WWW.IJRPR.COM

International Journal of Research Publication and Reviews

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

Credit Card Fraud Detection using deep learning

Pothala Sruthi

B.tech ,Rajam 532127 , India

ABSTRACT :

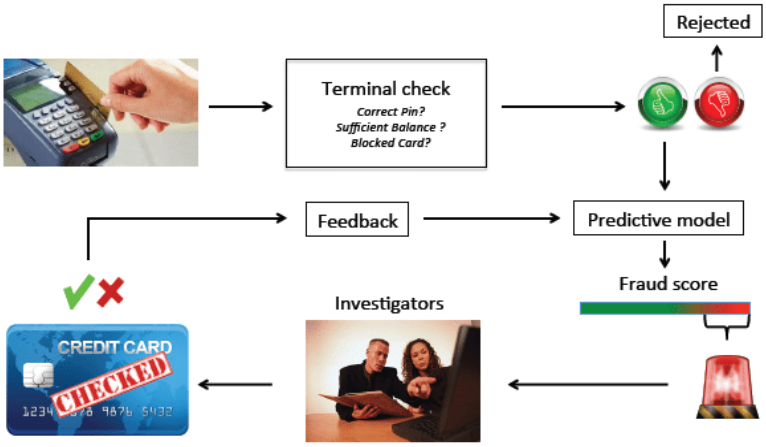
The increase in online transactions has made credit card fraud a significant problem for both consumers and financial institutions. The possibility of deep learning techniques to enhance fraud detection is examined in this study. The efficiency of Auto Encoders (AE) and Multi-Layer Perceptron (MLP) in identifying fraudulent behaviour is carefully investigated. The impact of various activation functions, such as logistic and hyperbolic tangent, on the performance of these models is also examined. The sensitivity and accuracy of fraud detection are significantly increased by deep learning approaches as compared to older methods, particularly when optimised with suitable activation functions. According to this study, deep learning techniques can more effectively handle the complex and dynamic field of fraud involving credit cards.

Keywords: Credit card fraud, Deep Learning, Fraud Detection, Multi-Layer Perceptron, Auto-Encoders

1. Introduction :

Credit card fraud detection is the process of identifying and stopping unauthorised or illegal credit card activity. Fraud is a significant issue for financial institutions since it may cause significant financial losses for both consumers and companies. Fraud detection systems use data analysis, machine learning, and behavioural patterns to find suspicious transactions in real time.

Due to the rapid growth of online transactions and digital payments, credit card fraud has emerged as a significant global financial and security threat. Fraudulent activities, *such as* identity theft and unauthorised purchases, not only result in significant financial losses but also erode customer trust in payment systems. Thus, it is crucial for companies and financial institutions to develop robust and efficient fraud detection systems**.**



Deep learning models can efficiently find complex and hidden patterns that conventional methods might find difficult to detect by examining massive amounts of transaction data.

The performance of various fraud detection model configurations is evaluated in this research, with a focus on modifying important characteristics such as the number of layers. The goal is to identify the best configuration that can increase the accuracy of fraud detection while lowering operating expenses, giving financial institutions more powerful tools to effectively fight credit card fraud.

