

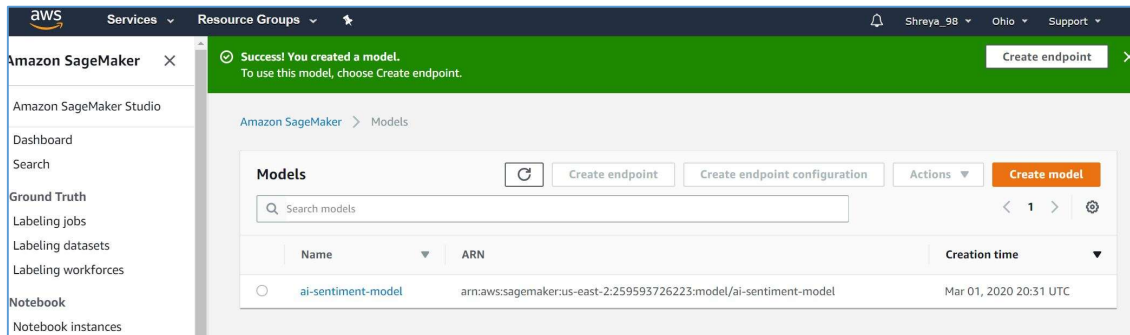
Assignment 5: Model Deployment

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

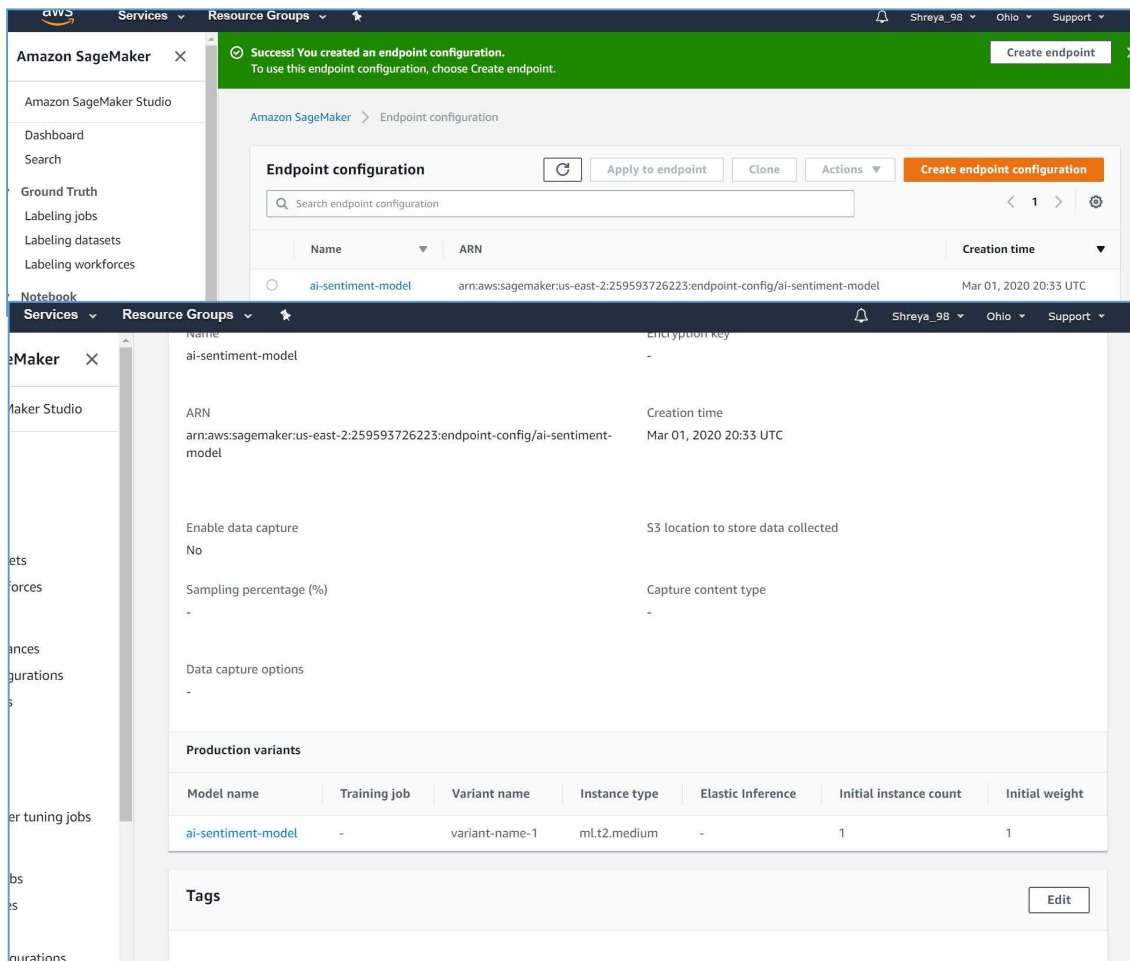
Deploying model on AWS using the following model: <https://aiops-2020-public.s3.us-east-2.amazonaws.com/model.tar.gz>

