Credit Card Fraud Detection Literature review

Author: Harshita Loomba

PRN: 23070243021 **Email: 23070213021@sig.ac.in**

Domain: Application of Data Science in Financial Systems

Topic: Credit Card Fraud Detection

Problem Statement: Because illegal operations constantly change, credit card fraud remains a significant threat in the banking sector. Despite the potential of machine learning-based fraud detection systems for overcoming these difficulties, more research is required to fully understand their efficacy, sensitivity, flexibility, and cost-effectiveness. The RaKShA model offers an exciting remedy and is renowned for its unique qualities and possible uses beyond detecting credit card fraud. This review of the literature aims to discuss the essential working, advantages, and limitations of the RaKShA model in credit card fraud, contrasting various machine learning approaches, analysing the unique characteristics of the RaKShA model, evaluating its potential for broader applications, and identifying any gaps or limitations in the existing literature. This literature review will ultimately offer insights into the effectiveness and suitability of the RaKShA model as a credit card fraud detection solution.

Research Question:

- How do machine learning-based fraud detection strategies, with their various approaches, address the evolving challenges of credit card fraud?
- What distinguishes the RaKShA model's sensitivity, adaptability, and cost-effectiveness, and how does it compare to other models in addressing credit card fraud?
- How does the RaKShA model's potential for application in areas like tax evasion impact its suitability as a credit card fraud detection solution?

Objective:

- Review the literature to provide an overview of the evolving challenges and trends in credit card fraud.
- Identify and categorise the various machine learning-based approaches and strategies for detecting credit card fraud.
- Analyse the effectiveness of different machine learning models in addressing specific challenges of credit card fraud, such as detecting advanced fraud techniques.
- Conduct a comprehensive literature review to understand the key attributes of the RaKShA model in credit card fraud detection.
- Compare the sensitivity of the RaKShA model with other machine learning models commonly used in credit card fraud detection.

Literature Review:

Both people and businesses face severe financial risk from credit card theft. Traditional fraud detection systems frequently cost money, take time, and have difficulty keeping up with the fraudsters' constantly evolving strategies. This study evaluates the efficacy of various categorisation, data preparation, and outlier detection strategies in detecting credit card fraud. The findings show that

the RaKShA model, which combines Long Short-Term Memory (LSTM) networks with Explainable Artificial Intelligence (XAI) approaches, provides a sensitive, flexible, and affordable solution. Our research suggests that credit card fraud detection systems might be improved by using the RaKShA model, which might result in significant cost savings for organizations and people and lessen the impact of credit card theft on society.

Below, we are inserting a table comparing different works we went through to make a comparative study.

S.No.	Author	Dataset Used	Problems and challenges	Algorithims Tested	Conclusion
1	S P Maniraj	Kaggle Dataset	Class Imbalance, Dynamic and massive Data Processing	Isolation Forest Algorithim	Isolation Forest algorithim Performs much better than the local outlier factor
2	Ong Shu Yee et al.	Dummy Data Set created based on behavioural studies	Sensitive Nature of the Data(Dummy Data creation based on behavioural studies using WEKA Tool)	True Augmented naïve Bayes	True Augmented naïve Bayes perform better in unprocessed data as well as processed data.
3	Kuldeep Randhawa et al.	benchmark data, Real World Data	Real-world credit card fraud detection data is very complex and highly variable, class imbalance	Majority Voting + Ada Boost	The majority voting combination method performed best overall with the highest MCC scores in the study.
4	Haichao Du et al.	confidential	Class Imbalance	AED-LGB	AED-LGB algorithm has higher performance in ACC, MCC indexes
5	Emilija Strelcenia et al.	Credit Card Fraud Detection	Class Imbalance	KC-GAN	The results suggest that K-CGAN, among other resampling methods, is effective at improving the classification of credit card fraud. This approach has the potential to capture the underlying structure and features of the data in a high-dimensional space.
6	Jay Raval et al.	financial resource is collected from the user entity	Enhancing Financial Ecosystem via Efficient Fraud Detection Portal.	X-LSTM	X-LSTM, enhances the power of the model in detecting credit card fraud, making the scheme scalable and adaptable to various situations. After 500 epochs, there was a 17.41% improvement over a standard LSTM model.
7	Ibomoiye Domor et al.	European cardholders dataset	presence of redundant and irrelevant features in most real-world credit card data degrades the performance of ML classifiers.	GA-ELM	GA-ELM method excelled, validated on two credit card datasets, proving robustness.
8	Zhaorui Meng et al.	Kaggle Data Set	noise expansion, over fitting ,class imbalance,	GAN-TabularAttention	Proposed GAN-TabularAttention excels in credit card fraud detection on imbalanced datasets.
9	Emanuel Mineda Carneiro et al.	Kaggle Dataset	High Cardinality and Scalability	VCCA+FNN	this approach can enhance detection quality while keeping costs manageable.
10	Bharat Kumar Padhi et al.	not provided	High Cardinality and Scalability	RHSOFS	RHSOFS improved accuracy but may cause data loss by removing noise.