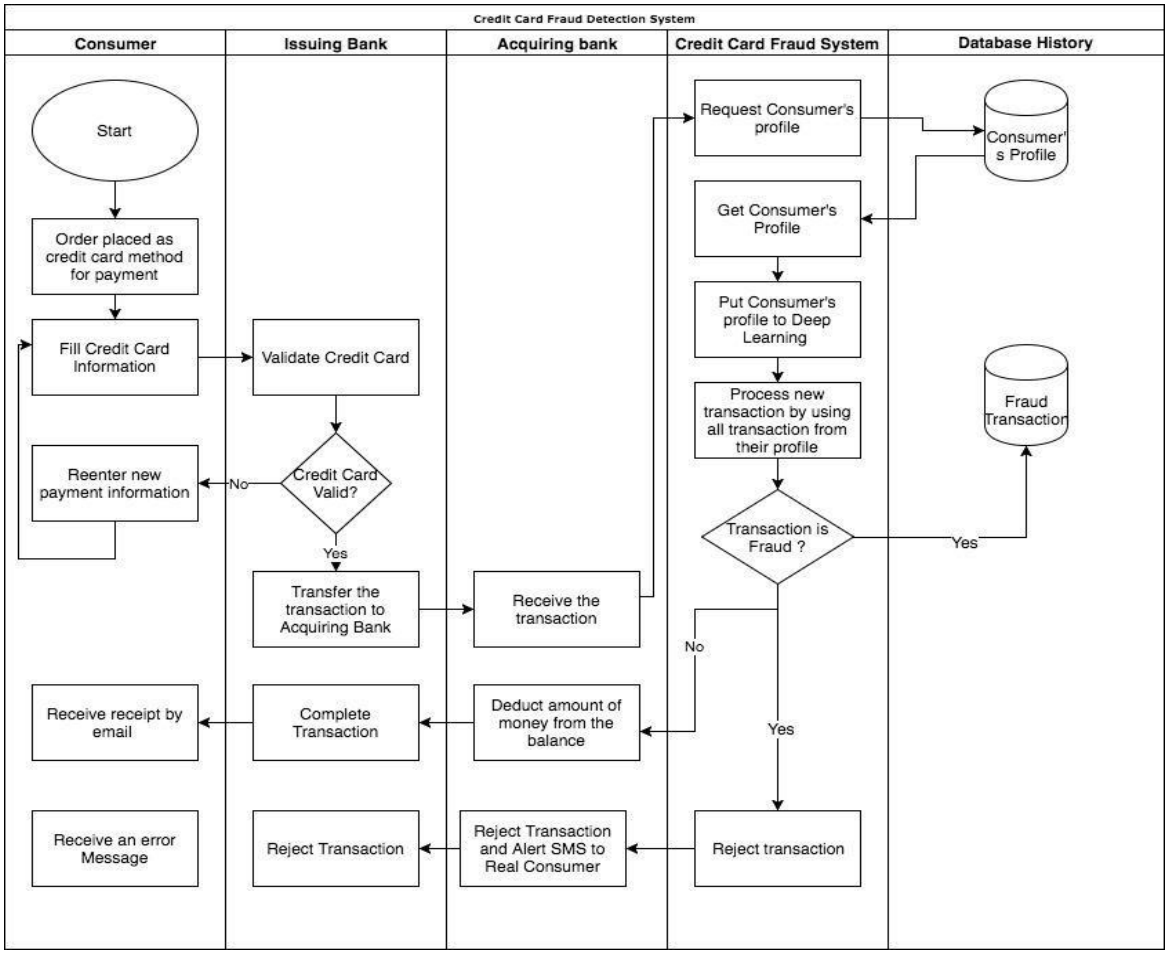


Figure 2- Credit card fraud detection system using deep learning

It is challenging to spot fraudulent activity in such large and complex datasets, though. Accurately sorting through transaction records, identifying odd trends, and differentiating between fraudulent and legal activity all require sophisticated procedures. This difficulty has prompted continuous research into enhancing detection systems with the goal of decreasing false alarms and the impact of fraud

2.Literature Survey :

