

# Suspicious Transaction Detection Using AWS ML & Data Pipeline

A CLOUD-BASED FINANCIAL INTELLIGENCE  
PROJECT BY NIALL CLEUR

# Project Overview

- ▶ Objective: Detect potentially fraudulent financial transactions using AWS on a large dataset.
- ▶ Supports business compliance, reduces false positives.
- ▶ Automated low code, scalable, secure, cloud-native solution for financial intelligence.

# Tools & AWS Services Used

- ▶ Amazon S3: Store raw & processed data
- ▶ AWS Glue (DataBrew) – ETL: Clean & transform dataset
- ▶ SageMaker Autopilot: Train ML model without coding
- ▶ Amazon QuickSight: Visualize suspicious activity and model results

# Architecture

CSV Dataset

Amazon S3  
Storage

AWS Glue  
DataBrew ETL

SageMaker  
Autopilot

Amazon QuickSight  
Visualisation

CSV Dataset → S3 → SageMaker → Autopilot → AWS QuickSight

# Dataset

- ▶ Data Source: Kaggle, Credit Card Fraud Detection (~284K rows)
- ▶ Features: Time, V1–V28 (PCA), Amount, Class
- ▶ We will use Class column as dependent variable for predictions.
- ▶ Other columns are used as features.

Amazon S3

General purpose buckets

Directory buckets

Table buckets

Access Grants

Access Points

Object Lambda Access Points

Multi-Region Access Points

Batch Operations

IAM Access Analyzer for S3

Block Public Access settings for this account

Storage Lens

Dashboards

Storage Lens groups

AWS Organizations settings

Feature spotlight 11

nialldata

Info

Objects

Properties

Permissions

Metrics

Management

Access Points

Objects (2)

Copy S3 URI

Copy URL

Download

Open

Delete

Actions

Create folder

Upload

Objects are the fundamental entities stored in Amazon S3. You can use [Amazon S3 inventory](#) to get a list of all objects in your bucket. For others to access your objects, you'll need to explicitly grant them permissions. [Learn more](#)

Find objects by prefix

< 1 >

<input type="checkbox"/>	Name	Type	Last modified	Size
<input type="checkbox"/>	<a href="#">cleaned/</a>	Folder	-	-
<input type="checkbox"/>	<a href="#">creditcard.csv</a>	csv	June 14, 2025, 12:20:11 (UTC+10:00)	143.8 MB

AWS Marketplace for S3

CloudShell

Feedback

© 2025, Amazon Web Services, Inc. or its affiliates.

Privacy

Terms

Cookie preferences

Amazon S3

General purpose buckets

Directory buckets

Table buckets

Access Grants

Access Points

Object Lambda Access Points

Multi-Region Access Points

Batch Operations

IAM Access Analyzer for S3

Block Public Access settings for this account

Storage Lens

Dashboards

Storage Lens groups

AWS Organizations settings

Feature spotlight 11

cleaned-creditcard-data\_14Jun2025\_1749869181071/

Info

Objects

Properties

Objects (1)

Copy S3 URI

Copy URL

Download

Open

Delete

Actions

Create folder

Upload

Objects are the fundamental entities stored in Amazon S3. You can use [Amazon S3 inventory](#) to get a list of all objects in your bucket. For others to access your objects, you'll need to explicitly grant them permissions. [Learn more](#)

Find objects by prefix

< 1 >

<input type="checkbox"/>	Name	Type	Last modified	Size	Storage class
<input type="checkbox"/>	<a href="#">cleaned-creditcard-data_14Jun2025_1749869181071_part00000.csv</a>	csv	June 14, 2025, 12:46:40 (UTC+10:00)	144.0 MB	Standard

AWS Marketplace for S3

CloudShell

Feedback

© 2025, Amazon Web Services, Inc. or its affiliates.

Privacy

Terms

Cookie preferences

# Model Training with SageMaker Autopilot

- ▶ No-code ML with target column: Class
- ▶ Input data from Glue-cleaned CSV in S3
- ▶ AutoML builds & tunes classification models
- ▶ Predict data and displays results.





# SageMaker Autopilot --> Select Data

My models > Fraud-model-niall > Version 1

+ Create new version

Select dataset

You can import a tabular dataset or choose one that has already been imported. Your dataset must contain at least one input column and a target column.

