

CHETAN MITTAL · UPDATED 24 DAYS AGO

24

<> Code

Download

Credit card Fraud data

Fraud Detection credit card

Data Card

Code (5)

Discussion (0)

Suggestions (0)

About Dataset

This dataset contains credit card transaction records used to identify fraudulent transactions. The primary goal is to build and evaluate machine learning models capable of accurately distinguishing between legitimate and fraudulent activities.

The dataset includes transactions made by credit card users, where each transaction is labeled as fraud (1) or non-fraud (0)

Usability

10.00

License

CC0: Public Domain

Expected update frequency

Quarterly

Tags

We will solve this problem creating a ML model to predict credit card fraud data, this dataset is disbalanced.

First you need to create and execute a compute instance.

Prompt flow

Assets

Data

Jobs

Components

Pipelines

Environments

Models

Endpoints

Manage

Compute

Monitoring

Data Labeling

Linked Services

Connections

Default Directory > mltestresource > Compute

Compute

Compute instances

Compute clusters

Kubernetes clusters

Attached computes

Serverless instances

Choose from a selection of CPU or GPU instances preconfigured with popular tools such as VS Code, JupyterLab, Jupyter, and RStudio, ML packages, deep learning frameworks, and GPU drivers. [Learn more about compute instances](#)

+ New

Refresh

Start

Stop

Restart

Schedule and idle shutdown

Delete

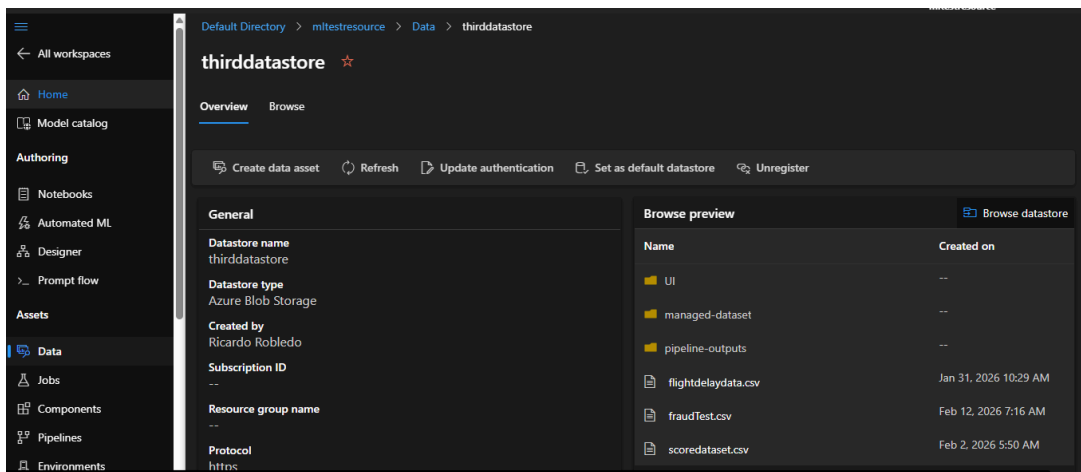
Reset view

Search

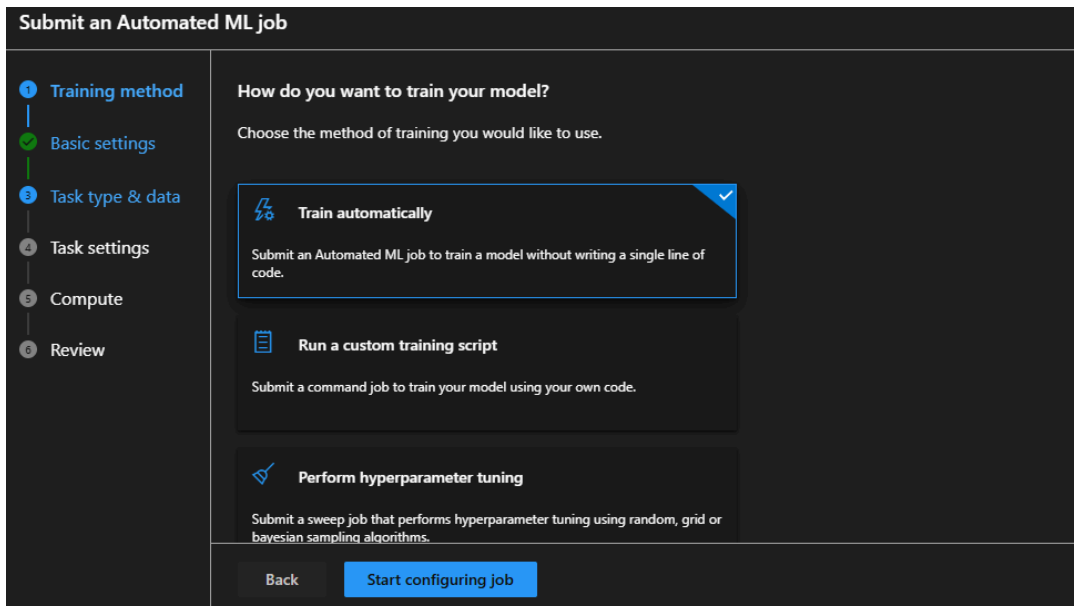
Filter

Name ↑	☆	State	Idle shutdown ↓ ⓘ	Applications ↓ ⓘ
samplecomputeinstance		Running	1 hour	JupyterLab Jupyter VS Code (Web) ...
samplecomputeinstance2		Stopped	1 hour	JupyterLab Jupyter VS Code (Web) ...

We create a datastore and put our dataset.



Then create the Train automatically option.



We put a name to the job and the experiment.

**Submit an Automated ML job**

1 Training method

2 Basic settings

3 Task type & data

4 Task settings

5 Compute

6 Review

Job name \* ⓘ

test

Experiment name \*

☐ Select existing ☒ Create new

New experiment name \* ⓘ

experiment

Description

Tags

Name

:

Value

Add

Back

Next

Select the task type, in this case classification. Then push create to create a dataset.

