Credit Card Fraud Detection System

Agenda

- Objective
- Background
- Key Insights
- Cost Benefit Analysis
- Appendix:
  - Data Attributes
  - Data Methodology

Objective

- Getting in place a credit card fraud detection system to save costs incurred.
- Huge costs are being incurred due to frauds and lack of latest financial technologies to track data breaches timely.

Background

- A machine learning model has been built to detect frauds early and mitigate losses.
- A cost benefit analysis has been done for the deployment of the same.

Key Insights

- Transaction amount, category and gender are the most important variables.
- Gas and transport, grocery and shopping are the top three categories.

	Varname	Imp
0	amt	0.876627
13	category_kids_pets	0.028953
8	category_gas_transport	0.023618
12	category_home	0.013925
18	category_shopping_pos	0.011293
19	category_travel	0.010896
10	category_grocery_pos	0.010110
15	category_misc_pos	0.008930
7	category_food_dining	0.004284
17	category_shopping_net	0.003880
1	gender	0.003284
3	age_at_trans	0.001899
2	city_pop	0.001509
11	category_health_fitness	0.000416
9	category_grocery_net	0.000194
4	lat_dist	0.000141
6	trans_month	0.000043
5	long_dist	0.000000
14	category_misc_net	0.000000
16	category_personal_care	0.000000

Current Incurred Losses

- 77,183 credit card transactions per month.
- 402 fraudulent transactions per month.
- $530.66 amount per fraud transaction.
- Total costs incurred from fraud transactions is $213,392.

After New Model Deployment

- 8,607 fraudulent transactions detected by the model.
- $ 1.5 cost to provide customer support to these transactions that is $ 12,910.81 in total.
- 27 fraudulent transactions not detected by model which amounts to $ 14,394.15 loss.
- Total cost incurred after new model deployment is $ 27,304.96.
- Final savings after new model deployment is $186,086.69 that is reduction in losses by ~87%

Appendix:
Data Attributes

Snapshot of the Data

- index = Unique Identifier for each row
- transdatetrans_time = Transaction Date Time
- cc_num = Credit Card Number of Customer
- merchant = Merchant Name
- category = Category of Merchant
- amt = Amount of Transaction
- first = First Name of Credit Card Holder
- last = Last Name of Credit Card Holder

- gender = Gender of Credit Card Holder
- street = Street Address of Credit Card Holder
- city = City of Credit Card Holder
- state = State of Credit Card Holder
- zip = Zip of Credit Card Holder
- lat = Latitude Location of Credit Card Holder
- long = Longitude Location of Credit Card Holder
- city_pop = Credit Card Holder’s City Population
- job = Job of Credit Card Holder
- dob = Date of Birth of Credit Card Holder
- trans_num = Transaction Number
- unix_time = UNIX Time of Transaction
- merch_lat = Latitude Location of Merchant
- mech_long = Longitude Location of Merchant
- is_fraud = Fraud Flag ←Target Class

Appendix:
Data Methodology

- A random forest classifier built on top of a Kaggle simulated dataset.
- Class imbalance adjusted using Adaptive Synthetic (ADASYN) sampling method
- Manual hyperparameter tuning done due to extensive computational times when using Grid Search Cross Validation.


Thank you!