A R DIGITECH
International Journal Of Engineering, Education And Technology (ARDIJEET)
www.ardigitech.in ISSN 2320-883X, VOLUME 13 ISSUE 01, 01/01/2025 pub: 15/03/2025

A Systematic Approach To Identifying Credit Card Fraud Detection From The Banker's Perspective Using Machine Learning

[1]Aarti Bharat Jadhav, [2]Prof. Pallavi P. Rane, [3]Prof. Nilesh N. Shingne, [4]Prof. Pravin S. Rane
[1]aartibj1996@gmail.com, [2]koltepallavi200@gmail.com, [3]shingne.nilesh236@gmail.com,
[4]pravin28.rane@gmail.com
[2,4]Assistant professor, Rajarshi Shahu College of Engineering, Buldhana, Maharashtra
[3]Assistant professor, Sanmati engineering College, Washim, Maharashtra

Abstract: Pre-issuance credit card fraud detection is crucial for minimizing losses and protecting customers. This paper explores various methods employed by banks to identify potentially fraudulent credit card applications *before* cards are issued. We examine identity verification techniques, including KYC checks, biometric authentication, and data consistency analysis. The paper also discusses risk assessment strategies, such as credit history checks, application data analysis, and fraud scoring models. Furthermore, we investigate the role of machine learning in anomaly detection and predictive modelling for identifying suspicious applications. Finally, we address the challenges of balancing effective fraud prevention with minimizing false positives and maintaining data privacy. This paper aims to provide a comprehensive overview of pre-issuance fraud detection techniques, highlighting best practices and future research directions for enhancing security in the banking sector.

Keywords: Fraud, Identity, Cards, Risk, Modelling

1. Introduction

Detecting credit card fraud *before* issuing cards is a proactive approach that can save banks and customer's significant trouble. Since you're dealing with applications rather than transactions, the focus shifts to verifying identity, assessing risk, and identifying potentially fraudulent applications. [5] Here's a breakdown of methods used:

1.1 Identity Verification and Authentication:

- **KYC (Know Your Customer) Checks:** These are fundamental. Banks verify the applicant's identity using government-issued IDs, address proofs, and other documents. Cross-referencing with databases is crucial.
- **Biometric Verification:** Increasingly used, biometrics (fingerprints, facial recognition) add a strong layer of authentication, making it harder for fraudsters to use stolen identities.
- **Social Media Verification:** While not always conclusive, checking publicly available social media profiles can sometimes reveal inconsistencies in the application information.
- **Phone and Email Verification:** Validating phone numbers and email addresses through confirmation codes helps ensure the applicant's contact information is genuine.

1.2 Risk Assessment and Scoring:

- **Credit History Check:** Accessing credit bureaus (like Equifax, Experian, TransUnion) provides a credit score and history, indicating the applicant's creditworthiness and potential risk. A lack of credit history can also be a red flag.
- **Application Data Analysis:** Banks analyze the application data for red flags:
  - Inconsistent information (e.g., different addresses, mismatched employment details).
  - Unusual spending patterns or income levels compared to the applicant's profile.
  - Requests for unusually high credit limits.
  - Use of temporary or virtual addresses.
- **Fraud Scoring Models:** Sophisticated algorithms analyze application data and assign a fraud score. These models are trained on historical data of fraudulent applications to identify patterns and predict risk.
- **Geographic Risk:** Applications from high-risk areas known for fraud might trigger further scrutiny.

1.3 Machine Learning Techniques:

- **Anomaly Detection:** Machine learning algorithms can identify anomalous patterns in application data that deviate from typical profiles, potentially indicating fraud.
- **Predictive Modeling:** Models can be trained to predict the likelihood of an application being fraudulent based on various factors.
- **Clustering:** Clustering algorithms can group similar applications together, making it easier to identify suspicious clusters that might indicate organized fraud rings. [6]

2. Project Objectives:

- To demonstrate the generation of credit card number and its comprise result for attacks
- To simulate the attacks through multiple profile
- To identify the difference of previous comprise of card numbers and current
- To generate the comprise numbers and stored for record in cloud through encryption process

3. Literature Review

Credit card fraud detection has drawn a lot of research interest and a number of techniques, with special emphasis on data mining, have been suggested. Gosh and Reilly [1] have developed fraud detection system with neural network. Their system is trained on large sample of labelled credit card account transactions. These transactions contain example fraud cases due to lost cards, stolen cards, application fraud, counterfeit fraud, mail-order fraud and non-receive issue(NRI) fraud.

