# Credit Card Fraud Detection System

**Developed By**
***Richa Jhunjhunwala***
***Rakhi Chandrasekhar***
***Ashish Kumar Jha***

# Agenda

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

# Objective

- Establishing an efficient credit card fraud detection system to mitigate financial losses.
- Substantial expenditures arise from fraudulent activities and the absence of up-to-date financial technologies for timely data breach detection.

# Background

- Developed a machine learning model for early fraud detection, aimed at minimizing losses.
- Conducted a comprehensive cost-benefit analysis to evaluate the deployment implications of the aforementioned model.

# Key Insights

- The key variables of significance include the transaction amount, category, and gender.
- The three primary categories are gas and transport, grocery, and shopping.

| | 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

- There are 77,183 credit card transactions on average every month.
- Among these, 402 transactions are identified as fraudulent.
- The average monetary loss per fraudulent transaction is $530.66.
- The cumulative expenses attributed to fraudulent transactions amount to $213,392.

# After New Model Deployment

- The model identified 8,607 fraudulent transactions, resulting in a total customer support cost of $12,910.81.
- Additionally, 27 fraudulent transactions went undetected by the model, resulting in a loss of $14,394.15.
- The cumulative cost post-model deployment is $27,304.96.
- As a result of implementing the new model, the final savings amount to $186,086.69, reflecting an impressive reduction of approximately 87% in losses

# 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

- Utilizing a Kaggle-simulated dataset, a random forest classifier was developed.
- To address class imbalance, the Adaptive Synthetic (ADASYN) sampling method was employed. Due to the resource-intensive nature of Grid Search Cross Validation, manual hyperparameter tuning was conducted.


Thank you!