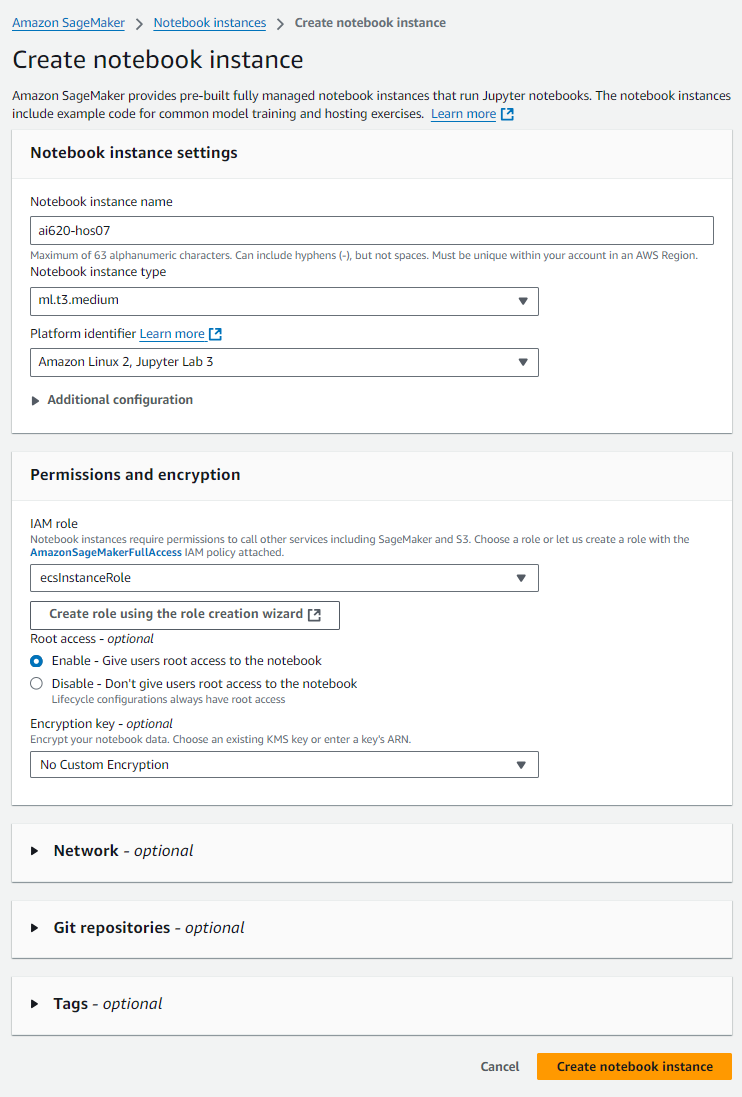
HOS07A Neural Topic Model (NTM) in SageMaker