1. Sagemaker Inference

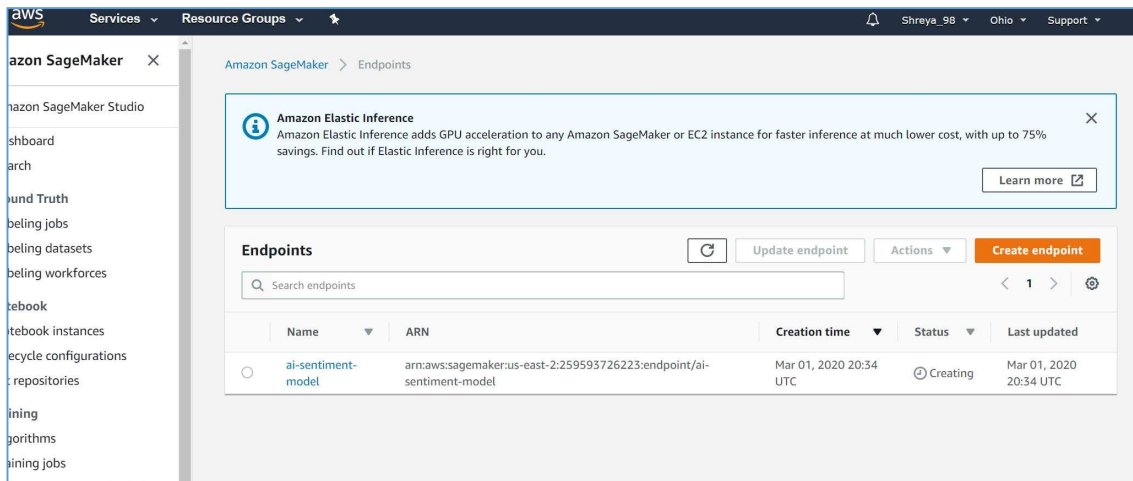
- Creating Model: ai-sentiment-model



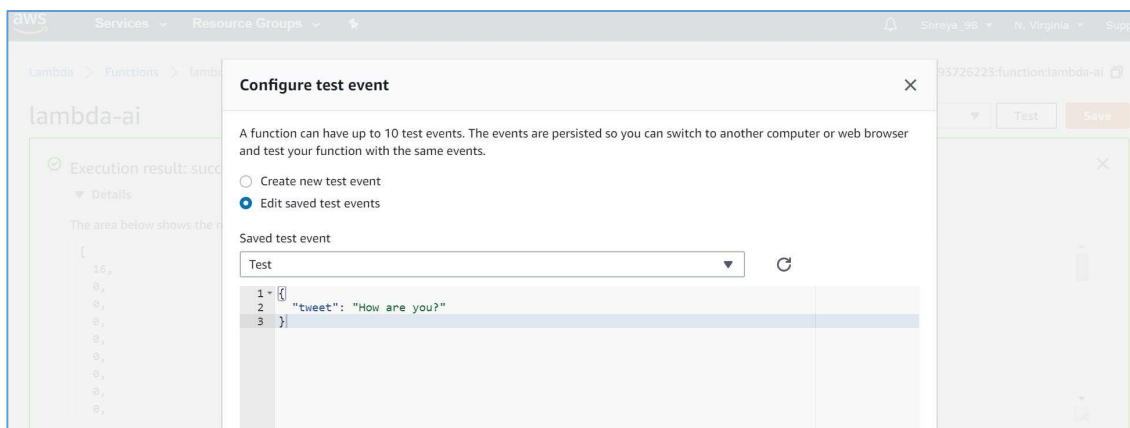
- Creating end-point configuration with mk.t2.medium instance



