Nishal Dbritto

A009

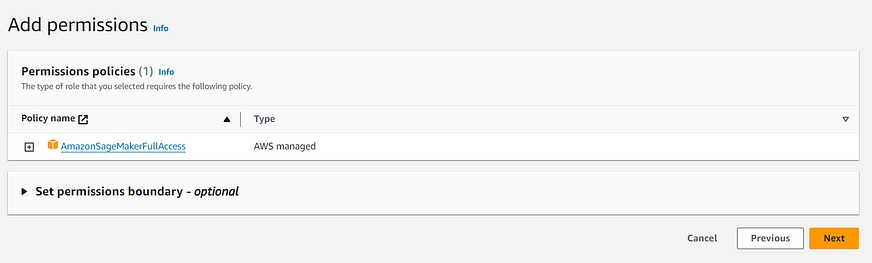
MSc SDS

**STEPS TO BUILD-TRAIN-DEPLOY MODEL IN AWS SAGEMAKER:**

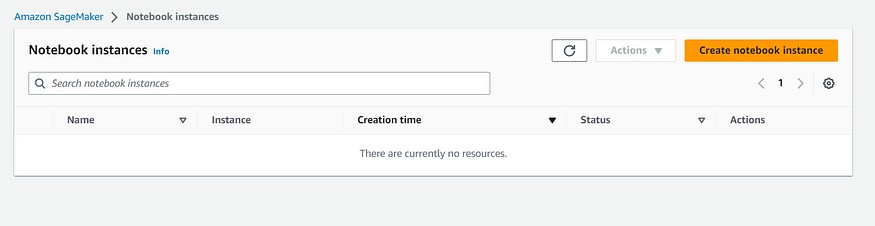
**STEP 1: CREATE IAM ROLE AND NOTEBOOK INSTANCE**

Create IAM role, choose Create a new role, and then choose Create role.

1. This IAM role automatically gets permissions to access any S3 bucket that has sagemaker in the name. It gets these permissions through

the AmazonSageMakerFullAccess policy, which SageMaker attaches to the role.

1. Go to Amazon SageMaker console.
2. Select “Notebook instances” and click “Create notebook instance.”

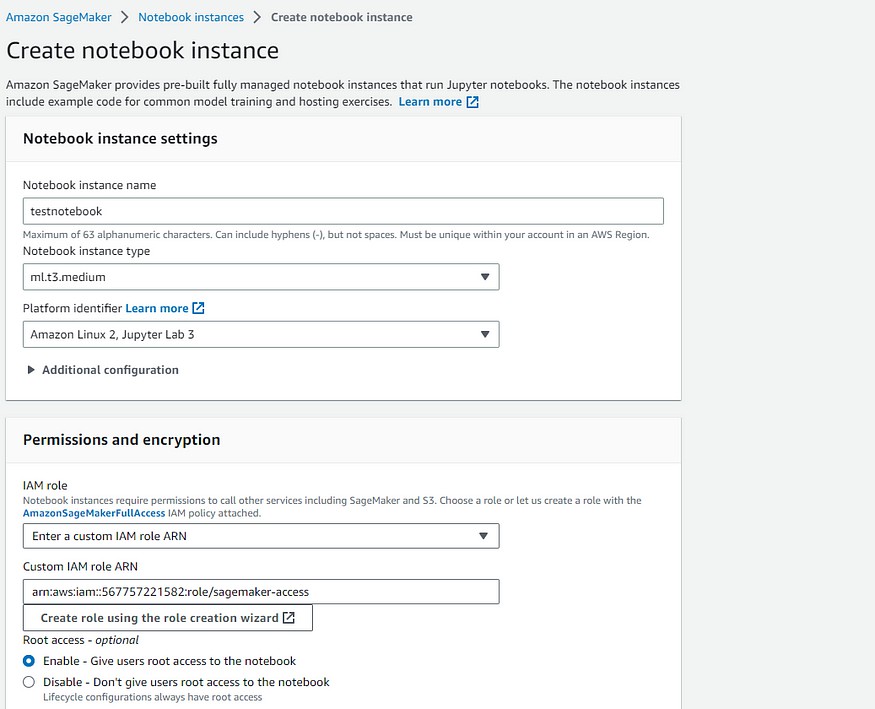


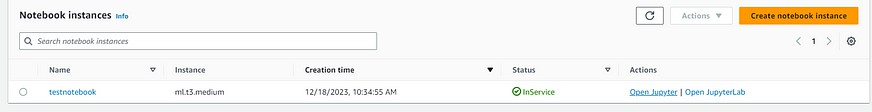
1. Fill in details:

Notebook instance name: Choose a name.

