CREDIT CARD FRAUD DETECTION

SUBMITTED BY

ADITYA VIKRAM

AGENDA

- Background
- Objective
- Solution Approach
- Key Insights
  - Important Variables
  - Financial Implications Before Model
  - Financial Implications After Model
- Appendix
  - Data Attributes
  - Attached Files

BACKGROUND

- Fraud transactions has increased drastically around the globe
- Retaining high profitable customers is the most important business goal
- Rise in digital payment channels is directly proportional to the number of fraudulent transactions
- The Federal Trade Commission (US) has estimated that around 10 million people become victims of credit card theft each year
- Credit card companies lose close to $50 billion per year to fraud
- Fraud detection using Machine Learning algorithm is a “MUST”

OBJECTIVE

- To develop a machine learning model to detect fraudulent transactions based on the historical transactional data of customers with a pool of merchants
- Data set captures the data from Jan 1st 2019 – Dec 31st 2020
- ~1000 customers, ~800 merchants
- ~9651 out of 18,52,394 (0.52%) are fraud transactions
- Perform cost-benefit analysis

SOLUTION APPROACH

- Data Understanding, Data Preparation and EDA :
  - Firstly , we checked the shape and datatype of the dataset and then assigned correct datatype and deleted unwanted columns. Thereafter, we checked for null values and outliers and no outlier or null values were found.
  - Further, we checked for data Imbalance and skewness and binary mapped categorical columns with two values and created dummy variables for rest of categorical columns
- Train/Test Data Splitting :
  - We split train and test dataset into X and y

SOLUTION APPROACH

- Model Building or Hyperparameter Tuning:
  - We used ADASYN to oversample the dataset
  - We used Random Forest model for model building
- Model Evaluation:
  - We evaluate the model performance using appropriate evaluation metric such as Precision, Recall and F-1 Score
- Cost-Benefit Analysis :
  - We checked for cost implications before the model was built and after the model was built

KEY INSIGHTS

KEY VARIABLES

- Transaction amount, category and gender are the most important variables
- Gas and transport, kids_pet and home are the top three categories
- Transaction month, longitude and miscellaneous category are the least important variables

	Varname	Imp
0	amt	0.877397
13	category_kids_pets	0.028940
8	category_gas_transport	0.023524
12	category_home	0.013879
18	category_shopping_pos	0.010981
19	category_travel	0.010898
10	category_grocery_pos	0.010155
15	category_misc_pos	0.008859
7	category_food_dining	0.004213
17	category_shopping_net	0.003882
1	gender	0.003158
3	age_at_trans	0.001953
2	city_pop	0.001433
11	category_health_fitness	0.000412
9	category_grocery_net	0.000195
4	lat_dist	0.000097
6	trans_month	0.000015
5	long_dist	0.000008
14	category_misc_net	0.000000

FINANCIAL IMPLICATIONS BEFORE MODEL

- Average 77,183 credit card transactions per month
- Average 402 fraudulent transactions per month
- Average $ 531 amount per fraud transaction
- Total costs incurred per month from fraud transactions before the model was deployed is $ 213,392

FINANCIAL IMPLICATIONS AFTER MODEL

- 8,745 fraudulent transactions detected by the model
- $ 1.5 cost to provide customer support to these transactions that is $ 13,117 in total
- 27 fraudulent transactions not detected by model which amounts to $ 14,284 loss
- Total cost incurred after new model deployment is $ 27,401
- Final savings after new model deployment is $185,992 that is reduction in losses by ~87%

APPENDIX: DATA ATTRIBUTES

index - Unique Identifier for each row
trans_date_trans_time - Transaction DateTime
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
merch_long - Longitude Location of Merchant
is_fraud - Fraud Flag <--- Target Class

APPENDIX: ATTACHED FILE

- Cost Benefit Analysis
  - Cost+Benefit+Analysis.xlsx
- Random Forest Machine Learning Model
  - Capstone Project I.ipynb
- Presentation File
  - Capstone_Project_ CC_Fraud_Detection.pptx
- Video explanation

THANK YOU !!!