Credit Card Fraud

Group 6: Sally Mei, Debolina Bhaumik, Cole Barnes, Kanu Madhok

Topic Overview

- According to Data Breach Index, more than 5 million records are stolen on a daily basis
- Fraud-> common for both Card-Present and Card-not present type of payments

- <u>Credit Card Fraud Detection</u>
  - Identify & reject fraudulent purchase attempts
  - Significance - prevent fraudulent purchase attempt charges

- Matplotlib & ML models used to determine trends in fraudulent purchases

Data Overview

- Kaggle Credit Card Fraud Detection dataset: https://www.kaggle.com/datasets/dhanushnarayananr/credit-card-fraud

- 1,000,000 entries
- 8 features

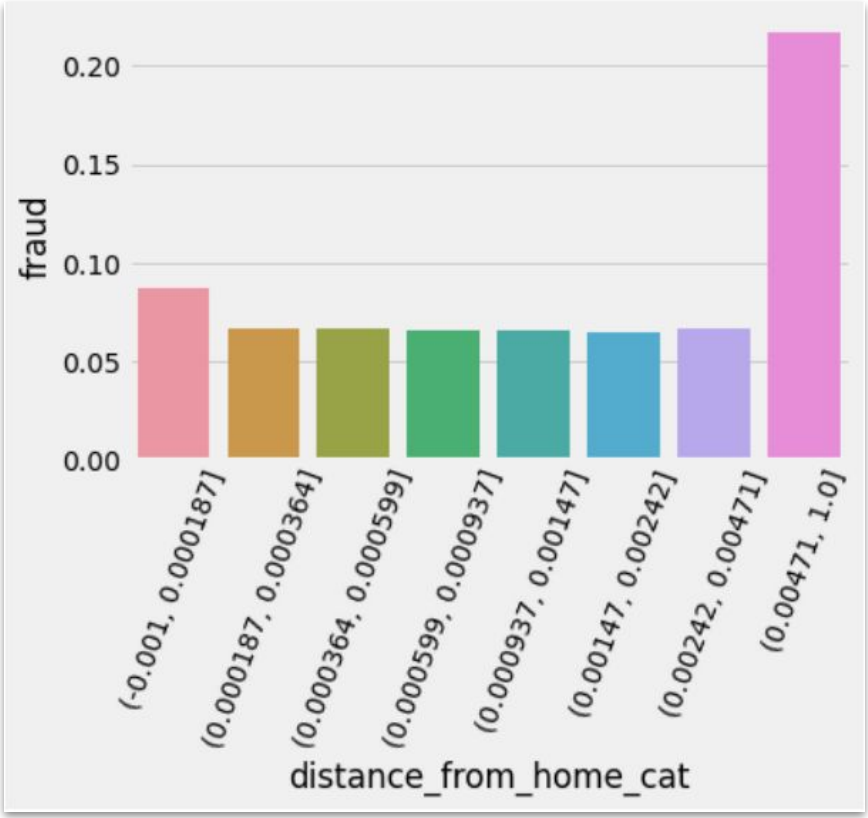
distance_from_home	distance_from_last_transaction	ratio_to_median_purchase_price	repeat_retailer	used_chip	used_pin_number	online_order	fraud