**Submit an Automated ML job**

1 Training method

2 Basic settings

3 Task type & data

4 Task settings

5 Compute

6 Review

Task type & data

Choose the type of task that you would like your model to perform and the data to use for training. [Learn more](#)

Select task type \* ⓘ

Classification

Select data

Make sure your data is preprocessed into a supported format.

+ Create

Refresh

☒ Show supported data assets only

Reset view

Search

Filter

Name

Type

Created on

Modified

Back

Next

Select Azure Storage from our storage account.

Create data asset

Data type

Data source

Source storage type

Storage path

Settings

Schema

Review

Choose a source for your data asset

Choose the data source you want to create your asset from. A data source can be from a local storage location on your computer, from an attached datastore, from Azure storage, or from a publicly available web location.

From Azure storage

Create a data asset from registered data storage services including Azure Blob Storage, Azure file share, and Azure Data Lake.

From local files

Create a data asset by uploading files from your local drive.

From SQL databases

Create a dataset from Azure SQL database and Azure PostgreSQL database.

From web files

Create a data asset from a single file located at a public web URL.

From Azure Open Datasets

Create a dataset with one-click from pre-made data sets. These data sets are created by the general public and published as Azure Open Datasets.

Back

Next

Cancel

Select datastore

Create data asset

Data type

Data source

Source storage type

Storage path

Settings

Schema

Review

Select a datastore

Choose a storage type and a datastore that contains your data. You can also create a new datastore for your data first.

Datastore type

Azure Blob Storage

Create new datastore

Search datastore

Name	Storage name	Created
thirddatastore	sampleblobstorage	Feb
seconddatastore2	sampleblobstorage	Jan

Page 1 of 1

25/Page

Back

Next

Confirm data

Create data asset

✔ Data type

✔ Data source

✔ Source storage type

✔ Storage path

⚙ Settings

📄 Schema

🔍 Review

Settings

These settings determine how the data is parsed. The initial settings are automatically detected; you can change them as needed to reparse the data.

