Enhancing Fraud Detection for NullFraud Bank

Presented By: Bolt UBC Team 35 (Helen Meng, Guojun Ma, Kylie Seto)

Presented To: NullFraud Bank

Executive Summary

ISSUES	Increased Online Fraudulent Transactions	High False Positive Rate	Low operational efficiency
OBJECTIVE	Reduce fraud and false positives to boost operational efficiency and increase customer satisfaction		
RECOMMENDATION	Push Usage of Physical Cards + Chips	Logistic Regression Fraud Prediction Model	Enhanced Transaction Verification Process
IMPACT	Increase Business Revenues and Enhance Customer Satisfaction		

Problems to solve

1. Increased digital transactions
2. Sophisticated cyber threats
3. Changing customer habits

→

- High false positive rate
- Rising operational costs
- Risk to customer trust

Company Overview

NullFraud Bank is leading the charge in combating fraud within the digital finance landscape, utilizing cutting-edge technologies to enhance security in its payment system. As the shift towards cashless transactions accelerates, there's a growing demand for **secure**, **efficient**, and **sustainable** payment solutions. With its extensive network of cardholders and top-notch customer service, NullFraud Bank is poised to revolutionize fraud management, setting a new standard in secure digital payments.

Primary Project Objective

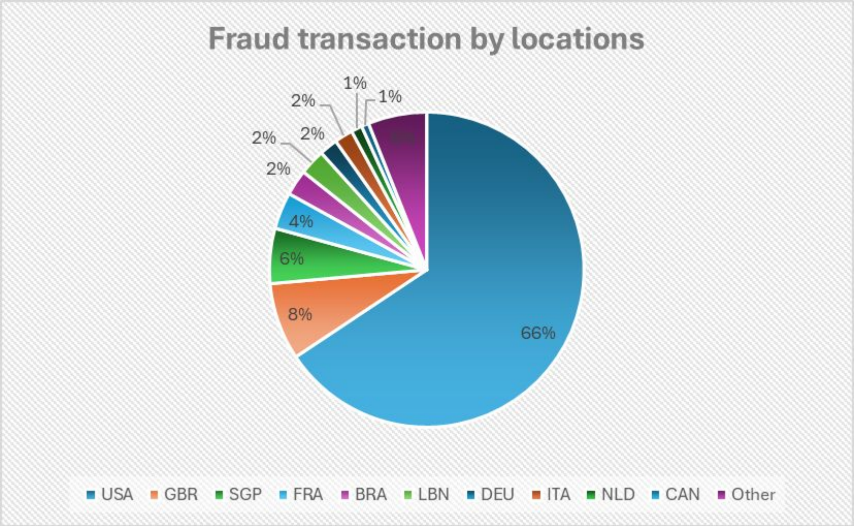
Design a solution that **reduces fraud**, **decreases false positives**, and cements NullFraud Bank's reputation as a pioneer in **secure transactions**.

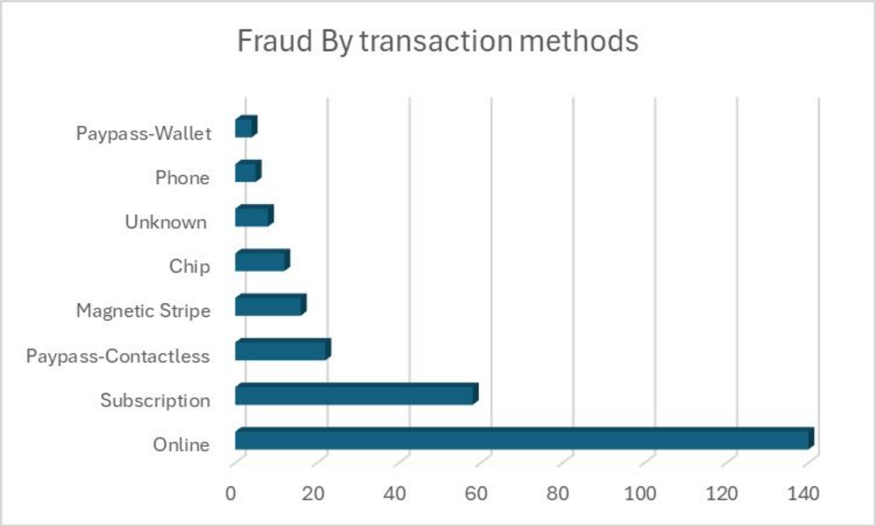
- Enhance customer loyalty
- Reduce fraud-related costs
- Boost operational efficiency

