International Journal of Research Publication and Reviews

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

Enhanced Credit Card Fraud Detection Using Ensemble Learning

Surru Hruthik, Yalamati Purusotham Chowdary, Dudekula Saidu Mastan Bi, Sakhineti Rakesh Varama

Information Technology, GMR Institute of Technology, Rajam
Hruthiksurru366@gmail.com

ABSTRACT:

Ensemble learning is a powerful machine learning approach that combines multiple models to improve prediction accuracy and robustness. In the context of credit card fraud detection, it effectively identifies fraudulent transactions in highly imbalanced datasets. By utilizing ensemble techniques such as Random Forest, Gradient Boosting, and Stacking, we significantly enhance the detection capabilities while minimizing false positives, a critical challenge in fraud detection systems. Our approach incorporates various preprocessing strategies, including addressing class imbalance and optimizing feature selection, to ensure improved performance across models. Extensive experiments demonstrate that ensemble learning not only increases fraud detection accuracy but also reduces false alarms, providing a more reliable solution for real-time fraud prevention. The ensemble model's ability to adapt and generalize across different transaction patterns ensures robustness, making it well-suited for deployment in dynamic financial environments. In this paper, we present a detailed analysis of ensemble learning techniques and their application to credit card fraud detection, showcasing their potential in mitigating financial losses caused by fraudulent activities

Keywords: Ensemble Learning, Credit Card Fraud Detection, Machine Learning, Imbalanced Data, Real-time Detection

Introduction:

Credit card fraud detection is a very critical task nowadays because millions of transactions take place per day. Traditional detection methods cannot keep pace with ever-changing fraud tactics. Ensemble learning, a powerful technique using the aspect of machine learning, combines several models to obtain the better accuracy and reliability in terms of identifying fraud. Ensemble methods can identify suspicious patterns in credit card transactions more effectively using diverse algorithms, thereby reducing false positives and catching real fraud cases much better. This approach hence calls in a stronger structure of fraud detection systems, making online transactions safer and more secure for the users and businesses

I. Literature Review

It is the main point of a literature review to look into so far the most recent strides in credit card fraud detection, which might include establishing traditional statistical techniques as contrasts to machine learning approaches. Most important information is derived from using ensemble learning strategies, like Random Forests, Gradient Boosting, or CatBoost, which show higher accuracy and more strength in identifying fraudulent transactions against the separate models of SVMs or decision trees. Ensemble methods are quite well suited for dealing with emerging fraud strategies, boosting detection rates while reducing false positives. Another possible route for boosting the model's performance is to linearly combine a mix of different classifiers, such as SMOTE-ENN resampling and LSTM neural networks applied within an AdaBoost framework, which shows real promise in dealing with imbalanced datasets and hence improving the performance metrics of sensitivity, specificity, and AUC.

The visualization methods are presented as one of the main tools to assess complex financial data research, which emphasizes nowadays a significant role of visual within fraud detection, especially in finding patterns and intuitive analysis of large and complex datasets. While simple classifiers such as naive Bayes provide useful exploratory analysis, they often lack flexibility to adapt to evolving fraudulent behaviours. Meanwhile, a constant limitation in this area is the reliance sometimes on synthetic datasets which take place, thus yielding overfitting that does not mirror the real nature of transactions. This gap, therefore, entails the need for future research into optimal feature selection, representative datasets development, and model improvement for interpretability and applicability toward real-world conditions.

II. Methodology

1. Hybrid Ensemble Classification for Fraud Detection:

Data Collection and Sources:

Used two datasets: one real from European credit card transactions and the other synthetic by Sparkov.

Data Preprocessing:

Dropped unnecessary columns such as customer's name and place

Encoded categorical variables. Standardized numeric values.

Re-balanced class by oversampling.

Classification Algorithms:

Naive Bayes, Random Forest, XGBoost are used as classifiers

Model Optimization:

The three classes were combined to further improve the prediction accuracy of the fraudulent transactions.

Performance Metrics:

Improving detection with F1-score, ROC-AUC, and accuracy

Applications:

Real-Time Fraud Detection Banking Risk Management

2. Fraud Detection Framework Outline:

Data Collection:

Gather insights from the financial fields through collection from banks, credit card companies, financial institution planners, and other net-surfers keen on monetary analysis.

Model Selection for Fraud Detection:

Select prediction models such as Logistic regression, Decision-Tree classifier, Random forest, for any sort of classification algorithm on labeled data containing examples of known fraud cases.

Data Visualization Techniques:

Graph the landscape between different entities (like accounts, transactions, and users).

Analysis:

Use clustering and outlier detection to track real-time transactions and visualize outliers so they can be flagged.

Visualization Tools:

Use Tableau, Power BI, or any of the Python libraries (such as Matplotlib, Seaborn, Portly) to make an interactive dashboard.

Validation:

Work with a fraud analyst to interpret and validate the visualizations and thus improve our model according to practical insights.

Applications:

Louisiana-NCUS 2011. Banking Money Laundering Detection. Insurance Fraud Detection.

3.Credit Card Fraud Detection Workflow:

All this can be categorized as:

Data Preprocessing:

Raw data was cleaned to remove irrelevant or redundant features.

Feature Engineering:

Relevant features were engineered to improve the quality of input data.

Neural Network Ensemble:

Multiple neural networks were trained independently on the dataset.

Ensemble Aggregation:

Title is given

This aggregation aimed to leverage the strengths of each individual model to improve overall accuracy and robustness.

Result Interpretation:

The result is: Neural network ensemble-with the aid of feature engineering-was far superior compared to standard single neural network models in detecting fraudulent transactions by means of credit card.