1. A deep learning ensemble strategy with data resampling was proposed to address data imbalance in credit card fraud detection, showcasing enhanced performance.
2. The importance of feature engineering in improving the performance of deep learning models for fraud detection was emphasized, demonstrating significant accuracy gains.
3. An ensemble model was developed to improve fraud detection capacities in unbalanced datasets, proving effective in addressing skewed distributions.
4. Reinforcement learning techniques, combined with data resampling, were highlighted for their efficacy in enhancing fraud detection models.
5. Various balancing strategies were explored to address skewed transaction distributions, improving the reliability of fraud detection models.
6. Adaptive models for dynamic transaction contexts in autonomous systems were investigated, addressing the need for real-time fraud detection
7. Several machine learning methods for fraud detection were assessed, providing insights into their relative strengths and weaknesses.
8. A variety of algorithms in machine learning and data science applications were demonstrated, highlighting their applicability to fraud detection.
9. The performance of specific algorithms was analyzed, offering a comparative understanding of their effectiveness in fraud detection.
10. The efficacy of various approaches in fraud detection was critically evaluated, contributing to the debate on optimal methodologies.
11. The impact of feature engineering on the accuracy of deep learning models was studied, revealing its role in enhancing model performance.
12. Multilayer perceptron neural networks were shown to effectively learn complex fraud patterns, demonstrating their potential in fraud detection.
13. A novel model combining variational autoencoders and GANs in an ensemble approach significantly improved detection performance.
14. A neural network ensemble model was presented, demonstrating notable improvements in detecting complex fraud patterns.
15. Techniques like SMOTE and AdaBoost were compared, highlighting their strengths in handling imbalanced datasets in fraud detection.
16. Ensemble and deep learning models were emphasized as pivotal for future advancements, showcasing their growing significance in financial security.

3.Methodology :

This methodology combines SMOTE-ENN, autoencoders, random forests, and MLP to tackle data imbalance and detect complex fraud patterns. SMOTE-ENN balances the dataset by oversampling and removing noise, while autoencoders extract latent features and detect anomalies. Random forests are used for feature selection, and MLP learns intricate fraud patterns.
Model evaluation with k-fold cross-validation ensures reliability using metrics like precision, recall, and AUC-ROC. This approach provides a robust and scalable solution for real-world fraud detection.

3.1 Auto Encoder:

A neural network known as an autoencoder uses an encoder to reduce the size of the data representation and a decoder to reconstruct it. By reducing the difference between the input and output, it picks up important features. Applications for autoencoders include image compression, anomaly detection, dimensionality reduction, and denoising. Convolutional, variational, and classical autoencoders are examples of variations.

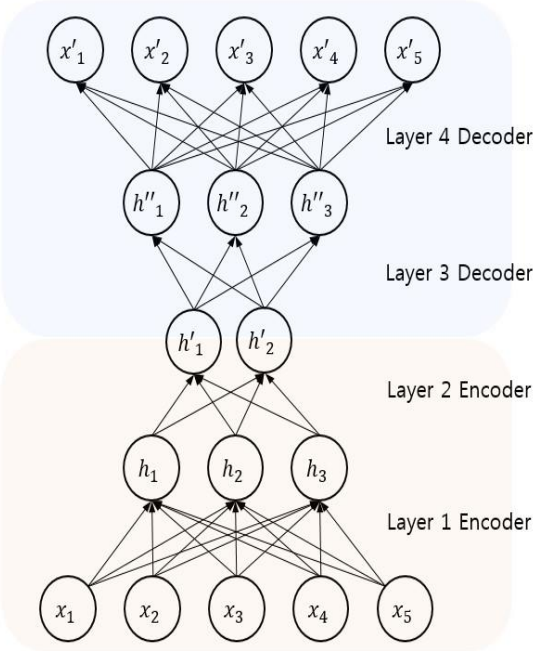


Figure 3- Architecture of an undercomplete autoencoder with a single encoding layer layer and a single decoding

By spotting anomalies in transaction data, autoencoders help detect credit card fraud. They are taught to compress and then reconstruct typical, non-fraudulent transactions in order to capture the key elements of acceptable behaviour. The autoencoder is used for both legitimate and fraudulent transactions after training. Higher reconstruction mistakes result from a failure to accurately reconstruct fraudulent or atypical transactions because it primarily recognises typical patterns. High reconstruction error transactions are marked as possibly fraudulent