* + Notebook Instance type: Use “ml.t2.medium” (or “ml.t3.medium” if unavailable).
  + Platform Identifier: Choose a platform type for OS and JupyterLab version.

1. Leave other fields with default values.
2. Complete the creation process based on preferences.



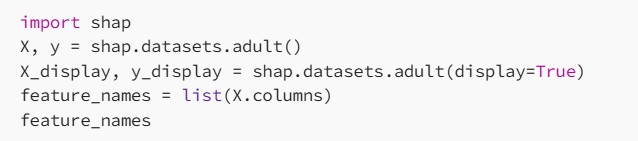
**STEP 2: CREATE A JUPYTER NOTEBOOK**

1. Open Jupyter or JupyterLab according to the interface needed.
2. Go to File menu->Choose New-> Notebook.
3. Select Kernel as ‘conda\_python3’

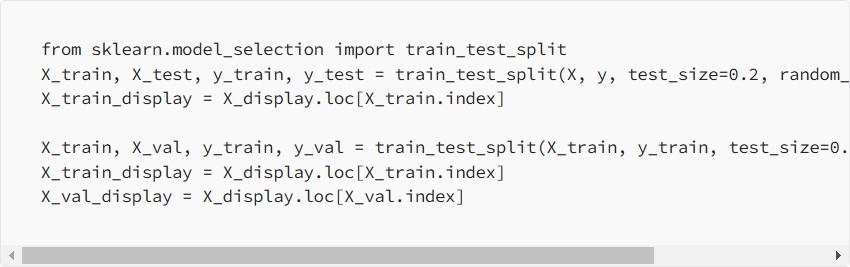
**STEP 3: DOWNLOAD, EXPLORE AND TRANSFORM DATA**

1. Load Adult Census dataset using SHAP

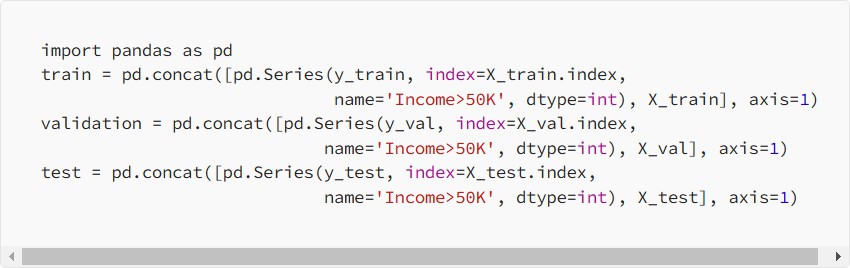
*(The dataset:* [*https://archive.ics.uci.edu/dataset/2/adult*](https://archive.ics.uci.edu/dataset/2/adult)*)*