File format

Delimited

Delimiter

Comma

Example

Field1,Field2,Field3

Encoding

UTF-8

Column headers

All files have same headers

Skip rows

None

☐ Dataset contains multi-line data

Note: Processing tabular files with multi-line data is slower because multiple CPU cores cannot be used to ingest the data in parallel. Checking this option may result in slower Data previewing times.

sn	trans_d...	cc_num	merchant	category	amt	first	last	gender	street	city	state	zip
0	2020-06...	2.291.16...	fraud_Ki...	persona...	2.86	Jeff	Elliott	M	351 Darl...	Columbia	SC	2920
1	2020-06...	3.573.03...	fraud_S...	persona...	29.84	Joanne	Williams	F	3638 M...	Altonah	UT	8400
2	2020-06...	3.598.22...	fraud_S...	health_fi...	41.28	Ashley	Lopez	F	9333 Val...	Bellmore	NY	1171

Back

Next

Review

Cancel

Include and change values needed. In this case we use: *trans\_date\_trans\_time*, *category*, *amt*, *gender*, *city*, *state*, *zip*, *lat*, *long*, *city\_pop*, *job*, *dob*, *unix\_time*, *merch\_lat*, *merch\_long*, *is\_fraud*.

delete: *trans\_num*, *street*, *first*, *last*, *cc\_num*, *sn*, *merchant*.

Create data asset

✔ Data type

✔ Data source

✔ Source storage type

✔ Storage path

✔ Settings

⚙ Schema

🔍 Review

Schema

Column types are auto-detected based on the initial subset of the data and can be updated here. Values not aligning with the specified column type will fail conversion and would be either null-filled or replaced with error value. Any conversions preview errors are non-blocking and you can proceed.

Search column name

Include	Column name	Type	Example values	Date format	Properties
<input type="checkbox"/>	Path	String		Not applicable to s...	Not applicable t...
<input type="checkbox"/>	sn	Integer	0, 1, 2	Not applicable to s...	Not applicable t...
<input checked="" type="checkbox"/>	trans_date_trans_time	Date	2020-06-21 12:14:00, 202...	%d-%m-%Y %H:%M	None
<input type="checkbox"/>	cc_num	Decimal (dot '.')	2291160000000000, 3573...	Not applicable to s...	Not applicable t...

Back

Next

Cancel

Verify and create dataset.

Create data asset

- Data type
- Data source
- Source storage type
- Storage path
- Settings
- Schema
- Review

**Review**  
Review the settings for your data asset and make any changes as needed.

<b>Data type</b>	<b>Schema</b>
<b>Name</b> creditcardfraud	<b>trans_date_trans_time</b> Date
<b>Description</b> --	<b>category</b> String
<b>Type</b> tabular	<b>amt</b> Decimal
<b>Data source</b>	<b>gender</b> String
<b>Type</b> AzureStorage	<b>city</b> String
<b>Storage</b>	(showing 5 of 24 columns)
<b>Datastore type</b>	

Back Create Cancel

We'll use `is_fraud` like our target column and `AveragePrecisionScoreWeighted` like metric.

Default Directory > mltestresource > Training job

Submit an Automated ML job

- Training method
- Basic settings
- Task type & data
- Task settings
- Compute
- Review

**Target column \***  
is\_fraud (Integer)

**Classification settings**

☐ Enable deep learning

☒ View additional configuration settings

**Validate and test**

You can choose a validation type and select

**Validation type**  
Automatic

Back Next

**Additional configuration**

**Primary metric**  
AveragePrecisionScoreWeighted

☒ Explain best model

☐ Enable ensemble stacking

☒ Use all supported models

**Blocked models**  
A list of models that Automated ML will not use during training.

**Positive class label**  
Positive class label

Save Close

Select your compute instance.

Default Directory > mltestresource > Training job

### Submit an Automated ML job

✓ Training method

✓ Basic settings

✓ Task type & data

✓ Task settings

5 Compute

6 Review

#### Compute

Select and configure the compute resource for executing your training job.

