

H2O FlowによるIsolation Forestの実行


# Data Load

- データの読み込み

データパスの指定

```
importFiles
```

17ms

 Import Files

Search: `./data/my_sample/creditcard.csv`

Search Results: (All files added)

Selected Files: 1 file selected: [Clear All](#)

✕ `./data/my_sample/creditcard.csv`

Actions: 


Import


**◀ PUSH**

CS

```
importFiles [ "./data/my_sample/creditcard.csv" ]
```

221ms

 1 / 1 files imported.

Files  `./data/my_sample/creditcard.csv`

Actions 

Parse these files...

**◀ PUSH**

# Data Load

- データの読み込み：カラム型の指定

CS

setupParse source\_frames: [ "nfs://./data/my\_sample/creditcard.csv" ]

624ms

### Setup Parse

PARSE CONFIGURATION

Sources nfs://./data/my\_sample/creditcard.csv

ID creditcard.hex

Parser CSV

Separator ','

Column Headers ☐ Auto  
☒ First row contains column names  
☐ First row contains data

Options ☐ Enable single quotes as a field quotation character  
☒ Delete on done

EDIT COLUMN NAMES AND TYPES

Search by column name...

	1	Time	Numeric	0	0	1	1	2	2	4
2	V1	Numeric	-1.3598071336738	1.19185711131486	-1.35835406159823	-0.966271711572087	-1.15823309349523	-0.425965884412454	1.229	
3	V2	Numeric	-0.0727811733098497	0.26615071205963	-1.34016307473609	-0.185226008082898	0.877736754848451	0.960523044882985	0.141	
4	V3	Numeric	2.53634673796914	0.16648011335321	1.77320934263119	1.79299333957872	1.548717846511	1.14110934232219	0.045	
5	V4	Numeric	1.37815522427443	0.448154078460911	0.379779593034328	-0.863291275036453	0.403033933955121	-0.168252079760302	1.202	
6	V5	Numeric	-0.338320769942518	0.0600176492822243	-0.503198133318193	-0.0103088796030823	-0.407193377311653	0.42098688077219	0.191	
7	V6	Numeric	0.462387777762292	-0.0823608088155687	1.80049938079263	1.24720316752486	0.0959214624684256	-0.0297275516639742	0.272	
8	V7	Numeric	0.239598554061257	-0.0788029833323113	0.791460956450422	0.23760893977178	0.592940745385545	0.476200948720027	-0.00	
9	V8	Numeric	0.0986979012610507	0.0851016549148104	0.247675786588991	0.377435874652262	-0.270532677192282	0.260314333074874	0.081	
10	V9	Numeric	0.363786969611213	-0.255425128109186	-1.51465432260583	-1.38702406270197	0.817739308235294	-0.56867137571251	0.464	
11	V10	Numeric	0.0907941719789316	-0.166974414004614	0.207642865216696	-0.0549519224713749	0.753074431976354	-0.371407196834471	-0.09	
12	V11	Numeric	-0.551599533260813	1.61272666105479	0.624501459424895	-0.226487263835401	-0.822842877946363	1.34126198001957	-1.41	
13	V12	Numeric	-0.617800855762348	1.06523531137287	0.066083685268831	0.178228225877303	0.53819555014995	0.359893837038039	-0.15	
14	V13	Numeric	-0.991389847235408	0.48909501589608	0.717292731410831	0.507756869957169	1.3458515932154	-0.358090652573631	-0.75	
15	V14	Numeric	-0.311169353699879	-0.143772296441519	-0.165945922763554	-0.28792374549456	-1.11966983471731	-0.137133700217612	0.167	

← Previous page→ Next page

Parse

PUSH

# Data Load

- データの読み込み

```
CS parseFiles
source_frames: ["nfs://./data/my_sample/creditcard.csv"]
destination_frame: "creditcard.hex"
parse_type: "CSV"
separator: 44
number_columns: 31
single_quotes: false
column_names:
["Time","V1","V2","V3","V4","V5","V6","V7","V8","V9","V10","V11","V12","V13","V14","V15","V16","V17","V18","V19","V20","V21","V22","V23","V24","V
25","V26","V27","V28","Amount","Class"]
column_types:
["Numeric","Numeric","Numeric","Numeric","Numeric","Numeric","Numeric","Numeric","Numeric","Numeric","Numeric","Numeric","Numeric","Numeric","Num
eric","Numeric","Numeric","Numeric","Numeric","Numeric","Numeric","Numeric","Numeric","Numeric","Numeric","Numeric","Numeric","Numeric","Numeric",
,"Numeric","Numeric"]
delete_on_done: true
check_header: 1
chunk_size: 2356736
```

☰ Job

Run Time 00:00:01.408

Remaining Time 00:00:00.0

Type Frame

Key 🔍 [creditcard.hex](#)

Description Parse

Status DONE

Progress 100%

Done.