Factors in Fraudulent Transactions

Most fraudulent charges come from online purchases

Total fraud transaction: 265/100,000 =0.265%
Total transaction value: $25,439

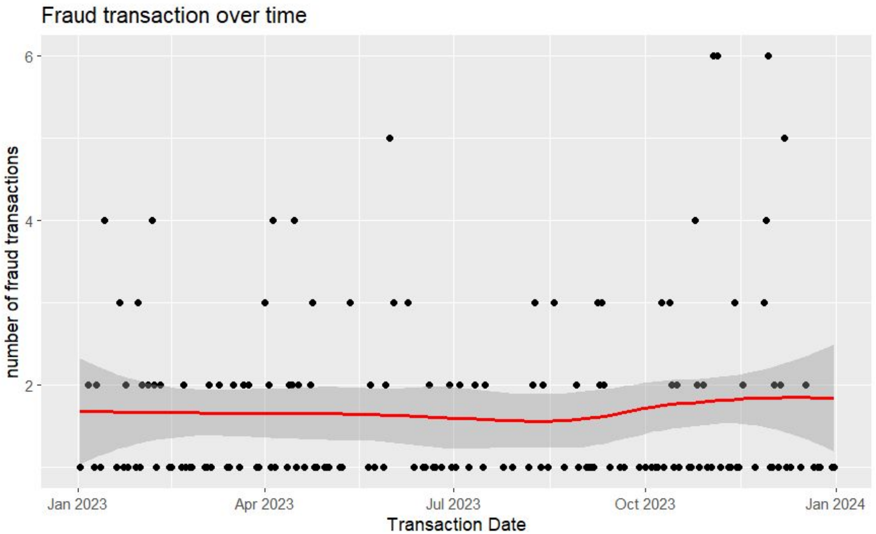


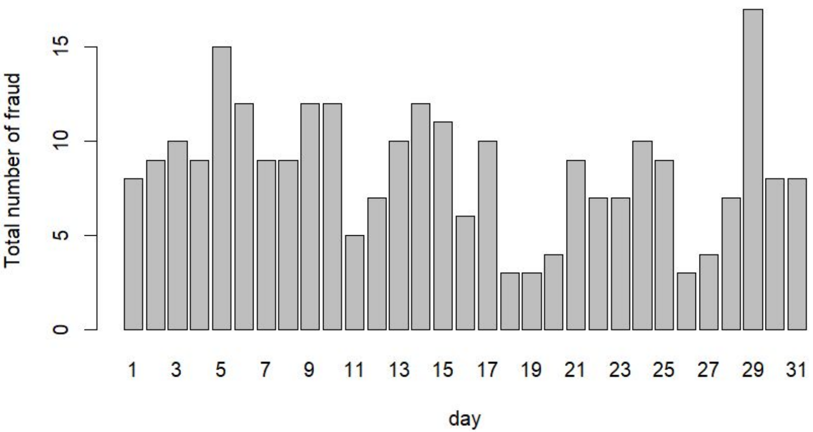


USA is where the most fraudulent transactions take place.

Most fraud happen through **online** and **subscription** payments

Fraud Rate by time





Fraud rate **increases** in the **fourth quarter** due to **increased transactions** because of **holidays** such as Christmas.

