Assignment 5: Model Deployment

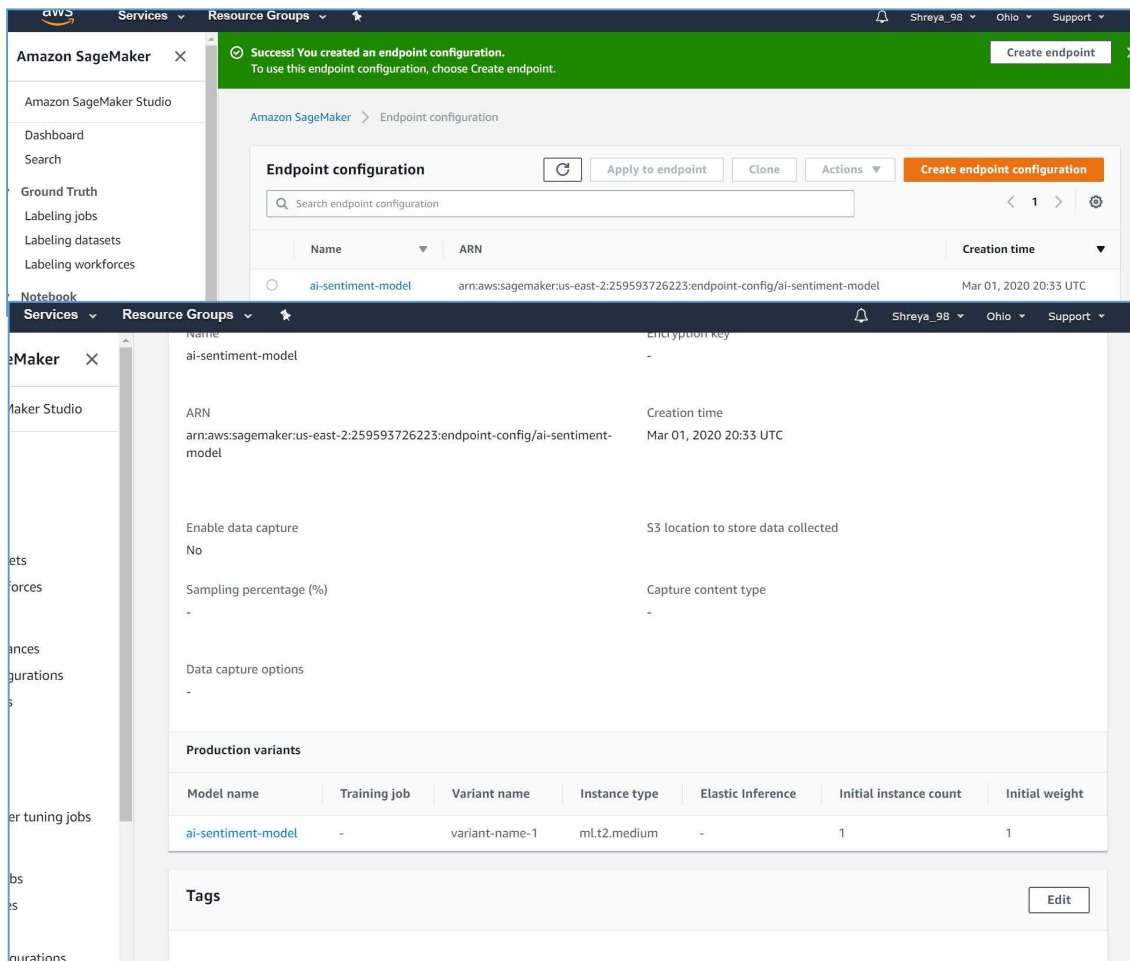Github repository link: https://github.com/Shreya98-code/assign5-aiops