Select compute type

Compute instance

Select Azure ML compute instance \*

samplecomputeinstance, Running, Standard\_DS11\_v2, 2 vCPUs (cores), 14 GB, 28 GB (storage)

samplecomputeinstance - Running

Standard\_DS11\_v2

2 vCPUs (cores), 14 GB, 28 GB (storage), \$0.18/hr

simplecomputeinstance2 - Stopped

Standard\_A1\_v2

1 vCPU (core), 2 GB, 10 GB (storage), \$0.04/hr

Back

Next

Review and create.

Default Directory > mltestresource > Training job

### Submit an Automated ML job

✓ Training method

✓ Basic settings

✓ Task type & data

✓ Task settings

✓ Compute

6 Review

#### Review

Review or make changes to your job before submission.

##### Basic settings

Name  
epic\_jicama\_c2ldq2xxmm

Experiment name  
test

Timeout (hours)  
--

##### Task type & data

Task type  
Classification

Data  
creditcardfraud

##### Task settings

Target column  
is\_fraud

Max trials: 3

Max concurrent runs: 1

Max nodes: 1

Metric score  
Area Under the Curve (AUC)

Experiment timeout (hours)  
1

Iteration timeout (minutes)  
15

Enable deep learning  
No




Back

Submit training job







Cancel

After this is completed we can verify the results.


Default Directory > mltestresource > Jobs > test > epic\_jicama\_c2ldq2xmmm

epic\_jicama\_c2ldq2xmmm    Completed

Overview Data guardrails Models + child jobs Outputs + logs Child jobs

 Refresh  Edit and submit (preview)  Register model  Cancel  Delete |  Compare

**Properties**

**Status**  
 Completed

**Created on**  
Feb 12, 2026 8:29 AM

**Start time**  
Feb 12, 2026 8:29 AM

**Duration**  
17m 28.31s

**Compute duration**  
17m 28.31s

**Inputs**

**Input name:** training\_data  
Data asset: [creditcardfraud:1](#)

Asset URI: [azureml:creditcardfraud:1](#)




**Outputs**

**Output name:** best\_model  
Model:  
[azureml\\_epic\\_jicama\\_c2ldq2xmmm\\_1\\_output\\_mlflow\\_log\\_model\\_209276562:1](#)

Asset URI: [azureml:azureml\\_epic\\_jicama\\_c2ldq2xmmm\\_1\\_output\\_mlflow\\_l...](#)

These are the results.

Default Directory > mltestresource > Jobs > test > epic\_jicama\_c2ldq2xmmm

epic\_jicama\_c2ldq2xmmm    Completed

Overview Data guardrails Models + child jobs Outputs + logs

**Compute duration**  
17m 28.31s

**Compute target**  
[samplecomputeinstance](#)

**Name**  
epic\_jicama\_c2ldq2xmmm

**Script name**  
--

**Created by**  
Ricardo Robledo

**Job type**  
Automated ML

**Experiment**  
[test](#)

**Run Metrics**

**Accuracy**  
0.99916

**AUC macro**  
0.99820

**AUC micro**  
0.99998

**AUC weighted**  
0.99820

**Average precision score macro**  
0.97032

**Average precision score micro**  
0.99998




**Average precision score weighted**  
0.99976

**Balanced accuracy**  
0.90675

**F1 score macro**  
0.94078



Default Directory > mltestresource > Jobs > test > epic\_jicama\_c2

epic\_jicama\_c2ldq2xxmm    Completed

Overview

Data guardrails

Models + child jobs

Outputs + logs

Compute target

samplecomputeinstance

Name

epic\_jicama\_c2ldq2xxmm

Script name

--

Created by

Ricardo Robledo

Job type

Automated ML

Experiment

test

Arguments

None

See all properties

Log loss

0.0029130

Matthews correlation

0.88474

Norm macro recall

0.81351

Precision score macro

0.98113

Precision score micro

0.99916

Precision score weighted

0.99914

Recall score macro

0.90675

Recall score micro

0.99916

Recall score weighted

0.99916

Weighted accuracy

0.99988

Close

That's all!, this is the model finished.