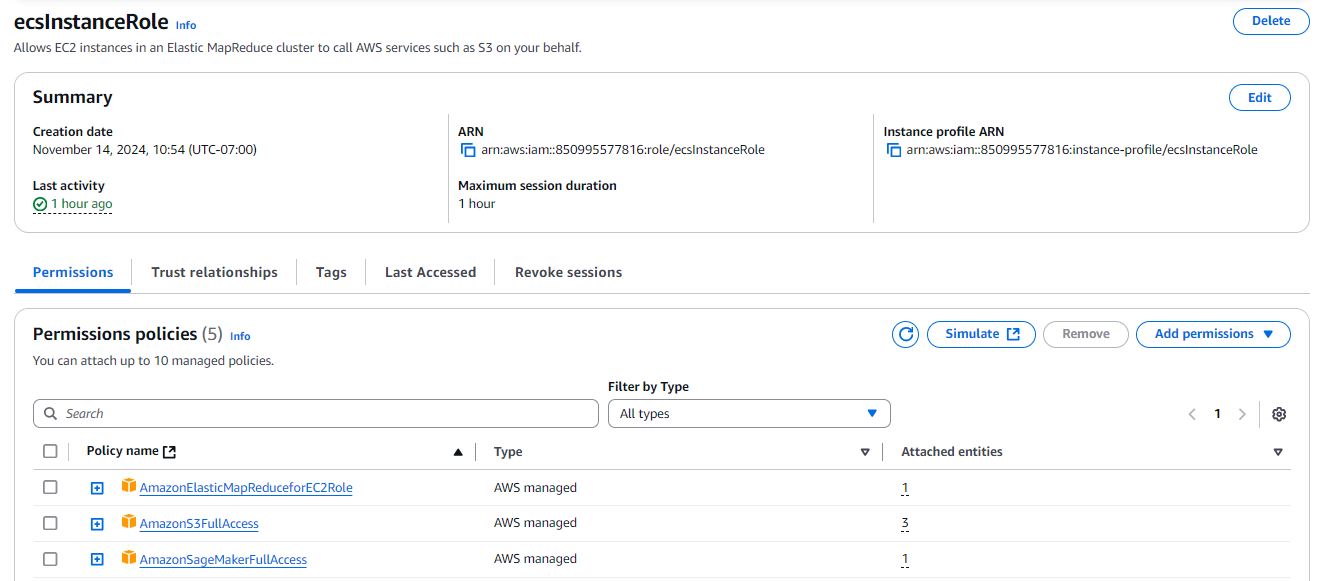
Fall 2024 by Verónica Elze

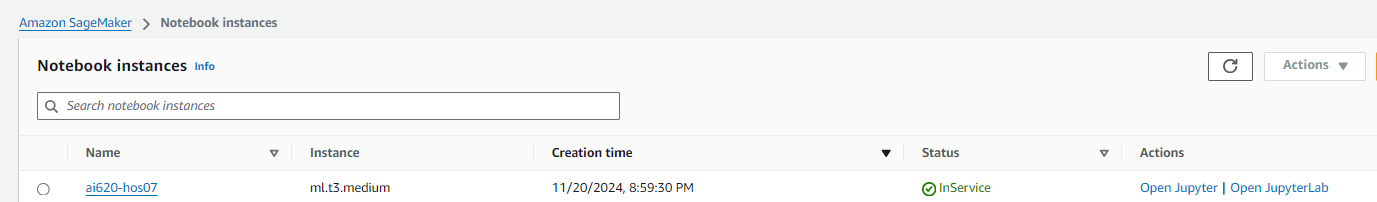
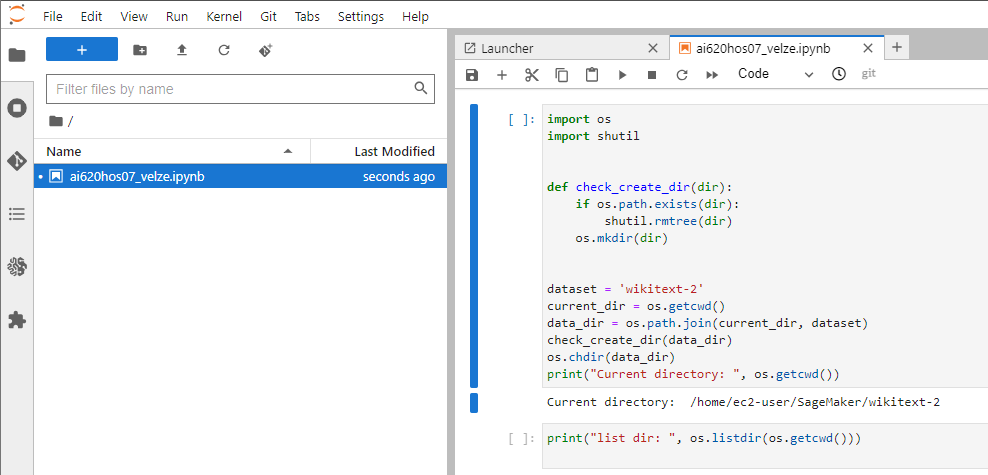
# Setup Notebook for NTM

1. Create Notebook Instance

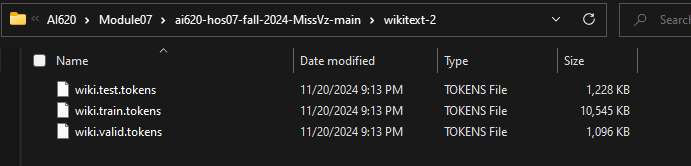
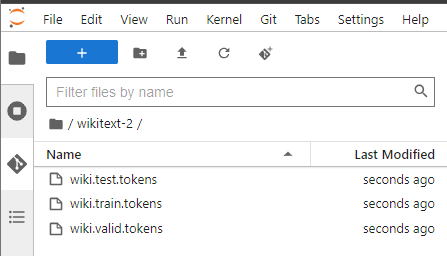