Applications:

Mobile Payment and Digital Wallet Security.

Real-Time Transaction Monitoring.

4. Model Training and Evaluation:

Data Collection and Processing Collected a dataset of credit card transactions, including features related to transaction characteristics.

Feature Engineering: Provided features from the raw transaction data in order to boost model performance.

LightGBM model: Selected LightGBM for its high computational speed and capacity to handle a large and imbalanced dataset with high accuracy.

Hyperparameter Optimization: Applied optimization techniques such as grid search or Bayesian optimization to obtain sets of hyperparameters for tuning the LightGBM.

Model Training and Evaluation: Split the data into training and testing sets for the purpose of evaluating model performance.

Results and comparison: Compared the results of LightGBM with those of other models (Logistic Regression, Random Forest) to show that it performed effectively.

III. RESULTS & DISCUSSION

It establishes a complete system against online transaction scams, using some of the most ubiquitous and vital cyber security techniques: tokenization, digital signatures, and multi-factor authentication-which together create a better protection for certain transactions. Tokenization uses a RNG algorithm and protects sensitive data by substituting it with randomly generated tokens. Digital signatures based on a hash function set guarantee the integrity of the message and serve to verify the identity of the sender. Eventually, MFA entire TOTP executes a very important layer of verification. These techniques fuse to develop a sturdy architecture that strongly reduces the vulnerability to online transaction cyber-attacks for a safe arena for payments.

CONCLUSION

Ensemble learning proves to be an effective and robust approach for credit card fraud detection, significantly improving both accuracy and reliability in identifying fraudulent transactions. By leveraging techniques like Random Forest, Gradient Boosting, and Stacking, we enhance model performance, particularly in handling imbalanced datasets, which is a common challenge in fraud detection. The integration of preprocessing strategies such as class imbalance correction and optimized feature selection further strengthens the model's ability to minimize false positives, offering a more reliable real-time solution for fraud prevention. The adaptability of the ensemble model to various transaction patterns ensures its effectiveness in dynamic financial environments, making it an essential tool for mitigating financial losses caused by fraud. This research highlights the potential of ensemble learning in delivering robust, scalable, and precise fraud detection systems for the evolving challenges of the financial industry.

REFERENCES

[1]. Jemai, Jaber, Anis Zarrad, and Ali Daud. "Identifying Fraudulent Credit Card Transactions using Ensemble Learning." IEEE Access (2024).

[2]. M. E. Lokanan, ‘‘Financial fraud detection: The use of visualization techniques credit card fraud and money laundering domains,’’ J. Money Laundering Control, vol. 26, no. 3, pp. 436–444, Apr. 2024

[3]. Esenogho, E.; Mienye, I.D.; Swart, T.G.; Aruleba, K.; Obaido, G. A neural network ensemble with feature engineering for improved credit card fraud detection. IEEE Access 2022, 10, 16400–16407

[4]. Taha, A.A.; Malebary, S.J. An intelligent approach to credit card fraud detection using an optimized light gradient boosting machine. IEEE Access 2020, 8, 25579–25587

[5]. F. K. Alarfaj, I. Malik, H. U. Khan, N. Almusallam, M. Ramzan and M. Ahmed,"Credit card fraud detection using state-of-the-art machine learning and deep learning algorithms", IEEE Access, vol. 10, pp. 39700-39715, 2022.

[6]. 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,’’ IEEE Access, vol. 10, pp. 16400–16407, 2022.

[7]. F. K. Alarfaj, I. Malik, H. U. Khan, N. Almusallam, M. Ramzan, and M. Ahmed,''Credit card fraud detection using state-of-the-art machine learning and deep learning algorithms,'' IEEE Access, vol. 10, pp. 39700–39715, 2022.

[8]. S.Makki,Z.Assaghir,Y.Taher,R.Haque,M.-S.Hacid,andH.Zeineddine, An experimental study with imbalanced classification approaches for credit card fraud detection, IEEE Access, vol. 7, pp. 9301093022, 2019

[9]. A. A. Taha and S. J. Malebary, An intelligent approach to credit card fraud detection using an optimized light gradient boosting machine, IEEE Access, vol. 8, pp. 2557925587, 2020.

[10]. J. Yang, J. Qu, Q. Mi, and Q. Li, A CNN-LSTM model for tailings dam risk prediction, IEEE Access, vol. 8, pp. 206491206502, 2020.

[11]. H. Tingfei, C. Guangquan, and H. Kuihua, Using variational auto encoding in credit card fraud detection, IEEE Access, vol. 8, pp. 149841149853, 2020.

[12]. Sikarwar, R., Yadav, P., & Dubey, A. (2020, April). A Survey on IOT enabled cloud platforms. In 2020 IEEE 9th International Conference on Communication Systems and Network Technologies (CSNT) (pp. 120-124). IEEE.

[13]. M. Ramzan, H. U. Khan, S. M. Awan, A. Ismail, M. Ilyas, and A. Mahmood, A survey on state-of- the-art drowsiness detection techniques, IEEE Access, vol. 7, pp. 6190461919, 2019, doi: 10.1109/ACCESS.2019.2914373.

[14]. H. Tingfei, C. Guangquan, and H. Kuihua, Using variational auto encoding in credit card fraud detection, IEEE Access, vol. 8, pp. 149841149853, 2020, doi: 10.1109/ACCESS.2020.3015600.

[15]. 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,'' IEEE Access, vol. 10, pp. 16400–16407, 2022.Challenges. IEEE Communications Magazine, 55(1), 26-33.