Data Mining presentation content:

"Credit card fraud poses a significant challenge to the banking and finance sector, resulting in billions of dollars in losses annually for banks, financial institutions, and individuals. Nearly all credit card firms rely on transaction-level fraud detection methods to mitigate this threat. This project aims to address not only the challenge of identifying fraudulent transactions within a highly imbalanced dataset but also to minimize false positives, which often lead to customer dissatisfaction and financial disruptions.

Unlike many existing approaches that struggle with high false-positive rates, this project combines multiple machine learning models and targeted evaluation metrics to balance fraud detection accuracy with practical usability. By prioritizing interpretability, our model provides clear, actionable reasons for each fraud flag, making it easier for financial analysts to understand and verify fraud predictions. This transparency fosters trust among end-users, aids compliance with regulatory requirements, and enables improved customer communication. With a focus on high detection rates, reduced false positives, and user-friendly explanations, this project aims to deliver a fraud detection model that is both effective and adaptable for real-world deployment in financial institutions."’’

Diagram of a diagram of a fraud detection process

Description automatically generated

Source : https://web.instantaccept.com/wp-content/uploads/2022/11/FRAUD-Process-Flow-V4.2023.01.13.jpg

A graph of a company

Description automatically generated with medium confidence

Source: Federal Trade Commission (<https://www.ftc.gov/reports/consumer-sentinel-network-data-book-2023>)

Key Findings: - facts

* 60% of U.S. credit card holders have been victimized by fraud, and 45% have experienced fraud multiple times.
* 52 million Americans had fraudulent charges on their credit or debit cards last year, with unauthorized purchases exceeding $5 billion.
* Only 7% of fraudulent charges involved stolen or lost credit cards; the rest accessed personal data and account information remotely.

-Source: Security.org Last Updated Jul 26, 2024(https://www.security.org/digital-safety/credit-card-fraud-report/)

What is credit card fraud :

Credit card fraud is the unauthorized use of a credit card or credit card information to make purchases or withdraw funds without the cardholder’s consent. It’s a type of identity theft and a widespread financial crime that leads to significant losses for individuals and financial institutions alike.

**Types of Credit Card Fraud:**

1. **Card-Not-Present (CNP) Fraud**:
   * Occurs when a transaction is made without the physical card, often online or over the phone. CNP fraud is the most common type, as it only requires card information, which can be stolen through phishing, data breaches, or other methods.
2. **Card-Present Fraud**:
   * Involves physically using a stolen or counterfeit card. This can happen when a card is lost, stolen, or “skimmed” (copied) at an ATM or point of sale.
3. **Application Fraud**:
   * Occurs when a fraudster uses stolen personal information to apply for a new credit card in someone else's name.
4. **Account Takeover**:
   * In this case, the fraudster gains access to an existing account (through hacking, phishing, etc.) and makes unauthorized transactions or withdraws funds.
5. **Counterfeit Card Fraud**:
   * Using cloned or fake credit cards, which can be created using stolen credit card data from a magnetic strip or chip.

A computer security system with pictures of people

Description automatically generated with medium confidence

Source : Fintra (https://fintra.co.in/blog/credit-card-fraud)

**How Credit Card Fraud Occurs:**

Credit card fraudsters often obtain card information through phishing scams, hacking databases, using malware, or purchasing stolen data on the dark web. In cases of physical theft, fraudsters may also steal cards directly from individuals or use skimming devices to capture card data at ATMs or point-of-sale terminals.

How the project is different from other projects?

This project is user centric! most fraud detection projects prioritize accuracy or improving fraud recall on imbalanced data, but they don’t necessarily emphasize **interpretability** and **end-user usability**. Your project specifically addresses how the model’s decisions can be **explained and trusted by financial analysts** and other stakeholders.

The focus on balancing high detection rates with understandable outputs positions the model as a **scalable, real-world solution**, setting it apart from projects that focus purely on technical metrics without end-user considerations.

In summary, by addressing not only the detection of fraud but also how fraud detection can be communicated clearly, the project provides **a user-centered approach** to fraud detection.

**Evaluation Metrics**

For credit card fraud detection, particularly with imbalanced data, focus on metrics that capture the model’s ability to identify fraud accurately while minimizing false alarms:

1. **Confusion Matrix**: Tracks True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN).

2. **Precision**: Measures the accuracy of fraud predictions (TP / (TP + FP)) – useful to reduce false alarms.

3. **Recall**: Measures the model’s ability to capture actual fraud cases (TP / (TP + FN)).

4. **F1-Score**: Balances precision and recall; ideal for handling both false positives and false negatives.

5. **ROC-AUC**: Evaluates the model’s ability to distinguish between fraud and non-fraud cases across all thresholds.

6. **Precision**-Recall AUC: Highlights the trade-off between precision and recall, especially useful in imbalanced datasets.

7. **Specificity**: Tracks correctly identified legitimate transactions, helping prevent legitimate flags.

8. **FPR** **&** **FNR**: False Positive Rate (FP / (FP + TN)) reduces legitimate flags, while False Negative Rate (FN / (FN + TP)) helps maximize fraud detection.

Choosing and optimizing these metrics helps balance fraud detection accuracy with customer trust by minimizing unnecessary transaction flags.

**Anticipated Hurdles**

1. **Class Imbalance**: Fraud cases are rare compared to legitimate transactions, leading to a highly imbalanced dataset that can bias the model towards predicting non-fraud.
2. **High False Positives**: Misclassifying legitimate transactions as fraud can lead to customer dissatisfaction and potential loss of trust in the system.
3. **Data Quality and Noise**: Transaction data may contain inaccuracies or irrelevant information that can affect model performance and accuracy.
4. **Limited Contextual Features**: Without rich contextual information (e.g., customer behavior over time), it can be challenging to differentiate between legitimate anomalies and actual fraud.
5. **Evolving Fraud Patterns**: Fraud techniques frequently change, meaning models need regular updates or retraining to stay effective.
6. **Real-Time Processing Requirements**: Detecting fraud in real time is computationally demanding, especially with large volumes of transactions.
7. **Scalability**: Ensuring the model can handle large datasets and high transaction volumes efficiently as data grows.
8. **Model Interpretability**: Stakeholders and regulatory bodies often require transparent models that provide understandable predictions.
9. **Privacy and Security Concerns**: Handling sensitive financial data requires strict adherence to data privacy regulations, which can limit the data available for model training.

**Challenges and Solutions in Fraud Detection**

**1. Challenge: Highly Imbalanced Dataset**

* **Solution:** Apply **SMOTE** (Synthetic Minority Over-sampling Technique) to balance classes and enhance model learning.

**2. Challenge: Risk of High False Positives**

* **Solution:** Use **Precision-Recall AUC** and **threshold tuning** to optimize the model’s precision, reducing false alarms.

**3. Challenge: Limited Contextual Features**

* **Solution:** Implement **PCA** for dimensionality reduction, focusing on key features to retain useful variance for fraud detection.

**4. Challenge: Model Interpretability for Stakeholders**

* **Solution:** Utilize **Decision Trees** and **SHAP values** to increase transparency, helping stakeholders understand predictions.