

Detect Fraud from Customer Transactions

Introduction

Lots of people use credit card nowadays, there may be millions of online transactions per day. At the same time, fraud happens from time to time, and it cause millions of dollars loss. So online transaction fraud detection is important for banks and customers.

We need a system to detect the credit card fraud automatically in time to prevent further damage. This will help the banks to save money, help the government to crack down economic crimes, and this protect customers benefits.

So, the problem is "can we automatically detect the online transaction fraud in time based on an ongoing transaction"? This is a binary classification problem, with the historical data of online transaction, we can use machine learning algorithms to solve this problem.

Data description

IEEE-CIS Fraud Detection is a completion in kaggle.com, IEEE-CIS works across a variety of AI and machine learning areas, including deep neural networks, fuzzy systems, evolutionary computation, and swarm intelligence. Today they're partnering with the world's leading payment service company, Vesta Corporation, seeking the best solutions for fraud prevention industry. In this report, I will use the dataset from this completion.

The dataset includes following csv files. The data comes from Vesta's real-world e-commerce transactions and contains a wide range of features from device type to product features.

- `train_{transaction, identity}.csv` - the training set
- `test_{transaction, identity}.csv` - the test set

The data is broken into two files `identity` and `transaction`, which are joined by `TransactionID`. Not all transactions have corresponding identity information. The target is predicting the probability that an online transaction is fraudulent, as denoted by the binary target `isFraud`.

```
train_transaction.head()
```

:

	TransactionID	isFraud	TransactionDT	TransactionAmt	ProductCD	card1	card2	card3	card4	card5	...	V330	V331	V332	V333
0	2987000	0	86400	68.5	W	13926	NaN	150.0	discover	142.0	...	NaN	NaN	NaN	NaN
1	2987001	0	86401	29.0	W	2755	404.0	150.0	mastercard	102.0	...	NaN	NaN	NaN	NaN
2	2987002	0	86469	59.0	W	4663	490.0	150.0	visa	166.0	...	NaN	NaN	NaN	NaN
3	2987003	0	86499	50.0	W	18132	567.0	150.0	mastercard	117.0	...	NaN	NaN	NaN	NaN
4	2987004	0	86506	50.0	H	4497	514.0	150.0	mastercard	102.0	...	0.0	0.0	0.0	0.0

5 rows × 394 columns