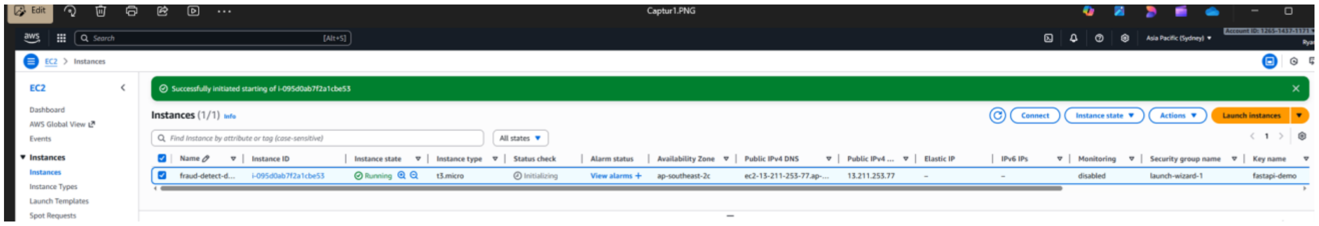
End-to-End Fraud Detection Model Deployment on AWS using FastAPI and Docker

This document provides a summary of the deployment of a fraud detection LSTM model as a REST API
using FastAPI, Docker, and AWS EC2. The screenshots included serve as evidence of the successful end-to-end deployment,
testing, and cloud cost management process.

1. AWS EC2 Instance Setup

An AWS EC2 t3.micro instance was configured in the Asia Pacific (Sydney) region using Amazon Linux 2023.
Security groups were set up to allow inbound HTTP (port 80) and SSH (port 22) traffic.
The instance was used to host the containerized FastAPI application.



Successfully initiated starting of i-095d0ab7f2a1cbe53

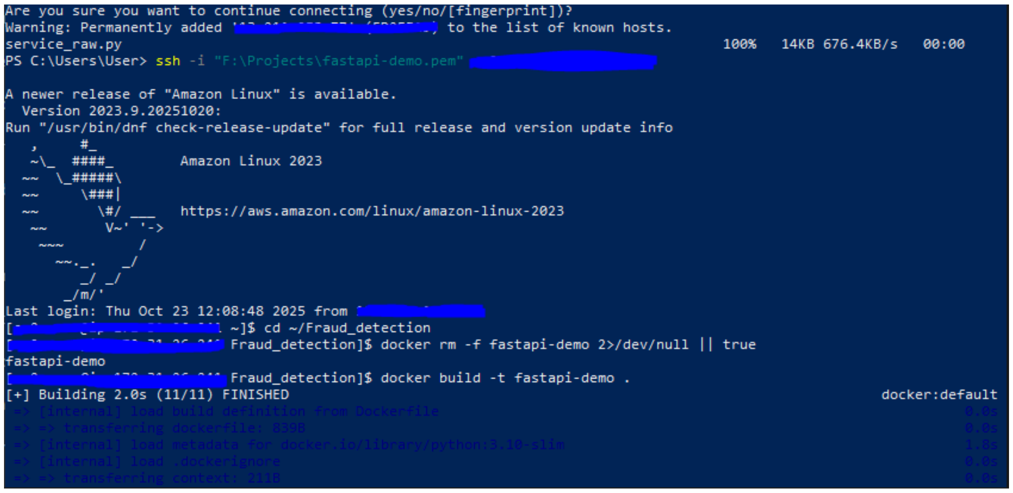
Instances (1/1) Info

Find Instance by attribute or tag (case-sensitive)

Name	Instance ID	Instance state	Instance type	Status check	Alarm status	Availability Zone	Public IPv4 DNS	Public IPv4 ...	Elastic IP	IPv6 IPs	Monitoring	Security group name	Key name
fraud-detect-d...	i-095d0ab7f2a1cbe53	Running	t3.micro	Initializing	View alarms +	ap-southeast-2c	ec2-13-211-253-77.ap-...	13.211.253.77	–	–	disabled	launch-wizard-1	fastapi-demo

2. SSH Connection and Docker Deployment

Deployment was performed through SSH access to the EC2 instance. Project files were transferred to the server, and the Docker image was built and run using the defined requirements and service scripts. The container exposed the FastAPI application on port 80 for public access.

