A Review on Machine Learning and Deep Learning Models for Credit Card Fraud Detection

Raj Kumar Ahirwar[1], Ms. Pooja Hardiya[2]
Research Scholar[1], Assistant Professor[2]
Department of CSE, SDBCT, Indore, India[1,2]

***Abstract*—** With increased internet usage, online transactions have been on the rise. One of the most prevalent problems faced is credit cards frauds. While web applications and mailing services are heavily spammed, the upsurge of handheld mobile devices has led to an outburst of heavy mobile credit card spamming. The matter is more severe in mobile devices due to lesser sophisticated filtering mechanisms in built in mobile operating systems. Recent advancements in electronic commerce and communication systems have significantly increased the use of credit cards for both online and regular transactions. However, there has been a steady rise in fraudulent credit card transactions, costing financial companies huge losses every year. The development of effective fraud detection algorithms is vital in minimizing these losses, but it is challenging because most credit card datasets are highly imbalanced. Traditional rule-based systems are often insufficient to handle the sophisticated and evolving techniques fraudsters use. Machine learning (ML) provides more dynamic, scalable, and effective methods for detecting fraudulent activities. This paper presents a comprehensive review on credit card datasets, imbalanced nature of datasets and existing baseline techniques in the domain.

Keywords*— *Credit Card Fraud Detection, Machine Learning, Feature Selection, Imbalanced Datasets, Classification Accuracy

I. INTRODUCTION

With increasing digitization, card usage has resulted in a continuous rise in fraudulent transactions. Rule-based filters operate based on predetermined rules, making them less suitable for some situations. The increasing prevalence of digital transactions and online commerce has provided convenience to consumers globally in recent years. Nevertheless, this technological revolution has also led to the emergence of advanced types of deception, especially in the domain of credit card transactions. Scammers always develop new strategies to avoid being detected by conventional approaches. Financial institutions face a significant problem in detecting fraudulent actions in real-time. Deep learning, a subfield of artificial intelligence, has emerged as a highly promising method for addressing the intricacies of credit card fraud detection. By utilizing sophisticated neural network structures, deep learning models possess the ability to analyze extensive volumes of transaction data, detect complex patterns, and accurately identify fraudulent behaviors.

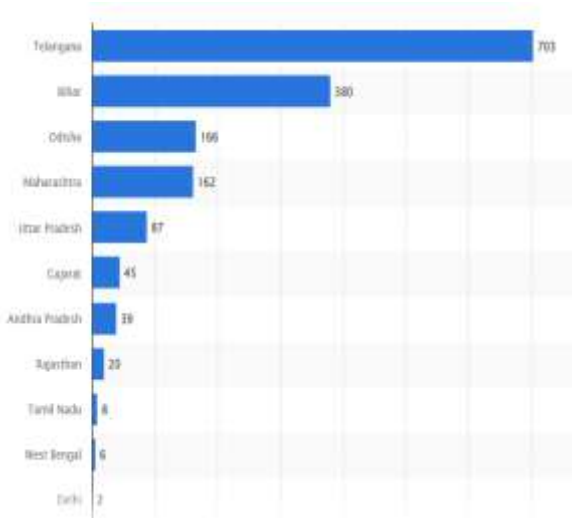

Fig.1 Credit Card Frauds in India (State Wise) (Source: Statista)

https://www.statista.com/statistics/1097927/india-number-of-credit-debit-card-fraud-incidents-by-leading-state/

Machine learning (ML) algorithms have the ability to analyze vast amounts of transaction data in real time, identifying patterns and anomalies that may indicate fraudulent activity. Unlike rule-based systems, which rely on predefined criteria, ML models can learn from historical data and adapt to new types of fraud as they emerge. This dynamic learning capability enhances the accuracy and effectiveness of fraud detection systems, reducing false positives and enabling quicker responses to potential threats.

II. MACHINE LEARNING MODELS FOR IDENTIFYING CREDIT CARD FRAUDS

Various machine learning algorithms are employed to detect credit card fraud, each with its unique strengths. Supervised learning algorithms, such as decision trees and support vector

machines, are trained on labeled datasets where examples of both fraudulent and non-fraudulent transactions are provided. This training enables the models to classify new transactions with a high degree of accuracy. On the other hand, unsupervised learning algorithms, like clustering and anomaly detection methods, do not require labeled data and can identify outliers in transaction data that may represent fraud. These algorithms are particularly useful for detecting novel fraud patterns that have not been previously encountered.

Decision trees: Decision trees are commonly used for fraud detection since they are straightforward and easy to understand. They operate by partitioning the dataset into subsets according to the input feature values, resulting in a hierarchical structure of decision nodes. Every node in the representation reflects a specific feature, each branch represents a decision rule, and each leaf represents an outcome. Decision trees possess the capability to process both numerical and categorical input, rendering them adaptable and comprehensible.