Deploying model on AWS using the following model: https://aiops-2020-public.s3.us-east-2.amazonaws.com/model.tar.gz
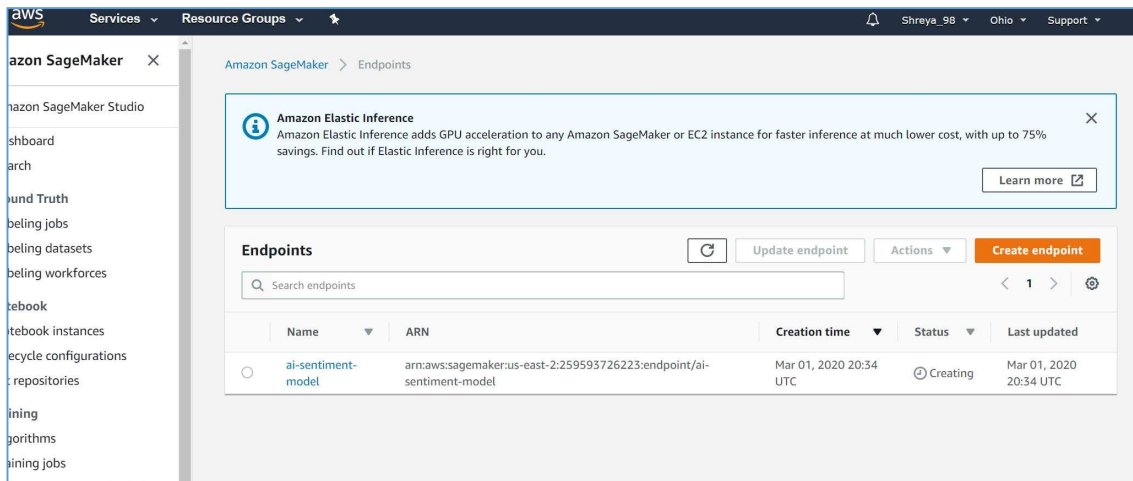
1. Sagemaker Inference
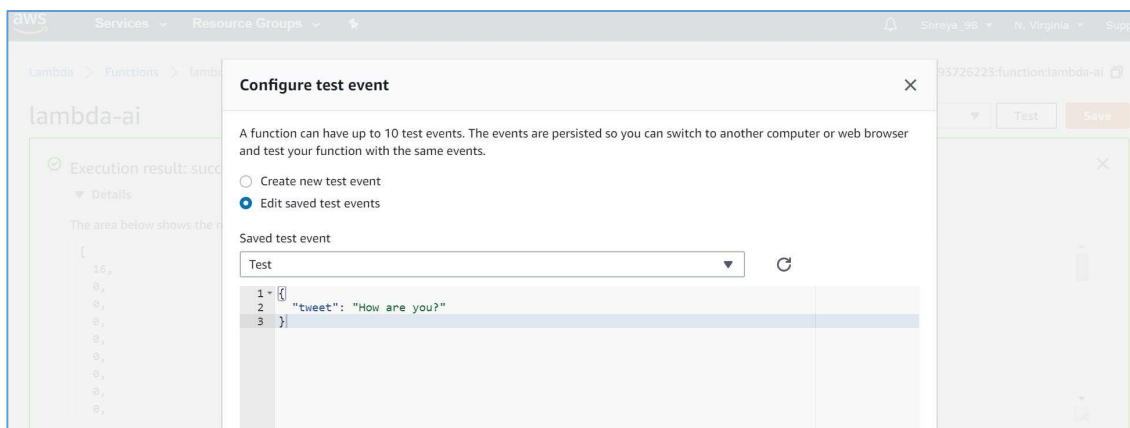
- Creating Model: ai-sentiment-model
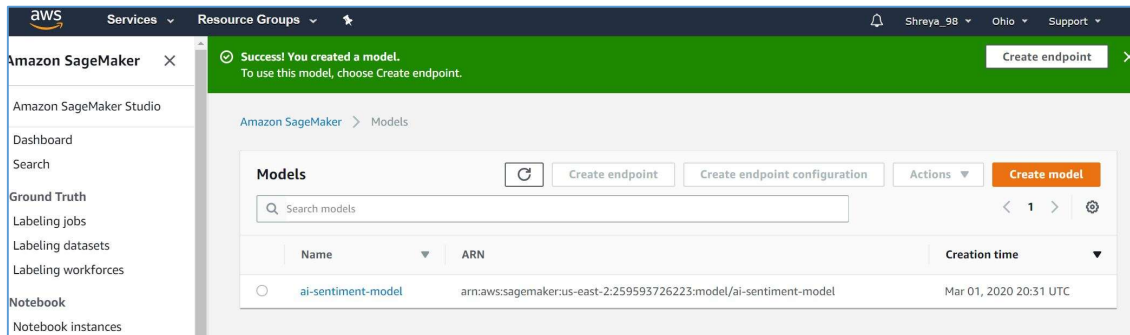
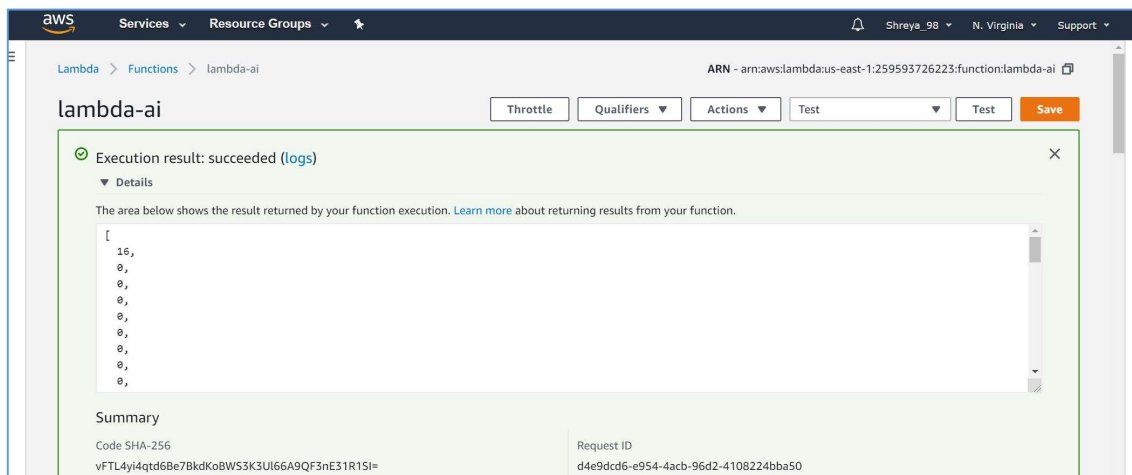- Creating end-point configuration with mk.t2.medium instance