Within the month, **fraud spikes** on the **5th** and **29th** day because of **payday** & **bills + rent due**

Cross-border transactions increases risk of fraud

Cross-border Transaction	Total number of transaction	Fraud	Ratio of fraud
Yes	14845	91	**0.00613**
No	85037	174	0.00205

A Higher Ratio of Fraud

Chi-square test confirms this difference in ratio is significant:

X-squared = 78.123, df = 1, p-value < 2.2e-16

The p-value rejects the null hypothesis that these two variables are independent.

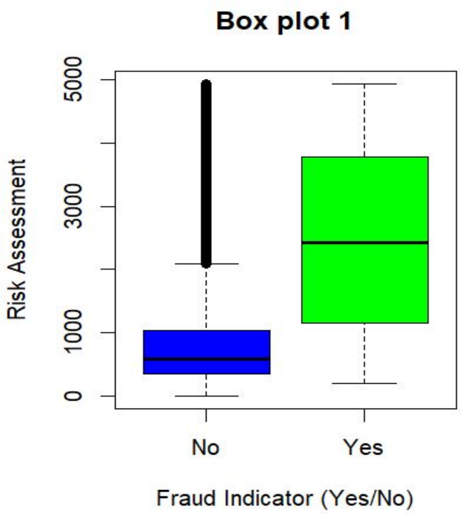
Physical card + chip payment method decreases fraud risk

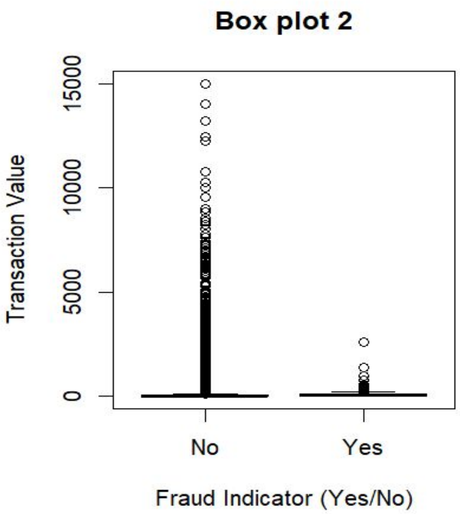
Card Present Status	Chip Usage	Fraud	Non-Fraud	Ratio of Fraud
No	No	211	49882	0.00421
No	Yes	0	9	0
Yes	No	17	4082	0.00415
Yes	Yes	37	45644	**0.000810**

A group with **both physical card and chip usage** significantly decrease the risk of fraud.

Using the physical card but not using the chip is not sufficient to decrease the risk of fraud.

Risk assessment doesn't accurately predict risk of fraud





- **Box plot 1:** Shows that the average assessment are different for two groups.
- However, there are many transactions in the non-fraud group were assigned high risk assessment
- **Box plot 2:** Shows that while the average transaction value is close for the two groups, the typical value is less than 5000 for the fraud group.

Logistic regression model for fraud detection

Logistic regression: the standard way of classifying binary variables.

- Similar to linear regression, logistic regression aism to find the linear relationship between the predictors and the log-odds of the response variable
    - Let Y_i denotes whether the i-th transaction is fraud or not. (Y_i = 1 if fraud and Y_i = 0 if not)
    - p(Y_i = 1) / p(Y_i = 0) is the odds of whether Y_i is fraud(between 0 and ∞)
    - log(p(Y_i = 1) / p(Y_i = 0)) ~ intercept + β * predictors
    - Estimate intercept and β using the training data

- Plug in the estimates to the new samples to predict the probability p(Y_(new) = 1)

- Decide whether Y_(new) is fraud or not based on certain decision threshold (Assign Y_(new) to fraud if p(Y_(new) = 1) > c) for c between 0 and 1. The standard is to set c= 0.5.

