



Deteção de Fraudes IEEE-CIS

Descrição

Imagine estar no caixa do supermercado com uma longa fila atrás de você, e o caixa anuncia em voz alta que seu cartão foi recusado. Nesse momento, provavelmente você não está pensando na ciência de dados que determinou seu destino. Envergonhado e certo de que tem fundos para cobrir tudo o que é necessário para uma festa épica de nachos para 50 dos seus amigos mais próximos, você tenta passar o cartão novamente.

Mesmo resultado. Enquanto você se afasta e permite que o caixa atenda o próximo cliente, você recebe uma mensagem de texto do seu banco. "Aperte 1 se você realmente tentou gastar \$500 em queijo cheddar." Embora possa ser inconveniente (e muitas vezes embaraçoso) no momento, esse sistema de prevenção de fraudes está economizando milhões de dólares por ano para os consumidores. Pesquisadores da Sociedade de Inteligência Computacional do IEEE (IEEE-CIS) desejam melhorar esse número, ao mesmo tempo em que melhoram a experiência do cliente. Com uma detecção de fraudes de maior precisão, você pode aproveitar seus chips sem complicações.

A IEEE-CIS trabalha em várias áreas de inteligência artificial e aprendizado de máquina, incluindo redes neurais profundas, sistemas difusos, computação evolutiva e inteligência de enxame. Hoje, eles estão se associando à principal empresa de serviços de pagamento do mundo, a Vesta Corporation, em busca das melhores soluções para a indústria de prevenção de fraudes, e agora você está convidado a participar do desafio.

Nesta competição, você avaliará modelos de aprendizado de máquina em um grande conjunto de dados desafiador. Os dados provêm de transações de comércio eletrônico do mundo real da Vesta e contêm uma ampla gama de recursos, desde o tipo de dispositivo até as características do produto. Você também tem a oportunidade de criar novos recursos para melhorar seus resultados. Se for bem-sucedido, você melhorará a eficácia dos alertas de transações fraudulentas para milhões de pessoas em todo o mundo, ajudando centenas de milhares de empresas a reduzir suas perdas com fraudes e aumentar sua receita. E, é claro, você salvará pessoas de festa, assim como você, do incômodo de falsos positivos.

A Vesta Corporation forneceu o conjunto de dados para esta competição. A Vesta Corporation é a precursora em soluções de pagamento de comércio eletrônico garantidas. Fundada em 1995, a Vesta foi pioneira no processo de transações de pagamento de cartão não presente totalmente garantidas

para a indústria de telecomunicações. Desde então, a Vesta expandiu firmemente suas capacidades de ciência de dados e aprendizado de máquina em todo o mundo e consolidou sua posição como líder em pagamentos de comércio eletrônico garantidos. Hoje, a Vesta garante mais de \$18 bilhões em transações anualmente. Foto do cabeçalho de Tim Evans no Unsplash

Avaliação

As submissões são avaliadas com base na área sob a curva ROC entre a probabilidade prevista e o alvo observado.

Arquivo de Submissão

Para cada TransactionID no conjunto de teste, você deve prever uma probabilidade para a variável **isFraud**. O arquivo deve conter um cabeçalho e ter o seguinte formato:

| TransactionID | isFraud |
|---------------|---------|
| 3663549 | 0.5 |
| 3663550 | 0.5 |
| 3663551 | 0.5 |

Descrição do Conjunto de Dados

Nesta competição, você está prevendo a probabilidade de que uma transação online seja fraudulenta, conforme indicado pelo alvo binário isFraud. Os dados estão divididos em dois arquivos, identity (identidade) e transaction (transação), que são unidos pelo TransactionID. Nem todas as transações têm informações de identidade correspondentes.

Recursos categóricos - transação

- **ProductCD**
- **card1 - card6**
- **addr1 addr2**
- **P_emaildomain**
- **R_emaildomain**
- **M1 - M9**

Características Categóricas - Identidade

- **DeviceType**
- **DeviceInfo**
- **id_12 - id_38**

O recurso **TransactionDT** é um timedelta de uma determinada data e hora de referência (não um carimbo de data/hora real). Você pode ler mais sobre os dados nesta postagem do anfitrião da competição.

Arquivos

- **train_{transaction, Identity}.csv** - o conjunto de treinamento
- **test_{transaction, Identity}.csv** - o conjunto de teste (você deve prever o valor isFraud para essas observações)
- **sample_submission.csv** - um arquivo de envio de amostra no formato correto

Citação

Addison Howard, Bernadette Bouchon-Meunier, IEEE CIS, inversão, John Lei, Lynn@Vesta, Marcus2010, Prof. Hussein Abbass. (2019). IEEE-CIS Fraud Detection. Kaggle.
<https://kaggle.com/competitions/ieee-fraud-detection>