+ Create dataset

Search datasets in Canvas

All Joined

Name	Columns	Rows	Cells	Created	Status
<input checked="" type="radio"/> datafraud	31	284,807	8,829,017	06/14/2025 2:23 PM	Ready
<input type="radio"/> canvas-sample-shipping-logs.csv	12	1,000	12,000	06/14/2025 2:15 PM	Ready
<input type="radio"/> canvas-sample-product-descriptions.csv	5	120	800	06/14/2025 2:15 PM	Ready
<input type="radio"/> canvas-sample-housing.csv	10	1,000	10,000	06/14/2025 2:15 PM	Ready
<input type="radio"/> canvas-sample-loans-part-2.csv	5	1,000	5,000	06/14/2025 2:15 PM	Ready
<input type="radio"/> canvas-sample-loans-part-1.csv	19	1,000	19,000	06/14/2025 2:15 PM	Ready
<input type="radio"/> canvas-sample-databricks-dolly-15k.csv	2	10,879	21,758	06/14/2025 2:15 PM	Ready
<input type="radio"/> canvas-sample-maintenance.csv	9	1,000	9,000	06/14/2025 2:15 PM	Ready

Select dataset

# SageMaker Autopilot --> Build Model



d-3cbzvfaeu1o.studio.ap-southeast-2.sagemaker.aws/canvas/default/models/Fraud-model-niall



Home



Amazon Q



Data Wrangler



Datasets



My Models



ML Ops



Ready-to-use



Help



Log out

My models > Fraud-model-niall > Version 1

+ Create new version



Select

Build

Analyze

Predict

Deploy

## Select a column to predict

Choose the target column. The model that you build predicts values for the column that you select.

Target column



data1  
Rando

- ☐ V21
- ☐ V22
- ☐ V23
- ☐ V24
- ☐ V25
- ☐ V26
- ☐ V27
- ☒ V28
- ☒ Amount
- ☒ Class

## Model type

SageMaker Canvas automatically recommends the appropriate model type for your analysis.

2 category prediction

Your model classifies Class into two categories.

[Configure model](#)

Standard build

[Preview model](#)



Manage columns



Manage rows



Time series

View all

[Data visualizer](#)



Feature type ⓘ

Missing ⓘ

Mismatched ⓘ

Unique ⓘ

Mode ⓘ



-

0.00% (0)

0.00% (0)

19734

0.36



-

0.00% (0)

0.00% (0)

19734

0.78



-

0.00% (0)

0.00% (0)

19734

-0.42



-

0.00% (0)

0.00% (0)

19734

3.14



-

0.00% (0)

0.00% (0)

19734

2.61



-

0.00% (0)

0.00% (0)

19734

-0.58



-

0.00% (0)

0.00% (0)

19734

-2.96

Total columns: 31

Total rows: 284,007

Total cells: 8,829,017

☒ Show dropped columns



# SageMaker Autopilot --> Create Model

Home

Amazon Q

Data Wrangler

Datasets

My Models

ML Ops

Ready-to-use

Help

Log out

d-3cbzovfaeu1o.studio.ap-southeast-2.sagemaker.aws/canvas/default/models/Fraud-model-nlall

VPN

My models > Fraud-model-nlall > Version 1

+ Create new version

Select

Build

Analyze

Predict

Deploy

Model status Quick build

Accuracy ⓘ F1 ⓘ Optimization metric

99.93% 0.759

The model predicts the correct Class 99.93% of the time. ⓘ

Predict Standard build Deploy

Overview

Scoring

Advanced metrics

Model leaderboard

Positive Class

F1 ⓘ Optimization metric

Accuracy ⓘ

Precision ⓘ

Recall ⓘ

AUC-ROC ⓘ

1 0

99.965%

99.93%

99.95%

99.98%

0.94

Metrics table

Confusion matrix

Precision recall curve

Actual values

0

1

True positive (TP)  
9957  
99.8% of predicted results

False negative (FN)  
2  
0.0% of predicted results

False positive (FP)  
5  
0.1% of predicted results

True negative (TN)  
11  
0.1% of predicted results

datafraud

Total columns: 31

Total rows: 284,807

Total cells: 8,829,017

Class

2 category prediction

Predict

# SageMaker Autopilot --> Model ready

My models > Fraud-model-niall > Version 1

Model status [Quick build](#)

Accuracy ① F1 ① Optimization metric

99.93% 0.759

The model predicts the correct Class 99.93% of the time. ①

Predict Standard build Deploy

Overview Scoring Advanced metrics

Model leaderboard

Column impact ① ↓

Search columns...

1	V4	22.23%
2	V14	8.333%
3	V6	6.053%
4	V15	5.291%
5	V23	4.057%
6	V12	3.666%

Impact of V4 on prediction of Class

Impact on prediction

V4

datafraud Total columns: 31 Total rows: 284,807 Total cells: 8,829,017 Class 2 category prediction

Predict

# SageMaker Autopilot --> Predict Data

Home

Amazon Q

Data Wrangler

Datasets

My Models

ML Ops

Ready-to-use

Help

Log out

Operations: Deployment / **canvas-Niall-Fraud-Model**

Update configuration

Details

Test deployment

Modify values to predict **Class** in real time.

