Credit Card Fraud Detection Using Machine Learning Algorithms

Upload your input CSV file

Drag and drop file here
Limit 200MB per file • CSV

Browse files

📄 card_transdata.csv 72.7MB   ✕

First 5 rows of the data:

	distance_from_home	distance_from_last_transaction	ratio_to_median_purchase_price	repeat_retailer
0	57.8779	0.3111	1.9459	1
1	10.8299	0.1756	1.2942	1
2	5.0911	0.8052	0.4277	1
3	2.2476	5.6	0.3627	1
4	44.1909	0.5665	2.2228	1

Shape of the data:

```
(1000000, 8)
```

Data types of the columns:

	0
distance_from_home	float64
distance_from_last_transaction	float64
ratio_to_median_purchase_price	float64
repeat_retailer	float64
used_chip	float64
used_pin_number	float64
online_order	float64
fraud	float64

Summary statistics of the data:

	distance_from_home	distance_from_last_transaction	ratio_to_median_purchase_price	repeat_retailer
count	1,000,000	1,000,000	1,000,000	1,000,000
mean	26.6288	5.0365	1.8242	0.8815
std	65.3908	25.8431	2.7996	0.3232
min	0.0049	0.0001	0.0044	0
25%	3.878	0.2967	0.4757	1
50%	9.9678	0.9987	0.9977	1
75%	25.744	3.3557	2.0964	1
max	10,632.7237	11,851.1046	267.8029	1

Missing values in the data:

	0
distance_from_home	0
distance_from_last_transaction	0
ratio_to_median_purchase_price	0
repeat_retailer	0
used_chip	0
used_pin_number	0
online_order	0
fraud	0

Correlation matrix of the data:

	distance_from_home	distance_from_last_transaction	ratio_to_median_purch
distance_from_home	1	0.0002	
distance_from_last_transaction	0.0002	1	
ratio_to_median_purchase_price	-0.0014	0.001	
repeat_retailer	0.1431	-0.0009	
used_chip	-0.0007	0.0021	
used_pin_number	-0.0016	-0.0009	
online_order	-0.0013	0.0001	
fraud	0.1876	0.0919	

