International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Review on Credit Card Fraud Detection

Hiba Fathima K P[1], Riji R[2]

[1]Dept.Computer Science and Engineering hibafathimakp2019@gmail.com
[2]Dept. Computer Science and Engineering rijir.kmr@gmail.com

DOI : https://doi.org/10.55248/gengpi.6.0225.0964

ABSTRACT—

In recent years, credit card theft has become a major concern, presenting financial hazards to consumers, businesses, and financial institutions. All parties concerned have suffered significant financial losses as a result of the unintentional rise in fraudulent activity brought on by the widespread use of credit cards. Machine learning has emerged as one of the most successful methods for identifying and stopping fraudulent transactions, which helps to lessen this increasing difficulty. An extensive analysis of the many machine learning methods used in credit card fraud detection is given in this research. It assesses these techniques using important performance indicators like specificity, accuracy, and precision. The study intends to determine the most effective strategies for improving fraud detec- tion systems, guaranteeing increased security and dependability in financial transactions, by evaluating and contrasting these metrics.

INDEX TERMS: Credit Card Fraud, Machine Learning, SMOTE, Variational autoencoding.

I. INTRODUCTION

The use of credit cards has grown dramatically over the last ten years due to the explosive growth of e-commerce, which has also resulted in an increase in fraudulent transactions. The financial sector has suffered billion-dollar losses as a result of this tendency. Machine learning (ML) has been widely used for credit card fraud detection in an effort to reduce these losses. However, the efficacy of conventional ML algorithms is hampered by issues like class imbalance in datasets, where legal transactions greatly outnumber fraudulent ones. The reliability of the models is limited because the skewed data distribution frequently leads to biased models that incorrectly categorize minority class samples.

Researchers have looked into data-level, algorithm-level, and hybrid approaches to address the problem of class im- balance. While algorithm-level techniques alter classifiers to give preference to the minority class, data-level techniques concentrate on balancing datasets through undersampling or oversampling. To get superior results, hybrid procedures inte- grate both approaches. Examples include Taha and Malebary's LightGBM approach with Bayesian hyperparameter optimiza- tion, which obtained high accuracy and precision, and Padmaja et al.'s method, which uses k-reverse nearest neighbor (KRNN) for outlier elimination and hybrid resampling.
Deep learning techniques have demonstrated potential in modeling sequential data for fraud detection, especially re- current neural networks (RNNs) like long short-term mem- ory (LSTM). LSTMs improve detection accuracy by better adapting to dynamic shopping trends than traditional ML algorithms. In this work, LSTM and the adaptive boosting (Ad- aBoost) algorithm are integrated to create a reliable credit card fraud detection system. The method addresses data imbalance by using SMOTE-ENN for feature engineering. By efficiently modeling sequential data and lowering false positives, the LSTM-AdaBoost ensemble improves detection and provides a more precise and flexible fraud protection solution.

II. RELATED WORK

A. Credit Card Fraud Detection Based on CatBoost And Deep Neural Network

Rooted in the family of Gradient Boosted Decision Trees (GBDTs), CatBoost has rapidly become a preferred algorithm for supervised classification tasks due to its ability to effec- tively address statistical challenges faced by existing state- of-the-art GBDT implementations. CatBoost tackles a critical issue termed prediction shift. This phenomenon occurs when there is a difference in the distribution of predictions, $F(x_k)$, for a training example x_k, and $F(x)$, for a test example x.

The authors identified this issue through the hypothesis that a dataset $D = \{(x_k, y_k)\}^n$ exists, where $x_k = (x^1, \dots, x^m)$ represents a random vector of m features, and $y_k \in \mathrm{R}$ is a binary target variable. The samples (x_k, y_k) are assumed to be independently and identically distributed according to a probability distribution $P(\cdot)$. The learning objective is to train a function $H : \mathrm{R}^m \rightarrow \mathrm{R}$ that minimizes the expected loss $L(F) = \mathrm{E}[L(y, F(x))]$, where $L(\cdot)$ is a smooth loss function and (x, y) represents the test data sampled from the training data D.

