**Anomaly Detection for Fraud Prevention in Financial Transactions**

***By Jasmine Cooper***

**Overview**

This project develops an anomaly detection model to identify potential fraud based on transaction data. The dataset contained two labels for fraud data, isFraud and isFlaggedFraud, that appears to be unreliable since only 535 transactions out of 999999 transactions were determined to be fraud.

In a real-world scenario, management might find this result suspicious and request a data scientist to evaluate whether the existing labels and model are flawed. This analysis aims to refine the fraud detection process and build a more robust model for detecting irregularities.

The project focuses on developing both unsupervised and supervised models to predict potential fraud, with the goal of improving the accuracy of fraud detection and addressing the apparent under-identification of fraudulent transactions.

**Data Collection & Preprocessing**

The transaction data was sourced from a Kaggle CSV file: [Paysim Dataset](https://www.kaggle.com/datasets/ealaxi/paysim1)

The key features selected:

* Step: 1 step = 1 hour
* Type: Payment type, converted to numerical values using label encoding
* Amount: Transaction amount
* OldbalanceOrig: Customer initial balance before the transaction
* NewBalanceOrig: Customer balance after the transaction
* OldbalanceDest: Recipient’s initial balance before the transaction
* NewbalanceDest: Recipient’s balance after the transaction

There were no missing values in the dataset.

**Methodology**

The low proportion of transactions identified as fraud are extremely low (<1%), raising concerns about the reliability of the existing fraud detection process. Management suspects this result may be inaccurate and has tasked myself with ensuring the robustness of the fraud detection model.

To address this, an unsupervised Isolation Forest model was developed to detect anomalies in the dataset. The identified anomalies were then used to train a supervised XGBoost model, utilizing the original features and anomaly labels to enhance fraud detection performance.

The models were evaluated using metrics such as precision, recall, F1-score, and AUC to assess their effectiveness. This hybrid approach combines unsupervised learning for anomaly detection with supervised learning for fraud classification improving performance, especially in the presence of class imbalance.

**Implementation**

To build a robust fraud detection system, the implementation combined unsupervised and supervised learning techniques using Python libraries and tools for machine learning, evaluation, and visualization.

The following libraries were used for model development:

* Scikit-learn: For anomaly detection, data splitting, and model evaluation
* XGBoost: For building and optimizing the supervised learning model
* RandomizedSearchCV: For hyperparameter tuning to improve model performance.
* Tableau: Used for visualizing and reporting insights derived from the models.

The following steps were undertaken for model training:

An Isolation Forest was employed for unsupervised anomaly detection to identify potential fraud in the dataset. The model predicts anomalies by isolating data points and converts them to binary labels where 1 represents fraud and 0 represents non-fraud. The predicted anomalies became the target variable, y, for the supervised learning model.

An XGBoost classifier was trained using the feature set, X, and the anomaly labels, y, from the Isolation Forest. The dataset was split into training and test sets (80/20 ratio) for model evaluation. The XGBoost model was initialized with parameters designed to address the class imbalance.

RandomizedSearchCV was applied to optimize XGBoost’s hyperparameters, exploring combinations like max\_depth, learning\_rate, and n\_estimators. The best parameters were identified and used to train the final model for improved performance.

Metrics such as precision, recall, F1-score, and AUC were used to assess the performance of both the baseline and tuned models. The tuned model demonstrated improved fraud detection, particularly in identifying rare fraudulent transactions.

**Results & Evaluation**

The original dataset’s isFraud label flagged less than 1% of transactions as fraudulent. In contrast, the hybrid model approach, combining an unsupervised Isolation Forest for anomaly detection and a supervised XGBoost model, identified approximately 20% of transactions as likely fraudulent.

The stark contrast between the original dataset’s <1% fraud rate and the model’s 20% prediction rate highlights potential flaws in the initial labeling process. These flaws may include underreporting of fraud due to outdated detection methods or insufficient identification capabilities in the original dataset. By incorporating anomaly detection, the hybrid approach identifies additional cases of potential fraud, suggesting that the original labels may not fully capture the extent of fraudulent activity.

To further interpret the results, a feature importance plot and SHAP plot were generated. The feature importance plot provides a global view of which features the model relies on most across all predictions. The SHAP plot illustrates the individual contribution of each feature to each prediction. Globally, step (a time variable, that represents 1 step = 1 hour), transaction amount, and recipient’s balance after the transaction have the greatest influence on fraud predictions.

The original dataset’s reliance on historical or inaccurate labeling methods likely missed fraudulent transactions that did not conform to prior patterns of fraud. The hybrid model’s ability to surface a broader range of suspicious transactions suggests that the original isFraud labels lack sufficient depth and coverage. These flaws could stem from bias, evolving fraud techniques, and/or incomplete data. Historical labels may only capture fraud that was easy to detect using older rules-based methods. Also, as fraud schemes become more sophisticated, older datasets may fail to reflect new tactics. In addition, limited access to external signals or secondary verification might result in many fraudulent transactions going undetected.

This approach demonstrates the value of integrating multiple methods to optimally identify hidden fraud cases, providing a more robust and comprehensive fraud detection mechanism. This underscores the need to reassess the original labeling process and incorporate modern techniques for fraud detection.

