**(Vietnamese)**

**Báo cáo phân tích chi tiết về Phát hiện Gian lận Thẻ Tín dụng**

**1. Tỷ lệ gian lận (Fraud Rate)**

Tỷ lệ gian lận trong dữ liệu thể hiện sự chênh lệch giữa số lượng tài khoản gian lận và không gian lận:

* **Số lượng tài khoản không gian lận (TARGET = 0):** 282,686 tài khoản chiếm **91.93%**.
* **Số lượng tài khoản gian lận (TARGET = 1):** 24,825 tài khoản chiếm **8.07%**.

Mặc dù số lượng tài khoản không gian lận vượt trội, nhưng tỷ lệ gian lận 8.07% là đáng kể trong bối cảnh phát hiện gian lận, vì con số này cao hơn mức trung bình trong ngành tài chính. Điều này cho thấy cần có các biện pháp kiểm soát rủi ro mạnh mẽ và cải thiện mô hình phát hiện gian lận để giảm thiểu tổn thất.

**2. Phân tích thu nhập khách hàng và tín dụng (AMT\_INCOME\_TOTAL vs. AMT\_CREDIT)**

Biểu đồ phân tán giữa tổng thu nhập khách hàng (AMT\_INCOME\_TOTAL) và tổng số tiền tín dụng (AMT\_CREDIT) cho thấy các khoản vay chạy trong các khoảng sau:

* **Tổng thu nhập của khách hàng (AMT\_INCOME\_TOTAL):** Trong khoảng từ **0 đến 5 triệu**.
* **Tổng số tiền tín dụng khách hàng đã vay (AMT\_CREDIT):** Trong khoảng từ **0 đến 3 triệu**.

Sự phân bổ này cho thấy các khoản vay lớn thường tập trung vào những người có mức thu nhập cao hơn. Điều này có thể gợi ý rằng rủi ro tín dụng và khả năng gian lận có mối liên hệ với mức tín dụng được phê duyệt. Tuy nhiên, các phân tích sâu hơn về xu hướng gian lận theo mức thu nhập và tín dụng sẽ giúp hiểu rõ hơn mối quan hệ này.

**3. Giờ xử lý đơn xin vay cao điểm (HOUR\_APPR\_PROCESS\_START)**

Trong một ngày, các giờ cao điểm khi xử lý đơn vay là từ 9 giờ đến 14 giờ, với lượng tài khoản vay nhiều nhất vào:

* **10 giờ**: 37,722 tài khoản
* **11 giờ**: 37,229 tài khoản
* **12 giờ**: 34,233 tài khoản
* **13 giờ**: 30,959 tài khoản
* **14 giờ**: 27,682 tài khoản

Các giờ cao điểm này cho thấy áp lực đến hệ thống xử lý đơn xin vay, đặc biệt là vào khoảng giữa ngày (10 giờ đến 14 giờ). Điều này có thể ảnh hưởng đến hiệu suất xử lý đơn vay và khả năng phát hiện các dấu hiệu gian lận.

**4. Phân tích loại thu nhập của khách hàng (NAME\_INCOME\_TYPE)**

Top 5 loại hình thu nhập có số lượng tài khoản nhiều nhất trong dataset bao gồm:

* **Working**: 143,550 tài khoản không gian lận và 15,224 tài khoản có gian lận.
* **Commercial Associate**: 66,257 tài khoản không gian lận và 5,360 tài khoản có gian lận.
* **Pensioner**: 52,380 tài khoản không gian lận và 2,982 tài khoản có gian lận.
* **State Servant**: 20,454 tài khoản không gian lận và 1,249 tài khoản có gian lận.
* **Unemployed**: 14 tài khoản không gian lận và 8 tài khoản có gian lận.

Loại thu nhập "Working" và "Commercial Associate" có số lượng tài khoản gian lận cao nhất, với lần lượt là 15,224 và 5,360 tài khoản gian lận. Điều này cho thấy cần phải tập trung nhiều hơn vào các nhóm này để nâng cao khả năng phát hiện gian lận.

**5. Phân tích loại hình tổ chức (ORGANIZATION\_TYPE)**

Top 5 loại hình tổ chức có số lượng tài khoản nhiều nhất:

* **Business Entity Type 3**: 61,669 tài khoản không gian lận và 6,323 tài khoản gian lận.
* **XNA**: 52,384 tài khoản không gian lận và 2,990 tài khoản gian lận.
* **Self-employed**: 34,504 tài khoản không gian lận và 3,908 tài khoản gian lận.
* **Other**: 15,408 tài khoản không gian lận và 1,275 tài khoản gian lận.
* **Medicine**: 10,456 tài khoản không gian lận và 737 tài khoản gian lận.

Loại hình tổ chức **Business Entity Type 3** có tỷ lệ tài khoản gian lận cao nhất. Việc có tới 6,323 tài khoản gian lận trong loại tổ chức này cho thấy các doanh nghiệp loại này có nguy cơ cao hơn về gian lận so với các tổ chức khác.

