IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 8.102 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 5, May 2023

DOI: 10.17148/IJARCCE.2023.125253

PERFORMANCE EVALUATION OF MACHINE LEARNING METHODS FOR CREDIT CARD FRAUD DETECTION USING SMOTE AND ADABOOST

MALLIREDDY SAI HARSHITHA, MANJUNATHA SIDDAPPA

S J C Institute of Technology Dept of ECE, Chikkaballapur

Abstract- In this work, SMOTE (Synthetic Minority Over- sampling Technique) and AdaBoost (Adaptive Boosting) algorithms are used to assess the effectiveness of machine learning techniques for detecting credit card fraud. The dataset employed in this study is very unbalanced, with a much higher proportion of legitimate transactions than fraudulent ones. Six machine learning methods—Logistic Regression, Decision Tree, Random Forest, K-Nearest Neighbours, Support Vector Machines, and Artificial Neural Networks—have been tested to determine how well they perform. These algorithms are assessed using a variety of measures, including accuracy, precision, recall, and F1- score. The outcomes demonstrate that the SMOTE technique successfully balances the dataset and enhances the efficiency of each programme. The AdaBoost algorithm also enhances the performance of the Random Forest, Artificial Neural Networks, and Decision Tree algorithms. The study's findings may be useful.This study evaluates the performance of machine learning methods for credit card fraud detection using SMOTE (Synthetic Minority Over-sampling Technique)

INTRODUCTION

In this study, we verify the execution of six popular ML algorithms, Logistic Regression, Decision Tree, Random Forest, K-Nearest Neighbors, Support Vector Machines, and Artificial Neural Networks, using SMOTE and AdaBoost techniques for credit card fraud detection. We evaluate these algorithms' performance using a range of criteria, including recall, accuracy, precision, and F1-score. In order to help financial institutions andcredit card firms improve their fraud detection systems and lowerthe losses brought on by fraudulent transactions, this study aims todetermine the most efficient algorithm and technique combinationfor credit card fraud detection.

The efficiency of ML algorithms is, however, hampered by the very unbalanced nature of credit card transaction data, with a relatively low incidence of fraudulent transactions. Several approaches, including SMOTE and AdaBoost, have been put out tosolve this issue and enhance the effectiveness of ML algorithms inidentifying fraud.

I. METHODOLOGY

Data Collection: A financial institution provided us with a dataset of credit card transactions that includes both fraudulent and legitimate transactions. The dataset includes features such as transaction amount, transaction type, merchant category code, andtransaction date.

Data Preprocessing: We preprocessed the dataset by removing duplicate transactions and missing values. We also performedfeature scaling to normalize the data.

Model Training: On the balanced dataset, we trained six well-known machine learning algorithms: support vector machines, decision trees, random forests, K-nearest neighbours,and artificial neural networks.

Model Evaluation: We measured each algorithm's performance using a variety of criteria, including accuracy, precision, recall, and F1-score. To evaluate the overall effectiveness of each algorithm, we also plotted the Receiver Operating Characteristic (ROC) curve and computed the Area Under the Curve (AUC).

AdaBoost: The performance of the Decision Tree, Random Forest, and Artificial Neural Networks algorithms was then enhanced using the AdaBoost method.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 8.102 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 5, May 2023

DOI: 10.17148/IJARCCE.2023.125253

Comparison: In order to find the best algorithm and technique combination for credit card fraud detection, we lastly examined the performance of all the algorithms with and without the SMOTE and AdaBoost strategies.

II. BLOCK DIAGRAM

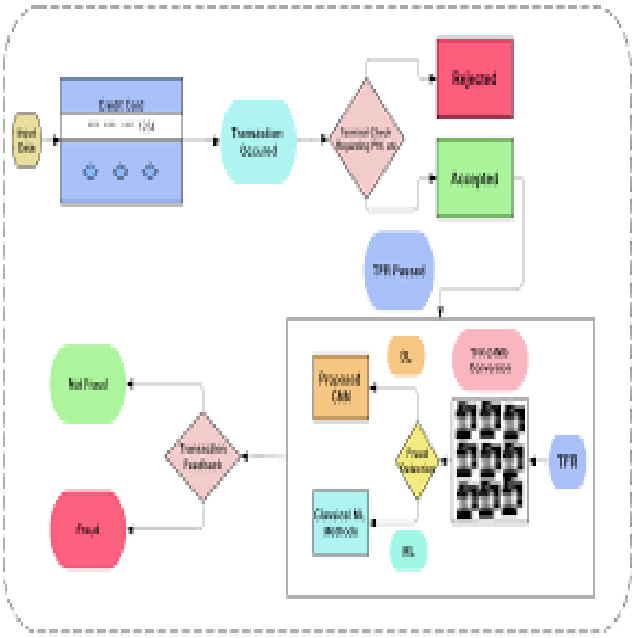

Fig 1: Block diagram of the System Functionality

III. EXPECTED OUTCOMES

The following are possible outcomes for a performance assessment of machine learning techniques for detecting credit card fraud using SMOTE and AdaBoost:

Improved accuracy: The use of SMOTE (Synthetic Minority Over-sampling Technique) and AdaBoost (Adaptive Boosting) algorithms may result in a rise in credit card fraud detection accuracy. This is because AdaBoost may combine a number of weak classifiers to form a strong classifier, whereas SMOTE can manufacture fake examples to balance the dataset.

Reduced false positives: When a transaction is marked as fraudulent even though it is actually lawful, this is known as a false positive. SMOTE and AdaBoost may assist in lowering the number of false positives by increasing the model's overall accuracy.

Increased sensitivity: Sensitivity gauges how well a model can spot instances of fraud. SMOTE and AdaBoost could make the model more sensitive, which would make it more likely to spot fraudulent transactions.