Gradient boosting constructs a sequence of approximations $F_t : \mathrm{R}^m \rightarrow \mathrm{R}$ iteratively in a greedy manner. At each iteration, the model F_t is derived from the previous approximation F_{t-1} using an additive process $F_t = F_{t-1} + \eta h_t$. Here, η denotes the step size, and $h_t : \mathrm{R}^m \rightarrow \mathrm{R}$ is the base predictor chosen to minimize the loss function:

$$h_t = \underset{h \in H}{\arg\min}\, L(F_{t-1}+h) = \arg\min \mathbb{E}[L(y, F_{t-1}(x)+h(x))]$$

Along with prediction shift, CatBoost also handles distribu- tion shifts that occur when category features are preprocessed. Target leakage may result from converting these features to their target statistics, which would further aggravate prediction shifts. The authors used the ordered target statistics approach

as inspiration for their new boosting algorithm, ordered boost- ing, to address issue. Another mode in CatBoost is called plain, and it is the conventional GBDT algorithm with integrated sorted target statistics. Their work describes the pseudocode used in CatBoost to build trees.

The model combines CatBoost and neural networks as base learners, integrating their predictions for overlapping and non- overlapping user segments into a unified output.

For non-overlapping users, a neural network architecture with an input layer sized to match the selected features, three hidden layers (512, 256, and 1 neurons), and a single-neuron output layer was employed.

CatBoost was utilized to enhance the prediction rate for overlapping users due to its strengths in classification tasks with large datasets, particularly those with abundant categori- cal features like the IEEE-CIS dataset. It employs target-based ordering, leveraging observed history for improved accuracy without the need for extensive preprocessing of categorical values. Additionally, CatBoost's robust performance requires minimal hyperparameter tuning while achieving high speed and accuracy.

B. *Synthetic minority oversampling technique and edited nearest neighbor (SMOTE-ENN)*

The study addresses the significant challenges posed by the highly imbalanced credit card dataset, which can lead to poor performance in traditional machine learning models[1]. Imbalanced datasets, where the majority class significantly outnumbers the minority class, often result in biased predictions that favor the majority class. To mitigate this issue, over- sampling techniques like the Synthetic Minority Oversampling Technique (SMOTE) and undersampling methods such as Edited Nearest Neighbor (ENN) are commonly used. SMOTE balances the class distribution by generating synthetic samples for the minority class, while ENN removes some majority class samples to achieve balance. However, each technique has its limitations. SMOTE can lead to overfitting by creating redundant data, and ENN may delete potentially valuable examples critical for learning. Additionally, ENN becomes less effective when the majority class is disproportionately larger than the minority class, as is the case in this study.

To overcome these challenges, the proposed credit card fraud detection model employs a hybrid approach combining SMOTE and ENN, known as the **SMOTE-ENN** method. This method benefits from both oversampling and undersam- pling, addressing the drawbacks of each individual technique. SMOTE generates synthetic samples to increase the size of the minority class, while ENN removes overlapping instances by applying a neighborhood cleaning rule. The ENN algorithm identifies and deletes examples that differ from at least two of their three nearest neighbors, thereby reducing noise and enhancing the quality of the dataset.

By using the SMOTE-ENN technique, the dataset achieves a better balance between classes, enabling more effective training of fraud detection models. This hybrid approach not only mitigates the risks of overfitting but also preserves essential data points for improved learning.

C. *Variational Auto Encoding*

The baseline method employs a deep learning[2] network consisting of three layers, with two ReLU activation functions between the layers. During training, the dataset (D0) passes through these layers to produce classification labels. Super- vised learning optimizes the model parameters, which are retained after training. The trained model outputs classification results when tested, and its performance is evaluated using related indicators.

The first enhanced model integrates the SMOTE module to replace the VAE (Method 1). Minority class samples (fraud cases) are input into SMOTE to generate synthetic data, which is combined with the original training set (D0) to create a balanced dataset (D3). This new dataset is then used to train the classifier, with the rest of the process aligning with the baseline method.