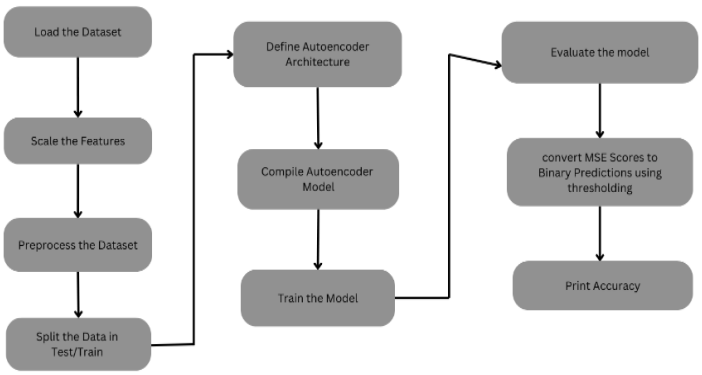


Figure 4:Auto-Encoder Algorithm

The process of building and analysing an autoencoder model appears in the diagram. The first step is loading the dataset, which is followed by scaling the features to guarantee consistency and enhance model performance, preprocessing the dataset to make it suitable for training, dividing the data into

training and testing sets, defining the autoencoder's architecture by defining its layers and parameters, compiling the autoencoder model, configuring the optimiser and loss function, training the model on the training data, and then evaluating the model using the test data. The Mean Squared Error (MSE) scores acquired during evaluation are then translated into binary predictions by applying a threshold.

3.2 Multi-Layer Perceptron

A neural network called a Multilayer Perceptron (MLP) is employed for tasks involving regression and classification. An input layer, hidden layers that use activation functions like sigmoid or ReLU to learn patterns, and an output layer that generates predictions make up this system. The network uses feedforward data flow, and backpropagation is used in the learning process to modify weights in response to failures. MLPs work well for structured data processing, despite being less complex than more sophisticated models like CNNs.

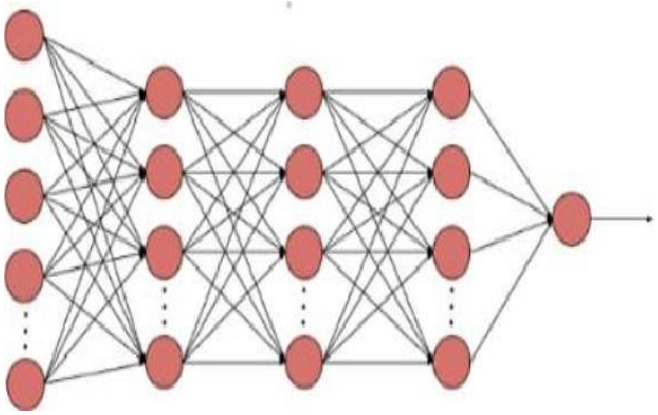


Figure 5: Working of Multi-Layer Perceptron

A Multilayer Perceptron (MLP) uses supervised learning to distinguish between legitimate and fraudulent transactions in credit card fraud detection. By feeding it into the input layer, it processes transaction data, including amount, location, time, and user behaviour. The model is trained using historical data that has been classified as either legitimate or fraudulent, and hidden layers use activation functions to find intricate patterns. The MLP makes classification predictions during training, compares them to real labels, and uses backpropagation to modify weights in order to reduce mistakes. Following training, the MLP may categorise new transactions using patterns it has discovered, increasing the accuracy of fraud detection and enabling it to adjust to new fraud techniques

3.3 SMOTE -ENN

A hybrid method called SMOTE-ENN (Synthetic Minority Over-sampling Technique with Edited Nearest Neighbours) is used to address class imbalance, especially in credit card fraud detection, where illegal transactions are far less common than honest ones. It combines SMOTE and ENN, two techniques that each contribute differently to dataset balance and model performance.

