CREDIT CARD FRAUD DETECTION MODEL

by Vladimir Nikonov

video link:
https://drive.google.com/file/d/1zu85JlEmFWc2HMAxdlYoxMnI_kJ5FYH4/view?usp=sharing

AGENDA

- Problem Statement
- Objective
- Key Findings
- Model Selection
- Cost-Benefit Analysis

PROBLEM STATEMENT

- Credit card fraud is any dishonest act or behavior to obtain information without the proper authorization of the account holder for financial gain.

- In the banking industry, detecting credit card fraud using machine learning is a necessity. Machine learning helps these institutions reduce time-consuming manual reviews, costly chargebacks and fees, and denial of legitimate transactions.

OBJECTIVE

- Build a Machine Learning Model which will successfully detect fraudulent transactions

- Cut the costs of the client

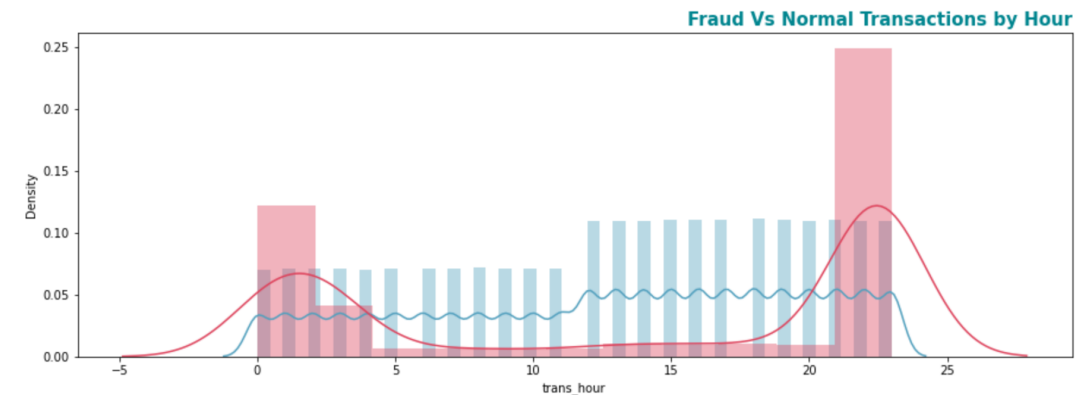- Improve customer experience and raise satisfaction with the business

KEY FINDINGS

- Categories with the most fraudulent transactions are **grocery** and **online shopping** (>40%). [Pic 1]

- Online shopping is also one of the costliest categories among fraudulent transactions **($1.7 million in total).** [Pic 2]

- The peak time for fraudulent transactions is **9 PM to 2 AM.** [Pic 3]
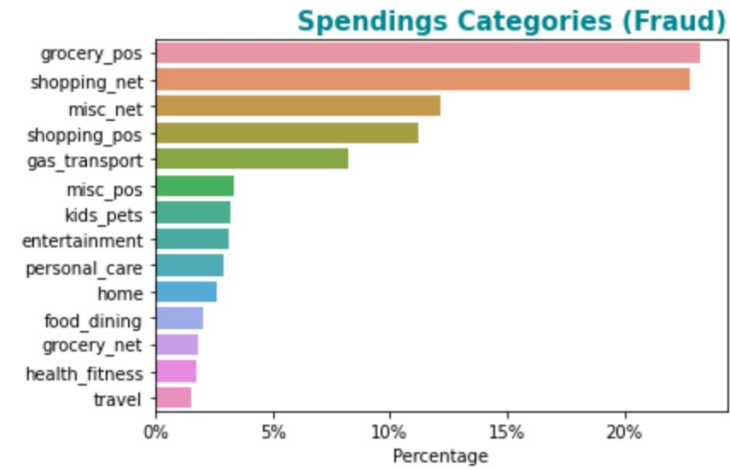
Pic 1
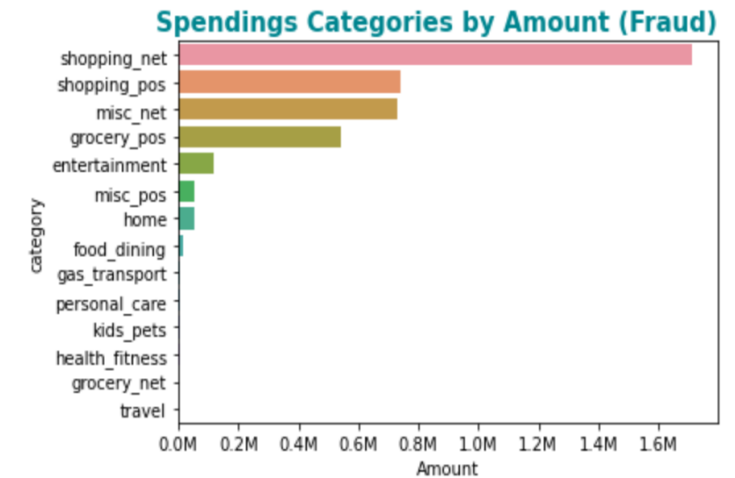
Pic 2

Pic 3

MODEL SELECTION

- Types of ML models attempted:

  - Linear Regression

  - Decision Tree

  - Random Forest

- As the dataset is highly imbalanced (only ~0.5% of all the transactions are labeled as fraudulent), all the models were trained on the data oversampled with SMOTE for better detection of fraudulent transactions

- Random Forest was chosen as the final model

- The features and their importance identified by the model can be seen in Table 1

Variable	Imp
amt	0.794439
trans_hour	0.068494
category_home	0.017801
category_shopping_pos	0.013082
category_gas_transport	0.012122
city_pop	0.010447
category_misc_pos	0.009886
gender	0.009780
category_shopping_net	0.009494
category_travel	0.009201
category_kids_pets	0.007808
category_grocery_pos	0.007806
category_misc_net	0.007549
category_food_dining	0.007513
category_health_fitness	0.004501
customer_age	0.004133
category_personal_care	0.004097
category_grocery_net	0.001076
trans_month	0.000433
distance_lat	0.000219
distance_long	0.000118

Table 1

COST-BENEFIT ANALYSIS

- With the average of *402* fraudulent transactions per month with the amount of *$530.66*, the losses of the client were around **$213,392.22**

- After the model has been deployed, it identified *2,810* transactions as fraudulent with *32* transactions misclassified as non-fraudulent

- Considering that providing customer support for transactions identified as fraudulent is *$1.5*, the average total cost of this service is **$4,215.69** per month

- The average monthly losses inflicted by the misclassified cases are **$17,025.39**

- Thus, the average monthly losses of the client after the model deployment are **$21,241.07** (*10% of the losses before the model deployment*)

- By deploying the model, the client can potentially save **$192,151.15** per month on average