E. Aleskerov et al. [2] present CARDWATCH, a database mining system used for credit card fraud detection. The system is based on a neural learning module and provides an interface to variety of commercial databases.

Dorronsoro et al. [3] have suggested two particular characteristics regarding fraud detection- a very limited time span for decisions and a large number of credit card operations to be processed. They have separated fraudulent operations from the normal ones by using Fisher's discriminant analysis. Syeda et al. [4] have used parallel granular neural network for improving the speed of data mining and knowledge discovery in credit card fraud detection. A complete system has been implemented for this purpose.

Chan et al. [5] have divided a large set of transactions into smaller subsets and then apply distributed data mining for building models of user behaviour. The resultant base models are then combined to generate a meta-classifier for improving detection accuracy. Chiu and Tsai [7] consider web services for data exchange among banks. A fraud pattern mining (FPM) algorithm has been developed for mining fraud association rules which give information regarding the new fraud patterns to prevent attacks.

4. Proposed Methodology:

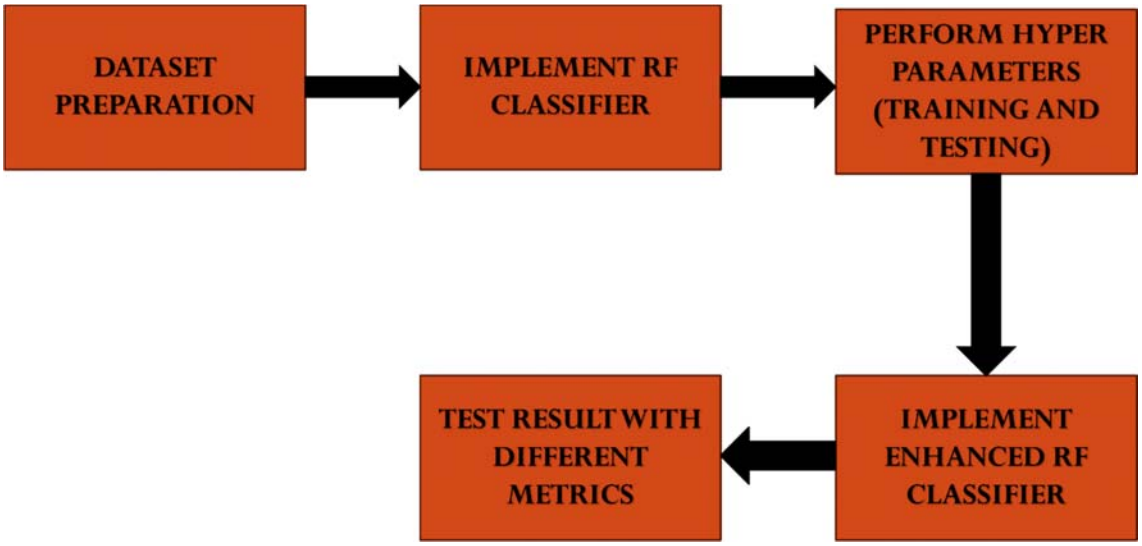

DATASET PREPARATION
IMPLEMENT RF CLASSIFIER
PERFORM HYPER PARAMETERS (TRAINING AND TESTING)
IMPLEMENT ENHANCED RF CLASSIFIER
TEST RESULT WITH DIFFERENT METRICS

Fig. 1 Flow of the proposed system

Step 1: (Data Preparation)