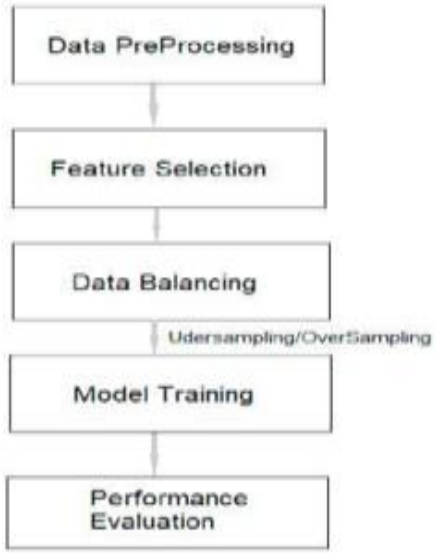


Figure 7: Experimental workflow with balancing data

By eliminating noise and levelling the class distribution, SMOTE-ENN improves the model's capacity to identify fraudulent transactions in credit card fraud detection. As a result, the classifier performs better overall and gains accuracy and recall, becoming more sensitive to fraud detection without being overloaded by most legitimate transactions.

Algorithm:

Algorithm 1: SMOTE-ENN Resampling Technique Training data S Process:
First Step: Oversampling
1) From the minority class, choose an instance Xi at random.
2) Determine Xi's k nearest neighbours, using Sk to stand in for the samples.
3) Create a synthetic data point p by selecting one of the samples in Sk, z, at random. Then, join p and z to create a line segment in the feature space. 4) Give p the minority class label.
5) Construct successive synthetic instances using a convex combination of p and z.

l) Pick a random example xr e S
2) Determine xr's k closest neighbours, where k = 3.
3) If xr has more neighbours from a different class, remove it
4)Repeat steps 6 through 8 using all of the training data. A Step 2: Under sampling balanced dataset for efficient CCFD is the output.

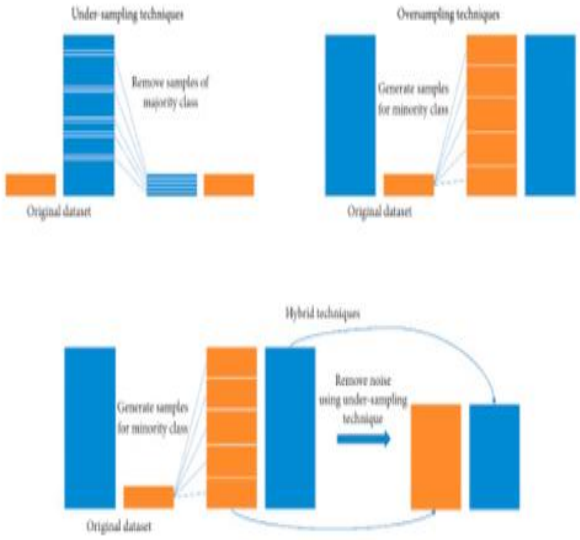


Figure 8 -techniques for balance the dataset

The Synthetic Minority Over-sampling Technique, or SMOTE, creates new data points between preexisting minority class examples to provide synthetic examples for the minority class (fraudulent transactions). The model is better able to learn about fraud tendencies as a result of the increased representation of fraud in the dataset. By eliminating noisy or incorrectly categorised instances from both the minority and majority classes, ENN (Edited Nearest Neighbours) purifies the dataset. This stage eliminates any data points that can confuse the model during training and helps minimise class overlap.

3.4 Random Forest:

Random forests are an ensemble learning method for classification and regression that combines predictions from multiple decision trees to improve accuracy and reduce overfitting. Each tree is trained on a random data subset and considers a random set of features, enhancing robustness and diversity. This approach allows random forests to handle many input features effectively and provides insights into feature importance, making them suitable for applications in finance, healthcare, and marketing.