1. Create an IAM role with Any S3 Bucket access

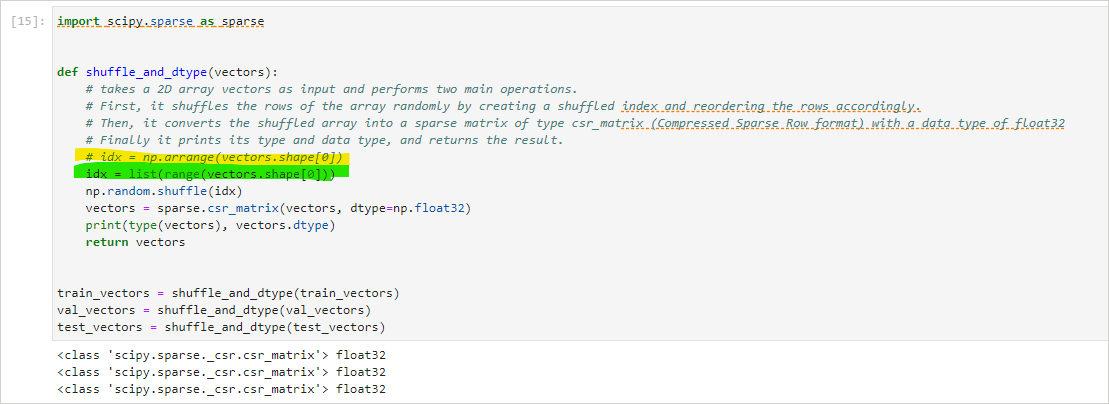
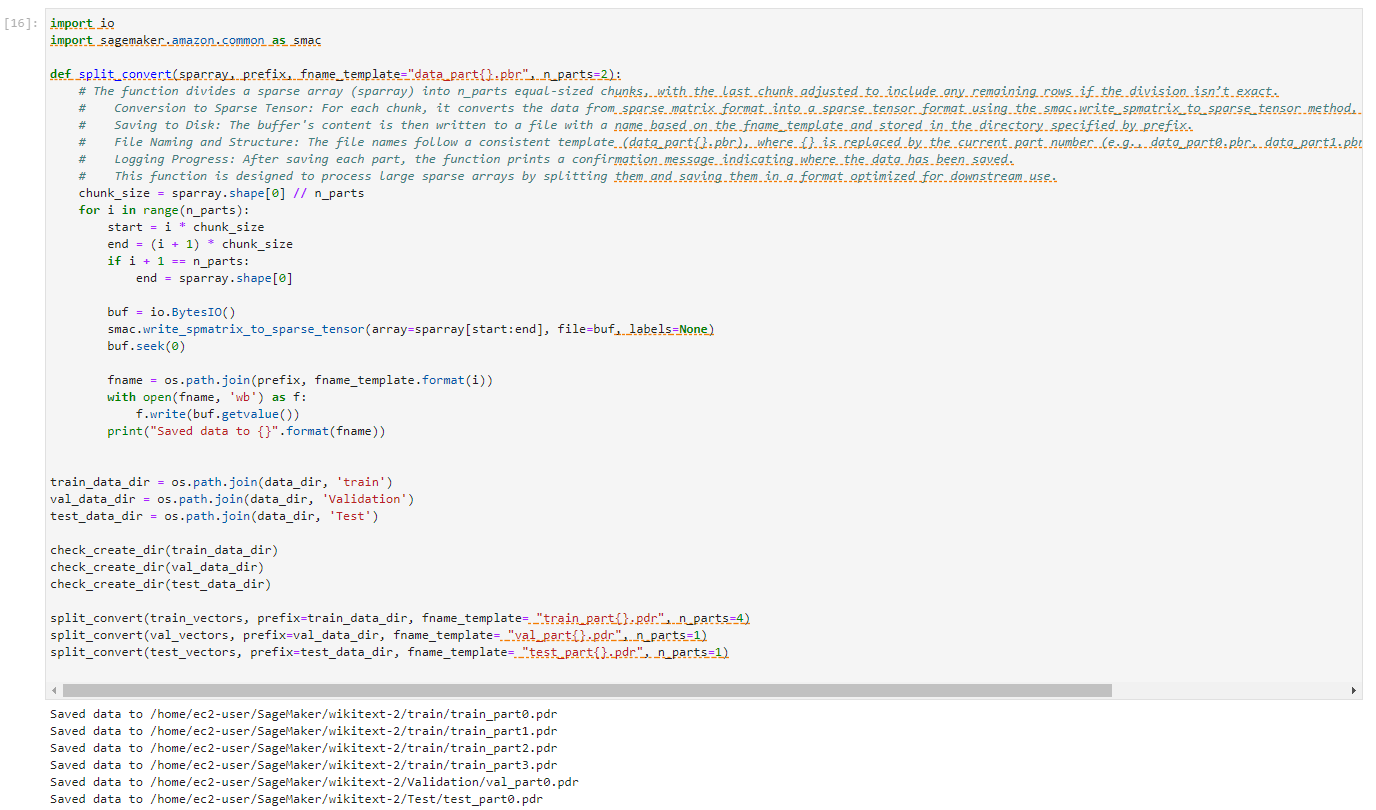