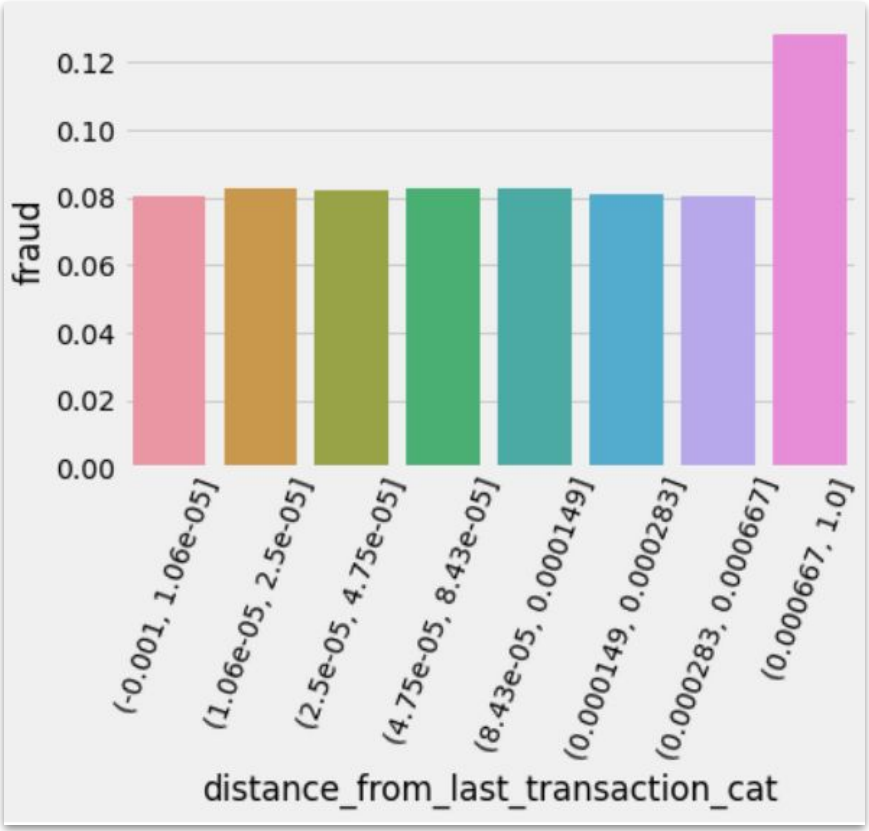
- Target variable: fraud

Pandas

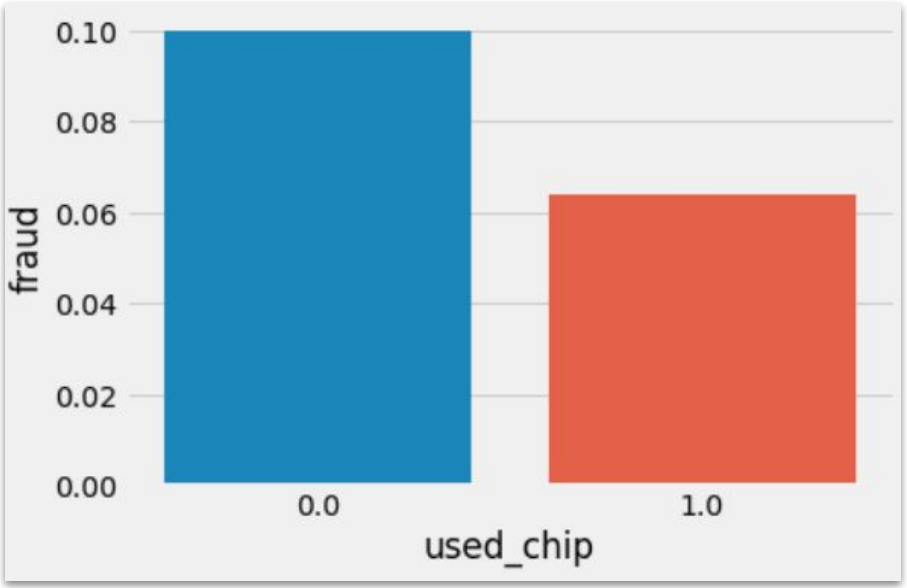
	distance_from_home	distance_from_last_transaction	ratio_to_median_purchase_price	repeat_retailer	used_chip	used_pin_number	online_order	fraud
0	0.005443	0.000026	0.007250	1.0	1.0	0.0	0.0	0.0
1	0.001018	0.000015	0.004816	1.0	0.0	0.0	0.0	0.0
2	0.000478	0.000068	0.001581	1.0	0.0	0.0	1.0	0.0
3	0.000211	0.000473	0.001338	1.0	1.0	0.0	1.0	0.0
4	0.004156	0.000048	0.008284	1.0	1.0	0.0	1.0	0.0