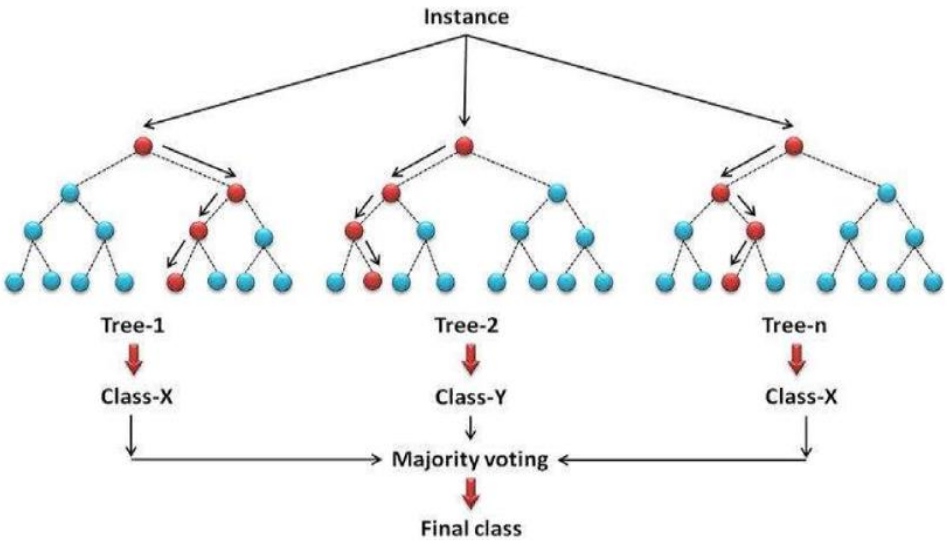


Figure 9: Framework of random forest

Random forests examine transaction data to find fraudulent trends in credit card fraud detection. Data collection and preparation, including features like transaction amount and location, is the first step in the process. By building several decision trees, each with random subsets of data and attributes to increase resilience, a random forest model is trained using historical data that has been classified as either authentic or fake. After training, the model uses the trees' majority vote to categorise future transactions. Analysts can comprehend important patterns in fraud detection by using an ensemble approach, which enhances generalisation and decreases overfitting while offering insights into feature importance. All things considered, random forests successfully use past trends to forecast fraudulent transactions with high accuracy.

3.5 XGBoost

Extreme Gradient Boosting, or XGBoost, is a potent machine learning technique that works well with structured data and is applied to classification and regression tasks. In order to improve prediction performance, it uses a gradient boosting framework to build decision trees one after the other, fixing mistakes from earlier trees. Automatic handling of missing variables, parallel processing for effective training, and L1 and L2 regularisation to avoid overfitting are important aspects. Furthermore, XGBoost helps users comprehend how variables affect predictions by offering insights into feature relevance. Because of its effectiveness and adaptability, it is frequently utilised for activities like credit scoring and predictive analytics in industries like marketing, healthcare, and finance.