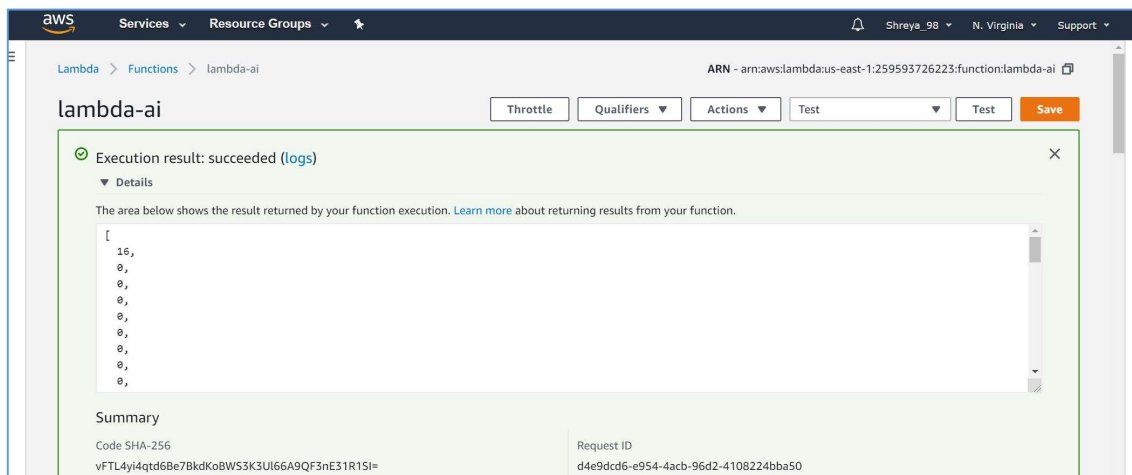
- Creating end-point



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:

Successfully updated the function **sentiment-lambda-a5**.

Lambda > Functions > sentiment-lambda-a5

ARN - arn:aws:lambda:us-east-2:259593726223:function:sentiment-lambda-a5

sentiment-lambda-a5 [Throttle] [Qualifiers] [Actions] [Test] [Test] [Save]

Execution result: succeeded (logs)

Details

The area below shows the result returned by your function execution. [Learn more](#) about returning results from your function.

```
{
  "predictions": [
    {
      0.807874203
    }
  ]
}
```

Summary

Code SHA-256	Request ID
6tpW6hvhJ/IN64YBN0pHul0xaBk9SMf4pTXPdgx4ovc=	0a0dd64c-781a-4329-ae35-a12b383412f7
Duration	Billed duration
594.22 ms	600 ms
Resources configured	Max memory used
128 MB	84 MB Init Duration: 453.02 ms

Log output

The section below shows the logging calls in your code. These correspond to a single row within the CloudWatch log group corresponding to this Lambda function. [Click here](#) to view the CloudWatch log group.

```
START RequestId: 0a0dd64c-781a-4329-ae35-a12b383412f7 Version: $LATEST
Inside
```

- Post-processing output using logic

Successfully updated the function **sentiment-lambda-a5**.

Lambda > Functions > sentiment-lambda-a5

ARN - arn:aws:lambda:us-east-2:259593726223:function:sentiment-lambda-a5

sentiment-lambda-a5 [Throttle] [Qualifiers] [Actions] [Test] [Test] [Save]

Execution result: succeeded (logs)

Details

The area below shows the result returned by your function execution. [Learn more](#) about returning results from your function.

```
{
  "sentiment": "positive"
}
```

Summary

Code SHA-256	Request ID
Ewfx/BpOpbeJku3Ouo9GDZQICwI+hZvYh/sQG9yqDgU=	3d94d7b0-38a9-42ec-9d62-37f2741783dc
Duration	Billed duration
341.58 ms	400 ms
Resources configured	Max memory used
128 MB	84 MB Init Duration: 443.20 ms