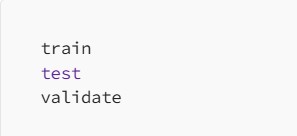


1. Split the dataset into train ,test and validation datasets

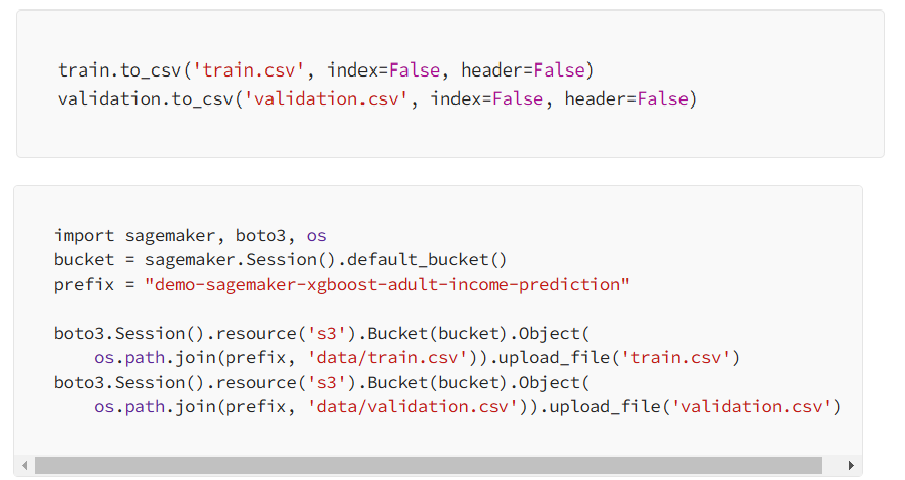


1. Aligning dataset by concatenating the numeric features with the true labels.



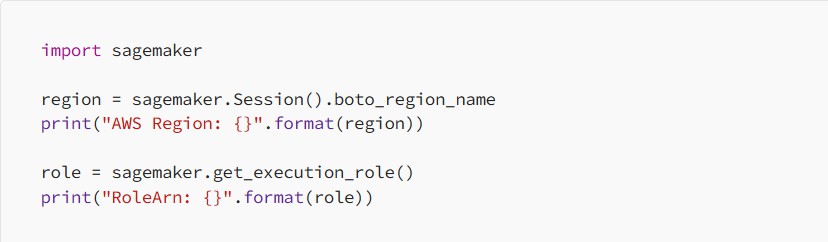
1. Check if the dataset is split properly.
2. Convert train, test and validation datasets to CSV

We need to convert these datasets to match the input file format for XGBoost algorithm which we will be using for model training and deployment and upload dataset to S3

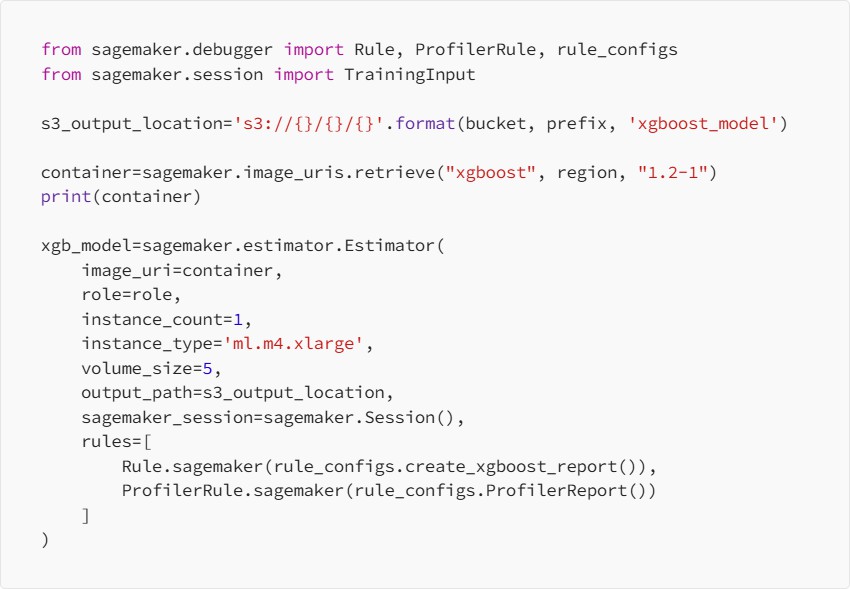


**STEP 4: TRAIN THE MODEL**

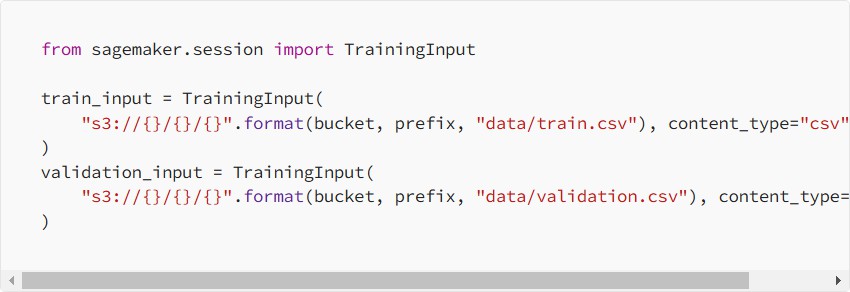
1. Retrieve region and role.



1. Create an XGBoost estimator.

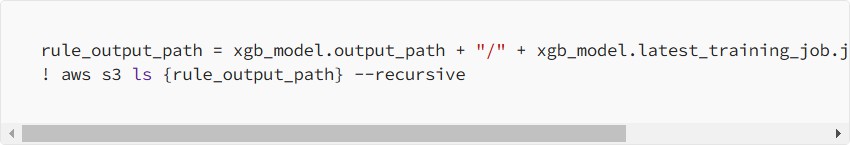
XGBoost, short for Extreme Gradient Boosting, is a machine learning algorithm that falls under the category of gradient-boosted decision trees.