Random Forests: Random forests use the idea of decision trees by employing a collection of numerous trees to enhance accuracy and resilience. The construction of each tree in a random forest involves using a random subset of the data, which helps to reduce overfitting and improve prediction performance. Random forests excel at detecting intricate fraud patterns in extensive datasets, providing exceptional accuracy and robustness against interference.

Logistic regression: Logistic regression is a statistical model that is specifically designed for binary classification tasks, allowing it to effectively differentiate between fraudulent and non-fraudulent transactions. The logistic function is used to evaluate the probability of a given input belonging to a specific class. Logistic regression is renowned for its simplicity, efficiency, and interpretability. It is particularly useful in cases where the relationships between features may be approximated as linear.

Support Vector Machines (SVM): Support vector machines (SVM) are robust classifiers that identify the most effective hyperplane for distinguishing between various classes in a space with many dimensions. Support Vector Machines (SVMs) are highly efficient in dealing with data that has a large number of dimensions. They are particularly valuable when the classes cannot be separated by a straight line, as they can employ kernel functions to transform inputs into spaces with even more dimensions. The capacity of SVMs to detect fraud makes them a highly advantageous option.

Artificial neural networks: Neural networks, particularly deep learning models, have become popular due to their capacity to acquire intricate patterns from extensive datasets. Neural networks are capable of representing complex connections between characteristics in fraud detection, enabling them to detect tiny deviations that are indicative of fraudulent activity. Convolutional neural networks (CNNs) and recurrent neural networks (RNNs) are utilized depending on the characteristics of the input and the specific demands of the task.

K-Nearest Neighbors (KNN): The K-nearest neighbors (KNN) technique is an instance-based learning method utilized for categorization. It identifies the 'k' most comparable transactions (neighbors) to a particular transaction. The class that has the highest number of occurrences among the neighbors is allocated to the new transaction. KNN is characterized by its simplicity and intuitiveness, yet it may incur high computing costs when dealing with extensive datasets. However, it is still efficient for smaller datasets and can yield rapid, easily understandable outcomes.

Despite the benefits, implementing machine learning for fraud detection comes with challenges. One major issue is the imbalance in datasets, where fraudulent transactions are significantly outnumbered by legitimate ones. This imbalance can skew the model's performance, making it less effective at identifying fraud. Techniques such as oversampling, undersampling, and synthetic data generation are often employed to address this issue. Additionally, ensuring the privacy and security of transaction data is paramount, as any breaches could have severe consequences for both consumers and financial institutions.

III. EXISTING CHALLENGS OF CLASS IMBALANCE

Class imbalance in the context of credit card fraud detection using deep learning refers to the unequal distribution of fraudulent and non-fraudulent transactions in the dataset. In most real-world scenarios, fraudulent transactions constitute only a tiny fraction of the overall transaction volume, while the majority of transactions are legitimate. This imbalance can pose challenges for machine learning models because they tend to be biased towards the majority class, leading to poor performance in identifying the minority class (fraudulent transactions). The imbalanced nature of the dataset can cause the model to prioritize accuracy at the expense of effectively detecting fraudulent transactions. As a result, the model may tend to classify most transactions as non-fraudulent, achieving high accuracy due to the dominance of the majority class but failing to detect fraudulent activities adequately.

Imbalanced Datasets

Imbalanced datasets pose significant challenges in credit card fraud detection, where the number of legitimate transactions far outweighs the instances of fraud. This imbalance can lead to biased models and hinder the effectiveness of fraud detection systems. Here are several challenges associated with imbalanced datasets in credit card fraud detection:

***Limited Representation of Fraudulent Cases**:* Imbalanced datasets often result in a scarcity of fraudulent transactions for model training. This limited representation makes it challenging for the algorithm to learn the patterns and characteristics of fraudulent activities, leading to a less accurate and robust model.

***Biased Model Performance**:*
Traditional machine learning algorithms are biased towards the majority class, in this case, non-fraudulent transactions. As a result, the model may prioritize accuracy on the majority class while neglecting the minority class (fraudulent transactions). This bias can lead to poor fraud detection performance.

***High False Negative Rates**:*
Imbalanced datasets can contribute to a higher rate of false negatives, where fraudulent transactions are incorrectly classified as non-fraudulent.

Dynamic Nature of Fraud Patterns:
Fraudulent activities evolve over time, and imbalanced datasets may not capture the latest patterns. As fraudsters adapt their tactics, models trained on imbalanced historical data may struggle to generalize to emerging fraud patterns.