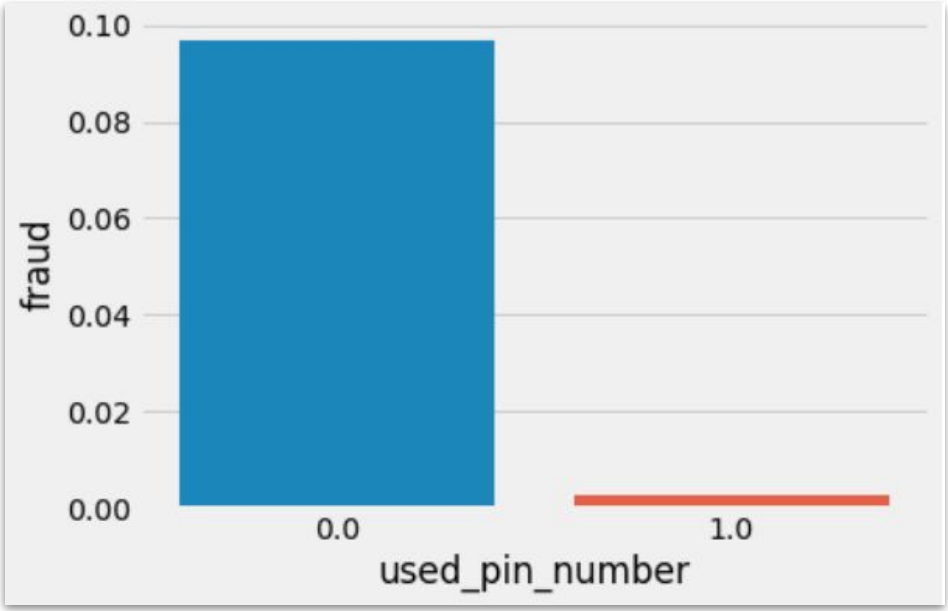
Dataset Overview: Matplotlib





Continued...