The second enhanced model replaces the VAE module with a GAN module (Method 2). GAN generates virtual fraud data by training a generator-discriminator network until Nash equilibrium is reached. The virtual fraud data is combined with the original dataset (D2) to create a balanced dataset (D3). This dataset is used to train the classifier, following the baseline approach.

The proposed method utilizes VAE, which outperforms GAN by generating data with more diversity and fewer restric- tions. VAE's encoding-decoding mechanism enables efficient data generation, particularly for text. Minority class samples are input into the VAE module to generate synthetic data, which is then merged with the original training set (D0). This balanced dataset (D3) is used to train the classifier, continuing as per the baseline method.

D. *Minority Oversampling-Based Generative Adversarial Net- works and Random Forest Algorithm*

The proposed Credit Card Fraud Detection Model (CCFDM) is constructed in three phases: data acquisition and preprocessing, data augmentation, and model construction. In the first phase, transaction-related features are extracted and normalized to ensure uniform scaling and to prevent skewed analyses. [3] The dataset contains 31 features, with 28 anonymized and reduced using Principal Component Analysis (PCA) to protect customer and merchant privacy. PCA ensures uncorrelated variables with maximized variance, making the dataset more interpretable. The remaining features, ”Amount” and ”Class,” are standardized, with fraudulent transactions comprising less than 1 percentage of the highly imbalanced dataset. The second phase addresses class imbalance using an ensemble approach combining SMOTE and GAN, called ESMOTE-GAN. While SMOTE is commonly used for over- sampling, it has limitations, such as oversampling noisy or overlapping data, generating less diverse samples, and focus- ing on local information. [9]GAN, on the other hand, learns the data distribution to produce realistic synthetic samples but

requires significant data for effective training. To overcome these challenges, multiple unbalanced subsets are created, ensuring that fraudulent samples comprise 10 percentage of each subset. SMOTE generates initial synthetic samples to balance these subsets, and GAN models further refine the data by enhancing diversity and better representing fraud cases. In the final phase, GAN’s generator network creates realistic fraud samples from latent space vectors, while the discriminator network learns to distinguish between real and generated samples. The training process stops when the dis- criminator cannot reliably differentiate between real and fake samples, indicating the GAN is ready. This ensemble approach improves data diversity, avoids overfitting, and produces a balanced dataset, enabling robust and effective fraud detection. The evaluation of the proposed ESMOTE-GAN-based Credit Card Fraud Detection Model (CCFDM) demon- strates its superior performance compared to related models.

[8] Among classifiers, Random Forest (RF) and XGBoost achieved the highest F-Measures of 92.31 and 92.44, respec- tively, due to their ability to handle non-linear and noisy datasets.[10] Logistic Regression (LR) had the lowest per- formance, attributed to its sensitivity to outliers and linear decision boundary limitations. Despite this, LR performed relatively well with unbalanced datasets, highlighting its adapt- ability to inherent class distributions.

Tree-based classifiers like RF and XGBoost excelled in high-dimensional and noisy data scenarios, surpassing artificial neural networks and deep learning models. Although the synthetic datasets created using SMOTE and GAN introduced noise, the ensemble approach ensured greater diversity and improved classification performance. While LR and SVM achieved better detection rates, their lower precision indicates challenges in handling non-linear decision boundaries in syn- thetic data. RF and XGBoost balanced detection rates and precision effectively, minimizing undetected frauds and false positives.

E. Credit Card Fraud Detection Using AdaBoost and Major- ity Voting

Majority voting and AdaBoost[4] are two important tech- niques used to improve classification performance, particularly in the context of credit card fraud detection. Majority vot- ing involves combining predictions from multiple classifiers, where each classifier votes for a target class based on its prediction for a test sample. The final output is determined by the class that receives the majority of the votes. For each classifier’s prediction, a binary function is used to represent the votes, and the class with the highest vote count is chosen as the final predicted class. AdaBoost, or Adaptive Boosting, enhances the performance of weak learners by assigning them weights based on their error rate. It combines the outputs of weak learners through a weighted sum to form a strong classifier.