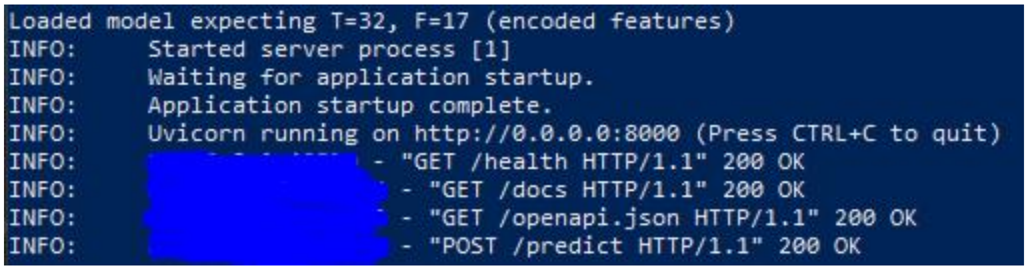


```
Are you sure you want to continue connecting (yes/no/[fingerprint])?
Warning: Permanently added [redacted] to the list of known hosts.
service_raw.py                                                        100%   14KB 676.4KB/s   00:00
PS C:\Users\User> ssh -i "F:\Projects\fastapi-demo.pem" [redacted]

A newer release of "Amazon Linux" is available.
  Version 2023.9.20251020:
Run "/usr/bin/dnf check-release-update" for full release and version update info
   ,     #_
   ~\_  ####_        Amazon Linux 2023
  ~~  \_#####\
  ~~     \###|
  ~~       \#/ ___   https://aws.amazon.com/linux/amazon-linux-2023
   ~~       V~' '->
    ~~~         /
      ~~._.   _/
         _/ _/
       _/m/'
Last login: Thu Oct 23 12:08:48 2025 from [redacted]
[[redacted] ~]$ cd ~/Fraud_detection
[redacted] Fraud_detection]$ docker rm -f fastapi-demo 2>/dev/null || true
fastapi-demo
[redacted] Fraud_detection]$ docker build -t fastapi-demo .
[+] Building 2.0s (11/11) FINISHED                                                      docker:default
 => [internal] load build definition from Dockerfile                                              0.0s
 => => transferring dockerfile: 839B                                                              0.0s
 => [internal] load metadata for docker.io/library/python:3.10-slim                               1.8s
 => [internal] load .dockerignore                                                                 0.0s
 => => transferring context: 211B                                                                 0.0s
```

3. FastAPI Server Startup Logs

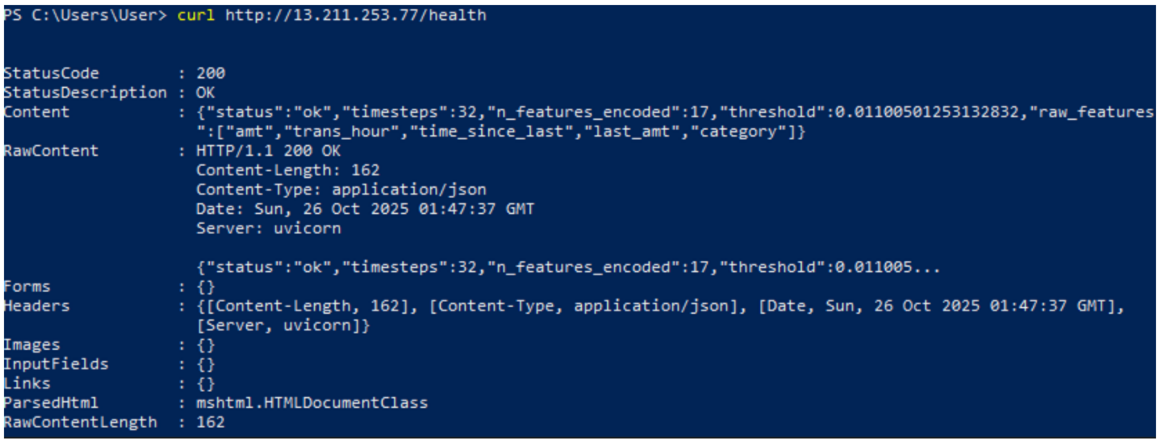
The Docker container logs confirm the successful startup of the FastAPI service and correct model initialisation.
The LSTM model and preprocessing pipeline were loaded successfully, with parameters T=32 (timesteps) and F=17 (encoded features).



```
Loaded model expecting T=32, F=17 (encoded features)
INFO:     Started server process [1]
INFO:     Waiting for application startup.
INFO:     Application startup complete.
INFO:     Uvicorn running on http://0.0.0.0:8000 (Press CTRL+C to quit)
INFO:     [redacted] - "GET /health HTTP/1.1" 200 OK
INFO:     [redacted] - "GET /docs HTTP/1.1" 200 OK
INFO:     [redacted] - "GET /openapi.json HTTP/1.1" 200 OK
INFO:     [redacted] - "POST /predict HTTP/1.1" 200 OK
```