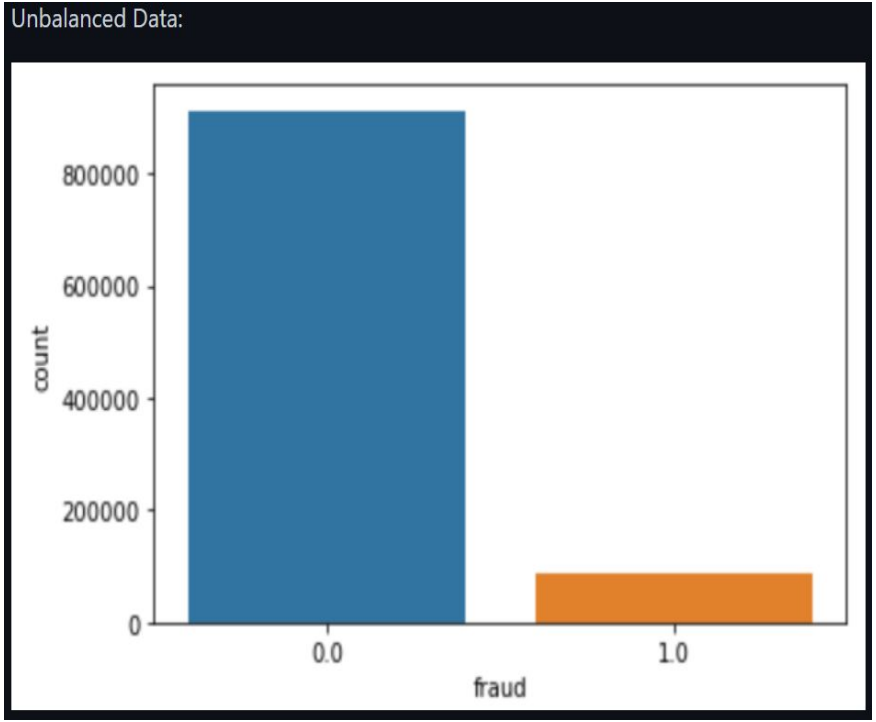

0.10
0.08
0.06
0.04
0.02
0.00
fraud
0.0
1.0
used_chip


0.10
0.08
0.06
0.04
0.02
0.00
fraud
0.0
1.0
used_pin_number

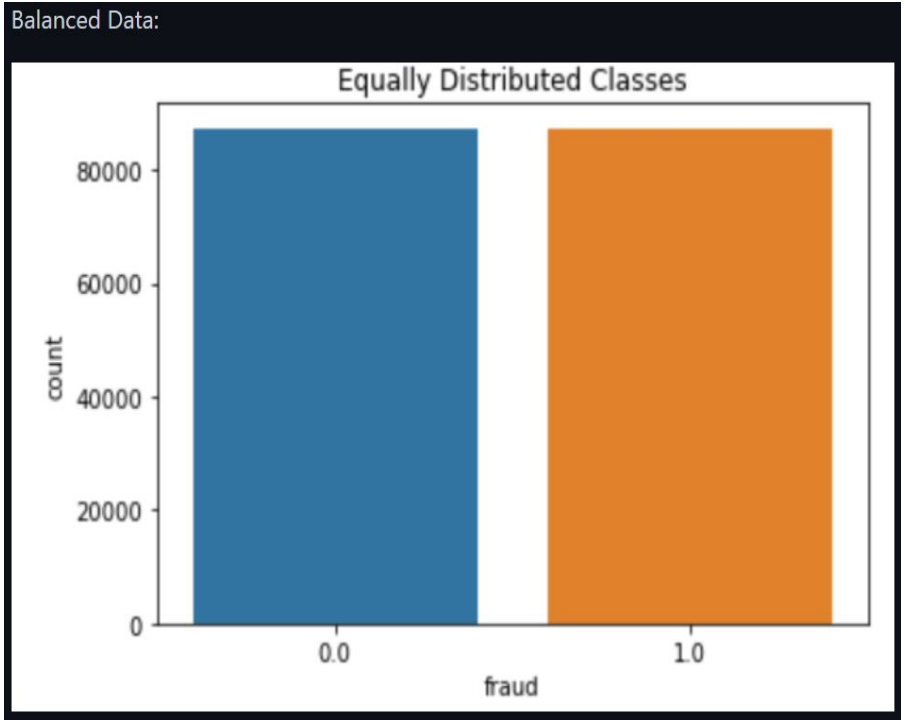
Balancing Data

SMOTE used

Unbalanced Data:



Balanced Data:



Machine Learning Model

Oversampling Accuracy & Precision

- Model Prediction
  - Logistic Regression
  - Random Forest Classifier
  - Decision Tree
- SMOTE -> oversampling
- High accuracy & precision results

Logistic Regression:

```
print("Accuracy - " + str(accuracy_score(y_test,prediction)))
print("Recall - " + str(recall_score(y_test,prediction)))
print("precision - " + str(precision_score(y_test,prediction)))
```

```
Accuracy - 0.9416435275992276
Recall - 0.9498059519742087
precision - 0.9346175065947139
```

Random Forest Classifier:

```
print("Accuracy - " + str(accuracy_score(y1_test,prediction1)))
print("Recall - " + str(recall_score(y1_test,prediction1)))
print("precision - " + str(precision_score(y1_test,prediction1)))
```

```
Accuracy - 0.9999934253636322
Recall - 0.9999868590501721
precision - 1.0
```

Decision Trees:

```
print("Accuracy - " + str(accuracy_score(y2_test,prediction2)))
print("Recall - " + str(recall_score(y2_test,prediction2)))
print("precision - " + str(precision_score(y2_test,prediction2)))
```

```
Accuracy - 0.9999726056918067
Recall - 0.9999744273554232
precision - 0.9999707742272523
```

Machine Learning Model

Undersampling Accuracy & Precision

- Model Prediction
  - Logistic Regression
  - Random Forest Classifier
  - Decision Tree

- High accuracy & precision results

Logistic Regression:

```
print("Accuracy - " + str(accuracy_score(y_test,prediction)))
print("Recall - " + str(recall_score(y_test,prediction)))
print("precision - " + str(precision_score(y_test,prediction)))
```

```
Accuracy - 0.9414214452427806
Recall - 0.9495286903999268
precision - 0.9343930115273775
```

Random Forest Classifier:

```
print("Accuracy - " + str(accuracy_score(y1_test,prediction1)))
print("Recall - " + str(recall_score(y1_test,prediction1)))
print("precision - " + str(precision_score(y1_test,prediction1)))
```

```
Accuracy - 1.0
Recall - 1.0
precision - 1.0
```

Decision Trees:

```
print("Accuracy - " + str(accuracy_score(y2_test,prediction2)))
print("Recall - " + str(recall_score(y2_test,prediction2)))
print("precision - " + str(precision_score(y2_test,prediction2)))
```

```
Accuracy - 0.9999427939437855
Recall - 1.0
precision - 0.9998856402241452
```

Webpage Development

- HTML & CSS
- Form uses ML model to reveal if purchase was fraudulent

Credit Card Fraud Detection

Credit card fraud is when someone who isn't authorized to use your card makes purchases that you didn't authorize.

86% of consumers were a victim of credit card fraud in 2020.

A model was created from a well prepared and balanced dataset that was trained optimally to indicate if a specific transaction was fraudulent.

Check Credit Card Fraud

Fill in the data below.

Distance from home (miles)

Distance from last transaction (miles)

Ratio to median purchase price

Repeat Retailer (0=no, 1=yes)

Used Chip (0=no, 1=yes)

Used Pin Number (0=no, 1=yes)

Online Order (0=no, 1=yes)

☐ I confirm that all data is correct

Register

Flask & Heroku Demo

- Undersampled Random Forest Classifier model chosen

- Flask app takes html form data and determines if the purchase was fraudulent

https://credit-card-fraud-predictor.herokuapp.com/

Questions?