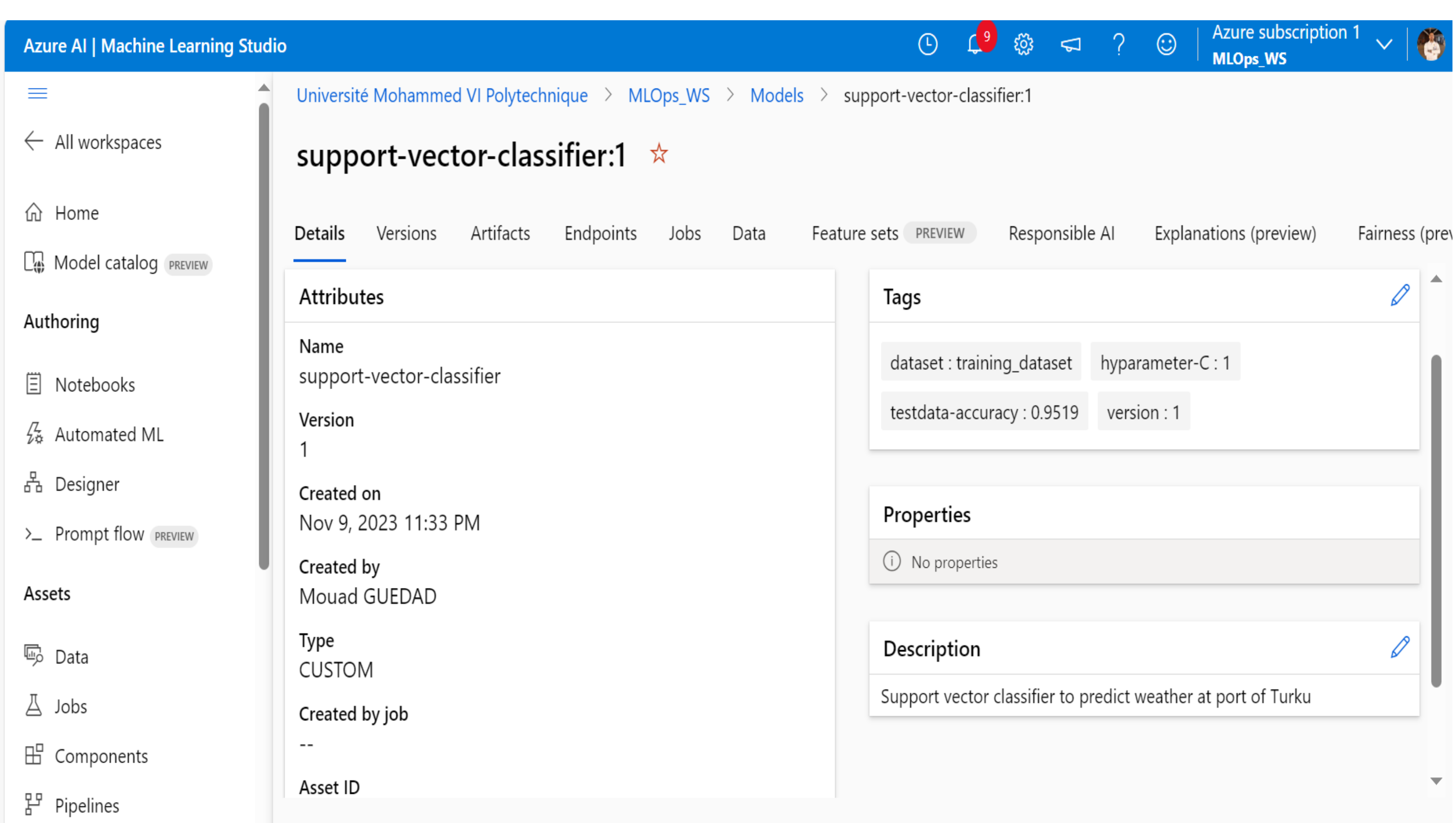
Once it is provisioned, I selected the JupyterLab option. JupyterLab is an open source web user interface. It comes with features like a text editor, code editor, terminal, and custom components built in an extensible manner. I used it as a programming interface connected to the provisioned compute to train the ML models.

After I cloned the respository to the Azure instance.





After running the notebook, i saved the model and the artifacts (parameters, version, etc.) on Azure machine learning studio by clicking on “Models”



4. Evaluating and packaging our models

The steps can be summarized as follows:

a. We must connect to the workspace and import the artifacts

We import the required packages, connect to the ML workspace using the Workspace() function, and then download the serialized scaler and model to make predictions. Scaler will be used to scale the input data into the same data scale used for training the model. The model file is serialized in ONNX format. Scaler and Model files are imported using the Model() function.

b. We load the artifacts for inference

We open and load the Scaler and Model files into variables that can be used for model inference. The scaler is read and loaded into a variable using pickle, and the ONNX runtime is used to load the ONNX file using InferenceSession() to make predictions

I followed exactly the instructions in the "Model_evaluation_packaging" folder and after I run the netobook I get :