4. API Health Endpoint Verification

The /health endpoint verified that the API was operational, confirming correct model artifacts, timesteps,
feature dimensions, and decision threshold. Both PowerShell and browser tests returned successful JSON responses.



```
PS C:\Users\User> curl http://13.211.253.77/health


StatusCode        : 200
StatusDescription : OK
Content           : {"status":"ok","timesteps":32,"n_features_encoded":17,"threshold":0.01100501253132832,"raw_features
                    ":["amt","trans_hour","time_since_last","last_amt","category"]}
RawContent        : HTTP/1.1 200 OK
                    Content-Length: 162
                    Content-Type: application/json
                    Date: Sun, 26 Oct 2025 01:47:37 GMT
                    Server: uvicorn

                    {"status":"ok","timesteps":32,"n_features_encoded":17,"threshold":0.011005...
Forms             : {}
Headers           : {[Content-Length, 162], [Content-Type, application/json], [Date, Sun, 26 Oct 2025 01:47:37 GMT],
                    [Server, uvicorn]}
Images            : {}
InputFields       : {}
Links             : {}
ParsedHtml        : mshtml.HTMLDocumentClass
RawContentLength  : 162
```

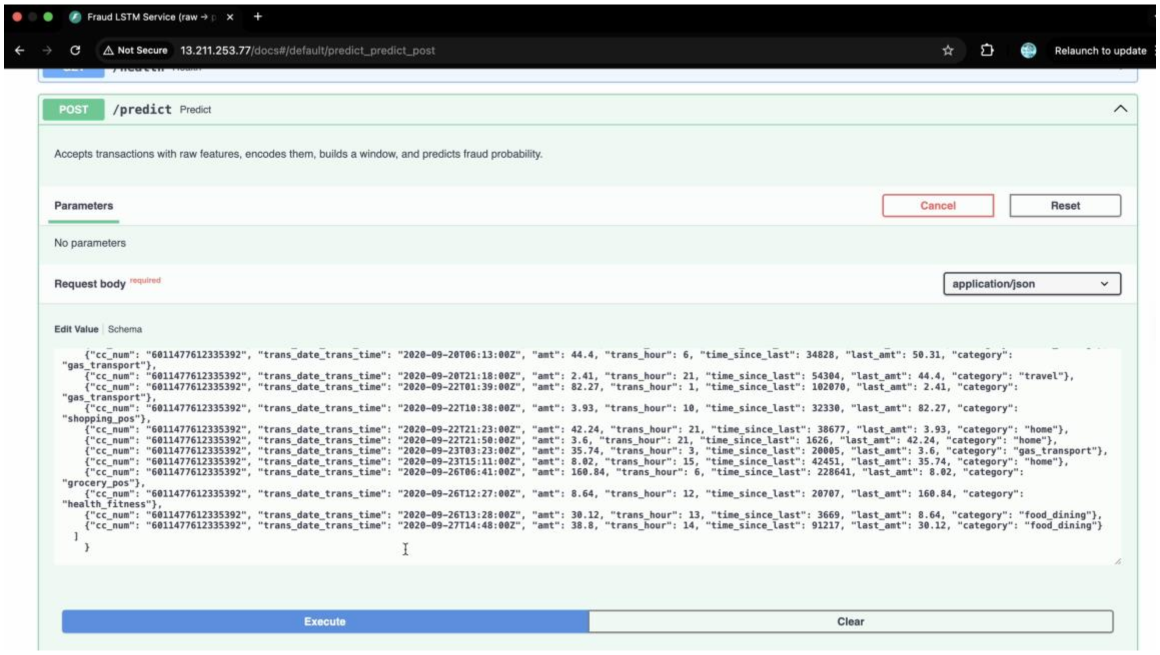
5. Environment and Version Verification

Inside the container, installed library versions were confirmed to match the training environment.
TensorFlow (2.20.0), scikit-learn (1.6.1), and FastAPI (0.119.0) were verified to ensure full compatibility
between development and production.