Class Imbalance Mitigation
Addressing class imbalance is crucial in credit card fraud detection to ensure that the model can effectively identify fraudulent transactions while minimizing false positives. Various techniques can be employed to mitigate class imbalance, including:

1. Resampling methods: This involves either oversampling the minority class (fraudulent transactions) to balance the class distribution or under sampling the majority class (non-fraudulent transactions) to reduce its dominance.
2. Algorithmic approaches: Some algorithms, such as ensemble methods like Random Forest or boosting algorithms like XGBoost, inherently handle class imbalance by adjusting the training process to give more weight to the minority class.
3. Cost-sensitive learning: Assigning different misclassification costs to different classes during model training to penalize misclassifying fraudulent transactions more severely can help mitigate class imbalance.
4. Synthetic data generation: Generating synthetic samples for the minority class using techniques like SMOTE (Synthetic Minority Over-sampling Technique) can help balance the class distribution and improve model performance.

By addressing class imbalance effectively, deep learning models for credit card fraud detection can achieve better sensitivity and specificity, thereby enhancing their ability to accurately detect fraudulent transactions while minimizing false alarms.

IV. EXISTING WORK

This section presents a review on the baseline approaches in the domain.

Wang et al. proposed an innovative method for identifying credit card fraud by combining the SMOTE-KMEANS technique with an ensemble machine learning model. The proposed model was benchmarked against traditional models such as logistic regression, decision trees, random forests, and support vector machines. Performance was evaluated using metrics, including accuracy, recall, and area under the curve (AUC). The results demonstrated that the proposed model achieved superior performance, with an AUC of 0.96 when combined with the SMOTE-KMEANS algorithm. This indicates a significant improvement in detecting fraudulent transactions while maintaining high precision and recall.

Wang et al. proposed that traditional methods often fail to capture the complex temporal dynamics and heterogeneous relationships inherent in financial transaction networks. To address these limitations, we propose TH-GCL (Temporal Heterogeneous Graph Contrastive Learning), a novel framework that integrates heterogeneous graph modeling, temporal pattern recognition, and contrastive learning for enhanced fraud detection. Our approach constructs a temporal heterogeneous graph incorporating multiple entity types including users, transactions, merchants, and devices, with time-aware edge weights to capture evolving behavioral patterns. We design a temporal-aware graph neural network architecture that learns hierarchical representations by jointly modeling structural dependencies and temporal evolution patterns.

Esenegho et al. proposed an efficient approach to detect credit card fraud using a neural network ensemble classifier and a hybrid data resampling method. The ensemble classifier is obtained using a long short-term memory (LSTM) neural network as the base learner in the adaptive boosting (AdaBoost) technique. Meanwhile, the hybrid resampling is achieved using the synthetic minority oversampling technique and edited nearest neighbor (SMOTE-ENN) method. The effectiveness of the proposed method is demonstrated using publicly available real-world credit card transaction datasets. The performance of the proposed approach is benchmarked against the following algorithms: support vector machine (SVM), multilayer perceptron (MLP), decision tree, traditional AdaBoost, and LSTM. The experimental results show that the classifiers performed better when trained with the resampled data, and the proposed LSTM ensemble outperformed the other algorithms

KS Adewole et al. proposed a unified framework is proposed for both spam message and spam account detection tasks. Authors utilized four datasets in this study, two of which are from SMS spam message domain and the remaining two from Twitter microblog. To identify a minimal number of features for spam account detection on Twitter, this paper studied bio-inspired evolutionary search method. Using evolutionary search algorithm, a compact model for spam account detection is proposed, which is incorporated in the machine learning phase of the unified framework. The results of the various experiments conducted indicate that the proposed framework is promising for detecting both spam message and spam account with a minimal number of features.

Aliaksandr Barushka et al. proposed a technique based on integrated distribution-based balancing approach for spam classification. The concept of deep neural networks is used in this paper. The major advantage of this approach is the distribution mechanism makes the computation of different parameters for classification simpler. Deep learning makes the classification accuracy higher.

Surendra Sedhai et al. proposed a technique that used semi-supervised approach for spam redirection classification mechanism. The concept used the training rules to be governed by supervised learning with an adaptive weight changing mechanism. However, the approach had the liberty of letting the weight adaptation fall into the purview of the training algorithm used.

Chao Chen et al. proposed a technique for the classification of drifted twitter spam based on statistical feature based classification. The major issues addressed in this paper, were the use of statistical features for spam classification. Drifted spam is often the result of several attached web links leading to the drifting mechanism of the tweets in social media applications with malicious URLs that can cause the spamming attacks on the web mails.

Nida Mirza et al. proposed a technique for spam classification based on hybrid feature selection. The major advantage of this approach was the fact that the hybrid parameters can be an amalgamation of both textual features and non-textual features. The evaluation of the performance of the proposed system was done on the basis of mean square error, hit rate and the accuracy. The performance of hybrid feature selection was shown to be better than the average features computation algorithms.