In the experimental setup, a credit card fraud detection dataset was used to evaluate the performance of various models. The dataset is highly imbalanced, with fraudulent transactions being much fewer than legitimate ones. To ad- dress this imbalance, under-sampling was used to balance the dataset. The experiments employed 10-fold cross-validation to minimize bias. Since the dataset is skewed, traditional accuracy measures might not be effective. The Matthews Correlation Coefficient (MCC) was used as the evaluation metric, as it provides a balanced measure that takes into account true positives, true negatives, false positives, and false negatives. An MCC value of 1 indicates perfect prediction, while -1 signifies total disagreement.

The best MCC score of 0.823 was achieved using the majority voting method. A real-world credit card dataset from a financial institution was also used for evaluation, where the AdaBoost and majority voting methods achieved a perfect MCC score of 1. To test the robustness of the hybrid models, noise levels ranging from 10 percentage to 30 percentage were added to the data. The majority voting method demonstrated the best performance, achieving an MCC score of 0.942 with 30 percentage noise, indicating its robustness in noisy conditions.

F. OptDevNet: A Optimized Deep Event-Based Network Framework

The proposed OptDevNet framework for credit card fraud detection utilizes a score deviation loss function combined with a Gaussian prior probability function for end-to-end optimization[5]. The fraud detection model is structured as a fusion of a feature extraction module and a scoring mecha- nism. The feature extraction module, represented by a neural network with hidden layers, extracts meaningful features from the input data. The scoring mechanism, a simple linear neural unit, calculates the fraud score based on these feature repre- sentations. This model provides a direct mapping from input data to fraud scores, enabling efficient training in an end-to- end manner.

To refine fraud score prediction, a reference score is com- puted as the average of fraud scores from a set of normal ob- jects.[6] Two approaches—data-driven and prior-driven—are used to determine the reference score, with the prior-driven approach preferred for its efficiency and interpretability. A Gaussian distribution is used to define the reference score, where fraudulent transaction scores are sampled. The deviation loss function then measures the deviation of the fraud score from the reference score, utilizing a Z-score. This is further modified into a contrastive loss function, which penalizes deviations based on a threshold to ensure fraudulent instances are adequately distinguished from normal data.

The training process involves initializing parameters ran- domly, sampling batches from the training data, and calculat- ing fraud scores from a normal distribution. The loss function is computed for each data point, and gradient descent is used to update the model parameters. The process is repeated across multiple epochs until the model is trained. The trained model is then evaluated using a test dataset to classify transactions as fraudulent or normal based on the predicted fraud scores.

In the classification and evaluation phase, the model calcu- lates fraud scores for test data points and applies a threshold to obtain binary classification labels (fraud or normal). Various metrics derived from the confusion matrix, such as true pos- itives, false positives, true negatives, and false negatives, are used to evaluate the performance of the OptDevNet framework in accurately identifying fraudulent transactions.

The model obtained results of 99.8 percentage with consid- erably fewer training iterations and epochs when compared to this model and its competing schemes.

G. Kolmogorov-Arnold Network Models

KAN (Kolmogorov-Arnold Networks) is a fraud detection framework that combines the strengths of neural networks and decision trees to provide interpretable predictions. The architecture of KAN is hierarchical, with multiple layers that represent various levels of abstraction. [7] This structure enhances interpretability by allowing the decision-making pro- cess to be visualized as a tree, where each node corresponds to decisions based on kernel activations. KAN's architecture includes key parameters such as the number of neurons in each layer (width), grid configuration for kernel activations, the order of the spline used in activation functions (k), and the computation device (typically set to 'gpu' for faster training). To improve the efficiency of KAN, especially for credit card fraud detection, several techniques were incorporated. These include regularization methods like L1 and L2 regularization to prevent overfitting, adaptive grid mechanisms that dynamically adjust kernel activations based on input data, and mini-batch training to enhance generalization and convergence.

A comparative analysis of KAN, its efficient variant, and a Multilayer Perceptron (MLP) was conducted using two datasets. The results consistently showed that KAN and its efficient variant outperformed MLP across various perfor- mance metrics, including accuracy, precision, recall, F1-score, AUC-ROC, and the PR curve. This performance advantage is attributed to KAN's hierarchical structure, which combines the predictive power of neural networks with the interpretability of decision-tree models. The efficient variant of KAN also stands out for optimizing computational resources without sacrificing performance, making it suitable for deployment in resource- constrained environments.