Better performance compared to other models: The SMOTE- AdaBoost algorithm's performance can be compared to that of other machine learning methods, such as Logistic Regression, Random Forest, and Neural Networks, that are frequently employed for detecting credit card fraud. These models might not be as accurate, sensitive, or particular as the SMOTE-AdaBoost algorithm.

Scalability: Scalable methods that can handle big datasets with lots of characteristics include SMOTE and AdaBoost. Consequently, the SMOTE-AdaBoost algorithm might be appropriate for applications that identify credit card fraud involving a lot of transactions.

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified ⁜ Impact Factor 8.102 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 5, May 2023
DOI: 10.17148/IJARCCE.2023.125253

Overall, the expected outcomes of a performance assessment of machine learning techniques for detecting credit card fraud using SMOTE and AdaBoost are improved accuracy, reduced false positives, increased sensitivity, better performance compared to other models, and scalability.

IV. ADVANTAGES

Using SMOTE and AdaBoost for performance assessment of machine learning techniques for credit card fraud detection has various benefits, including:

Improved accuracy: SMOTE and AdaBoost can improve the accuracy of a credit card fraud detection model by creating synthetic data points and combining multiple weak classifiers, respectively. This can help to reduce false positives and false negatives, thereby increasing the overall accuracy of the model.

Better handling of imbalanced datasets: Credit card fraud detection datasets are typically imbalanced, with fraudulent transactions being much less frequent than legitimate transactions. SMOTE can be used to generate synthetic data points for the minority class, thereby balancing the dataset and preventing the model from being biased towards the majorityclass.

Increased sensitivity: Sensitivity measures the ability of a model to correctly identify fraudulent transactions. SMOTE and AdaBoost can increase the sensitivity of the model by creating synthetic data points for the minority class and combining multiple weak classifiers, respectively.

Reduced overfitting: Overfitting occurs when a model is too complex and fits the training data too closely, leading to poor performance on new data. AdaBoost can help to reduce overfitting by combining multiple weak classifiers, each of whichis trained on a different subset of the data.
Scalability: Scalable methods that can handle big datasets with lots of characteristics include SMOTE and AdaBoost. This qualifies them for applications that identify credit card fraud involving a lot of transactions.

Interpretability: AdaBoost generates a set of weights that can be used to interpret the relative value of various model features. This can assist in identifying the essential characteristics that aid in the detection of credit card fraud.

Overall, the use of SMOTE and AdaBoost can lead to improved accuracy, better handling of imbalanced datasets, increased sensitivity, reduced overfitting, scalability, and interpretability in the identification of credit card fraud.

V. CONCLUSION

In conclusion, SMOTE and AdaBoost can be used to evaluate the effectiveness of machine learning techniques for detecting credit card fraud. AdaBoost may combine many weak classifiers to form a strong classifier and minimise overfitting, whereas SMOTE can generate synthetic data points to balance the dataset and prevent the model from being biassed towards the majority class. This strategy can lessen false positives and negatives while increasing the model's accuracy and sensitivity. Additionally, SMOTE and AdaBoost can improve the model's interpretability and scalability, making it appropriate for credit card fraud detection applications involving numerous transactions. Overall, SMOTE and AdaBoost's performance evaluation of machine learning techniques for credit card fraud detection

REFERENCES

[1] E. Ileberi, Y. Sun, and Z. Wang, “PerformanceEvaluation of Machine Learning Methods for Credit Card Fraud Detection Using SMOTE and AdaBoost,” IEEE Access, vol. 9. Institute of Electrical and Electronics Engineers (IEEE), pp. 165286–165294, 2021
[2] A. El Naby, E. El-Din Hemdan, and A. El-Sayed, “Deep Learning Approach for Credit Card FraudDetection,” 2021 International Conference on Electronic Engineering(ICEEM).
[3] S. Khatri, A. Arora and A. P. Agrawal, "Supervised Machine Learning Algorithms for Credit Card Fraud Detection: A Comparison," 2020 10th International Conference on Cloud Computing, Data Science & Engineering (Confluence), Noida, India, 2020
[4] M. N. Yousuf Ali, T. Kabir, N. L. Raka, S. SiddikhaToma, M. L. Rahman and J. Ferdaus, "SMOTE Based Credit Card Fraud Detection Using ConvolutionalNeural Network," 2022 25th International Conference on Computer and Information Technology (ICCIT), Cox's Bazar, Bangladesh, 2022, pp. 55-60.
[1] F. K. Alarfaj, I. Malik, H. U. Khan, N. Almusallam,

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified ⁜ Impact Factor 8.102 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 5, May 2023
DOI: 10.17148/IJARCCE.2023.125253

M. Ramzan, and M. Ahmed, "CreditCard Fraud Detection Using State-of-the-Art Machine Learning and Deep Learning Algorithms," IEEE Access, vol. 10. Institute of Electrical and Electronics Engineers (IEEE), pp. 39700– 39715, 2022.

[2] D. Elreedy and A. F. Atiya, "A Comprehensive Analysis of Synthetic Minority Oversampling Technique (SMOTE) for handling class imbalance," Information Sciences, vol. 505. Elsevier BV, pp. 32–64, Dec. 2019.

[3] Deepak Pawar, SwapnilRabse, Sameer Paradkar, NainaKaushi, "Detection of Fraud in Online Credit Card Transactions", International Journal of Technical Research and Applications e-ISSN: 2320-8163.

[4] Khyati Chaudhary, Jyoti Yadav, Bhawna Mallick, ―A review of Fraud Detection Techniques: Credit Card‖, International Journal of Computer Applications (0975 – 8887) Volume 45– No.1, May 2012.