Credit Card Fraud Detection

–Shivani Patil

Agenda

i. Background
ii. Objective
iii. Key Insights
iv. Cost Benefit Analysis
v. Data Attributes
vi. Data Methodology
vii. Recommendations

Background

- Finex is a leading financial service provider based out of Florida, US.
- It offers a wide range of products and business services to customers through different channels, ranging from in-person banking and ATMs to online banking.
- Over the last few years, Finex has observed that a significantly large number of unauthorized transactions are being made, due to which the bank has been facing a huge revenue and profitability crisis.
- Many customers have been complaining about unauthorized transactions being made through their credit/debit cards.
- It has been reported that fraudsters use stolen/lost cards and hack private systems to access the personal and sensitive data of many cardholders.
- They also indulge in ATM skimming at various POS terminals such as gas stations, shopping malls, and ATMs that do not send alerts or do not have OTP systems through banks.
- Such fraudulent activities have been reported to happen during non-peak and odd hours of the day leaving no room for suspicion.

Objective

- Building a fraud detection system using different machine learning techniques to identify fraudulent activities at the right time and prevent them from happening.
- To build a fraud detection model to help banks identify credit card frauds and be vigilant enough to reduce losses incurred due to such unauthorized transactions by the fraudsters.

Key Insights

- The category of grocery_pos and shopping_net showed the most number of fraud.
- Both male and female category revealed equal number of Credit Card Frauds.
- The most number of credit card frauds are reported in States OH,TX & LA.
- Most number of frauds occured in jobs of quantity surveyor then followed by naval architect and material engineer.
- The most number of frauds are reported in the cities of Dallas, Houston and Birmingham.

Key Insights

- Average number of transactions per month =77183
- Average number of fraudulent transactions per month=402
- Average amount per fraudulent transaction =530.66
- Cost incurred per month before the model was deployed= $ 213,392
- Cost incurred per month after the model is built and deployed= $ 271,9139
- Total cost of providing customer support per month for fraudulent transactions detected by the model= $ 13106.75
- Cost incurred due to these fraudulent transactions left undetected by the model =$ 14084.64
- The cost incurred per month after the model is built and deployed= $ 27191.39
- Final savings= $186200.83

Data Attributes

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1296675 entries, 0 to 1296674
Data columns (total 23 columns):
 #   Column                 Non-Null Count    Dtype
---  ------                 --------------    -----
 0   Unnamed: 0             1296675 non-null  int64
 1   trans_date_trans_time  1296675 non-null  object
 2   cc_num                 1296675 non-null  int64
 3   merchant               1296675 non-null  object
 4   category               1296675 non-null  object
 5   amt                    1296675 non-null  float64
 6   first                  1296675 non-null  object
 7   last                   1296675 non-null  object
 8   gender                 1296675 non-null  object
 9   street                 1296675 non-null  object
 10  city                   1296675 non-null  object
 11  state                  1296675 non-null  object
 12  zip                    1296675 non-null  int64
 13  lat                    1296675 non-null  float64
 14  long                   1296675 non-null  float64
 15  city_pop               1296675 non-null  int64
 16  job                    1296675 non-null  object
 17  dob                    1296675 non-null  object
 18  trans_num              1296675 non-null  object
 19  unix_time              1296675 non-null  int64
 20  merch_lat              1296675 non-null  float64
 21  merch_long             1296675 non-null  float64
 22  is_fraud               1296675 non-null  int64
dtypes: float64(5), int64(6), object(12)
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 555719 entries, 0 to 555718
Data columns (total 23 columns):
 #   Column                 Non-Null Count   Dtype
---  ------                 --------------   -----
 0   Unnamed: 0             555719 non-null  int64
 1   trans_date_trans_time  555719 non-null  object
 2   cc_num                 555719 non-null  int64
 3   merchant               555719 non-null  object
 4   category               555719 non-null  object
 5   amt                    555719 non-null  float64
 6   first                  555719 non-null  object
 7   last                   555719 non-null  object
 8   gender                 555719 non-null  object
 9   street                 555719 non-null  object
 10  city                   555719 non-null  object
 11  state                  555719 non-null  object
 12  zip                    555719 non-null  int64
 13  lat                    555719 non-null  float64
 14  long                   555719 non-null  float64
 15  city_pop               555719 non-null  int64
 16  job                    555719 non-null  object
 17  dob                    555719 non-null  object
 18  trans_num              555719 non-null  object
 19  unix_time              555719 non-null  int64
 20  merch_lat              555719 non-null  float64
 21  merch_long             555719 non-null  float64
 22  is_fraud               555719 non-null  int64
dtypes: float64(5), int64(6), object(12)
```

Data Methodology

- Dataset Loading
- EDA
- splitted the data set into training data and testing data in order to check the performance of your models with unseen data.
- Built Random Forest Model
- Adjusted Class Imbalance using ADASYN of Sampling Method
- Done with Hyperparameter Tuning because of extensive computational times.
- Model Evaluation

Recommendations

- The banking segments and local department needs to find out ways to curb these frauds in the state of OH,TX & LA as these states have shown highest number of credit card frauds.
- Based on the Analysis, it can be seen that the frauds are not biased towards any specific gender. Hence, security should not be gender specific.
- The Banking sector should moniter the online credit card transactions in the cities of Dallas, Houston & Birmingham as they show the highest number of frauds.
- Reconcilement of Baking accounts and transactions on daily basis.

Files Attached

1. Model Evaluation:
   Credit_Card_Fraud_Detection (1).ipynb

2. Cost Benefit Analysis:
   Cost Benefit Analysis (2).xlsx

THANK YOU