- Creating end-point

2. Lambda Functions: (Note: lambda-ai and sentiment-lambda-a5 have same functionality.)

- Output when lambda function performs preprocessing using earlier code on above tweet.

- Model Inference Output on another tweet:

- Post-processing output using logic

3. Payload Logging:

- Modifying Lambda function to display additional features such as processing time.

Testing it on a latest tweet by President Donald Trump.

- Modifying lambda function further to implement payload logging: Each request creating a unique json in the S3 directory
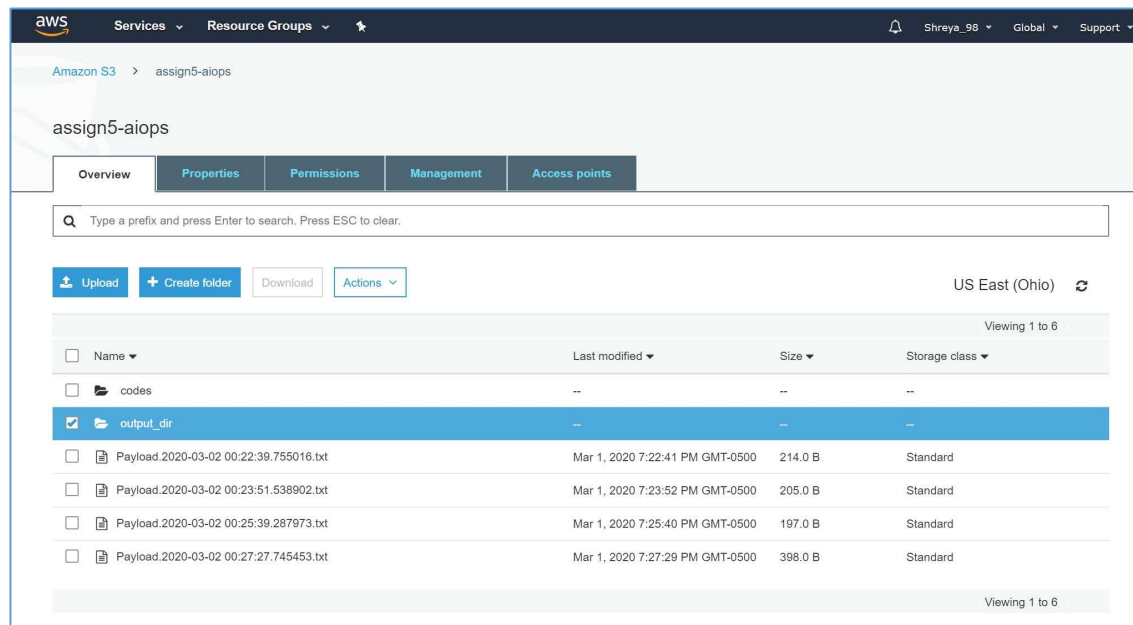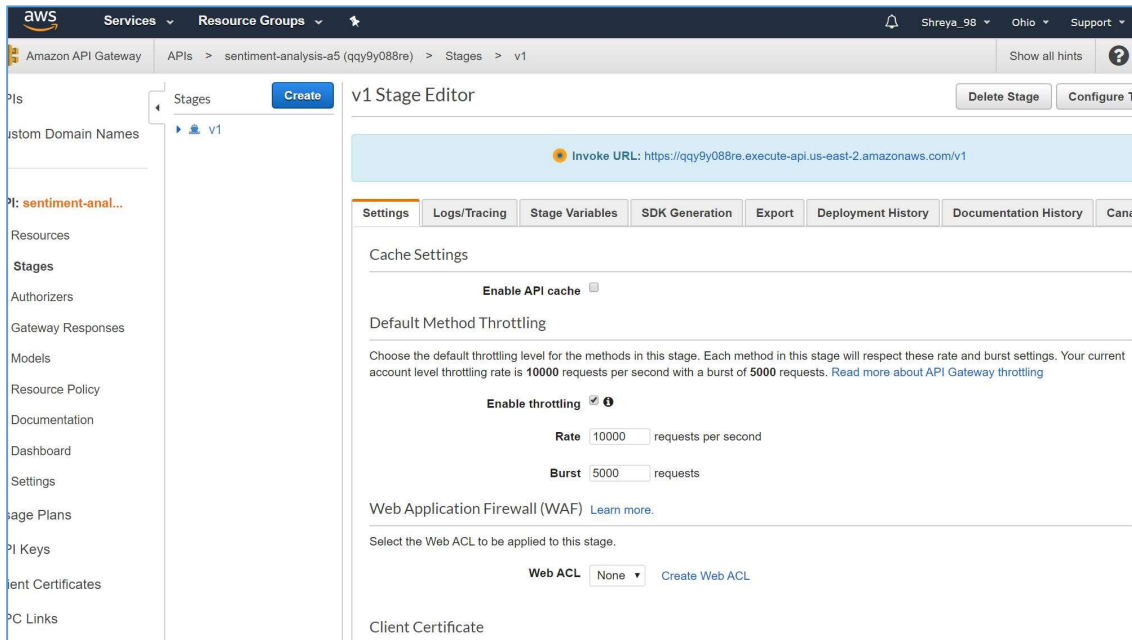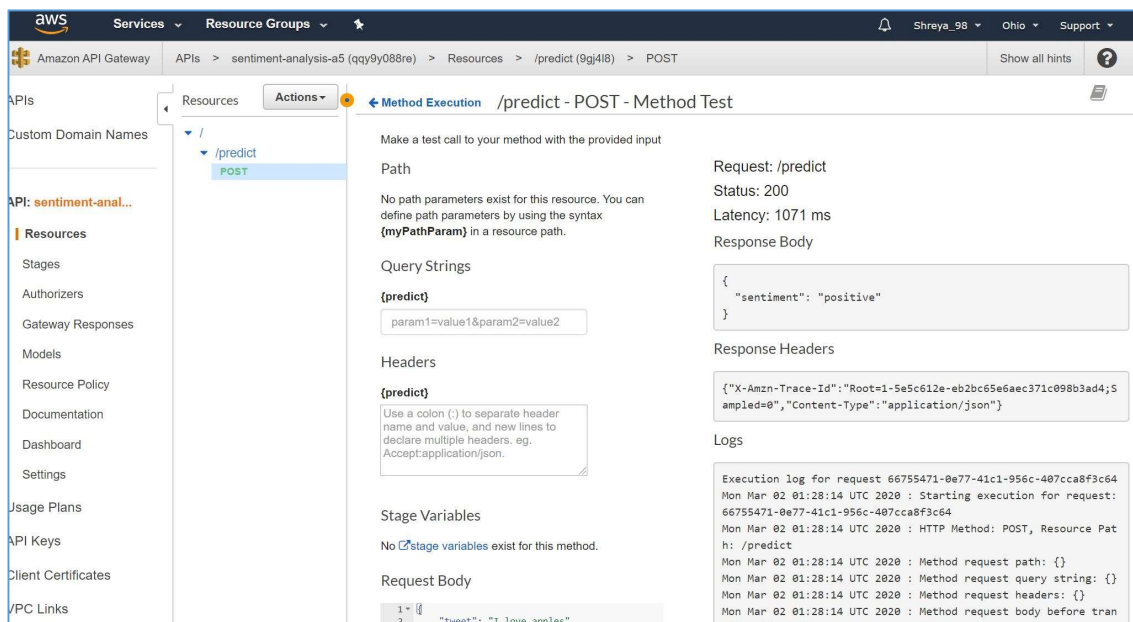