Fraudulent Transactions
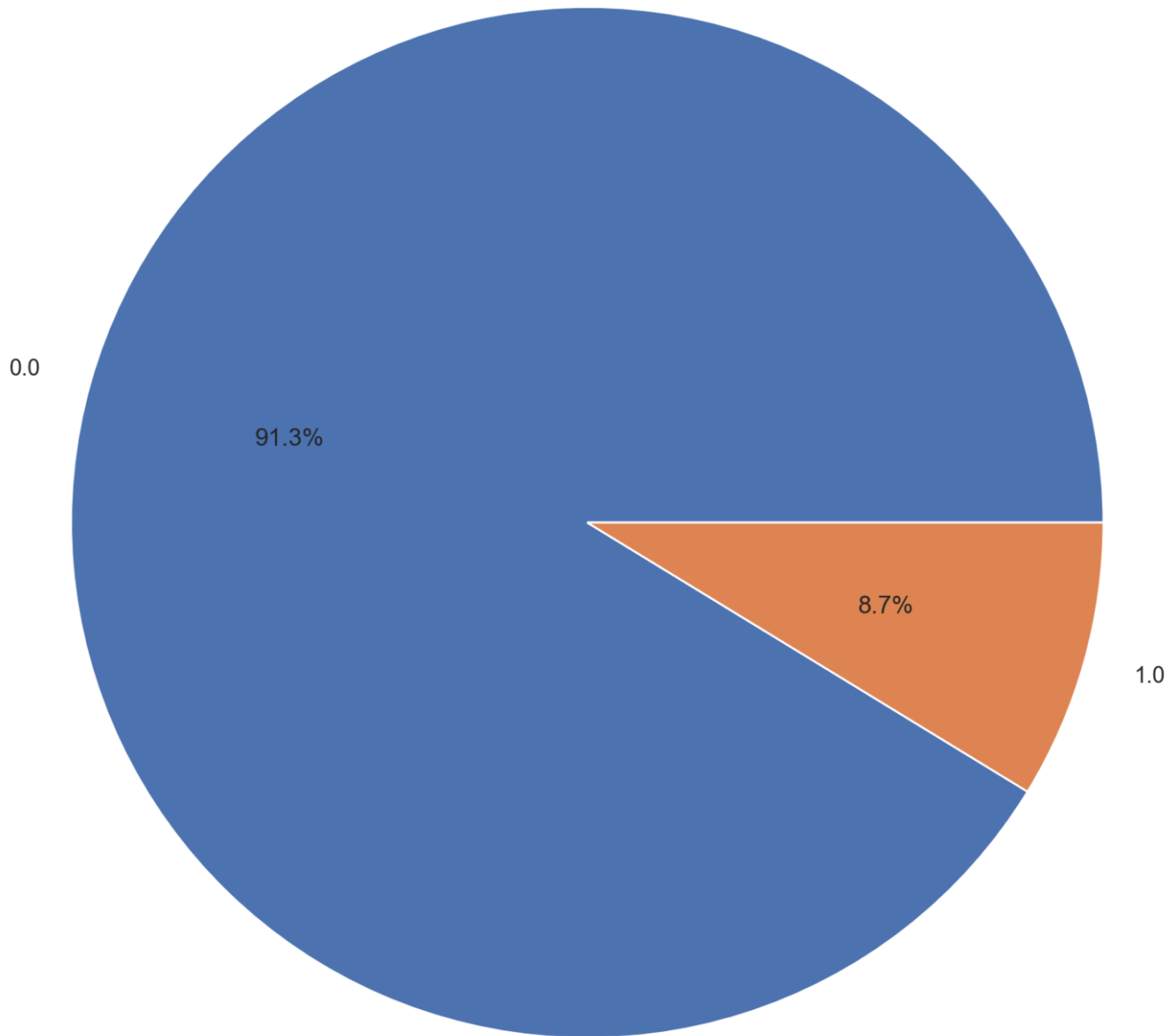
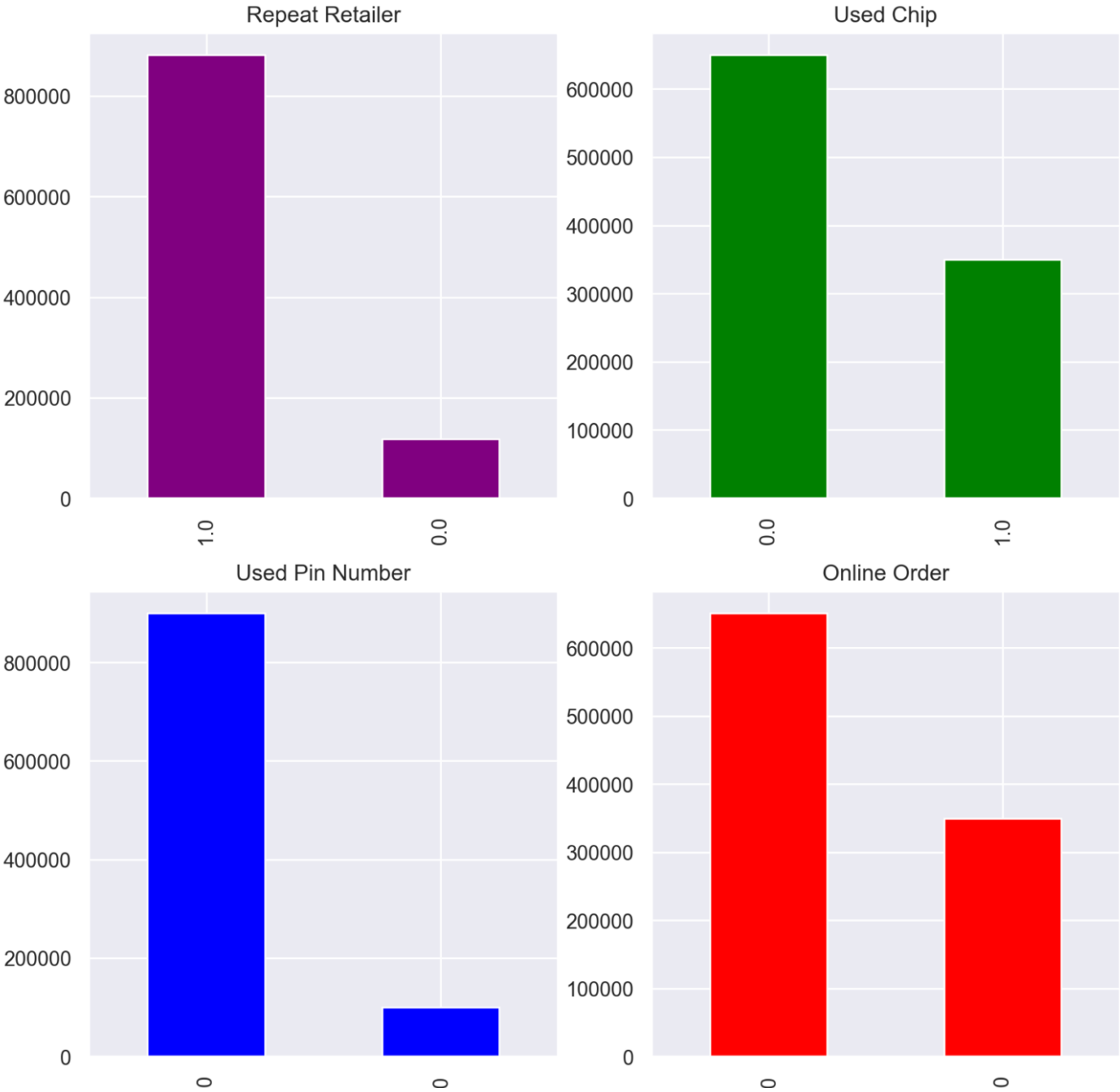# Percentage of Fraud and Non-Fraud Records