S.P. Maniraj's project on outlier detection using Local Outlier Factor (LOF) and Isolation Forest (I Forest) algorithms is a well-designed and informative study. It addresses important issues such as class imbalance, dynamic and massive data processing, and standardisation of time and amount columns. The project's key finding is that I Forest outperforms LOF significantly in evaluation metrics such as AUC, precision, and recall. This finding is significant, as it suggests that I Forest is a more effective algorithm for outlier detection in credit card fraud. The project's methodology is sound, and the results are well-presented. The authors have also discussed the limitations of their study and suggested directions for future research. Overall, S.P. Maniraj's project is a valuable contribution to credit card fraud detection. It provides empirical evidence that I Forest is a superior algorithm for outlier detection and suggests future research directions.

Ong Shu Yee and colleagues' study comparing algorithm performance on pre-processed and unprocessed data is a well-designed and informative study. It addresses an important issue in fraud detection: the impact of data preprocessing on the performance of fraud detection algorithms. The study's methodology is sound. The authors used a synthetic dataset generated based on behavioural studies, which allowed them to control the factors that could affect the performance of the algorithms. The authors also used various classifiers, allowing them to compare the performance of different algorithms. The study's key finding is that the Tree Augmented Naïve Bayes classifier performed better on unprocessed and pre-processed data. This finding is significant, as it suggests that Tree Augmented Naïve Bayes is a robust algorithm that can be used for fraud detection even when the data is not pre-processed. The authors also found that data preprocessing can improve the performance of fraud detection algorithms in some cases. For example, they found that PCA preprocessing improved the performance of the logistic and J48 classifiers. This finding is also significant, as it suggests that data preprocessing can be used to improve the performance of various fraud detection algorithms. Overall, Ong Shu Yee and colleagues' study is valuable to fraud detection. It provides empirical evidence that Tree Augmented Naïve Bayes is a robust algorithm for fraud detection, even when the data is not pre-processed. The study also suggests that data preprocessing can improve the performance of some fraud detection algorithms.

Kuldeep Randhawa and colleagues' study comparing individual algorithms against a hybrid credit card fraud detection model is a well-designed and informative study. It addresses an important issue in fraud detection: the class imbalance problem. The study's methodology is sound. The authors used both benchmark data and real-world data, which allowed them to assess the performance of the algorithms on a variety of different data sets. The authors also used various algorithms, including individual and hybrid models. The study's key finding is that the majority voting combination method demonstrated superior performance, achieving the highest MCC scores across the study. This finding is significant, as hybrid models can be more effective for fraud detection than individual algorithms. The authors also found that under-sampling techniques can effectively mitigate class imbalance in fraud detection data. This finding is also significant, as it suggests that there are practical ways to address the class imbalance problem in fraud detection. Overall, Kuldeep Randhawa and colleagues' study is valuable to fraud detection. It provides empirical evidence that hybrid models can be more effective for fraud detection than individual algorithms, and it suggests that under-sampling techniques can effectively mitigate class imbalance in fraud detection data.

Haichao Du and colleagues' study on the evaluation and comparison of AutoEncoder and LightGBM for credit card fraud detection is well-designed and informative. It addresses an important issue in fraud detection: the class imbalance problem. The study's methodology is sound. The authors used

various algorithms, including AutoEncoder, LightGBM, and other undisclosed algorithms. The authors also used a variety of evaluation metrics, including accuracy and MCC. The study's key finding is that the AED-LGB algorithm demonstrated superior performance in terms of accuracy and MCC compared to the other algorithms under consideration. This finding is significant, as it suggests that hybrid models incorporating deep learning algorithms can be more effective for fraud detection than individual models. The authors also found that SMOTE can effectively mitigate class imbalance in fraud detection data. This finding is also significant, as it suggests that there are practical ways to address the class imbalance problem in fraud detection. Overall, Haichao Du and colleagues' study is valuable to fraud detection. It provides empirical evidence that hybrid models incorporating deep learning algorithms can be more effective for fraud detection than individual models. It also suggests that SMOTE can effectively mitigate class imbalance in fraud detection data.