Hammad Afzal et al. in proposed a mechanism for the classification of bi-lingual tweets using machine learning algorithms. The methodology of the system was the use of natural language processing and thereafter the use of deep neural networks with multiple hidden layers. The learning rates were dependent on the differential changes in the architecture of the neural network used

CONCLUSION: Machine learning-based methods offer a significant advantage over traditional rule-based systems in detecting credit card fraud. Techniques ranging from supervised and unsupervised learning to hybrid approaches provide flexibility in handling different types of data and fraud patterns. However, challenges such as class imbalance, evolving fraud techniques, and the need for real-time analysis make fraud detection a complex task. By continually refining models and incorporating new data, machine learning can significantly enhance the ability to detect and prevent fraudulent transactions, helping financial institutions reduce losses and protect customers. This paper presents the salient aspects of machine learning based methods for credit card fraud detection and the existing literature in the domain.

References

[1] Y. Wang, "A Data Balancing and Ensemble Learning Approach for Credit Card Fraud Detection," IEEE Transactions on Dependable And Secure Computing, vol. 3, no. 4, 2025, pp. 386-390

[2] J. Wang, J. Liu, W. Zheng and Y. Ge, "Temporal Heterogeneous Graph Contrastive Learning for Fraud Detection in Credit Card Transactions," in IEEE Access, 2024, vol. 13, pp. 145754-145771

[3] E. Esenogho, I. D. Mienye, T. G. Swart, K. Aruleba and G. Obaido, "A Neural Network Ensemble With Feature Engineering for Improved Credit Card Fraud Detection," in IEEE Access, vol. 10, pp. 16400-16407, 2023

[4] KS Adewole, NB Anuar, A Kamsin, "SMSAD: a framework for spam message and spam account detection", Journal of Multimedia Tools and Applications, Springer 2022, vol. 78, pp. 78, 3925–3960.

[5] Aliaksandr Barushka, Petr Hajek, "Spam filtering using integrated distribution-based balancing approach and regularized deep neural networks", Springer 2020

[6] Surendra Sedhai, Aixin Sun, "Semi-Supervised Spam Detection in Twitter Stream", IEEE 2019

[7] Chao Chen, Yu Wang, Jun Zhang, Yang Xiang, Wanlei Zhou, Geyong Min, "Statistical Features-Based Real-Time Detection of Drifted Twitter Spam", IEEE 2018

[8] Nida Mirza, Balkrishna Patil ,Tabinda Mirza ,Rajesh Auti, "Evaluating efficiency of classifier for email spam detector using hybrid feature selection approaches",IEEE 2017

[9] Hammad Afzal ,Kashif Mehmood, "Spam filtering of bi-lingual tweets using machine learning",IEEE 2016

[10] Hailu Xu ,Weiqing Sun ,Ahmad Javaid," Efficient spam detection across Online Social Networks", IEEE 2016

[11] Nadir Omer Fadl Elssied,Othman Ibrahim ,Ahmed Hamza Osman," Enhancement of spam detection mechanism based on hybrid kkkk-mean clustering and support vector machine",SPRINGER 2015

[12] Tarjani Vyas , Payal Prajapati , Somil Gadhwal," A survey and evaluation of supervised machine learning techniques for spam e-mail filtering",IEEE 2015
[13] Nishtha Jatana ,Kapil Sharma," Bayesian spam classification: Time efficient radix encoded fragmented database approach", IEEE 2014
[14] Kamalanathan Kandasamy ,Preethi Koroth," An integrated approach to spam classification on Twitter using URL analysis, natural language processing and machine learning techniques", IEEE 2014
[15] Navneel Prasad ,Rajeshni Singh ,Sunil Pranit Lal," Comparison of Back Propagation and Resilient Propagation Algorithm for Spam Classification",IEEE 2013
[16] Wojciech IndykEmail author, Tomasz Kajdanowicz, Przemyslaw Kazienko,Slawomir Plamowski," Web Spam Detection Using MapReduce Approach to Collective Classification", SPRINGER 2013
[17] Ashwin Rajadesingan, Anand Mahendran," Comment Spam Classification in Blogs through Comment Analysis and Comment-Blog Post Relationships", Springer 2016.
[18] Lauret, P., Fock, E., Randrianarivonyh, R.N., Manicom-Ramsamy, J.-F. (2008), Bayesian neural network approach to short time load forecasting. Energy Convers. Managagement, vol.49, no.5. pp.1156–1166.
[19] E. Esenogho, I. D. Mienye, T. G. Swart, K. Aruleba and G. Obaido, "A Neural Network Ensemble With Feature Engineering for Improved Credit Card Fraud Detection," in IEEE Access, vol. 10, pp. 16400-16407, 2022