### Actions

**Q** View

**◀ PUSH**

# Loaded Data

- ロードされたデータ

各カラムの要約統計量の表示

CS

getFrameSummary "creditcard.hex"

602ms

creditcard.hex

Actions: 

View Data

Split

Build Model

Run All

Predict

Download

Export

Delete

Rows

 284807 

Columns

 31 

Compressed Size

 63MB

COLUMN SUMMARIES

Label	type	Missing	Zeros	+Inf	-Inf	min	max	mean	sigma	cardinality	Actions
Time	int	0	2	0	0	0	172792.0	94813.8596	47488.1460		Convert to enum
V1	real	0	0	0	0	-56.4075	2.4549	0.0	1.9587		
V2	real	0	0	0	0	-72.7157	22.0577	0.0	1.6513		
V3	real	0	0	0	0	-48.3256	9.3826	-0.0	1.5163		
V4	real	0	0	0	0	-5.6832	16.8753	0.0	1.4159		
V5	real	0	0	0	0	-113.7433	34.8017	0.0	1.3802		
V6	real	0	0	0	0	-26.1605	73.3016	0.0	1.3323		
V7	real	0	0	0	0	-43.5572	120.5895	-0.0	1.2371		
V8	real	0	0	0	0	-73.2167	20.0072	0.0	1.1944		
V9	real	0	0	0	0	-13.4341	15.5950	-0.0	1.0986		
V10	real	0	0	0	0	-24.5883	23.7451	0.0	1.0888		
V11	real	0	0	0	0	-4.7975	12.0189	0.0	1.0207		
V12	real	0	0	0	0	-18.6837	7.8484	-0.0	0.9992		
V13	real	0	0	0	0	-5.7919	7.1269	0.0	0.9953		
V14	real	0	0	0	0	-19.2143	10.5268	0.0	0.9586		
V15	real	0	0	0	0	-4.4989	8.8777	0.0	0.9153		
V16	real	0	0	0	0	-14.1299	17.3151	0.0	0.8763		
V17	real	0	0	0	0	-25.1628	9.2535	-0.0	0.8493		
V18	real	0	0	0	0	-9.4987	5.0411	0.0	0.8382		
V19	real	0	0	0	0	-7.2135	5.5920	0.0	0.8140		

Previous 20 Columns

Next 20 Columns

CHUNK COMPRESSION SUMMARY

FRAME DISTRIBUTION SUMMARY

# Run Algorithm (Model設定)

- モデル設定（ハイパーパラメータ等）の実施

モデルの選択  
("Isolation Forest")

利用しない変数の指定  
(正解ラベルである"Class"  
カラムを除外する)

CS

assist buildModel, null, training\_frame: "creditcard1.hex"

14ms

Build a Model

Select an algorithm: Isolation Forest

PARAMETERS

model\_id

isolationforest-6a0eb88c-d2e1-4f87-9246-5ab5bbf6cc12

Destination id for this model; auto-generated if not specified.

training\_frame

creditcard1.hex

Id of the training data frame.

ignored\_columns

Search...

Showing page 4 of 4. -30 ignored.

☒ Class

INT

☒ All

☐ None

← Previous 10

→ Next 10

Names of columns to ignore for training.

Only show columns with more than 0 % missing values.

ignore\_const\_cols

☒

Ignore constant columns.

ntrees

50

Number of trees.

max\_depth

8

Maximum tree depth.

min\_rows

1

Fewest allowed (weighted) observations in a leaf.

seed

-1

Seed for pseudo random number generator (if applicable)

mtries

-1

Number of variables randomly sampled as candidates at each split. If set to -1, defaults (number of predictors)/3.

sample\_size

256

Number of randomly sampled observations used to train each Isolation Forest tree. Only one of parameters sample\_size and sample\_rate should be defined. If sample\_rate is defined, sample\_size will be ignored.

sample\_rate

-1

Rate of randomly sampled observations used to train each Isolation Forest tree. Needs to be in range from 0.0 to 1.0. If set to -1, sample\_rate is disabled and sample\_size will be used instead.

