Proposal for Online Payment Fraud Detection Using Machine Learning

Presented by Eng/ Abdullah Nabil

To Eng/ Tarek Ziad

Project Overview

Online Payment Fraud Detection is critical in modern financial systems due to the increase in online transactions. The goal of this project is to develop a machine learning model that can accurately detect fraudulent transactions, ensuring the safety of online payments and protecting users from financial losses.

Objectives

1. Data Understanding and Visualization: Analyze and visualize the dataset to understand the underlying patterns and relationships between features.

2. Data Preprocessing:Clean and preprocess the data to make it suitable for model training.

3. Feature Engineering: Transform and create relevant features to improve the model's performance.

4. Model Training and Evaluation: Train various machine learning models and evaluate their performance using appropriate metrics.

5. Model Deployment: Develop a pipeline to deploy the model for real-time fraud detection.

Dataset

The dataset contains the following columns:

- `**step**`**:** Time unit of the transaction.

- `**type**`**:** Type of transaction.

- `**amount**`**:** Total amount of the transaction.

- `**nameOrig**`**:** Account initiating the transaction.

- `**oldbalanceOrg**`**:** Initial balance of the sender's account.

- `**newbalanceOrg**`**:** Balance of the sender's account after the transaction.

- `**nameDest**`**:** Account receiving the transaction.

- `**oldbalanceDest**`**:** Initial balance of the receiver's account.

- `newbalanceDest`: Balance of the receiver's account after the transaction.

- `isFraud`: Target variable indicating whether the transaction is fraudulent (1) or not (0)

Data Information:

```
Cleaned Dataset Information:
<class 'pandas.core.frame.DataFrame'>
Int64Index: 4319406 entries, 0 to 6259911
Data columns (total 11 columns):
 #   Column          Dtype
---  ------          -----
 0   step            int64
 1   type            object
 2   amount          float64
 3   nameOrig        object
 4   oldbalanceOrg   float64
 5   newbalanceOrig  float64
 6   nameDest        object
 7   oldbalanceDest  float64
 8   newbalanceDest  float64
 9   isFraud         int64
 10  isFlaggedFraud  int64
dtypes: float64(5), int64(3), object(3)
memory usage: 395.5+ MB
None
```

Distribution of Transaction Type
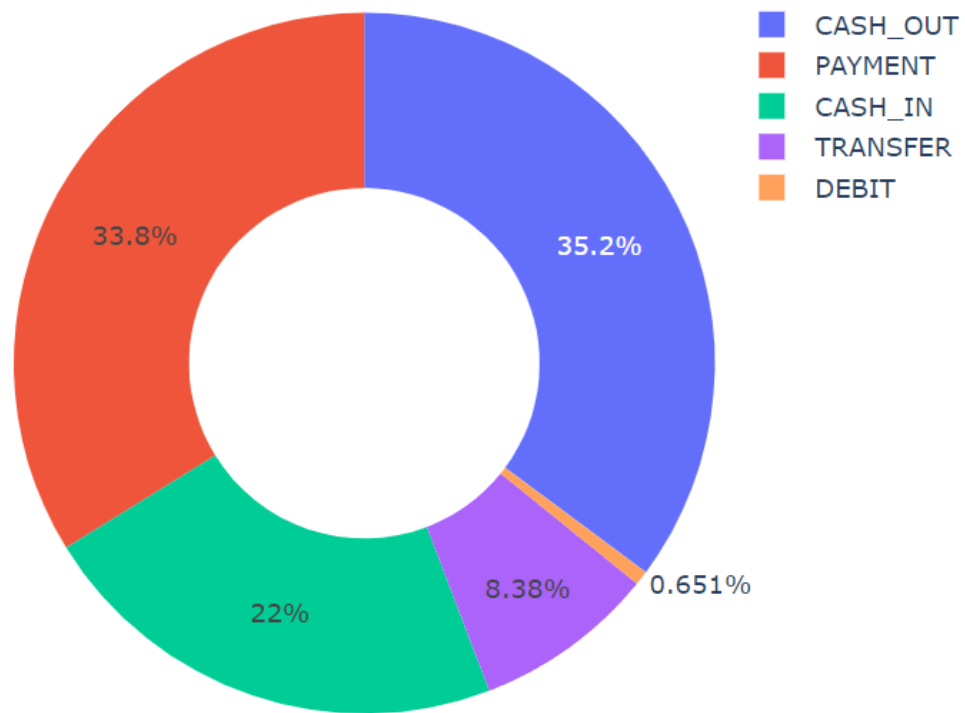
Data Information (Cont.):

```
In [5]:   # Exploring transaction type

          print(data.type.value_counts())
```

```
CASH_OUT    2237500
PAYMENT     2151495
CASH_IN     1399284
TRANSFER     532909
DEBIT         41432
Name: type, dtype: int64
```

Data overview:

	step	type	amount	nameOrig	oldbalanceOrg	newbalanceOrig	nameDest	oldbalanceDest	newbalanceDest	isFraud	isFlaggedFraud
0	1	PAYMENT	9839.64	C1231006815	170136.0	160296.36	M1979787155	0.0	0.0	0	0
1	1	PAYMENT	1864.28	C1666544295	21249.0	19384.72	M2044282225	0.0	0.0	0	0
2	1	TRANSFER	181.00	C1305486145	181.0	0.00	C553264065	0.0	0.0	1	0
3	1	CASH_OUT	181.00	C840083671	181.0	0.00	C38997010	21182.0	0.0	1	0
4	1	PAYMENT	11668.14	C2048537720	41554.0	29885.86	M1230701703	0.0	0.0	0	0

```
# Reference: The dataset used for this project is available at:
# https://www.kaggle.com/datasets/jainilcoder/online-payment-fraud-
detection/code
```