Logistic regression model results

Partitioned the data set into training and testing set (80%, 20%). Trained the logistic regression model on the training set.

```
Coefficients:
                                       Estimate Std. Error z value Pr(>|z|)    
(Intercept)                          -7.505e+00  1.480e-01 -50.699  < 2e-16 ***
`Risk Assessment`                     1.138e-03  5.203e-05  21.870  < 2e-16 ***
`Transaction Value`                  -3.325e-04  2.575e-04  -1.291   0.1966    
`Card Present Status`CP               6.506e-01  2.600e-01   2.502   0.0123 *  
`Chip Usage`Yes                      -1.320e+00  2.987e-01  -4.420 9.87e-06 ***
`Cross-border Transaction (Yes/No)`Yes  3.424e-01  1.418e-01   2.414   0.0158 *  
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- All predictors except the "transaction value" significantly associated with the risk of fraud.
- For example, the cross-border transaction increases the log odds of being fraud by 0.342 on average.

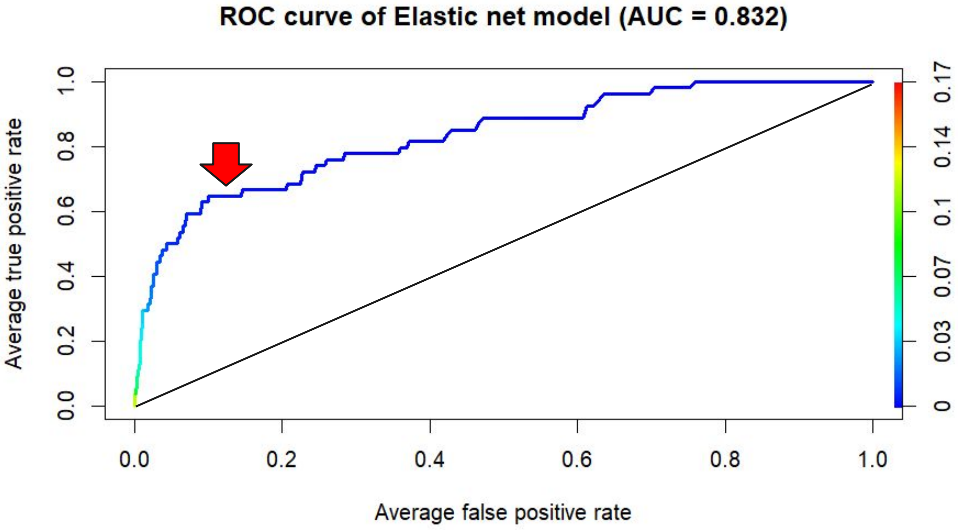
Logistic regression performance evaluation criteria

Evaluated the model's performance by the following metrics:

1. **Sensitivity:** the proportion of actual fraud which are correctly identified (True positive).
2. **Specificity:** the proportion of actual negatives which are correctly identified (True negative).
   (note: 1 - specificity is the false positive rate)
3. **AUC:** the area under the ROC curve, which plots the trade off between sensitivity and specificity(between 0.5 and 1). A higher AUC value indicates a better model performance.