Credit card information has been sent to the server over the network. The information is a 46 digit string formed by concatenates some values such as Credit card number, card verification value and expiry date of the card. For the purpose of security, the information has been encrypted before sent to server over the network. The encryption technique used for securing the cipher text is Caesar cipher. Caesar cipher has been used with shift value of three. Original string of 46 digits has been converted into encrypted string of same length. For the purpose of security of Credit card application, some of the network attacks have been examined. An attack has been examined on the credit card information being sent over the network. [8]

Step 2: Implement Rf Classifier

Random forest classifier creates a set of decision trees from a randomly selected subset of the training set. It is a set of decision trees (DT) from a randomly selected subset of the training set and then it collects the votes from different decision trees to decide the final prediction

Step 3: Training And Testing For Model Creation

The selected ML algorithm learns how to make predictions or categorize data using the training set. In this phase, the model refines its internal settings to best match the training set of data. Finding the optimal values for hyperparameters (parameters that govern the learning process) that are not learned from the data is known as "hyperparameter tuning." In order to enhance the performance of the model, we are experimenting with various hyperparameter settings using the validation set.

Step 4: Implement Enhanced Rf Classifier

In this step, we will prepare the data by standardizing it, separating features from labels, and then splitting it into training and validation sets for machine learning model development and evaluation.

Step 5: Test Result With Different Metrics

This stage allows us to identify the comprise value of credit card and stored on cloud so as to maintain the records for future references. [9]

5. Benefits

- **Feature Importance**: Random Forest can provide insights into which features (e.g., transaction amount, location, time of day, etc.) are most important in distinguishing between legitimate and fraudulent transactions. This helps analysts understand the characteristics of fraudulent activities better.
- **Ensemble Learning**: Random Forest is an ensemble learning method that combines multiple decision trees to make predictions. Each tree in the forest is trained on a random subset of the data and features, reducing the risk of overfitting and improving generalization performance.
- **Robustness to Overfitting**: Random Forest is less prone to overfitting compared to individual decision trees. By averaging the predictions of multiple trees, it can reduce variance and provide more reliable predictions, especially in scenarios with noisy or incomplete data.

- **Handling Imbalanced Data**: Credit card fraud detection datasets are often highly imbalanced, with the majority of transactions being legitimate. Random Forest can handle class imbalance by oversampling the minority class (fraudulent transactions) using techniques like SMOTE or by adjusting class weights during training. [10]
- **Scalability**: Random Forest is relatively scalable and can handle large datasets efficiently. It can be parallelized across multiple CPU cores, making it suitable for processing large volumes of credit card transaction data in real-time or batch processing scenarios.
- **Non-linear Relationships**: Random Forest can capture non-linear relationships and interactions between features effectively, making it suitable for detecting complex patterns and anomalies associated with fraudulent activities that may not be apparent with linear models.
- **Model Interpretability**: While Random Forest is not as interpretable as simpler models like logistic regression, it still provides insights into feature importance and decision-making processes, helping analysts understand why a particular transaction was flagged as fraudulent.
- **Incremental Learning**: Random Forest supports incremental learning, allowing the model to be updated with new data over time without retraining the entire model from scratch. This is useful in dynamic environments where fraud patterns evolve over time. [11]

