**Isak Khatua**

INTERNSHIP TASK

**FRAUD DETECTION MODEL**

1. **Data cleaning including missing values, outliers and multi-collinearity.**

**Ans**. Missing Values: Our dataset did not contain any missing values, which was verified using the isnull().sum() function. This ensured that all transactions had complete information.

Outliers: The amount column exhibited a significant number of outliers, which is expected in financial transactions. These outliers were retained to ensure the model could learn to identify potential frauds. Outliers were visualized using box plots .

Multi-Collinearity: These features were either removed or combined to simplify the model and improve predictive performance.

Unnecessary Columns: During data cleaning, we identified three columns that were not necessary for fraud detection. These columns were dropped to streamline the dataset and focus on the most relevant features.

1. **Describe your fraud detection model in elaboration.**

**Ans.** Our fraud detection model is based on a Random Forest classifier. Random Forest was chosen due to its robustness in handling large datasets and its ability to manage both numerical and categorical features. It also provides feature importance scores, which help in understanding which features are most influential in predicting fraud. The model was trained using a balanced dataset created by stratified sampling to ensure both fraudulent and non-fraudulent transactions were adequately represented. While hyperparameter tuning was not performed due to the computational constraints of the large dataset, stratified sampling was employed to ensure the training and test sets were representative of the overall data distribution, thereby enhancing the model’s reliability.

1. **How did you select variables to be included in the model?**

**Ans.** Variable selection for the fraud detection model was conducted through statistical techniques:-

Initial Consideration: All available features were initially considered for inclusion in the model to ensure no potentially important information was overlooked.

Correlation Analysis: We performed correlation analysis to detect multi-collinearity among features. This involved calculating correlation coefficients to identify pairs of features with high correlations. Highly correlated features can cause redundancy and instability in the model, so we addressed this by either combining these features or removing one of the correlated pairs.

1. **Demonstrate the performance of the model by using best set of tools.**

**Ans.** The model’s performance was evaluated using several metrics, including precision, recall, and F1-score. Precision was particularly emphasized as it measures the accuracy of the positive predictions, which is crucial in fraud detection. Cross-validation was used to ensure the robustness of the results. Additionally, a confusion matrix was plotted to visualize the true positives, true negatives, false positives, and false negatives. These tools collectively demonstrated the model’s ability to accurately detect fraudulent transactions while minimizing false positives.

1. **What are the key factors that predict fraudulent customer?**

**Ans.** The key factors identified by our model include:

Transaction Amount:

Higher Amounts: Large transactions are more likely to be fraudulent, as fraudsters aim to maximize their gains.

Unusual Patterns: Transactions that deviate significantly from the customer's usual spending patterns.

Old and New Balances:

Discrepancies: Significant differences between old and new balances, particularly zero balances after a transaction, can indicate fraud.

Zero Balances: Frequent zero balances immediately after transactions suggest attempts to deplete accounts quickly.

Transaction Type:

Specific Types: 'TRANSFER' and 'CASH\_OUT' transactions are more commonly associated with fraud.

Transfers: Transfers to and from rarely used accounts are often red flags.

Frequency of Transactions:

High Frequency: A high volume of transactions in a short period is indicative of fraud.

Burst Patterns: Sudden bursts of activity can suggest attempts to quickly drain or launder money through an account.

Transaction Location and Time:

Unusual Locations: Transactions from locations atypical for the user can signal fraud.

Time of Day: Transactions at odd hours, inconsistent with usual behavior, may indicate fraudulent activity.

Account Activity History:

Inconsistent Activity: Sudden changes in activity levels, such as a burst of transactions after a period of inactivity, can be suspicious.

Account Age: New accounts processing large or numerous transactions may be flagged as fraudulent.

These key factors help our model accurately identify potentially fraudulent transactions and mitigate financial losses. They are based on both statistical analysis and domain expertise, ensuring a robust and practical fraud detection system.

