EFFECTIVE FRAUD DETECTION

Leveraging Logistic Regression

BUSINFO 704 Quarter Three 2024

Sharan Srinivasan (ssri440)
Eric Jang (sjan666)
Yunxu Shen (yshe274)
Xioyue Tan (xtan923)

DISCLAIMER: The data was synthetically generated to make it clear it is not based on real ASB customer data.

According to the Ministry of Business, Innovation and Employment, Kiwis lost $198 million to scam in 2023.[1]

BUSINESS PROBLEM

- The increase in online and mobile banking has led to a rise of sophisticated fraud schemes.
- Bank A's current fraud detection system struggles to keep up with evolving fraud patterns. The consequences are:

Financial Loss | **Customer Mistrust**

THE DATASET

- 500,000 bank customers
- Transaction history (2/5/24 - 31/7/24)
- 13 variables involved

Eight variables were selected for the final analysis.

- Fraud Label (Dependent)
- Transaction Time (Recoded)
- Balance
- Transaction Amount
- Joint_Flag
- Agent (Recoded)
- Age
- Transaction Type

DATA PIPELINE

1.
   - Cleaned the raw data on RStudio
   - Recoded variables Transaction Time and Agent
2.
   - Analysed data associated with fraud
   - Plotted relevant graphs to find the best independent variables
3.
   - Data split of 75 (training) /25 (testing)
4.
   - Balanced the minority class of Fraud Label with upsampling
5.
   - Tested different classification algorithms based highest scores on performance metrics
6.
   - Chose the logistic regression model
   - Predicted fraud with unseen test data

WHY ADOPT LOGISTIC REGRESSION?

- Logistic regression is a classification algorithm technique used to model and predict binary outcomes (such as yes/no or fraud/no fraud) based on one or more predictor variables.

+2000
fraud transactions correctly identified, minimising financial loss

+80000
non-fraud transactions correctly identified, helping with customer retention

71.3%
Sensitivity
The model correctly identifies 71.3% of actual fraudulent transactions.

66.8%
Accuracy
The model correctly classifies 66.8% of all transactions, including both fraudulent and non-fraudulent cases.

Overall, a highly interpretable model can identify significant predictors, making it easier to design appropriate strategies for customers.

RESULTS

The following variables are significant predictors of fraud.

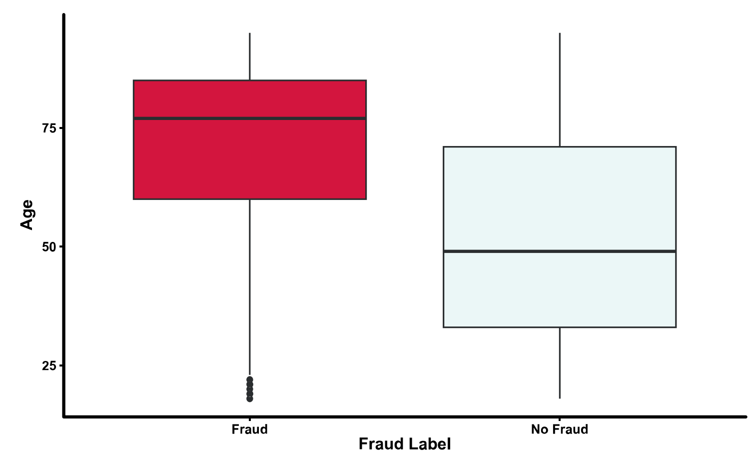
Age	Transaction Amount	Transaction Time	Balance
78.2%	10.3%	6.9%	1.7%

CONCLUSION

- The model with higher sensitivity successfully identifies a significant proportion of fraudulent transactions.
- The model supports the organisation's strategic priority of risk mitigation, ensuring that fraud detection efforts are robust and aligned with long-term financial protection goals.

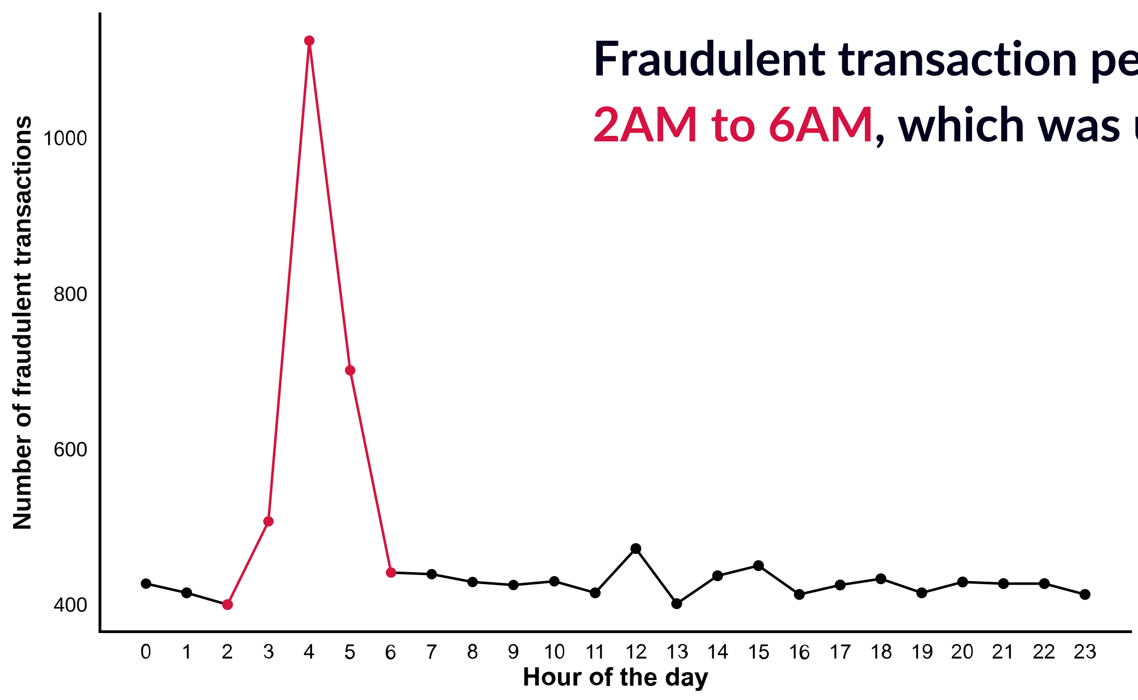
KEY TAKEAWAYS

As balance decreases ... the likelihood of **fraud** increases



As age increases ... the likelihood of **fraud** increases

Fraudulent transaction peaked from 2AM to 6AM, which was unusual.



FOUR STRATEGIC RECOMMENDATIONS

- Senior citizens over 70 are far more vulnerable to fraud. One possible implementation is a targeted awareness program for older customers and two-factor authentication.
- Send alerts via email and text message to customers when their balance goes negative.
- Implement stricter transaction limits and two-factor authentication between the hours of 2 and 6 AM to discourage fraudsters targeting innocent customers.
- Develop an alert system that classifies transactions into high, medium, and low categories according to a risk score for long-term fraud prevention.

1 Ministry of Business, Innovation & Employment. (2024, August 15). *$198 million dollars lost to scams in the last year*. MBIE. https://www.mbie.govt.nz/about/news/198-million-dollars-lost-to-scams-in-the-last-year
2 OpenAI. (2024). ChatGPT (GPT-4) [Large language model]. OpenAI. https://www.openai.com/