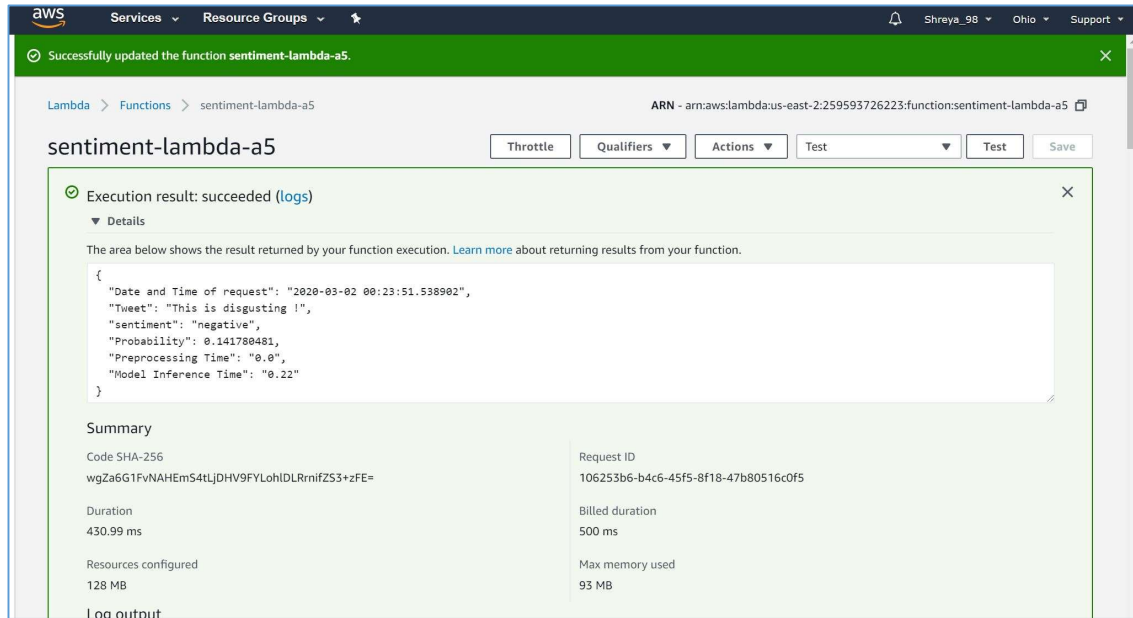
Log output

The section below shows the logging calls in your code. These correspond to a single row within the CloudWatch log group corresponding to this Lambda function. [Click here](#) to view the CloudWatch log group.

```
START RequestId: 3d94d7b0-38a9-42ec-9d62-37f2741783dc Version: $LATEST
```

3. Payload Logging:

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



Successfully updated the function **sentiment-lambda-a5**.

Lambda > Functions > sentiment-lambda-a5

ARN - arn:aws:lambda:us-east-2:259593726223:function:sentiment-lambda-a5

sentiment-lambda-a5 [Throttle] [Qualifiers] [Actions] [Test] [Test] [Save]

Execution result: succeeded (logs)

Details

The area below shows the result returned by your function execution. [Learn more](#) about returning results from your function.

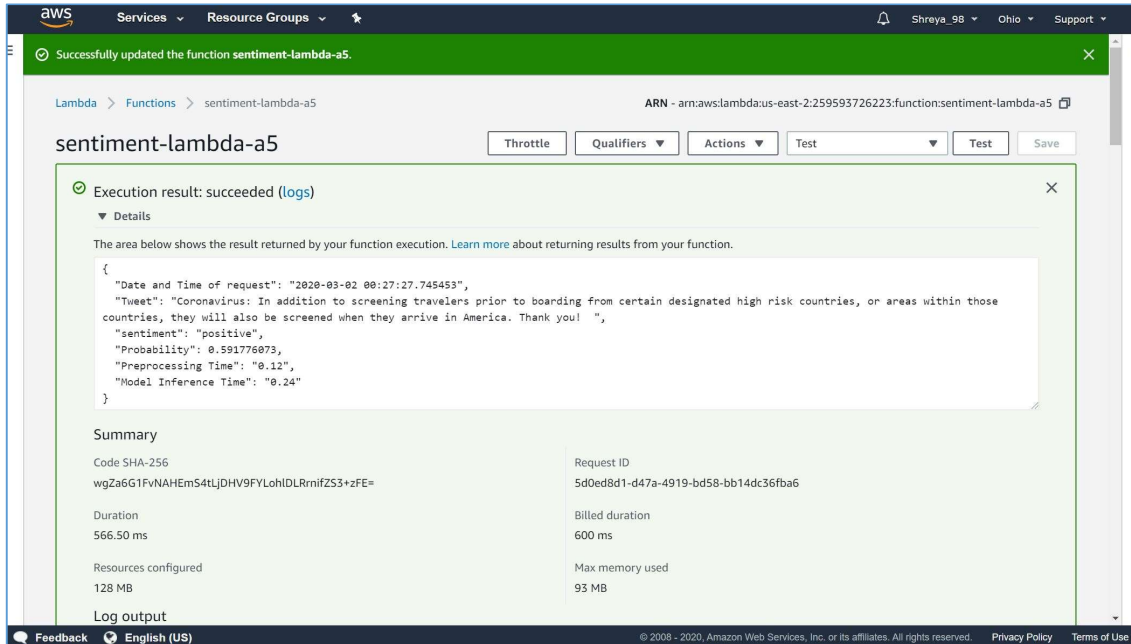
```
{
  "Date and Time of request": "2020-03-02 00:23:51.538982",
  "Tweet": "This is disgusting !",
  "sentiment": "negative",
  "Probability": 0.141780481,
  "Preprocessing Time": "0.0",
  "Model Inference Time": "0.22"
}
```