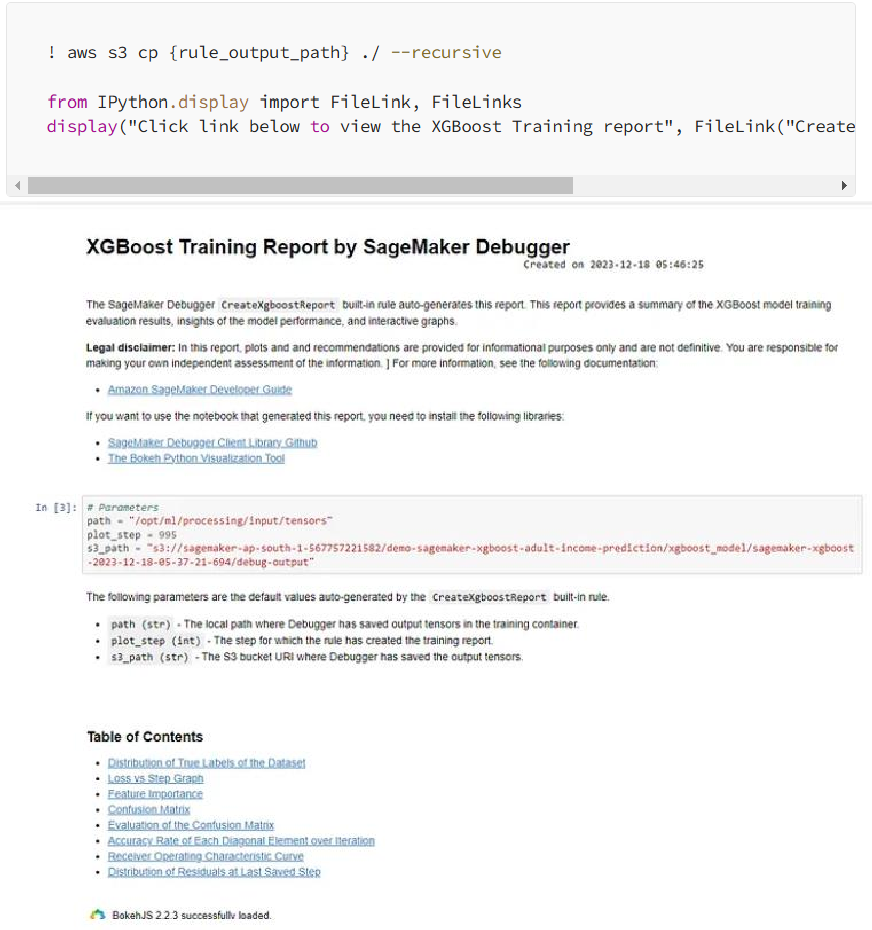
1. Set hyperparameters for XGBoost algorithm.
2. Using TrainingInput class to configure data input flow.

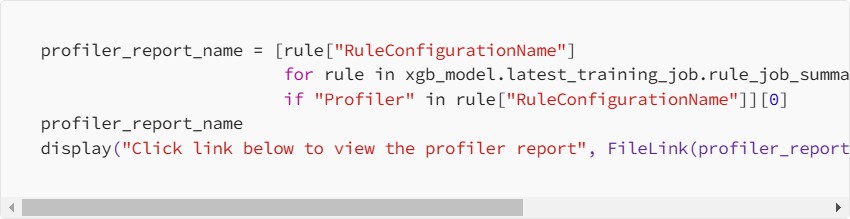


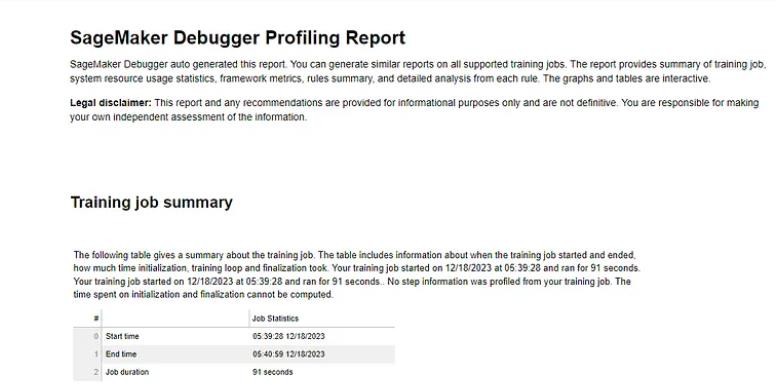
1. Train the model



1. Check the location of debugging reports generated.
2. Download the Debugger XGBoost reports and generate link.



1. Get profiling reports for instance resource utilization, system bottleneck etc.



**STEP 5: DEPLOY THE MODEL**



**STEP 6: EVALUATE THE MODEL**

