



Credit Card Fraud Detection Model:

by Sairaj Gupta G
VenakataSaiPrasad Manchala
Amaresh Dutta



Agenda:

- Objective
- Background
- Key Findings
- Recommendations
- Appendix:
 - Data Source
 - Data Methodology
 - Files Attached



Objective:

- To develop a machine learning model to detect fraudulent transactions based on the historical transactional data of customers



Background:

- The number of fraud transactions has increased drastically, due to which credit card companies are facing a lot of challenges
- Retaining high profitable customers is the most important business goal for banks



Key Insights:

- Transaction amount, Categories, Gender and Age group are important variables
- Gas transport, Grocery Net, Food dining, Shopping net are important categories

	Varname	Imp
7	gas_transport	0.263854
0	amt	0.142087
8	grocery_net	0.095810
6	food_dining	0.078509
11	home	0.074508
16	shopping_net	0.056139
18	travel	0.046024
17	shopping_pos	0.039712
13	misc_net	0.038209
15	personal_care	0.036302
9	grocery_pos	0.035776
12	kids_pets	0.024466
14	misc_pos	0.023575
10	health_fitness	0.022249
5	Old Age	0.008290
1	gender	0.006618
4	Middle Age	0.003321
2	diat	0.002509



Current Loss:

- Cost incurred per month before the model was deployed is \$2,13,856
- Average amount lost per fraud transaction \$530.66
- Average number of transaction per month - 77,184
- Average number of fraud transactions per month - 403



Final Savings after model deployed:

- Average number of transactions per month detected as fraudulent by the model - 3339
- cost to provide customer support to these transactions - $1.5 * 3339 = \$5009$
- Average number of transactions per month that are fraudulent but not detected by the model - 15 which results in loss of \$7960
- Cost incurred per month after the model is built and deployed - \$ 12969
- Final savings after new model deployment is \$ 2,00,887.383 that means loss is reduced by 94%



Appendix - Data Source

- We have used Credit Card Transactions Fraud Detection Dataset from Kaggle
- Data set contains:
 - Transaction details, card details
 - Transaction Amount, Card holder details
 - Transaction location, Category of purchase



Appendix - Methodology:

- XGBoost Classifier is built on top of Kaggle dataset
- By using ADASYN we have removed class imbalance in dataset
- By using Logistic Regression, Decision Tree, Random Forest Classifier we have built models on dataset but XGBoost model have high recall. So we consider XGBoost Classifier.



Attached Files:

- XGBoost Classifier Model:

https://drive.google.com/file/d/1U9o-6yizEF0BJAJWzVd1-Lk1j68xD0Np/view?usp=share_link

- Cost Benefit Analysis:

https://docs.google.com/spreadsheets/d/12nBUERFUVWCW2vjjWm03mkGvx7jrJovk1/edit?usp=share_link&ouid=100933778655232304793&rtpof=true&sd=true