1. **Do these factors make sense? If yes, How? If not, How not?**

**Ans.** Yes, these factors make sense. Fraudulent transactions often involve large sums of money to maximize the fraudulent gain. Discrepancies between old and new balances can indicate that the money was moved to a different account, which is a common tactic in fraud. Certain transaction types are more vulnerable to fraud due to the nature of how funds are transferred. High-frequency transactions can indicate attempts to bypass detection systems by overwhelming them with numerous small transactions.

1. **What kind of prevention should be adopted while company update its infrastructure?**

**Ans.** To prevent fraud during infrastructure updates, the company should adopt the following measures:

Implement Real-Time Fraud Detection Systems:

Machine Learning Models: Deploy advanced machine learning models that can analyze transactions in real-time, identifying and flagging suspicious activities as they occur.

Immediate Action: Ensure the system can take immediate action, such as temporarily freezing accounts or flagging transactions for further review, to prevent potential fraud from progressing.

Enhance Security Protocols:

Robust Encryption: Use strong encryption methods to protect sensitive data both at rest and in transit.

Multi-Factor Authentication (MFA): Implement MFA to add an additional layer of security, requiring users to verify their identity through multiple methods.

Regular Security Audits: Conduct frequent security audits to identify and address vulnerabilities in the system. Ensure compliance with industry standards and regulations.

Monitor Transaction Patterns:

Continuous Monitoring: Set up systems to continuously monitor transaction patterns and detect anomalies that may indicate fraud.

Model Updates: Regularly update fraud detection models to recognize and adapt to new fraud tactics and evolving threat landscapes.

Behavioral Analysis: Incorporate behavioral analysis to distinguish between legitimate and suspicious activities based on user behavior.

Employee Training:

Regular Training Sessions: Conduct regular training sessions for employees on fraud detection techniques, security best practices, and the latest trends in cyber threats.

Incident Response Training: Train employees on how to respond to suspected fraud incidents quickly and effectively to minimize damage.

Awareness Programs: Develop awareness programs that keep employees informed about the importance of security and their role in maintaining it.

Customer Awareness:

Educational Campaigns: Launch educational campaigns to inform customers about common fraud schemes and how to protect themselves.

Reporting Mechanisms: Encourage customers to report any suspicious activities or potential fraud incidents promptly.

Communication Channels: Provide clear and accessible communication channels for customers to reach out if they suspect fraud, ensuring they receive timely assistance.

By implementing these comprehensive measures, the company can significantly enhance its fraud prevention capabilities during infrastructure updates, protecting both its assets and its customers' interests.

1. **Assuming these actions have been implemented, how would you determine if they work?**

**Ans.** The effectiveness of these actions can be determined by:

Reduction in Fraudulent Transactions: Track the number of fraudulent transactions before and after implementation.

Customer Feedback: Collect feedback from customers regarding any suspicious activity and their overall satisfaction with the security measures.

Regular Audits: Conduct regular security audits to identify and address any vulnerabilities.

Performance Metrics: Use performance metrics such as precision, recall, and F1-score to evaluate the fraud detection system’s accuracy and efficiency.

Incident Reports: Monitor the number and severity of fraud incidents reported by customers and staff.

**Conclusion**

For effective fraud detection and prevention, a comprehensive approach is essential. Our fraud detection model, utilizing a Random Forest classifier, efficiently identifies key fraud indicators like transaction amounts, balance discrepancies, transaction types, and unusual frequencies. Stratified sampling ensures a balanced dataset for reliable model training.

Key fraud predictors include high transaction amounts, zero balances, specific transaction types, and unusual frequencies. These insights validate our model and pinpoint areas for additional preventive measures.

To bolster fraud prevention during infrastructure updates, the company should:

Implement Real-Time Fraud Detection

Enhance Security Protocols

Monitor Transaction Patterns

Provide Employee Training

Increase Customer Awareness

By adopting these strategies, the company can effectively protect its infrastructure and customer data against evolving fraud tactics, ensuring security and trust.

*Thank You*