A CASE STUDY ON COST SENSITIVE CARD FRAUD DETECTION BASED ON DYNAMIC RANDOM FOREST AND K-NEAREST NEIGHBOUR

Dr.R.Yadagiri Rao[1], **D. Nagaraju**[2], **G Reshma Reddy**[3], **Burre Nandhini**[4], **B Vishnu Kumar**[5], **B Amarnath**[6]

[1]Professor, Department of CSE, Sri Indu Institute of Engineering & Technology, Hyderabad

[2]Assistant professor, Department of CSE, Sri Indu Institute of Engineering & Technology, Hyderabad

[3,4,5,6] IVth Btech Student, Department of CSE, Sri Indu Institute of Engineering & Technology, Hyderabad

ABSTRACT

Payment card fraud leads to heavy annual financial losses around the world, thus giving rise to the need for improvements to the fraud detection systems used by banks and financial institutions. In the academe, as well, payment card fraud detection has become an important research topic in recent years. With these considerations in mind, we developed a method that involves two stages of detecting fraudulent payment card transactions. The extraction of suitable transactional features is one of the key issues in constructing an effective fraud detection model. In this method, additional transaction features are derived from primary transactional data. A better understanding of cardholders' spending behaviours is created by these features. After which the first stage of detection is initiated. A cardholder's spending behaviours vary over time so that new behaviour of a cardholder is closer to his/her recent behaviours. Accordingly, a new similarity measure is established on the basis of transaction time in this stage. This measure assigns greater weight to recent transactions. In the second stage, the dynamic random forest algorithm is employed for the first time in initial detection, and the minimum risk model is applied in cost-sensitive detection. We tested the proposed method on a real transactional dataset obtained from a private bank.

Key words: Dynamic Random forest, Fraudulent payment.

1. INTRODUCTION

The use of bank cards as efficient tools for bank transactions has increased with the development of technologies for e-commerce (Mahmoudi & Duman, 2015; Kundu, Panigrahi, Sural, & Majumdar, 2009). The consequent rise in the number of bank transactions has translated to increased fraud. According to the Nilson Report, the financial losses from card fraud totalled $21.84 billion in 2015, $24.71 billion in 2016 and $27.69 billion in 2017. It is also reported that the actual amount of losses will increase by 2020 (Robertson, 2017). In our country, the financial losses at the result of payment card fraud have also increased with the growth of electronic banking and existence of more than 400 million payment cards and 10 billion bank transactions (Vosough, Taghavifard, & Alborzi, 2015). Fraud is as old as human life and considered a multimillion-dollar business around the world. Fraudsters typically target bank cards because successful deception means a considerable amount of money in a very short time for defrauders. This problem highlights the need for all banks and financial institutions to develop effective fraud detection systems to reduce the damages caused by fraud and combat strategies that form with the development of knowledge (Bahnsen, Aouada, Stojanovic, & Ottersten, 2016; Van Vlasselaer et al., 2015). There are various forms of payment card frauds that are categorized into two groups: application and behavioural fraud (Bolton & Hand, 2001). In application fraud, fraudsters receive new cards from issuing companies using false information. Behavioural frauds consist of mail theft, stolen/lost card, counterfeit card and ‘cardholder not present’ fraud. In mail fraud, fraudsters intercept payment cards in mail before they reach the cardholders. In stolen/lost card fraud, fraudsters obtain cards through theft. A physical card is used to perpetrate counterfeit, stolen/lost card and mail theft frauds. In ‘Card holder not present’ fraud, payment card details are stolen and then are used to carry out a transaction without the physical card, typically online. Online fraud enables the anonymity, reach, and speed to perpetrate fraud around the world.

2. SYSTEM IMPLEMENTATION:

The architecture diagram for this system would typically include the following components:

Data Ingestion:

This component handles the ingestion of transaction data from various sources such as payment gateways, databases, or data feeds.

Data Preprocessing:

The transaction data is preprocessed to clean and transform it into a suitable format for further analysis.

Preprocessing tasks may include data cleaning, handling missing values, data normalization, and engineering.

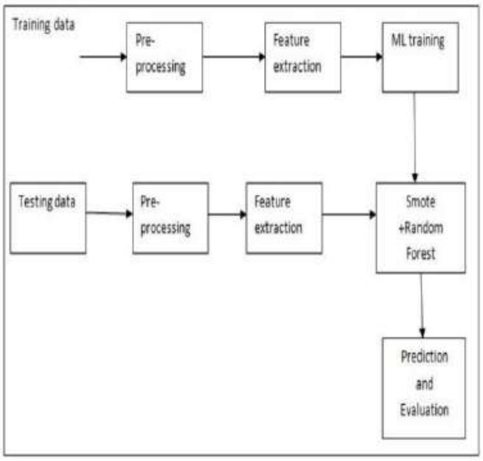

Figure 1: Evaluation of data

Random Forest Model:

The Random Forest model is a machine learning algorithm used for fraud detection. This component includes training the Random Forest model using the preprocessed transaction data and the extracted features. The trained model is used for prediction on new, unseen transactions.