6. Simulation Techniques for Evaluating Credit Card Attacks

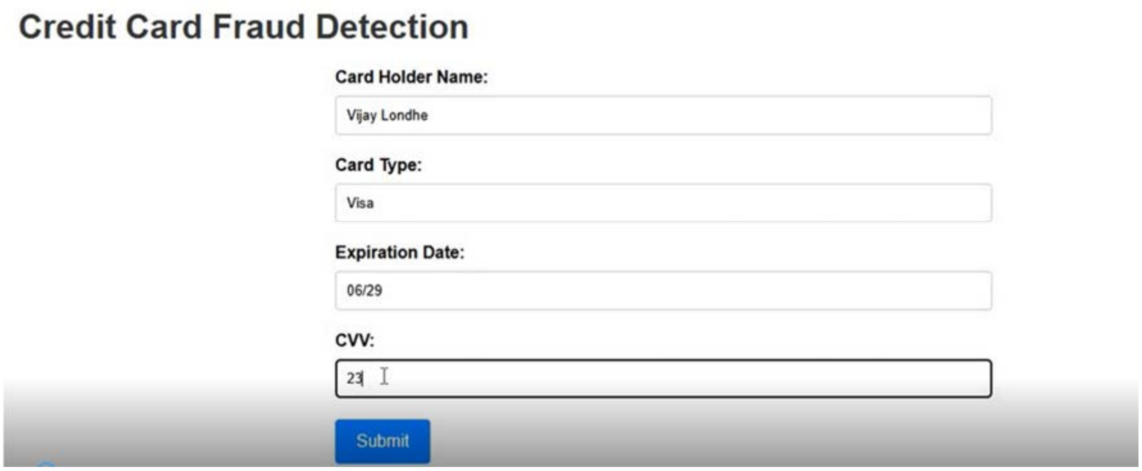
Simulating credit card processing failures in the cloud allows for proactive risk assessment and security control testing. Techniques include:

- **Threat Modeling:** This identifies potential threats, vulnerabilities, and attack vectors related to cloud-based credit card processing. Simulations based on these threats can test system resilience.
- **Fault Injection Tools:** These tools inject faults (e.g., network delays, memory errors) to observe their impact on credit card processing. Chaos Monkey and Gremlin are examples.
- **Security Testing Frameworks:** These frameworks simulate various attacks, including denial-of-service, data breaches, and unauthorized access. OWASP ZAP and Nessus are examples. [12]

7. Results

- Load the dataset containing credit card transaction data.
- Then, we separate the features (X) from the target variable (y).
- Next, we split the data into training and testing sets using **train_test_split**.
- We initialize a Random Forest classifier with 100 trees.

- The classifier is trained on the training data using the **fit** method.
- Predictions are made on the testing set using the **predict** method.
- Finally, we evaluate the model's performance using accuracy, confusion matrix, and classification report. [14]



Credit Card Fraud Detection

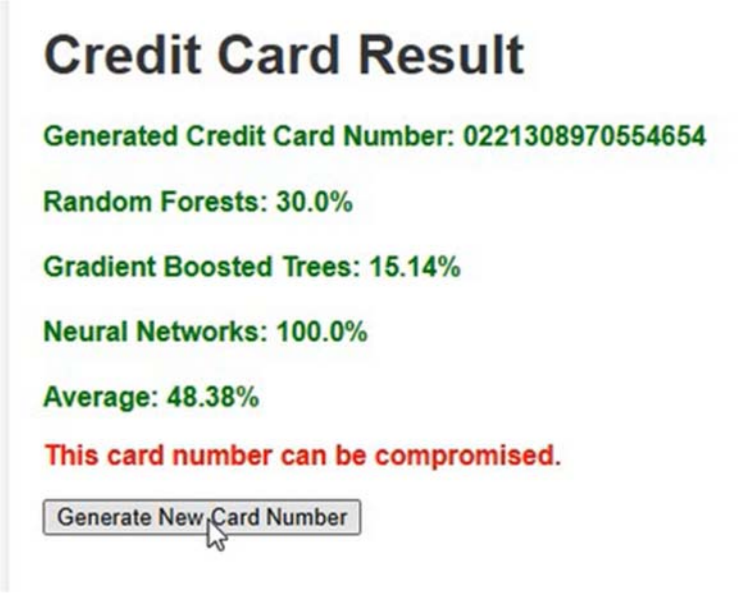
Card Holder Name: Vijay Londhe

Card Type: Visa

Expiration Date: 06/29

CVV: 23

Fig. 2 Input for the system



Credit Card Result

Generated Credit Card Number: 0221308970554654

Random Forests: 30.0%

Gradient Boosted Trees: 15.14%

Neural Networks: 100.0%

Average: 48.38%

This card number can be compromised.