**6. Tỷ lệ phê duyệt và khách hàng tốt**

Mặc dù không có thông tin trực tiếp về tỷ lệ phê duyệt cuối cùng và tỷ lệ phê duyệt khách hàng tốt trong dữ liệu này, chúng có thể được suy ra từ các yếu tố khác:

* **Tỷ lệ phê duyệt cuối cùng** có thể liên quan đến tỷ lệ giữa các tài khoản tín dụng đã được chấp thuận và tổng số tài khoản xin vay. Với tỷ lệ lớn các tài khoản không gian lận, tỷ lệ phê duyệt cuối cùng có thể tương đối cao.
* **Tỷ lệ phê duyệt khách hàng tốt** (tức là khách hàng không có gian lận) có thể được tính bằng tỷ lệ giữa số lượng tài khoản không gian lận (91.93%) so với tổng số tài khoản.

**7. Kết luận**

Dữ liệu phát hiện gian lận thẻ tín dụng này cho thấy tỷ lệ gian lận chiếm một phần nhỏ nhưng vẫn đáng kể (8.07%). Các yếu tố như loại thu nhập, loại hình tổ chức và giờ xử lý đơn xin vay đều có ảnh hưởng đến rủi ro gian lận. Việc phân tích sâu hơn các mối quan hệ này có thể giúp cải thiện mô hình phát hiện gian lận và tối ưu hóa quá trình xử lý tín dụng.

**(English)**

**Detailed Analysis Report on Credit Card Fraud Detection**

**1. Fraud Rate**

The fraud rate in the dataset highlights the difference between the number of fraudulent and non-fraudulent accounts:

* **Number of non-fraudulent accounts (TARGET = 0):** 282,686 accounts, accounting for **91.93%**.
* **Number of fraudulent accounts (TARGET = 1):** 24,825 accounts, accounting for **8.07%**.

While the majority of accounts are non-fraudulent, the fraud rate of 8.07% is significant in the context of fraud detection, as this figure is higher than the industry average. This indicates the need for strong risk controls and improved fraud detection models to minimize financial losses.

**2. Analysis of Customer Income and Credit (AMT\_INCOME\_TOTAL vs. AMT\_CREDIT)**

The scatter plot between total customer income (AMT\_INCOME\_TOTAL) and the total credit amount (AMT\_CREDIT) shows the following ranges:

* **Total customer income (AMT\_INCOME\_TOTAL):** Ranges from **0 to 5 million**.
* **Total credit amount (AMT\_CREDIT):** Ranges from **0 to 3 million**.

This distribution suggests that larger loans tend to be given to higher-income individuals. This could imply a relationship between credit risk and the approved credit amount. Further analysis on fraud trends based on income and credit levels could provide more insights into this relationship.

**3. Peak Loan Application Processing Hours (HOUR\_APPR\_PROCESS\_START)**

During a day, peak loan application processing hours are between 9 AM and 2 PM, with the highest number of accounts being processed at:

* **10 AM**: 37,722 accounts
* **11 AM**: 37,229 accounts
* **12 PM**: 34,233 accounts
* **1 PM**: 30,959 accounts
* **2 PM**: 27,682 accounts

These peak hours indicate a heavy load on the loan processing system, especially during the middle of the day (10 AM to 2 PM). This could potentially impact the efficiency of loan processing and the ability to detect fraud effectively.

**4. Customer Income Type Analysis (NAME\_INCOME\_TYPE)**

The top 5 income types with the highest number of accounts in the dataset are:

* **Working**: 143,550 non-fraudulent accounts and 15,224 fraudulent accounts.
* **Commercial Associate**: 66,257 non-fraudulent accounts and 5,360 fraudulent accounts.
* **Pensioner**: 52,380 non-fraudulent accounts and 2,982 fraudulent accounts.
* **State Servant**: 20,454 non-fraudulent accounts and 1,249 fraudulent accounts.
* **Unemployed**: 14 non-fraudulent accounts and 8 fraudulent accounts.

The "Working" and "Commercial Associate" income types have the highest number of fraudulent accounts, with 15,224 and 5,360 fraudulent accounts respectively. This suggests that these groups should be a focus for enhancing fraud detection efforts.

**5. Organization Type Analysis (ORGANIZATION\_TYPE)**

The top 5 organization types with the highest number of accounts are:

* **Business Entity Type 3**: 61,669 non-fraudulent accounts and 6,323 fraudulent accounts.
* **XNA**: 52,384 non-fraudulent accounts and 2,990 fraudulent accounts.
* **Self-employed**: 34,504 non-fraudulent accounts and 3,908 fraudulent accounts.
* **Other**: 15,408 non-fraudulent accounts and 1,275 fraudulent accounts.
* **Medicine**: 10,456 non-fraudulent accounts and 737 fraudulent accounts.

The organization type **Business Entity Type 3** has the highest number of fraudulent accounts. The fact that 6,323 accounts are fraudulent in this category suggests that this type of business carries a higher fraud risk compared to other types of organizations.

**6. Approval Rates and Good Customer Approval**

Although there is no direct information on the final approval rate and good customer approval rate in this dataset, these can be inferred from other factors:

* **Final approval rate** can be related to the proportion of approved credit accounts to the total loan applications. Given the large number of non-fraudulent accounts, the final approval rate is likely to be relatively high.
* **Good customer approval rate** (i.e., customers without fraud) can be estimated by the ratio of non-fraudulent accounts (91.93%) to the total number of accounts.

**7. Conclusion**

The credit card fraud detection data reveals that the fraud rate is relatively small but significant (8.07%). Factors such as income type, organization type, and peak loan application processing hours all impact the risk of fraud. A deeper analysis of these relationships could help improve fraud detection models and optimize the credit processing system.