Emilija Strelcenia and her team's study on developing K-CGAN for addressing data imbalance in credit card fraud detection is well-designed and informative. It addresses an important issue in fraud detection: the class imbalance problem. The study's methodology is sound. The authors used a variety of different oversampling techniques, including SMOTE, ADASYN, B-SMOTE, CGAN, Vanilla GAN, WS GAN, SDG GAN, NS GAN, LS GAN, and their novel approach, K-CGAN. The authors also used a variety of evaluation metrics, including accuracy and MCC. The study's key finding is that K-CGAN outperformed the other oversampling techniques regarding accuracy and MCC. This finding is significant, as it suggests that K-CGAN is a promising technique for addressing class imbalance in credit card fraud detection. The authors also found that K-CGAN enhanced credit card fraud classification by capturing the underlying structure and features of the data in a high-dimensional space. This finding is also significant, as it suggests that K-CGAN can learn more meaningful representations of the data than other oversampling techniques. Overall, Emilija Strelcenia and her team's study is valuable to credit card fraud detection. It provides empirical evidence that K-CGAN is a promising technique for addressing class imbalance and improving classification results.

Jay Raval and colleagues' study on enhancing credit card fraud detection with LSTM, XAI, and blockchain is well-designed and informative. It addresses an important issue in fraud detection: the need for more effective and accurate methods. The study's methodology is sound. The authors used various techniques, including data preprocessing, LSTM, XAI, and blockchain. The authors also used a variety of evaluation metrics, including accuracy and MCC. The study's key finding is that the X-LSTM model outperformed other fraud detection models regarding accuracy and MCC. This finding is significant, suggesting that the X-LSTM model is promising for improving credit card fraud detection. The authors also found that blockchain technology can store fraud detection results and ensure data integrity. This finding is also significant, as it suggests that blockchain technology can be used to improve the security and reliability of fraud detection systems. Overall, Jay Raval and colleagues' study is valuable to credit card fraud detection. It provides empirical evidence that the X-LSTM model is a promising approach for improving fraud detection accuracy. The study also suggests that blockchain technology can improve the security and reliability of fraud detection systems.

Ibomoiye Domor and their team's study on advancing credit card fraud detection through machine learning and feature selection techniques is well-designed and informative. It addresses an important issue in fraud detection: the need to remove redundant and irrelevant features from credit card data. The study's methodology is sound. The authors used various techniques, including data preprocessing, feature selection, and machine learning. The authors also used a variety of evaluation metrics,

including accuracy and MCC. The study's key finding is that the GA-ELM method outperformed other fraud detection methods regarding accuracy and MCC. This finding is significant, suggesting that the GA-ELM method is promising for improving credit card fraud detection. The authors also found that the GA-ELM method is robust and effective, as it achieved good results on two different credit card datasets. This finding is also significant, as it suggests that the GA-ELM method is a general-purpose approach that can improve fraud detection on various datasets. Ibomoiye Domor and their team's study contributes to credit card fraud detection. It provides empirical evidence that the GA-ELM method is promising for improving fraud detection accuracy. The study also suggests that the GA-ELM method is a robust and practical approach that can be used to improve fraud detection on various datasets. Overall, the study is well-written and informative. It provides a valuable contribution to the field of credit card fraud detection. We are particularly interested in the development of the GA-ELM method. It is exciting to see how GA and ELM can be combined to improve the performance of fraud detection systems. We are also interested in the potential of the GA-ELM method to improve fraud detection on various datasets, such as credit card, medical, and financial data. We believe that the GA-ELM method has the potential to be a powerful tool for improving fraud detection in a variety of different industries.

Zhaorui Meng and their team's study on enhancing fraud detection through synthetic data generation and a Multi-Head Attention mechanism is well-designed and informative. It addresses crucial issues in fraud detection, such as noise expansion, overfitting, and class imbalance. The study's methodology is sound. The authors used various techniques, including GANs, attention mechanisms, SMOTE, and ADASYN. The authors also used a variety of evaluation metrics, including accuracy and MCC. The study's key finding is that the GAN-Tabular Attention approach outperformed other fraud detection methods on imbalanced datasets. This finding is significant, as it suggests that the GAN-Tabular Attention approach is promising for improving fraud detection on imbalanced datasets. The authors also found that the GAN-Tabular Attention approach could reduce noise expansion and overfitting. This finding is also significant, as it suggests that the GAN-Tabular Attention approach can improve the overall performance of fraud detection models. Overall, Zhaorui Meng and their team's study is a valuable contribution to the field of fraud detection. It provides empirical evidence that the GAN-Tabular Attention approach is promising for improving fraud detection on imbalanced datasets. The study also suggests that the GAN-Tabular Attention approach can reduce noise expansion and overfitting.

Emanuel Mineda Carneiro and their team's study on enhancing credit card fraud detection through deep learning techniques for high-cardinality categorical attributes is well-designed and informative. It addresses an important issue in fraud detection: the challenge of handling high-cardinality categorical attributes. The study's methodology is sound. The authors used various techniques, including deep learning, value clustering, and feedforward neural networks. The authors also used a variety of evaluation metrics, including accuracy and MCC. The study's key finding is that the proposed approach outperformed other fraud detection methods on high-cardinality categorical attributes. This finding is significant, as it suggests that the proposed approach is promising for improving fraud detection on datasets with high-cardinality categorical attributes. The authors also found that the proposed approach managed costs effectively. This finding is also significant, as it suggests that the proposed approach can be used to develop fraud detection systems that are both effective and affordable. Emanuel Mineda Carneiro and their team's study is valuable to fraud detection. It provides empirical evidence that the proposed approach is promising for improving fraud detection on datasets with high-cardinality categorical attributes. The study also suggests that the

proposed approach can be used to develop fraud detection systems that are both effective and affordable.