Summary

Code SHA-256 wgZa6G1FvNAHEmS4tLjDHV9FYLoHDLRmifZS3+zFE=	Request ID 106253b6-b4c6-45f5-8f18-47b80516c0f5
Duration 430.99 ms	Billed duration 500 ms
Resources configured 128 MB	Max memory used 93 MB

Log output

Testing it on a latest tweet by President Donald Trump.



Successfully updated the function **sentiment-lambda-a5**.

Lambda > Functions > sentiment-lambda-a5

ARN - arn:aws:lambda:us-east-2:259593726223:function:sentiment-lambda-a5

sentiment-lambda-a5 [Throttle] [Qualifiers] [Actions] [Test] [Test] [Save]

Execution result: succeeded (logs)

Details

The area below shows the result returned by your function execution. [Learn more](#) about returning results from your function.

```
{
  "Date and Time of request": "2020-03-02 00:27:27.745453",
  "Tweet": "Coronavirus: In addition to screening travelers prior to boarding from certain designated high risk countries, or areas within those countries, they will also be screened when they arrive in America. Thank you! ",
  "sentiment": "positive",
  "Probability": 0.591776073,
  "Preprocessing Time": "0.12",
  "Model Inference Time": "0.24"
}
```

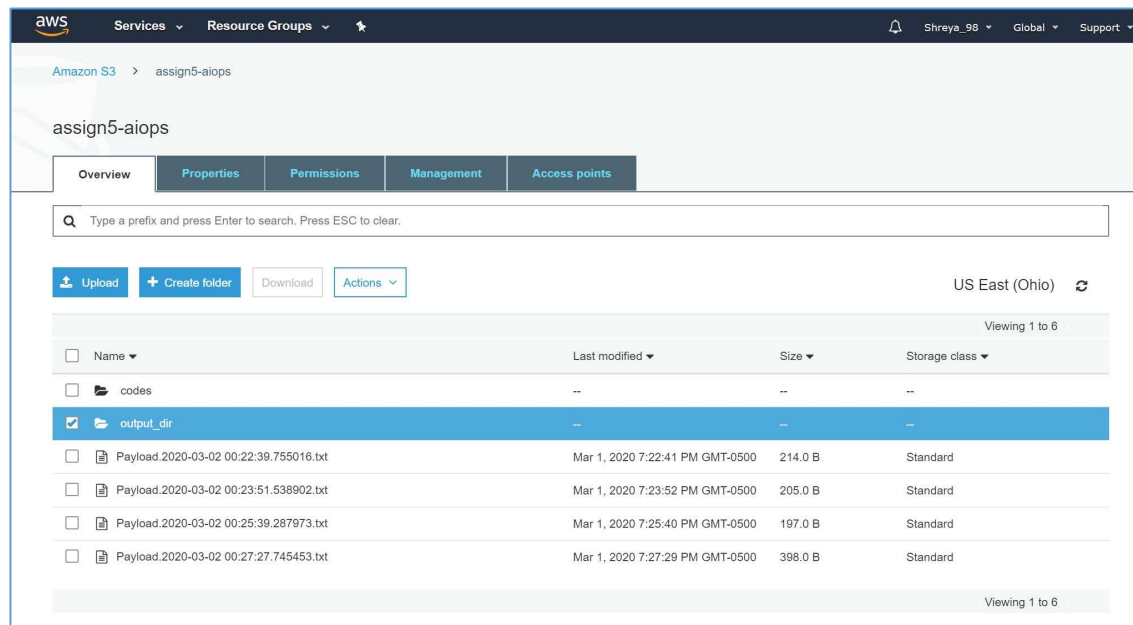
Summary

Code SHA-256 wgZa6G1FvNAHEmS4tLjDHV9FYLoHDLRmifZS3+zFE=	Request ID 5d0ed8d1-d47a-4919-bd58-bb14dc36fba6
Duration 566.50 ms	Billed duration 600 ms
Resources configured 128 MB	Max memory used 93 MB