The study highlighted the practical benefits of adopting KAN models, such as improved operational efficiency, reg- ulatory compliance, and scalability to manage large volumes of transaction data, making it an ideal solution for credit card fraud detection.

III. CONCLUSION

In conclusion, the reviewed literature highlights the ongoing advancements and various approaches to credit card fraud detection (CCFD) using machine learning and deep learning techniques. Several models, including traditional classifiers such as Naive Bayes (NB), Support Vector Machines (SVM), and Deep Learning (DL) networks, have been evaluated, showcasing their strengths and limitations in handling fraud detection tasks. Among these, hybrid models combining Ad- aBoost and Majority Voting methods have shown robust per- formance, particularly in noisy datasets, with Majority Voting offering notable results in terms of the Matthews Correlation Coefficient (MCC) score.

Moreover, more recent frameworks like OptDevNet and Kolmogorov-Arnold Networks (KAN) have introduced in- novative approaches, utilizing neural network structures tai- lored for fraud detection. OptDevNet, with its end-to-end optimization and Gaussian prior probability-based scoring mechanism, has demonstrated high efficiency in distinguishing fraudulent transactions. On the other hand, KAN's hierarchical structure, combining neural networks with decision-tree-like interpretability, has outperformed traditional models such as MLP, especially in terms of accuracy, precision, recall, and scalability. The efficient variants of KAN further optimize computational resources without compromising performance, making them ideal for real-time fraud detection in resource- constrained environments.

The review illustrates that while various models and tech- niques are available, there is no one-size-fits-all solution for CCFD. The hybrid approaches and optimization techniques emerging in the field are increasingly enhancing the effec- tiveness, interpretability, and efficiency of fraud detection systems. Future research should continue to focus on refining these models and exploring new methodologies to address challenges such as data imbalance, model interpretability, and deployment efficiency in real-world applications.

REFERENCES

[1] Nguyen, Nghia and Duong, Truc and Chau, Tram and Nguyen, Van-Ho and Trinh, Trang and Tran, Duy and Ho, Thanh " A proposed model for card fraud detection based on Catboost and deep neural network".

[2] Esenogho, Ebenezer and Mienye, Ibomoiye Domor and Swart, Theo G and Aruleba, Kehinde and Obaido, George "A neural network ensemble with feature engineering for improved credit card fraud detection ".

[3] Tingfei, Huang and Guangquan, Cheng and Kuihua, Huang, "Using variational auto encoding in credit card fraud detection ".

[4] Najadat, Hassan and Altiti, Ola and Aqouleh, Ayah Abu and Younes, Mutaz, "Credit card fraud detection based on machine and deep learn- ing".

[5] Taha, Altyeb Altaher and Malebary, Sharaf Jameel, ”An intelligent approach to credit card fraud detection using an optimized light gradient boosting machine”.

[6] Mienye, Ibomoiye Domor and Sun, Yanxia,”A deep learning ensemble with data resampling for credit card fraud detection”.

[7] Ghaleb, Fuad A and Saeed, Faisal and Al-Sarem, Mohammed and Qasem, Sultan Noman and Al-Hadhrami, Tawfik,”Ensemble Synthesized Minority Oversampling based Generative Adversarial Networks and Random Forest Algorithm for Credit Card Fraud Detection”.

[8] Randhawa, Kuldeep and Loo, Chu Kiong and Seera, Manjeevan and Lim, Chee Peng and Nandi, Asoke K,”Credit card fraud detection using AdaBoost and majority voting”.

[9] Adil, Muhammad and Yinjun, Zhang and Jamjoom, Mona M and Ullah, Zahid,”OptDevNet: A Optimized Deep Event-based Network Framework for Credit Card Fraud Detection”.

[10] Yeonjeong, Hwang and Kang, Hyoeun and Kim, Howon ,”Robust Credit Card Fraud Detection Based on Efficient Kolmogorov-Arnold Network Models”.