1. Wait until notebook instance Status is InService  
   
2. Upload starter notebook  
   

# Training NTM in SageMaker

1. Fetch data set  
     
   
2. Preprocessing  
   A screenshot of a computer program

   Description automatically generated
3. Install and import nltk  
   
4. Perform lemmatizing and counting text  
   A screenshot of a computer program

   Description automatically generated
5. Shuffle, convert to sparce format, and apply transformation to training, validation, and test datasets  
   
6. Function to Split and Save Sparse Matrices for Distributed Processing  
   
7. Save text file  
   A computer code with text

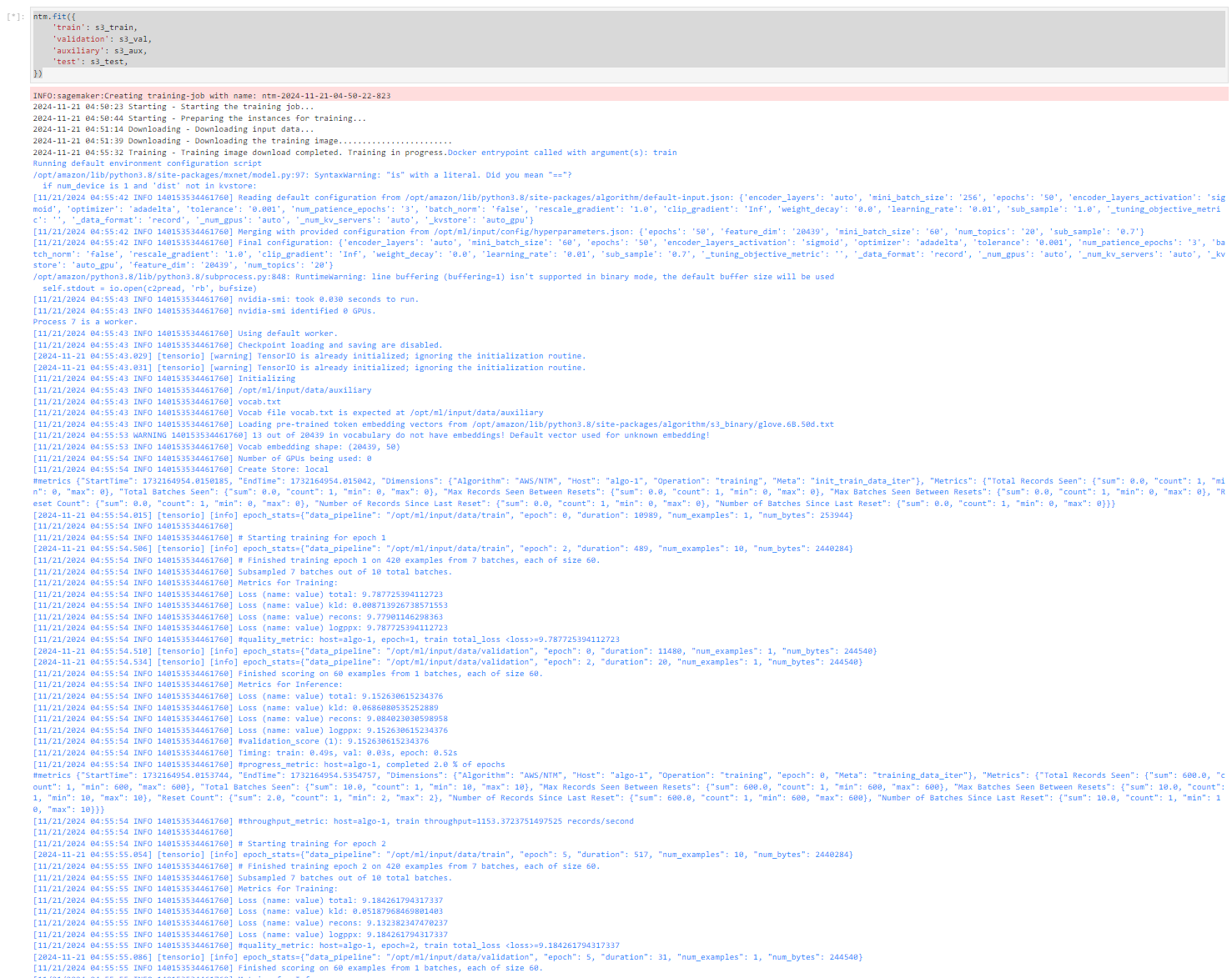
   Description automatically generated

# Store Data on S3

1. Specify data locations and access roles  
   
2. Upload Input Directories to S3  
   A screenshot of a computer

   Description automatically generated

# Model Training

1. Configure SageMaker training job to use NTM algorithm  
   
2. Run training job  
     
   A screen shot of a computer

   Description automatically generated
3. View training job status in AWS SageMaker Console  
     
   A screenshot of a computer

   Description automatically generated

A screenshot of a computer

Description automatically generated