Model Integration:

This component integrates the trained Random Forest and KNN models into the fraud detection system.

The models may be deployed in a production environment and made accessible for real-time or batch processing of new transactions.

Fraud Detection:

The integrated models are used to analyse new transactions and predict whether they are fraudulent or not.

2.1 FEATURE EXTRACTION:

This component extracts relevant features from the preprocessed transaction data. Features can include transaction amount, time of day, location, transaction frequency, and other domain-specific features.

The features of a cost-sensitive card fraud detection project based on random forest and k-nearest neighbour algorithm can include:

1. **Cost-sensitive framework:** Develop a framework that incorporates the cost sensitive approach, considering the asymmetric costs associated with misclassification errors in fraud detection. This allows the system to prioritize the accurate detection of fraudulent transactions while minimizing the financial cost of misclassification.
2. **Random forest algorithm:** Utilize the random forest algorithm, which is an ensemble learning method that combines multiple decision trees to make predictions. Random forest can handle high-dimensional data, capture complex relationships, and provide robustness against overfitting.
3. **K-nearest neighbour algorithm:**

   The k-nearest neighbour (KNN) algorithm is another machine learning algorithm used for fraud detection.

   This component includes training the KNN model using the preprocessed transaction data and the extracted features.

   The trained model is used for prediction on new transactions.

4. Implement the k-nearest neighbour algorithm, which is a non-parametric classification method that assigns labels to samples based on their proximity to the k-nearest training examples. K-nearest neighbour can be effective in detecting fraud by considering the similarities between transactions.
5. **Imbalanced data handling:** Address the imbalanced nature of credit card transaction data by employing techniques such as oversampling, undersampling, or hybrid approaches. These techniques aim to balance the representation of the minority class (fraudulent transactions) to avoid bias towards the majority class.
6. **Feature engineering and selection:** Perform feature engineering to extract meaningful information from the credit card transaction data. Select relevant features

that have a significant impact on distinguishing between fraudulent and legitimate transactions. Feature engineering techniques can include statistical measures, transaction metadata, and behavioural patterns.

7. **Model evaluation and optimization:** Evaluate the performance of the fraud detection system using appropriate evaluation measures such as precision, recall, F1-score, and area under the ROC curve (AUC-ROC). Optimize the hyperparameters of the random forest and k-nearest neighbor algorithms to achieve the best possible performance.
8. **Adaptability to new fraud patterns:** Implement mechanisms to monitor and detect emerging fraud patterns. Update the model periodically using new data to ensure that the system can adapt to evolving fraud techniques and maintain its effectiveness over time.
9. **Real-time or near-real-time processing:** Design the system to process credit card transactions in real-time or near-real-time to enable timely detection and prevention of fraudulent transactions. This involves optimizing the computational efficiency of the algorithms and ensuring efficient data processing pipelines.
10. **Visualization and reporting:** Provide visualizations and reports to present the results of the fraud detection system. This includes visualizing the classification outcomes, highlighting suspicious transactions, and generating comprehensive reports for stakeholders.
11. **Integration with existing systems:** Design the project to integrate seamlessly with existing card payment systems or fraud detection systems, ensuring smooth deployment and operation within the existing infrastructure.

3. PROTOTYPE EVALUATION:

In credit card fraud detection, we frequently deals with highly imbalanced datasets. For the chosen dataset from Kaggle, we shows that our proposed algorithms are able to detect fraud transactions with very high accuracy and low false positive rate. Hence for better performance, our result shows that classification of algorithms done by preprocessing data rather than raw data. Because of applying preprocessing technique and K-Means algorithm on the dataset, output of algorithms is with high accuracy and give best results. Hence comparison was done and it was concluded that K-Nearest Neighbour gives the best results. This was established using accuracy, precision and recall. Balancing of dataset and feature selection is important in achieving significant results. In future, to enhance the system, other machine learning algorithms or artificial neural networks approaches can be used to detect frauds in credit card.

In this study, we created a unique fraud detection technique that groups clients according to their transactions. and analyse behaviour to create a profile for each cardholder. Following the application of various classifiers to three distinct groups, rating scores are produced for each type of classifier. The system adapts as a result of these dynamic changes in the parameters. Prompt response to new cardholder's transactional behaviours. A feedback system is then used to address the issue of notion drift. We The Matthews Correlation Coefficient was shown to be the superior metric for handling imbalance datasets. It wasn't only MCC. solution. We attempted to balance the dataset by using SMOTE and discovered that the classifiers were performing better than before. The use of one-class classifiers, such as one-class SVM, is an alternative method for addressing imbalance datasets. Finally, we found that the algorithms that produced the best outcomes were random forest, decision tree, and logistic regression.