Correlation Matrix (Heat Map)
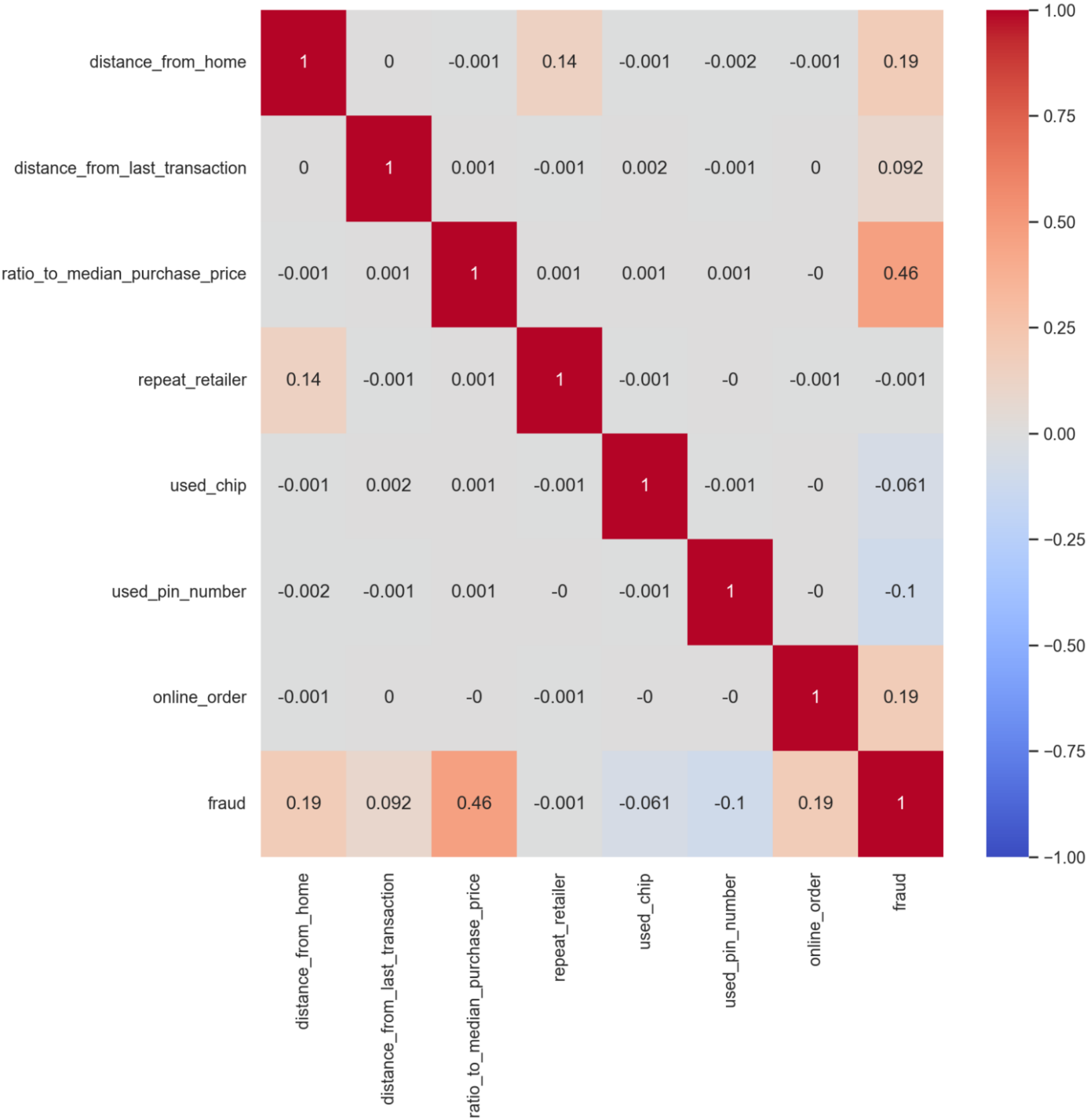
Fraud Data EDA