- Depending on the context, one can increase the sensitivity by decreasing the threshold c. However, false positive rate can increase as well. Consider the extreme case where we classify every new samples to fraud. Then the sensitivity is 1 but also the false positive rate is 1.
- The goal is to find a good balance between sensitivity and false positive rate -> select the model with high AUC.

Logistic regression performance evaluation

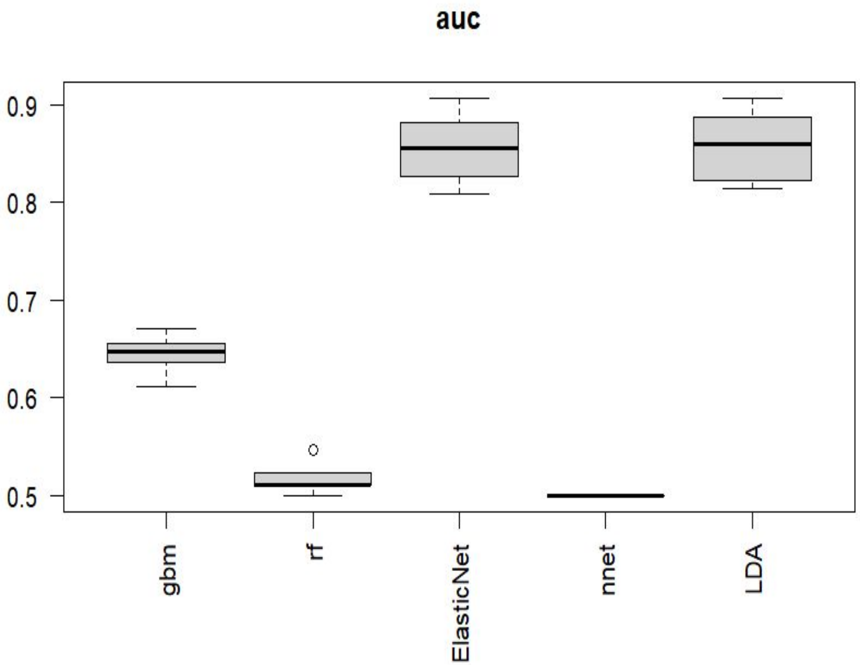


The true and false positive rate are evaluated on the test set.

The color gradient indicates the decision threshold c used to classify the samples.

For example, if we want to obtain a true positive rate of 0.6 and the false positive rate of 0.1, then we need to set the decision threshold c ~= 0.03.

Evaluation of other classification model alternatives



Using **5-fold stratified cross-validation** on the training set, we also evaluate the performance of following model:

1. Linear discriminant analysis (LDA)
2. Elastic Net (EDA)
3. Random Forest (RF)
4. Tree-based gradient boosting (GBA)
5. Feed-forward neural network (FNN)

We found **elastic net** and **LDA** have a **similar performance**. The **other models** have poor performance.

Increase the use of physical card and chips

Recommendation #1

According to logistic regression analysis, the highest reduction in fraudulent transactions occurs with an increase in the use of physical cards and chip technology.