**Model Refinement & Next Steps**

*Refining Fraud Detection Process*

To improve fraud detection and address the shortcomings identified in the analysis, the following steps are recommended:

1. Expand Feature Set: Include additional features such as transaction metadata, geolocation data, or device identifiers to provide a more comprehensive view of transactions.
2. Improve Data Quality: Enhance the dataset by combining it with external data sources like credit bureau information or customer profiles to verify transactions. This information can be regularly updated to the dataset to reflect the evolving nature of fraud schemes.
3. Enhance Labeling Accuracy: implement semi-supervised learning techniques to iteratively improve fraud labels using a mix of human verification and model predictions.

*Deploying the Model*

1. Integration with Real-Time Systems:
   * Deploy the hybrid model in a production environment by integrating it with real-time transaction systems, ensuring timely identification of potential fraud.
   * Set up API’s to communicate between the model and financial systems for seamless fraud flagging during transaction processing.
2. Monitoring and Feedback Loops:
   * Continuously monitor the model’s performance using real-world data and update it as necessary to maintain accuracy.
   * Establish feedback loops where flagged transactions can be reviewed and used to retrain the model, ensuring continuous improvement.

*Future Enhancements*

Model performance can be further enhanced by exploring alternative anomaly detection methods, such as Autoencoders or One-Class SVMs, which may capture complex fraud patterns more effectively. Experimenting with ensemble models that combine multiple fraud detection techniques could further improve robustness by leveraging the strengths of different approaches. Additionally, tailoring the model to specific industries or regions by incorporating localized patterns and domain-specific features can increase accuracy and relevance, ensuring the model addresses fraud behaviors effectively.

**Dashboard Design and Visualization**

To ensure the dashboard loaded efficiently, the final dataset with predictions was reduced to 10% of its original size, enabling smoother performance for visualization.

There are 4 key visualizations in the dashboard, each providing different insights into fraud patterns in transactions. Below are the visualizations and corresponding insights:

*Fraud Transactions by Type*

This bar chart reveals that Cash In transactions are most frequently flagged as fraudulent. The dataset contains different transaction types including Cash In, Cash Out, Payment, Debit, and Transfer. Fraud tends to be more associated with cash inflows rather than outflows, payments, or transfers which suggest a potential pattern of fraudulent activity targeting deposits or incoming funds.

*Transaction Amount vs. Customer Balances*

This scatter plot visualizes the difference in customer balances before and after transactions, where the Y-axis shows the change in customer balance and the X-axis represents the transaction amount.

The non-fraudulent transactions tend to cluster around the $0M, indicating little to no significant change in customer balance. Fraudulent transactions are often flagged when there are larger negative or positive balance changes. Specifically, these transactions are flagged when both the transaction amount and the balance amount become larger. The pattern suggests that transactions with higher amounts and greater deviations in customer balances (whether positive or negative) are more likely to be identified as fraudulent. Fraudulent transactions tend to occur when the customer balance decreases significantly (negative balance change) and the transaction amount is large relative to the balance. For example, fraudulent transactions are more frequently identified when transactions amounts reach $1M or higher and customer balance changes approach positive or negative $2M. This indicates that both high transaction amounts and large shifts in customer balances are strong indicators of fraud risk.

*Transaction Amount vs. Recipient Balance Changes*

This scatter plot visualizes the difference in the recipient’s balance before and after transactions, where the Y-axis shows in the change in recipient balance and the X-axis represents the transaction amount.

Non-fraudulent transactions for recipients tend to hover around and above the $0M mark in terms of balance change. A few outliers show low transaction amounts and low recipient balances but are still flagged as fraudulent. This suggests that fraud can sometimes be detected with small or unassuming transactions if the patterns align with other factors. As transaction amounts and recipient balance changes increase, the likelihood of a transaction being flagged as fraud also increases.

*Fraud Trends Over Time*

This bar chart visualizes fraud trends over a 45-hour period showing how fraud peaks at certain times of the day.

Fraudulent transactions peak around 9 am each day and continue steadily throughout the day until about 7 pm (hour 19). After that, the trend starts to decline. Fraud starts to pick back up the next day at 9 am and follows a similar pattern, peaking again around 9 am and tapering off after 8 pm. This suggests that there may be a time-related trend in fraudulent activities, possibly linked to higher transaction volumes during certain hours or more vulnerable periods of the day.

A screenshot of a graph

Description automatically generated

**Conclusion**

This project aimed to enhance fraud detection in financial transactions by addressing the limitations of an existing system that flagged less than 1% of transactions as fraudulent. Using a hybrid approach that combined an unsupervised Isolation Forest for anomaly detection with a supervised XGBoost model, the analysis revealed that approximately 20% of transactions were likely fraudulent.

Key insights from the dashboard visualizations provided further clarity on patterns associated with fraudulent transactions. The analysis indicated that non-fraudulent transactions generally involved minimal changes in customer and recipient balances. In contrast, fraudulent transactions were often flagged when both the transaction amount and balance deviations (either positive or negative) were significantly higher.

The methodology improved fraud detection performance through advanced modeling techniques and a comprehensive feature set, allowing the system to flag fraudulent transactions more effectively. This approach not only increased the detection rate to around 20% but also provided a more actionable framework for identifying fraud risks in financial transactions.