Fraud Data EDA

Fraud Detection App

Dataset

	distance_from_home	distance_from_last_transaction	ratio_to_median_purchase_price	repeat_retailer
0	57.8779	0.3111	1.9459	1
1	10.8299	0.1756	1.2942	1
2	5.0911	0.8052	0.4277	1
3	2.2476	5.6	0.3627	1
4	44.1909	0.5665	2.2228	1
5	5.5864	13.2611	0.0648	1
6	3.724	0.9568	0.2785	1
7	4.8482	0.3207	1.273	1
8	0.8766	2.5036	1.517	0
9	8.839	2.9705	2.3617	1

Feature Columns

```
▼ [
    0 : "distance_from_home"
    1 : "distance_from_last_transaction"
```

```
    2 : "ratio_to_median_purchase_price"

    3 : "repeat_retailer"

    4 : "used_chip"

    5 : "used_pin_number"

    6 : "online_order"

]
```

Target Variable

fraud

Training and Testing Sets

X_train: `(800000, 7)`

y_train: `(800000,)`

X_test: `(200000, 7)`

y_test: `(200000,)`

Accuracy of logistic regression classifier on test set:

0.95800

Classification Report

```
   precision    recall  f1-score    support


      0.0   0.962066  0.993127  0.977350     182460
```

```
       1.0    0.892351  0.592645  0.712255      17540

   accuracy                        0.958005     200000
  macro avg    0.927208  0.792886  0.844802     200000
weighted avg   0.955952  0.958005  0.954101     200000
```

Accuracy of decision tree classifier on test set:

0.99999

Classification Report

```
   precision    recall  f1-score   support

       0.0    1.000000  0.999984  0.999992    182460
       1.0    0.999829  1.000000  0.999914     17540

   accuracy                        0.999985     200000
  macro avg    0.999914  0.999992  0.999953     200000
weighted avg   0.999985  0.999985  0.999985     200000
```

Accuracy of Random Forest Classifier on test set:

0.99999

Classification Report

```
   precision    recall  f1-score   support
```

```
        0.0   0.999995  1.000000  0.999997    182460
        1.0   1.000000  0.999943  0.999971     17540

   accuracy                       0.999995    200000
  macro avg   0.999997  0.999971  0.999984    200000
weighted avg  0.999995  0.999995  0.999995    200000
```