4. CONCLUSION:

Payment card fraud is a massive problem for the Banking sector. Hence, an effective fraud detection system for card payments is needed by any bank or financial institution to reduce the damages caused by fraudulent activities. In this research, we assumed that deviation from the normal behaviour of the cardholder could serve as the basis for fraud detection. Our experiments showed that the calculation of the similarity between existing transactions in a cardholder’s profile and test transactions could be used for the efficient detection of payment card frauds. Moreover, our results showed that recent transactions exert considerable influence on evaluations of transactions as fraudulent or legal. We also realized that external causes such as a change in income and lifestyle of a cardholder might change cardholders’ spending habits over time. Tree ensembles (such as dynamic random forests, random forests, gradient boosted trees, etc.) learn signals from both classes due to their hierarchical structure and have become very popular in solving problems with imbalanced data such as payment card fraud detection. Also, the use of DRF algorithm is appropriate when there are many input features, and it is known as an accurate and fast learning algorithm that runs efficiently on large datasets. We used DRF to re-examine suspicious transactions and to prevent the occurrence of false positives. Our research results confirmed the effectiveness of DRF in payment card fraud detection.

5. FUTURE SCOPE:

Future enhancements for a cost-sensitive card fraud detection system based on Random Forest and k-nearest neighbour (KNN) algorithms can focus on improving various aspects of the system's performance, functionality, and adaptability. Here are some potential areas for future enhancement:

Feature Engineering:

Explore advanced feature engineering techniques to enhance the representation of transaction data, including temporal features, aggregated features, and behavioural patterns.

Incorporate external data sources, such as customer information, geographical data, or device information, to augment the fraud detection capabilities.

Ensemble Methods:

Investigate the use of ensemble methods, such as stacking or boosting, to combine the predictions of multiple Random Forest and KNN models.

Ensemble methods can potentially improve the overall performance by leveraging the strengths of different models and reducing individual model biases.

Deep Learning Approaches:

Explore the use of deep learning models, such as deep neural networks or recurrent neural networks, to capture complex patterns and dependencies in transaction data. Deep learning models have shown promising results in various domains and may provide additional insights and accuracy for fraud detection.

Online Learning and Adaptive Systems:

Develop an online learning framework that can continuously update and adapt the fraud detection models as new data arrives.

Implement adaptive systems that can dynamically adjust the model's parameters and thresholds based on evolving fraud patterns and changing business requirements.

Explainability and Interpretability:

Enhance the system's explainability and interpretability by incorporating techniques such as feature importance analysis, model interpretability algorithms, or generating human-readable explanations for the system's decisions.

This can help build trust and understanding in the system's fraud detection capabilities.

6. REFERENCES

Here is a bibliography of references related to the design of a cost-sensitive card fraud detection system based on Random Forest and k-nearest neighbour (KNN) algorithms:

[1] Bhasin, S., Kaur, H., & Malhotra, R. (2020). Cost-sensitive Random Forest algorithm for credit card fraud detection. Expert Systems with Applications, 143, 113053.

[2] Raza, A., Shafiq, M. Z., & Hafeez, A. (2020). A cost-sensitive ensemble approach for credit card fraud detection. Future Generation Computer Systems, 108, 876889.

[3] Alnazzawi, M., & Alani, M. (2020). An efficient credit card fraud detection system using k-nearest neighbour algorithm. Proceedings of the 2020 International Conference on Computer Science and Software Engineering, 53-58.

[4] Perera, P., & Bandara, K. (2021). Cost-sensitive fraud detection in credit card transactions using machine learning. In 2021 Moratuwa Engineering Research Conference (MERCon), 280-285.

[5] Li, Y., & Zhang, X. (2018). A cost-sensitive k-nearest neighbour classification algorithm for credit card fraud detection. In 2018 International Conference on Identification, Information and Knowledge in the Internet of Things (IIKI), 86-89.

[6] Ghalwash, M. F., & Khan, M. S. (2021). Improving cost-sensitive credit card fraud detection using random forest and hybrid sampling technique. Journal of Ambient Intelligence and Humanized Computing, 12(5), 6173-6187.

[7] Yu, X., & Zhang, Y. (2019). An efficient cost-sensitive credit card fraud detection model based on KNN algorithm. In 2019 4th International Conference on Mechanics,

Materials and Structural Engineering (ICMMSE), 144-148.

[8] Li, B., Han, F., & Yu, L. (2018). Credit card fraud detection using cost-sensitive random forests and AdaBoost. In 2018 14th International Conference on Natural

Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD), 10331037.

[9] Chen, Y., & Zhang, X. (2020). Cost-sensitive K-nearest neighbour method for credit card fraud detection. In 2020 International Conference on Artificial Intelligence in Information and Communication (ICAIIC).