Log output

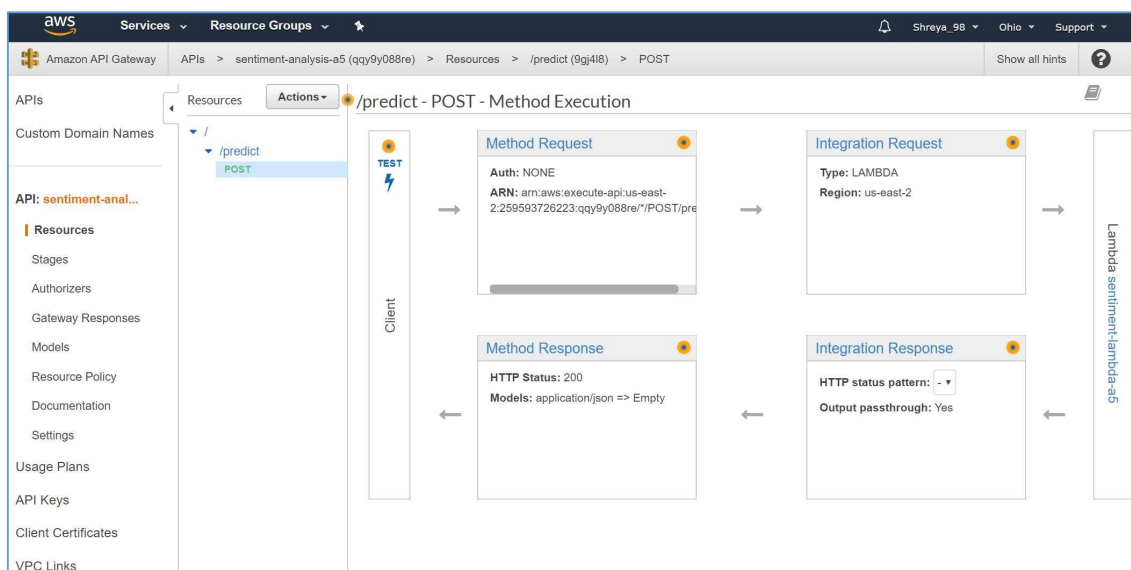
Feedback English (US) © 2009 - 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved. Privacy Policy Terms of Use

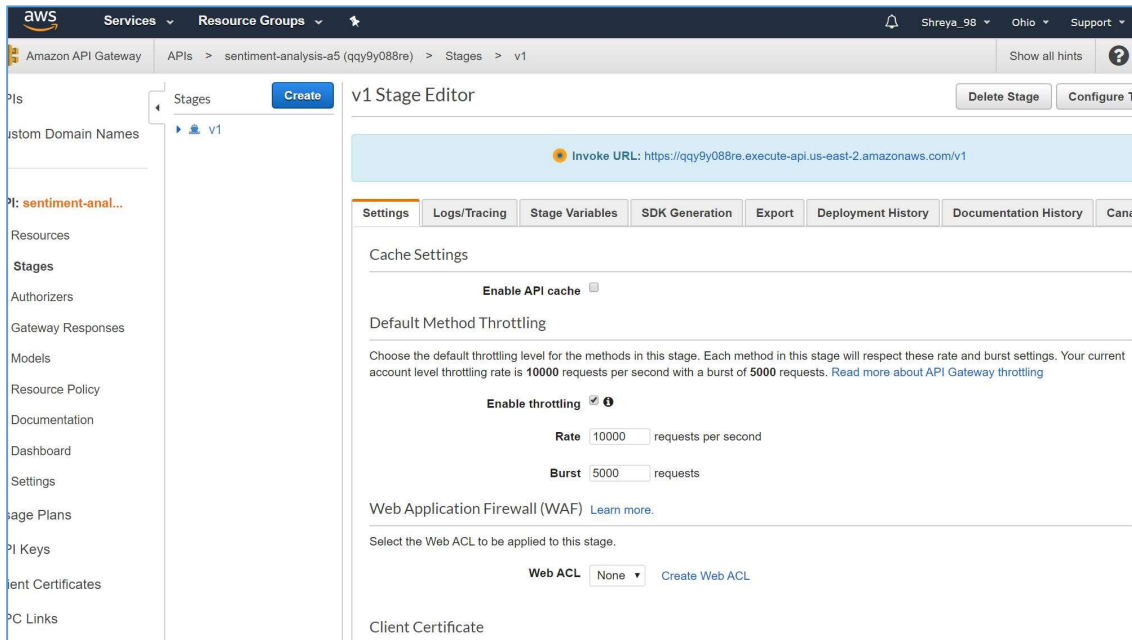
- Modifying lambda function further to implement payload logging: Each request creating a unique json in the S3 directory