6. Successful Fraud Prediction Request

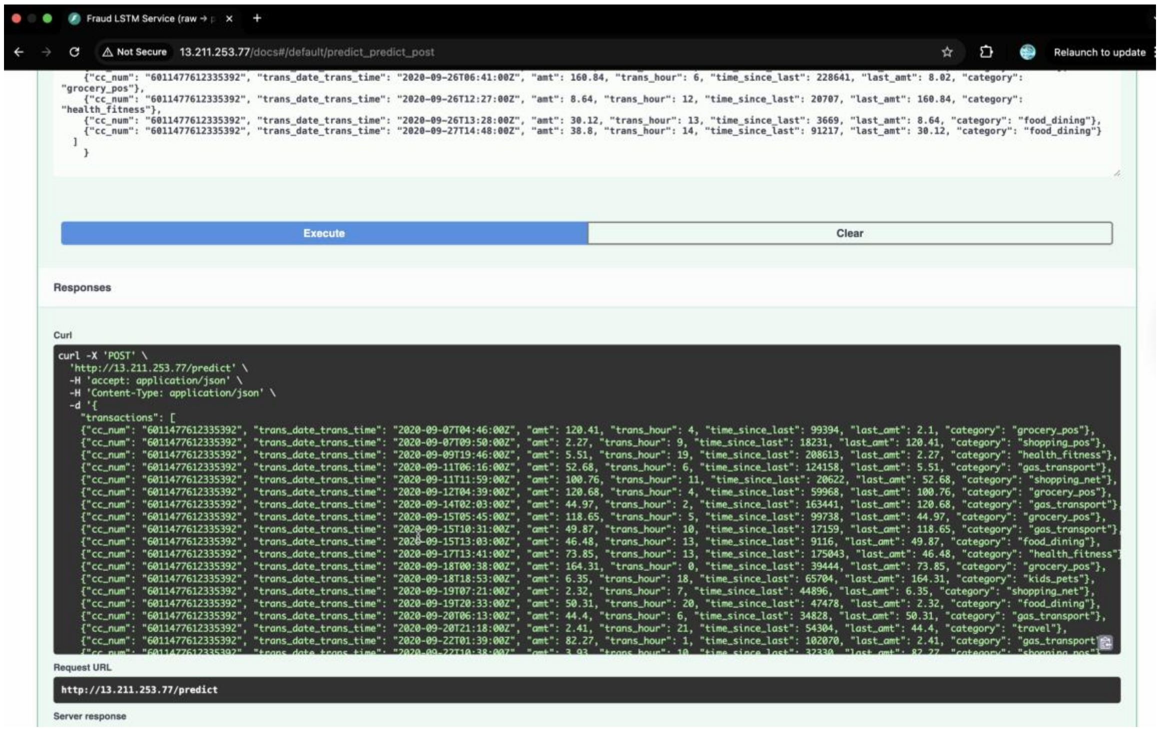
A real transaction payload was sent via Swagger UI and PowerShell to the /predict endpoint. The API successfully encoded inputs, built the input sequence, and returned a JSON response containing the predicted fraud probability, binary decision, and threshold value.