Bharat Kumar Padhi and their research team's study on enhancing credit card fraud detection through a novel feature selection algorithm, RHSOFS, is well-designed and informative. It addresses an important issue in fraud detection: the challenge of handling high cardinality and scalability. The study's methodology is sound. The authors used various techniques, including RHSOFS, Naive Bayes, K-Nearest Neighbours, Support Vector Machines, and Decision Trees. The authors also used a variety of evaluation metrics, including accuracy and MCC. The study's key finding is that RHSOFS outperformed other feature selection algorithms on high-cardinality and scalable datasets. This finding is significant, as it suggests that RHSOFS is a promising approach for improving fraud detection on datasets with these challenges. The authors also found that RHSOFS improved the accuracy of fraud detection models. This finding is also significant, as it suggests that RHSOFS can be used to develop fraud detection models that are more accurate. Overall, Bharat Kumar Padhi and their research team's study is a valuable contribution to the field of fraud detection. It provides empirical evidence that RHSOFS is a promising approach for improving fraud detection on datasets with high cardinality and scalability. The study also suggests that RHSOFS can be used to develop fraud detection models that are more accurate.

The AED-LGB deep learning algorithm uses Extreme Gradient Boosting (XGBoost) and Attention Mechanism. It is a relatively new algorithm but has shown promising credit card fraud detection results. The True Augmented Naïve Bayes algorithm is based on the Naive Bayes classifier. It is a robust algorithm that can detect fraud in various datasets. The Majority Voting + Ada Boost ensemble is a hybrid model combining the strengths of two algorithms, majority voting and Ada Boost. It is a robust algorithm that can achieve high accuracy in fraud detection. The GAN-Tabular Attention algorithm is a novel algorithm that combines the strengths of Generative Adversarial Networks (GANs) and Attention Mechanism. It is a promising algorithm that can improve the performance of fraud detection systems. The Isolation Forest algorithm and the VCCA+FNN algorithm are two algorithms that showed relatively lower performance in this evaluation. The Isolation Forest algorithm is a tree-based algorithm used for outlier detection. It is not explicitly designed for fraud detection, so it may not be as effective as other algorithms designed for this task. The VCCA+FNN algorithm is a hybrid model that combines the strengths of Value Clustering for Categorical Attributes (VCCA) and Feedforward Neural Networks (FNN). It is a promising algorithm but may not be as effective as other algorithms designed explicitly for fraud detection. Overall, the table shows that there are a variety of algorithms that can be used for credit card fraud detection. The best algorithm will depend on the specific dataset and the requirements of the fraud detection system. However, the algorithms mentioned in the table are all promising algorithms that can be used to improve the accuracy of fraud detection systems.

REFERENCES:

1. https://www.researchgate.net/publication/336800562_Credit_Card_Fraud_Detection_using_Machine_Learning_and_Data_Science
2. https://www.researchgate.net/publication/326986162_Credit_Card_Fraud_Detection_Using_Machine_Learning_As_Data_Mining_Technique
3. https://ieeexplore.ieee.org/abstract/document/8292883

4. https://eds.s.ebscohost.com/eds/pdfviewer/pdfviewer?vid=0&sid=ddfa59a6-bd5a-40e0-b369-d2aaa5d33756%40redis
5. https://eds.s.ebscohost.com/eds/pdfviewer/pdfviewer?vid=0&sid=a26882a5-7dd4-4fc6-b1df-28af3381cadd%40redis

6. https://eds.s.ebscohost.com/eds/pdfviewer/pdfviewer?vid=0&sid=12beca5d-af14-4d46-976d-f0699a4e4662%40redis

7. https://eds.s.ebscohost.com/eds/pdfviewer/pdfviewer?vid=0&sid=2090d938-3bc1-4136-a2b1-90074e8558b3%40redis

8. https://eds.s.ebscohost.com/eds/pdfviewer/pdfviewer?vid=0&sid=ef15c8c3-c3a0-4bc6-80bd-a2e9ae605210%40redis

9. https://eds.s.ebscohost.com/eds/pdfviewer/pdfviewer?vid=0&sid=d63e0135-8cdf-4c7a-a521-bbc62c11b594%40redis

10. https://eds.s.ebscohost.com/eds/pdfviewer/pdfviewer?vid=0&sid=591736c0-bcf2-4691-9a43-dad53366cdd1%40redis