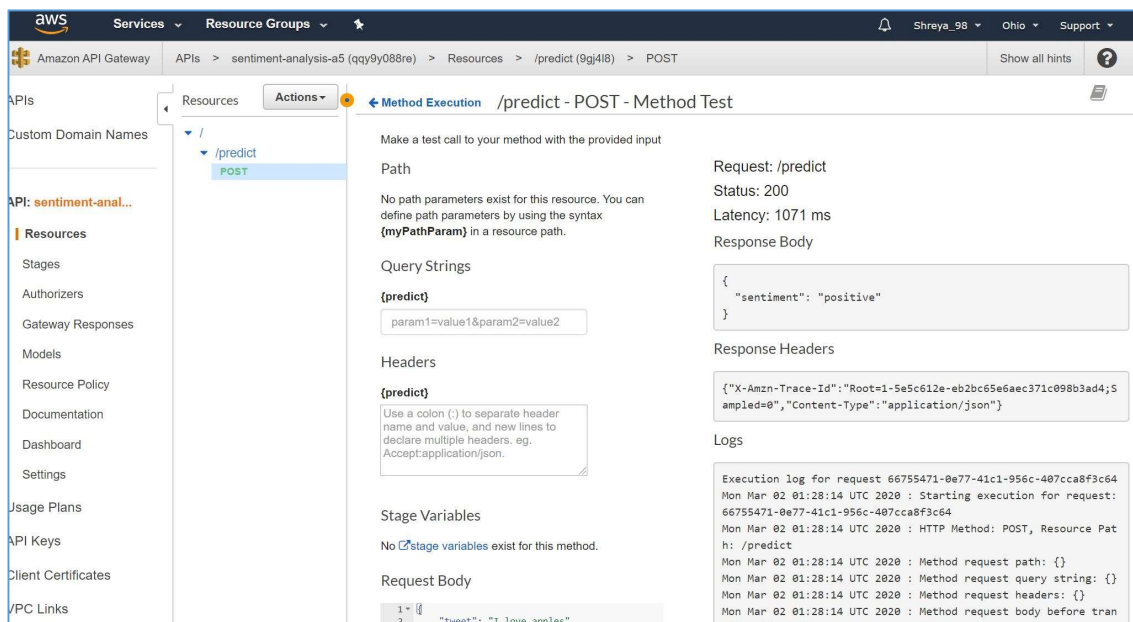
4. REST API

- 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:

```
C:\Shreya\Columbia\Spring 2020 Semester\AI at Scale on Cloud\Assignment5\lambda_dir>curl -X POST https://qqy9y088re.execute-api.us-east-2.amazonaws.com/v1/predict --header "Content-Type:application/json" --data '{"tweet": "I love apples"}'
{"sentiment": "positive"}
C:\Shreya\Columbia\Spring 2020 Semester\AI at Scale on Cloud\Assignment5\lambda_dir>
```

```
C:\Shreya\Columbia\Spring 2020 Semester\AI at Scale on Cloud\Assignment5\lambda_dir>
C:\Shreya\Columbia\Spring 2020 Semester\AI at Scale on Cloud\Assignment5\lambda_dir>
C:\Shreya\Columbia\Spring 2020 Semester\AI at Scale on Cloud\Assignment5\lambda_dir>curl -X POST https://qqy9y088re.execute-api.us-east-2.amazonaws.com/v1/predict --header "Content-Type:application/json" --data '{"tweet": "I love apples"}'
{"Date and Time of request": "2020-03-02 01:48:55.465423", "Tweet": "I love apples", "sentiment": "positive", "Probability": 0.98763895, "Preprocessing Time": "0.0", "Model Inference Time": "0.06"}
C:\Shreya\Columbia\Spring 2020 Semester\AI at Scale on Cloud\Assignment5\lambda_dir>
```