Using Swagger UI to send transaction data



Accepts transactions with raw features, encodes them, builds a window, and predicts fraud probability.

```
    {"cc_num": "6011477612335392", "trans_date_trans_time": "2020-09-20T06:13:00Z", "amt": 44.4, "trans_hour": 6, "time_since_last": 34828, "last_amt": 50.31, "category":
"gas_transport"},
    {"cc_num": "6011477612335392", "trans_date_trans_time": "2020-09-20T21:18:00Z", "amt": 2.41, "trans_hour": 21, "time_since_last": 54304, "last_amt": 44.4, "category": "travel"},
    {"cc_num": "6011477612335392", "trans_date_trans_time": "2020-09-22T01:39:00Z", "amt": 82.27, "trans_hour": 1, "time_since_last": 102070, "last_amt": 2.41, "category":
"gas_transport"},
    {"cc_num": "6011477612335392", "trans_date_trans_time": "2020-09-22T10:38:00Z", "amt": 3.93, "trans_hour": 10, "time_since_last": 32330, "last_amt": 82.27, "category":
"shopping_pos"},
    {"cc_num": "6011477612335392", "trans_date_trans_time": "2020-09-22T21:23:00Z", "amt": 42.24, "trans_hour": 21, "time_since_last": 38677, "last_amt": 3.93, "category": "home"},
    {"cc_num": "6011477612335392", "trans_date_trans_time": "2020-09-22T21:50:00Z", "amt": 3.6, "trans_hour": 21, "time_since_last": 1626, "last_amt": 42.24, "category": "home"},
    {"cc_num": "6011477612335392", "trans_date_trans_time": "2020-09-23T03:23:00Z", "amt": 35.74, "trans_hour": 3, "time_since_last": 20005, "last_amt": 3.6, "category": "gas_transport"},
    {"cc_num": "6011477612335392", "trans_date_trans_time": "2020-09-23T15:11:00Z", "amt": 8.02, "trans_hour": 15, "time_since_last": 42451, "last_amt": 35.74, "category": "home"},
    {"cc_num": "6011477612335392", "trans_date_trans_time": "2020-09-26T06:41:00Z", "amt": 160.84, "trans_hour": 6, "time_since_last": 228641, "last_amt": 8.02, "category":
"grocery_pos"},
    {"cc_num": "6011477612335392", "trans_date_trans_time": "2020-09-26T12:27:00Z", "amt": 8.64, "trans_hour": 12, "time_since_last": 20707, "last_amt": 160.84, "category":
"health_fitness"},
    {"cc_num": "6011477612335392", "trans_date_trans_time": "2020-09-26T13:28:00Z", "amt": 30.12, "trans_hour": 13, "time_since_last": 3669, "last_amt": 8.64, "category": "food_dining"},
    {"cc_num": "6011477612335392", "trans_date_trans_time": "2020-09-27T14:48:00Z", "amt": 38.8, "trans_hour": 14, "time_since_last": 91217, "last_amt": 30.12, "category": "food_dining"}
  ]
    }
```

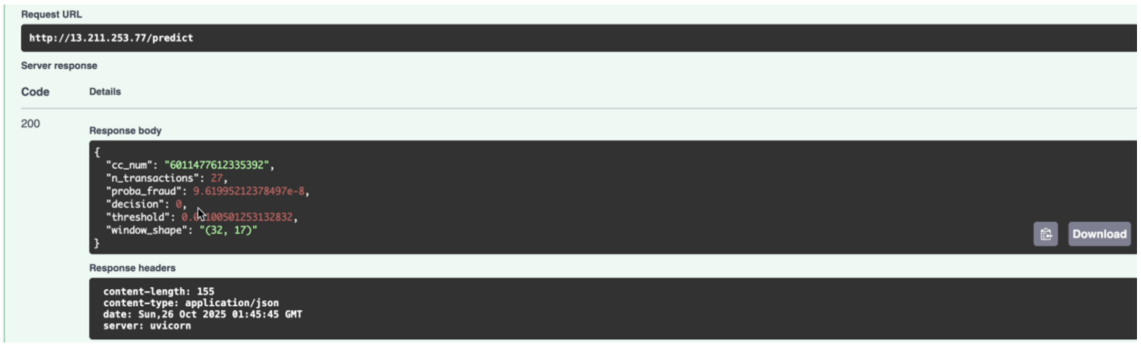


```
    {"cc_num": "6011477612335392", "trans_date_trans_time": "2020-09-26T06:41:00Z", "amt": 160.84, "trans_hour": 6, "time_since_last": 228641, "last_amt": 8.02, "category":
"grocery_pos"},
    {"cc_num": "6011477612335392", "trans_date_trans_time": "2020-09-26T12:27:00Z", "amt": 8.64, "trans_hour": 12, "time_since_last": 20707, "last_amt": 160.84, "category":
"health_fitness"},
    {"cc_num": "6011477612335392", "trans_date_trans_time": "2020-09-26T13:28:00Z", "amt": 30.12, "trans_hour": 13, "time_since_last": 3669, "last_amt": 8.64, "category": "food_dining"},
    {"cc_num": "6011477612335392", "trans_date_trans_time": "2020-09-27T14:48:00Z", "amt": 38.8, "trans_hour": 14, "time_since_last": 91217, "last_amt": 30.12, "category": "food_dining"}
  ]
    }
```

```
curl -X 'POST' \
  'http://13.211.253.77/predict' \
  -H 'accept: application/json' \
  -H 'Content-Type: application/json' \
  -d '{
   "transactions": [
    {"cc_num": "6011477612335392", "trans_date_trans_time": "2020-09-07T04:46:00Z", "amt": 120.41, "trans_hour": 4, "time_since_last": 99394, "last_amt": 2.1, "category": "grocery_pos"},
    {"cc_num": "6011477612335392", "trans_date_trans_time": "2020-09-07T09:50:00Z", "amt": 2.27, "trans_hour": 9, "time_since_last": 18231, "last_amt": 120.41, "category": "shopping_pos"},
    {"cc_num": "6011477612335392", "trans_date_trans_time": "2020-09-09T19:46:00Z", "amt": 5.51, "trans_hour": 19, "time_since_last": 208613, "last_amt": 2.27, "category": "health_fitness"},
    {"cc_num": "6011477612335392", "trans_date_trans_time": "2020-09-11T06:16:00Z", "amt": 52.68, "trans_hour": 6, "time_since_last": 124158, "last_amt": 5.51, "category": "gas_transport"},
    {"cc_num": "6011477612335392", "trans_date_trans_time": "2020-09-11T11:59:00Z", "amt": 100.76, "trans_hour": 11, "time_since_last": 20622, "last_amt": 52.68, "category": "shopping_net"},
    {"cc_num": "6011477612335392", "trans_date_trans_time": "2020-09-12T04:39:00Z", "amt": 120.68, "trans_hour": 4, "time_since_last": 59968, "last_amt": 100.76, "category": "grocery_pos"},
    {"cc_num": "6011477612335392", "trans_date_trans_time": "2020-09-14T02:03:00Z", "amt": 44.97, "trans_hour": 2, "time_since_last": 163441, "last_amt": 120.68, "category": "gas_transport"},
    {"cc_num": "6011477612335392", "trans_date_trans_time": "2020-09-15T05:45:00Z", "amt": 118.65, "trans_hour": 5, "time_since_last": 99738, "last_amt": 44.97, "category": "grocery_pos"},
    {"cc_num": "6011477612335392", "trans_date_trans_time": "2020-09-15T10:31:00Z", "amt": 49.87, "trans_hour": 10, "time_since_last": 17159, "last_amt": 118.65, "category": "gas_transport"},
    {"cc_num": "6011477612335392", "trans_date_trans_time": "2020-09-15T13:03:00Z", "amt": 46.48, "trans_hour": 13, "time_since_last": 9116, "last_amt": 49.87, "category": "food_dining"},
    {"cc_num": "6011477612335392", "trans_date_trans_time": "2020-09-17T13:41:00Z", "amt": 73.85, "trans_hour": 13, "time_since_last": 175043, "last_amt": 46.48, "category": "health_fitness"
    {"cc_num": "6011477612335392", "trans_date_trans_time": "2020-09-18T00:38:00Z", "amt": 164.31, "trans_hour": 0, "time_since_last": 39444, "last_amt": 73.85, "category": "grocery_pos"},
    {"cc_num": "6011477612335392", "trans_date_trans_time": "2020-09-18T18:53:00Z", "amt": 6.35, "trans_hour": 18, "time_since_last": 65704, "last_amt": 164.31, "category": "kids_pets"},
    {"cc_num": "6011477612335392", "trans_date_trans_time": "2020-09-19T07:21:00Z", "amt": 2.32, "trans_hour": 7, "time_since_last": 44896, "last_amt": 6.35, "category": "shopping_net"},
    {"cc_num": "6011477612335392", "trans_date_trans_time": "2020-09-19T20:33:00Z", "amt": 50.31, "trans_hour": 20, "time_since_last": 47478, "last_amt": 2.32, "category": "food_dining"},
    {"cc_num": "6011477612335392", "trans_date_trans_time": "2020-09-20T06:13:00Z", "amt": 44.4, "trans_hour": 6, "time_since_last": 34828, "last_amt": 50.31, "category": "gas_transport"},
    {"cc_num": "6011477612335392", "trans_date_trans_time": "2020-09-20T21:18:00Z", "amt": 2.41, "trans_hour": 21, "time_since_last": 54304, "last_amt": 44.4, "category": "travel"},
    {"cc_num": "6011477612335392", "trans_date_trans_time": "2020-09-22T01:39:00Z", "amt": 82.27, "trans_hour": 1, "time_since_last": 102070, "last_amt": 2.41, "category": "gas_transport
```

```
http://13.211.253.77/predict
```

Receiving a prediction from the model



Request URL

```
http://13.211.253.77/predict
```

Server response

Code Details

200

Response body

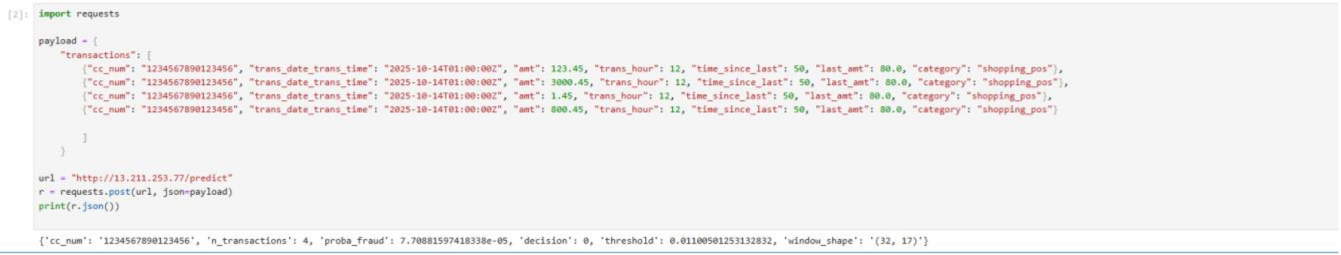
```
{
  "cc_num": "6011477612335392",
  "n_transactions": 27,
  "proba_fraud": 9.61995212378497e-8,
  "decision": 0,
  "threshold": 0.0[illegible]100501253132832,
  "window_shape": "(32, 17)"
}
```

Response headers

```
content-length: 155
content-type: application/json
date: Sun,26 Oct 2025 01:45:45 GMT
server: uvicorn
```

7. Python Client Validation

A Python client was also used to send a POST request to the /predict endpoint. The response matched the
results from Swagger UI and PowerShell, confirming consistent model inference across interfaces.



```
[2]: import requests

     payload = {
         "transactions": [
             {"cc_num": "1234567890123456", "trans_date_trans_time": "2025-10-14T01:00:00Z", "amt": 123.45, "trans_hour": 12, "time_since_last": 50, "last_amt": 80.0, "category": "shopping_pos"},
             {"cc_num": "1234567890123456", "trans_date_trans_time": "2025-10-14T01:00:00Z", "amt": 3000.45, "trans_hour": 12, "time_since_last": 50, "last_amt": 80.0, "category": "shopping_pos"},
             {"cc_num": "1234567890123456", "trans_date_trans_time": "2025-10-14T01:00:00Z", "amt": 1.45, "trans_hour": 12, "time_since_last": 50, "last_amt": 80.0, "category": "shopping_pos"},
             {"cc_num": "1234567890123456", "trans_date_trans_time": "2025-10-14T01:00:00Z", "amt": 800.45, "trans_hour": 12, "time_since_last": 50, "last_amt": 80.0, "category": "shopping_pos"}

             ]
         }

     url = "http://13.211.253.77/predict"
     r = requests.post(url, json=payload)
     print(r.json())

     {'cc_num': '1234567890123456', 'n_transactions': 4, 'proba_fraud': 7.70881597418338e-05, 'decision': 0, 'threshold': 0.01100501253132832, 'window_shape': '(32, 17)'}
```

8. EC2 Instance Shutdown

After deployment validation, the EC2 instance was stopped to prevent further billing. This follows cloud cost management best practices, ensuring AWS compute charges apply only while the instance is active.

9. Conclusion

The fraud detection model was successfully deployed and tested on AWS EC2 using FastAPI and Docker. The project demonstrates real-time model inference, environment consistency, and cost-efficient cloud deployment. This workflow reflects key MLOps principles, including version control, containerization, and scalable API hosting.