Filter columns

Column ↑	Value
Amount	1
Time	156037
V1	2.04021105776632
V10	-0.351075101407957
V11	0.329650883701029
V12	0.183508120624656

Class Prediction

Copy

No

Last prediction

No 99.983% ⓘ

Yes 0.017% ⓘ

Invocation result

Status	Execution length (ms) ⓘ <a href="#">Learn more</a>	Request time
Successful	101.706	2025-06-14 02:59:25 PM

# SageMaker Autopilot --> Deploy model

Home

Amazon Q

Data Wrangler

Datasets

My Models

ML Ops

Ready-to-use

Help

Log out

d-3cbzovfaeu1o,studio.ap-southeast-2.sagemaker.aws/canvas/default/models/Fraud-model-niall/deployments/am%3Aaws%3...

Update configuration

canvas-Niall-Fraud-Model

DetailsTest deployment

Deployment name	Status	Deployment type	Model
canvas-Niall-Fraud-Model	<span>✓ In service</span>	Real-time	Fraud-model-niall v1
Created	Average predictions per day	Last prediction	
06/14/25 02:51 PM	--	--	
Instance type	Instance count	Inference response content	Input format
ml.t2.medium	2	predicted_label, probability, probabilities, labels	text/csv

Deployment URL [Learn how to invoke a real-time endpoint](#)

<https://runtime.sagemaker.ap-southeast-2.amazonaws.com/endpoints/canvas-Niall-Fraud-Model/invocations>

View sample code

# Model Performance

- ▶ F1 Score: 99%
- ▶ Precision: 99%
- ▶ Recall: 99%
- ▶ Confusion Matrix used to assess prediction quality
- ▶ Good balance of fraud detection vs false alarms

# AWS QuickSight Visualizations

- ▶ Exported results to CSV and QuickSight
- ▶ Visualized fields: Class, Probability
- ▶ Dashboard to visualise data and anomalies.



Data

Dataset 100%

SPICE 8589bd75-874b-4...

Search fields

+ CALCULATED FIELD

# Class

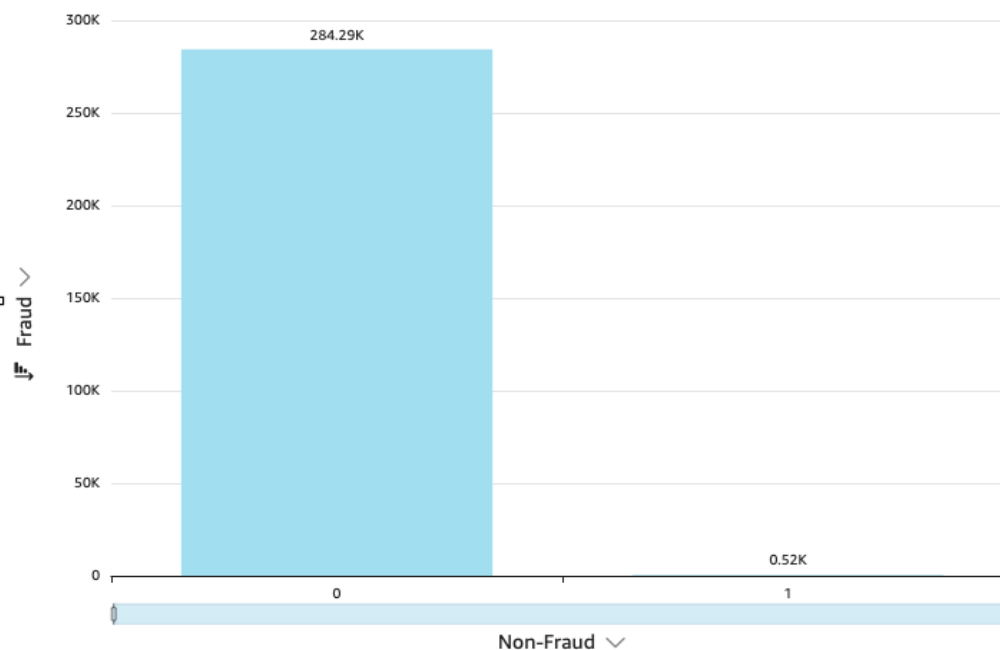
# probability

Fraud vs Non-Fraud

Top 100 Most Confident Fraud Predictions



Count of Class by Class



Properties

Visual Interactions

Display Settings

CARD TITLE

Edit title

Edit subtitle

Alt text

CARD STYLE

Background 100%

Border 100%

Selection 100%

☒ Loading animation

SIZE AND POSITION

W 736 H 528

X 0 Y 0

X-axis

Y-axis

Reference lines

Data labels ☐





# Business Impact Summary

- ▶ Enables fraud detection and reporting
- ▶ Predictions and auto ML generation
- ▶ Reduces manual effort, improves accuracy
- ▶ Useful for financial oversight and compliance

# Key Takeaways

- ▶ Built a complete AWS ML pipeline for fraud detection on a dataset
- ▶ Demonstrated ETL, AutoML, and BI integration
- ▶ Committed to cloud-native data roles



Thank You