GRID?

ADVANCED

score\_each\_iteration

☐

Whether to score during each iteration of model training.

GRID?

export\_checkpoints\_dir

Automatically export generated models to this directory.

EXPERT

build\_tree\_one\_node

☐

Run on one node only; no network overhead but fewer cpus used. Suitable for small datasets.

col\_sample\_rate\_change\_per\_level

1

Relative change of the column sampling rate for every level (must be > 0.0 and <= 2.0)

GRID?

Build Model

◀ PUSH

# Algorism Built (Model学習完了)

- モデル学習完了の表示

CS

```
buildModel 'isolationforest', {"model_id":"isolationforest-6a0eb88c-d2e1-4f87-9246-5ab5bbf6cc12","training_frame":"creditcard1.hex","ignored_columns":["Class"],"ignore_const_cols":true,"ntrees":50,"max_depth":8,"min_rows":1,"seed":-1,"mtries":-1,"sample_size":256,"sample_rate":-1,"score_each_iteration":false,"score_tree_interval":0,"max_runtime_secs":0,"col_sample_rate_per_tree":1,"categorical_encoding":"AUTO","stopping_rounds":0,"stopping_metric":"AUTO","stopping_tolerance":0.01,"build_tree_one_node":false,"col_sample_rate_change_per_level":1}
```

2.7s

☰ Job

Run Time 00:00:02.347

Remaining Time 00:00:00.0

Type Model

Key  isolationforest-6a0eb88c-d2e1-4f87-9246-5ab5bbf6cc12

Description IsolationForest

Status DONE

Progress 100%

Done.

Actions  View  **PUSH**

# Model Check

- 学習済みモデルの詳細確認

CS | `getModel "isolationforest-6a0eb88c-d2e1-4f87-9246-5ab5bbf6cc12"` 387ms

### Model

Model ID: isolationforest-6a0eb88c-d2e1-4f87-9246-5ab5bbf6cc12  
Algorithm: Isolation Forest

Actions: [Refresh](#) [Predict...](#) [PUSH](#) [Download Model Deployment Package \(MOJO\)](#) [Export](#) [Inspect](#) [Delete](#) [Download Gen Model](#)

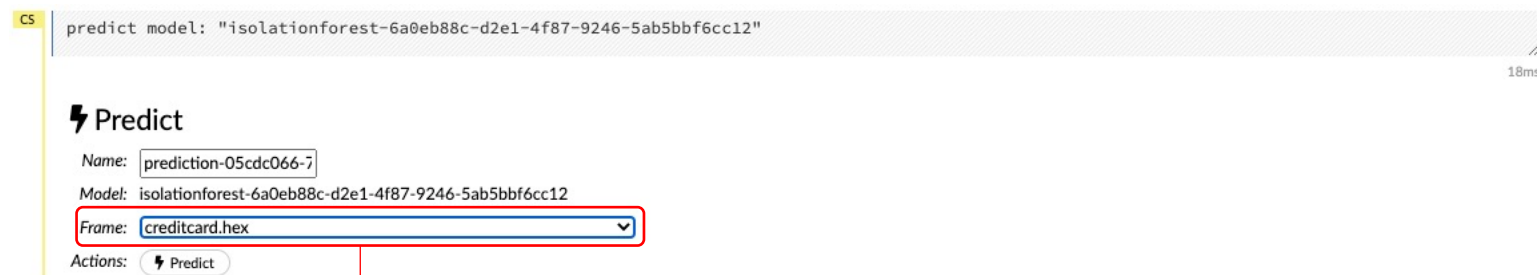
- ▶ MODEL PARAMETERS
- ▶ OUTPUT
- ▶ NAMES
- ▶ COLUMN TYPES
- ▶ DOMAINS
- ▶ OUTPUT - MODEL SUMMARY
- ▶ OUTPUT - SCORING HISTORY
- ▶ OUTPUT - TRAINING METRICS
- ▶ OUTPUT - HELP
- ▶ OUTPUT - VARIABLE SPLITS