4. REST API

- Creating an API gateway to expose lambda function sentiment-lambda-a5. Implements /predict resource with POST method
- API Gateway url: https://qqy9y088re.execute-api.us-east-2.amazonaws.com/v1
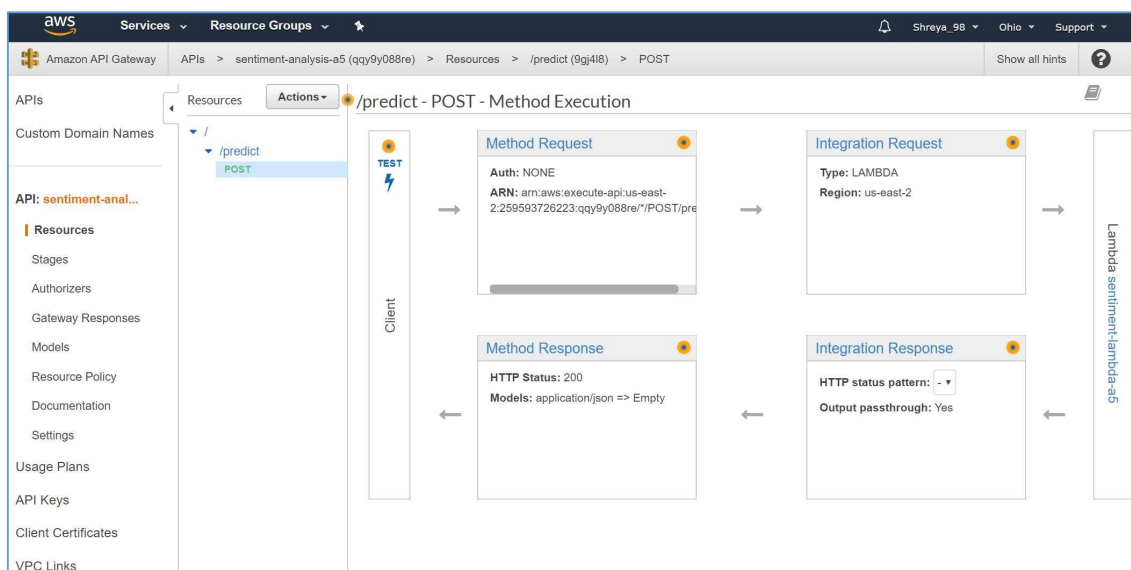
Testing online:

Deploying under v1 stage locally: