Detecting Credit Card Fraud Using Supervised Learning and Anomaly Detection Techniques

By Bhavna Kakkar
M.A. Clinical Psychology

Can we accurately detect and prevent fraudulent credit card transactions using historical transaction data and machine learning techniques?

The rapid growth of digital transactions challenges financial institutions with continually escalating rates of credit card fraud. Since fraudulent activity results in significant financial losses and damages customer trust, a **robust and adaptive solution** for real-time detection is essential.

The core objective of this proposal is to develop a predictive model that identifies potentially fraudulent transactions based on historical data, enabling timely intervention to protect institutional and customer assets.

Methodology

1. Analytic Approach: Hybrid Model Implementation

To create a robust and comprehensive system for detecting credit card fraud, this methodology utilizes a **Hybrid Analytic Model** that integrates **Supervised Learning** with **Anomaly Detection**. This combination enhances the accuracy and adaptability of the fraud detection system by leveraging the distinct strengths of both techniques.

The design of the hybrid approach is crucial: it reduces the probability of **false negatives** (missed fraud) while simultaneously controlling **false positives** (legitimate transactions flagged as suspicious). Furthermore, the system is designed to adapt to new fraud patterns through anomaly detection while still leveraging the high classification accuracy of supervised learning for known, established patterns.

The system operates using a **sequential two-stage screening process**:

- **Initial Screening with Supervised Learning:** Supervised learning models serve as the first line of defense. These models are trained extensively on historical transaction data, where each transaction is labeled as either "fraudulent" or "non-fraudulent." When a new transaction occurs, the model rapidly classifies it based on the patterns it has learned. This initial screening successfully identifies transactions that match known fraud characteristics, allowing for immediate intervention.

- **Second Screening with Anomaly Detection:** In the second stage, anomaly detection, in combination with behavior analysis, monitors and learns the typical transaction patterns of each cardholder, specifically their usual spending amounts, geographical locations, and transaction times. When a new transaction occurs, the system rigorously compares it against the cardholder's established behavior profile. If the transaction significantly **deviates from the norm**—such as a large purchase in a foreign country for a user who typically shops locally—it is flagged as an **anomaly**. This approach allows the system to identify potentially fraudulent activities that may not match known, labeled fraud patterns but still appear suspicious, providing a critical additional layer of security after the initial supervised learning screening.

2. Data Requirements and Preparation

For the implementation of the hybrid model combining supervised learning and anomaly detection, several types of data may be required:

- **Historical Transaction Data**: This include details of both fraudulent and non-fraudulent transactions, with features such as transaction amounts, timestamps, locations, merchant categories, and customer behavioural patterns.
- **Real-Time Transaction Data**: Continuous real-time data is necessary to monitor ongoing transactions and detect anomalies as they occur.
- **Supplementary Data:** External data sources, such as geolocation data and merchant information, is needed to enrich the dataset and provide additional context for analysis.

Data Understanding and Preparation Steps:

- **Data Exploration:** The process begins with diving into the collected data to understand the story it tells. The primary goal is to analyze the **distribution** of different transaction types, identify recurring **patterns**, and quickly spot any inconsistencies or initial anomalies. **Descriptive statistics and visualizations** (such as charts and graphs) are used to clarify how various features relate to fraud. For example, visualizations can immediately show that fraudulent transactions often occur more frequently at specific times of the day or involve unusually large amounts. By uncovering these key trends, the exact features most relevant for detection are identified.

- **Data Cleaning:** Next, the essential task of cleaning the data is performed. This involves systematically managing **missing values**, correcting any internal inconsistencies, and carefully handling **outliers** that could unfairly skew the models. For instance, transactions with missing values will require **imputation** from comparable, legitimate transactions. Conversely, outliers—like exceptionally large purchases—must be thoroughly examined; if they are likely to compromise the model's accuracy, they will be either adjusted or excluded from the training set.
- **Data Preprocessing:** The final step is preparing the data into the precise format required for modeling. This involves **normalizing transaction amounts** to ensure that large purchases do not disproportionately influence the model's learning process. Additionally, descriptive **categorical features** (such as merchant categories or transaction locations) will be **encoded** into numerical values. This critical step guarantees that the data is in the required structure for both the supervised learning models and the anomaly detection system to analyze everything effectively.

3. Modeling and Development

Once the data has been thoroughly understood and prepared, the next step is to develop the models that power the fraud detection system. Given the hybrid approach of combining **Supervised Learning** with **Anomaly Detection**, two distinct types of models are required:

Supervised Learning Models:

The supervised learning models are the first line of defense.

- These models are trained on the historical transaction data, where each transaction is explicitly labeled as either "**fraudulent**" or "**non-fraudulent**."
- The process begins by selecting highly suitable algorithms, such such as **Random Forest** and **Gradient Boosting Machines** (e.g., **XGBoost** or **LightGBM**). These are

highly recommended due to their strong performance in complex and **imbalanced datasets**—like those typically seen in fraud detection.

- The models analyze patterns in the labeled data to learn the specific characteristics that typically differentiate fraudulent transactions from legitimate ones. For instance, the models learn that fraud often involves unusually large amounts, occurs at odd hours, or is made in locations far from the cardholder's usual area.
- Once trained, these models are used to evaluate new, incoming transactions. Each transaction is assessed based on the learned patterns and is classified as either "**potentially fraudulent**" or "**non-fraudulent**."

Anomaly Detection Models:

An anomaly detection model is employed to identify transactions that deviate significantly from a cardholder's usual behavior.

- This model does not rely on labeled data; instead, it focuses solely on detecting **outliers**—transactions that appear highly unusual compared to the customer's established transaction history.
- For example, if a cardholder typically makes small, local purchases, a sudden large purchase in a foreign country is immediately flagged as suspicious by the anomaly detection model.
- The system uses specific techniques, such as **clustering** or **Isolation Forests**, which group transactions into "normal" behavior clusters and efficiently identify outliers that do not fit within these established norms.

System Integration and Lifecycle:

The **hybrid system** integrates both models in a sequential process to maximize coverage:

1. When a new transaction occurs, it is first evaluated by the supervised learning model. If classified as potentially fraudulent, it is immediately flagged for review.
2. If the transaction is not flagged by the supervised model, it then passes through the anomaly detection model, where it is compared against the cardholder's typical behavior. If the transaction is found to be an **anomaly**, it is flagged for further investigation.

Following initial deployment, the models are **continuously monitored** and updated as more data becomes available, allowing them to adapt to emerging fraud patterns. **Regular retraining cycles** are implemented to maintain high accuracy and responsiveness in detecting fraud.

4. Evaluation

After the models are fully developed, **rigorous evaluation** is essential to confirm they perform effectively in real-world scenarios. This final stage involves the following critical steps:

Performance Metrics:

The models are assessed using metrics specifically tailored for the imbalanced classification problem: **Accuracy, Precision, Recall, and the F1 Score**. For the supervised learning models, these metrics directly indicate how well the models distinguish between fraudulent and non-fraudulent transactions.

- A **high Precision** is crucial because it suggests the model accurately identifies fraud with minimal **false positives** (avoiding the flagging of legitimate customer transactions).
- A **high Recall** is necessary because it indicates that most actual fraudulent transactions are successfully detected (minimizing financial losses).

Testing on Real-World Data:

The models are tested on a separate set of real-world transaction data that was not used during the training phase. This dedicated test data provides a **realistic assessment** of how the models perform in a live environment, which is vital for identifying any potential issues such as **overfitting** (poor generalization) or **underfitting** (poor learning).

Behavioral Validation:

For the anomaly detection model, **behavioral validation** is conducted by comparing transactions flagged as anomalies against known patterns of both legitimate and fraudulent transactions. This validation step is essential for refining the model’s ability to detect genuinely unusual activity while systematically reducing the chances of incorrectly flagging legitimate behavior as suspicious.

Iterative Refinement:

Based on the comprehensive evaluation results, the models are **iteratively refined**. This crucial process involves adjusting the features used, retraining the models with additional data to address specific weaknesses, or modifying the internal thresholds for what constitutes an anomaly. The primary goal is to **continuously improve** the models until they meet the desired level of accuracy, reliability, and responsiveness required for operational deployment.

By carefully following these steps, a robust and effective hybrid fraud detection solution is ensured, capable of protecting both the financial institution and its customers from fraudulent activities.

Conclusion

The proposed methodology establishes a robust framework for an advanced fraud detection system. By leveraging a **Hybrid Analytic Model**—integrating the precision of Supervised Learning for known patterns with the adaptability of Anomaly Detection for emerging threats—the system is uniquely positioned to address the complexity and imbalance inherent in real-time transaction data. Through continuous monitoring and an **Iterative Refinement** lifecycle, this solution is capable of achieving the necessary high recall and precision metrics required to effectively mitigate financial risk and safeguard customer trust.