学習済みモデルを用いたスコアリング



# Predict New Data

- スコアリングを実施するデータの指定



スコアリング用データの選択  
(今回実施のモデルは教師無し学習(Isolation Forest).  
モデル学習に用いたデータをスコアリングする)


# Predicted Dataset

- スコアアリング終了の表示

CS

predict model: "isolationforest-6a0eb88c-d2e1-4f87-9246-5ab5bbf6cc12", frame: "creditcard.hex", predictions\_frame: "prediction-05cdc066-7b79-44f0-a463-25936a1ef403"


1.1s


 Prediction

Actions: Inspect

▼ PREDICTION

model	isolationforest-6a0eb88c-d2e1-4f87-9246-5ab5bbf6cc12
model_checksum	-6814786020382712704
frame	creditcard.hex
frame_checksum	-9153354249803528524
description	Isolation Forest Metrics
model_category	AnomalyDetection
scoring_time	1622102647286
predictions	prediction-05cdc066-7b79-44f0-a463-25936a1ef403
MSE	NaN
RMSE	NaN
nobs	284807
custom_metric_name	.
custom_metric_value	0
mean_score	6.806073
mean_normalized_score	0.048971

 Combine predictions with frame

 PUSH

スコアリング結果をスコアリング用データへ結合し、表示

# Check Data

- スコアアリング結果データ

データの表示

スコアリング結果カラム

predict: 正規化された異常値スコア (0~1の範囲)  
mean\_length: 分岐数の平均

データのダウンロード

bindFrames "combined-prediction-05cdc066-7b79-44f0-a463-25936a1ef403", [ "prediction-05cdc066-7b79-44f0-a463-25936a1ef403", "creditcard.hex" ]

376ms

Frames Combined

The specified frames were combined successfully.

View Frame

getFrameSummary "combined-prediction-05cdc066-7b79-44f0-a463-25936a1ef403"

406ms

combined-prediction-05cdc066-7b79-44f0-a463-25936a1ef403

Actions: View Data Split Build Model Run AutoML Predict Download Export Delete

Rows

Columns

Compressed Size

284807

33

64MB

COLUMN SUMMARIES

label	type	Missing	Zeros	+Inf	-Inf	min	max	mean	sigma	cardinality	Actions
predict	real	0	34652	0	0	0	1.0	0.0490	0.0669	*	*
mean_length	real	0	0	0	0	3.0400	7.0	6.8061	0.2648	*	*
Time	int	0	2	0	0	0	172792.0	94813.8596	47488.1460	*	Convert to enum
V1	real	0	0	0	0	-56.4075	2.4549	0.0	1.9587	*	*
V2	real	0	0	0	0	-72.7157	22.0577	0.0	1.6513	*	*
V3	real	0	0	0	0	-48.3256	9.3826	-0.0	1.5163	*	*
V4	real	0	0	0	0	-5.6832	16.8753	0.0	1.4159	*	*
V5	real	0	0	0	0	-113.7433	34.8017	0.0	1.3802	*	*
V6	real	0	0	0	0	-26.1605	73.3016	0.0	1.3323	*	*
V7	real	0	0	0	0	-43.5572	120.5895	-0.0	1.2371	*	*
V8	real	0	0	0	0	-73.2167	20.0072	0.0	1.1944	*	*
V9	real	0	0	0	0	-13.4341	15.5950	-0.0	1.0986	*	*
V10	real	0	0	0	0	-24.5883	23.7451	0.0	1.0888	*	*
V11	real	0	0	0	0	-4.7975	12.0189	0.0	1.0207	*	*
V12	real	0	0	0	0	-18.6837	7.8484	-0.0	0.9992	*	*
V13	real	0	0	0	0	-5.7919	7.1269	0.0	0.9953	*	*
V14	real	0	0	0	0	-19.2143	10.5268	0.0	0.9586	*	*
V15	real	0	0	0	0	-4.4989	8.8777	0.0	0.9153	*	*
V16	real	0	0	0	0	-14.1299	17.3151	0.0	0.8763	*	*
V17	real	0	0	0	0	-25.1628	9.2535	-0.0	0.8493	*	*

Previous 20 ColumnsNext 20 Columns

CHUNK COMPRESSION SUMMARY

FRAME DISTRIBUTION SUMMARY