Fig. 3 Credit card generated with compromised result

8. Conclusion:

Simulating cloud attacks proactively strengthens security. By mimicking real-world attacks, organizations can identify vulnerabilities, evaluate security controls, and pinpoint areas needing improvement. Simulations of credit card fraud reveal weaknesses in security configurations, access controls, and system design before attackers can exploit them, allowing for timely fixes. Analysing simulated attacks provides insights into attacker tactics and motivations, leading to more targeted defences. It offers a safe environment to test security control for banking system effectiveness, ensuring they can detect and respond to threats. Empirical data from simulations empowers organizations to make informed decisions about security investments and resource allocation. [15][16]

References:

[1] Sumathy, K. L., and M. Chidambaram. "Text mining: concepts, applications, tools, and issuesan overview." International Journal of Computer Applications 80, no. 4 (2013). pg. 23

[2] Aggarwal, Charu C., and Haixun Wang. "Text mining in social networks." In Social network data analytics, pp. 353-378. Springer, Boston, MA, 2011.

[3] Mostafa, Mohamed M. "More than words: Social networks' text mining for consumer brand sentiments." Expert Systems with Applications 40, no. 10 (2013): 4241-4251.

[4] Netzer, Oded, Ronen Feldman, Jacob Goldenberg, and Moshe Fresko. "Mine your own business: Market-structure surveillance through text mining." Marketing Science 31, no. 3 (2012): 521-543.

[5] Fuller, Christie M., David P. Biros, and Dursun Delen. "An investigation of data and text mining methods for real-world deception detection." Expert Systems with Applications 38, no. 7 (2011): 8392- 8398.

[6] Othman, Rohana, Nooraslinda Abdul Aris, Ainun Mardziyah, Norhasliza Zainan, and Noralina Md Amin. "Fraud detection and prevention methods in the Malaysian public sector: Accountants' and internal auditors' perceptions." Procedia Economics and Finance 28 (2015): 59-67.

[7] Dong, Wei, Shaoyi Liao, and Liang Liang. "Financial Statement Fraud Detection using Text Mining: A Systemic Functional Linguistics Theory Perspective." In PACIS, p. 188. 2016.

[8] Fu, Kang, Dawei Cheng, Yi Tu, and Liqing Zhang. "Credit card fraud detection using convolutional neural networks." In International Conference on Neural Information Processing, pp. 483-490. Springer, Cham, 2016.

[9] Rawte, Vipula, and G. Anuradha. "Fraud detection in health insurance using data mining techniques." In Communication, Information & Computing Technology (ICCICT), 2015 International Conference on, pp. 1-5. IEEE, 2015.

[10] Dilla, William N., and Robyn L. Raschke. "Data visualization for fraud detection: Practice implications and a call for future research." International Journal of Accounting Information Systems 16 (2015): 1-22.

[11] Kanapickienė, Rasa, and Živilė Grundienė. "The model of fraud detection in financial statements by means of financial ratios." Procedia-Social and Behavioral Sciences 213 (2015): 321-327.

[12] West, Jarrod, and Maumita Bhattacharya. "Some Experimental Issues in Financial Fraud Mining." In ICCS, pp. 1734-1744. 2016.

[13] Kim, Yeonkook J., Bok Baik, and Sungzoon Cho. "Detecting financial misstatements with fraud intention using multi-class cost-sensitive learning." Expert systems with applications 62 (2016): 32-43. pg. 24

[14] Olszewski, Dominik. "Fraud detection using self-organizing map visualizing the user profiles." Knowledge-Based Systems 70 (2014): 324-334.

[15] Albrecht, Chad, Daniel Holland, Ricardo Malagueño, Simon Dolan, and Shay Tzafrir. "The role of power in financial statement fraud schemes." Journal of Business Ethics 131, no. 4 (2015): 803-813.

[16] West, Jarrod, Maumita Bhattacharya, and Rafiqul Islam. "Intelligent financial fraud detection practices: an investigation." In International Conference on Security and Privacy in Communication Systems, pp. 186-203. Springer, Cham, 2014.