```
Coefficients:
                                          Estimate Std. Error
(Intercept)                              -7.505e+00  1.480e-01
`Risk Assessment`                         1.138e-03  5.203e-05
`Transaction Value`                      -3.325e-04  2.575e-04
`Card Present Status`CP                   6.506e-01  2.600e-01
`Chip Usage`Yes                          -1.320e+00  2.987e-01
`Cross-border Transaction (Yes/No)`Yes    3.424e-01  1.418e-01
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' '
```

Chip usage decreases the log odds of being fraud by 1.32 on average.

Demonstrates NullFraud's close attention to customers' needs and financial security, enhancement of customer loyalty and trust

NullFraud Bank can Increase the Use of Physical Cards + Chips By:

- Increasing customer education efforts regarding security-related benefits of using physical cards with chips
- Launching programs that incentivize customers (e.g. pretty card designs, cashback, or loyalty points)

Utilize logistic model to accurately classify future transactions and predict fraudulent activity

Recommendation #2

Adjust the decision threshold based on transaction values:

- Prioritize high sensitivity for high-value transactions.
- Minimizing false positives for low-value transactions.

Minimizing the operational cost of addressing further negative impacts of fraudulent transactions.

NullFraud Bank can Accurately Classify Future Transactions & Predict Fraudulent Activity By:

- Applying the logistic model to a larger data set and evaluate its performance to a greater extent. Adjust the model accordingly.
- Creating action plans that tailor the classification result of transactions and solutions to find fraudulent transactions.

Enhance customer verification process over fraud-susceptible transactions

Recommendation #3

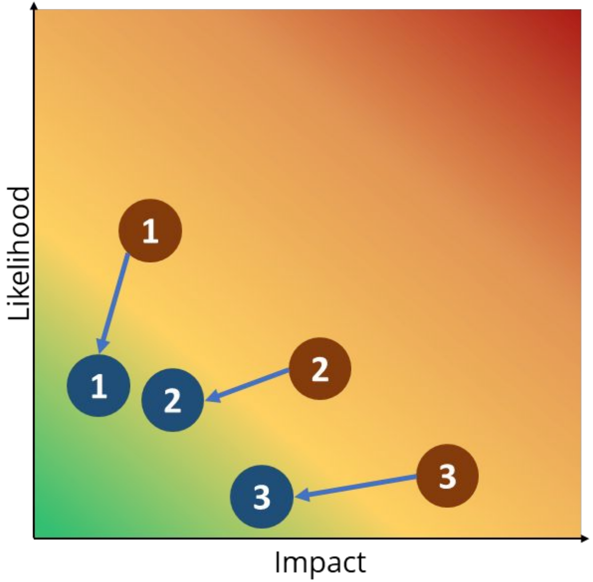
Strengthen the customer identification verification process of high-risk transactions which are emphasized based on factors including countries (e.g. USA), cross-border transactions, transaction methods (online and subscription), transaction value (e.g. less than 5000), and more.

Increases operational efficiency by the enhanced surveillance over potentially fraudulent transactions and directly approaches the issue of increased digital transactions.

NullFraud Bank can Enhance Customer Verification Process over Fraud-susceptible Transactions By:

- For the transactions that are more likely to be fraud based on multiple factors, customers need to verify their identities through official documents or technical services. Examples of additional verification are biometric authentication, one-time passcodes sent via SMS or email, and preset questions.
- Implement Know Your Customer (KYC) procedures, including the customer identification program (CIP), customer due diligence (CDD), and enhanced due diligence (EDD).

Risk and Mitigation



Risks

1

The use of physical cards and chips may increase counterfeit cards

2

Uncertainty of the logistic regression model

3

Lowered customer satisfaction due to secure transaction process

Mitigation

1

> Implement EMV (chip) technology that provides dynamic data for each transaction
> Seek assistance from legal legislation and agencies to prevent counterfeit cards

2

> Constantly revise & update model based on new data and other fraud indicators
> Regularly conduct tests and evaluations of the model's performance

3

> Raise awareness regarding consequences of fraudulent transactions
> Develop user friendly and effective verification process.

Implementation Timeline

Recommendation	Task	Q1 Y1	Q2 Y1	Q3 Y1	Q4 Y1	Q1 Y2	Q2 Y2	Q3 Y2	Q4 Y2
1. Increase Physical Card + Chip Usage	Develop Educational Campaign	■							
	Launch Educational Campaign		■						
	Outreach for Card Design Collabs (Sanrio, Disney, etc.)		■	■					
	Themed card production				■				
	Themed card launch					■	■	■	■
	Cashback + loyalty points roll out						■	■	■
2. Logistic Regression Model for Fraud Prediction	Develop Logistic Regression Model	■	■						
	Test Logistic Regression Model on Test Dataset			■					
	Integrate Model into Fraud Detection Software				■				
	Continuously check data accuracy and update data					■	■	■	■
3. Enhance Transaction Verification Process	Research make/buy options for two-factor authorization softwares	■							
	Integrate software into customer ID program		■						
	Implement KYC, CIP, CDD, and EDD procedures into program			■					
	Beta launch program				■				
	Launch program to public					■	■	■	■

KPI #1: Increase Physical Card + Chip Usage by 10% by Q2 Y2

KPI#2: Ensure 95% Prediction Accuracy by Q4Y1

KPI#3: Ensure high transaction verification

Conclusion

ISSUES	**Increased Online Fraudulent Transactions**	**High False Positive Rate**	**Low operational efficiency**
OBJECTIVE	**Reduce fraud and false positives to boost operational efficiency and increase customer satisfaction**		
RECOMMENDATION	**Push Usage of Physical Cards + Chips**	**Logistic Regression Fraud Prediction Model**	**Enhanced Transaction Verification Process**
IMPACT	**Increase Business Revenues and Enhance Customer Satisfaction**		