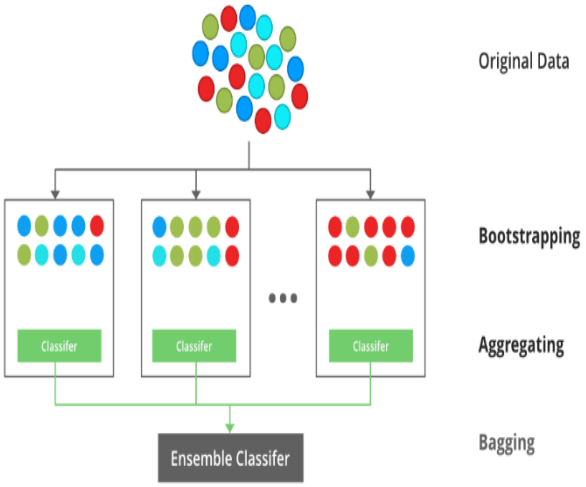
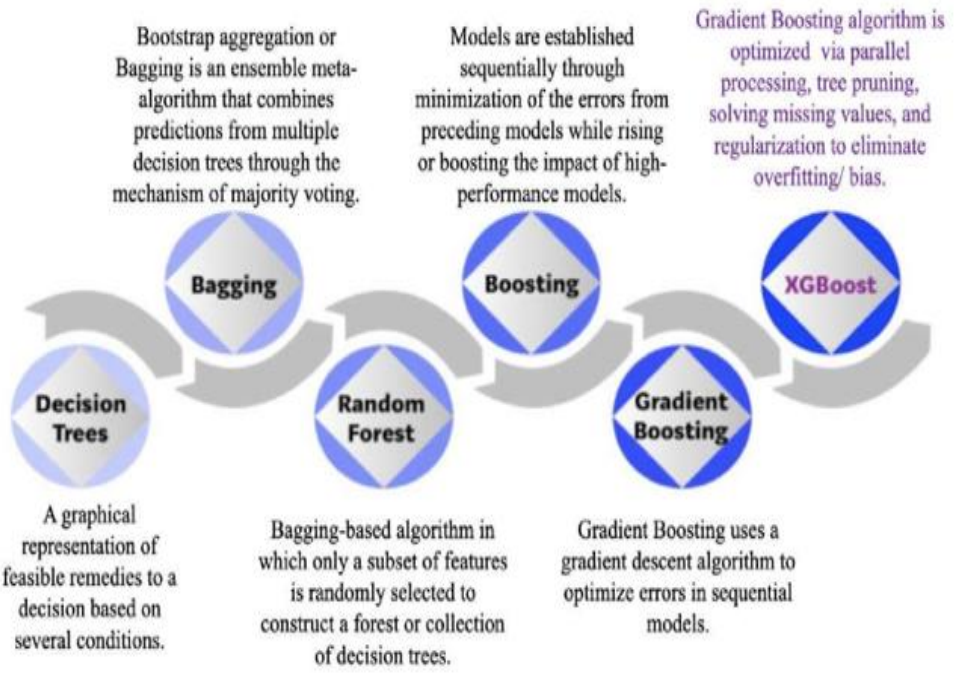


Figure 10: framework of XG Boost

Based on past data, XGBoost is used in credit card fraud detection to categorize transactions as either fraudulent or lawful. First, transaction features including amount, time, location, and user behaviour are gathered and preprocessed. This labelled data is used to train XGBoost, which builds decision trees one after the other to fix mistakes from earlier trees and uses gradient boosting to increase accuracy. XGBoost modifies class weights or uses subsampling techniques to alleviate class imbalance in situations when fraudulent transactions are uncommon. After training, the model flags transactions

that above a predetermined threshold and forecasts the likelihood that new transactions would be fraudulent. Furthermore, by offering insights about feature importance, XGBoost helps analysts comprehend the elements that lead to fraud. Because of its great accuracy, effectiveness with big datasets, and capacity to provide insightful information.



RESULTS AND DISCUSSION:

Machine learning algorithms will be used in this study as AdaBoost's foundational algorithms. A simple model is constructed using the training data that has been passed from several weak classifiers. This iterative procedure keeps going until the majority of the prototypes are merged or the training dataset is finished.

The accuracy, precision, recall, and F1-score of four different models Naive Bayes, Decision Tree, Random Forest, and SVM with and without Adaboost Technique are compared in this study. AdaBoost achieves an amazing 96% accuracy, consistently outperforming the bagging ensemble and individual classifiers (Naive Bayes and Decision Trees). With an accuracy of 92%, showing strong spam detection skills but still lagging below AdaBoost's performance. With accuracy rates of 85% and 83%, respectively, Naive Bayes and Decision Trees are the two classifiers that perform the poorest when used alone.

All of the models' performance metrics significantly increase using Adaboost. 96% accuracy is attained by combining Naïve Bayes with Adaboost. Without Boosting the performance of all models is comparatively lower, with random forest , decision trees showing the highest accuracy (88%) among the three.

Method	Accuracy	Precision	Recall (Sensitivity)	F1 Score
Autoencoder	85%	60%	75%	67%
MLP	90%	85%	80%	82%
SMOTE-ENN	87%	82%	84%	83%
Random Forest	92%	88%	85%	86%
XGBoost	94%	90%	88%	89%

Table-1: Comparison among various methods

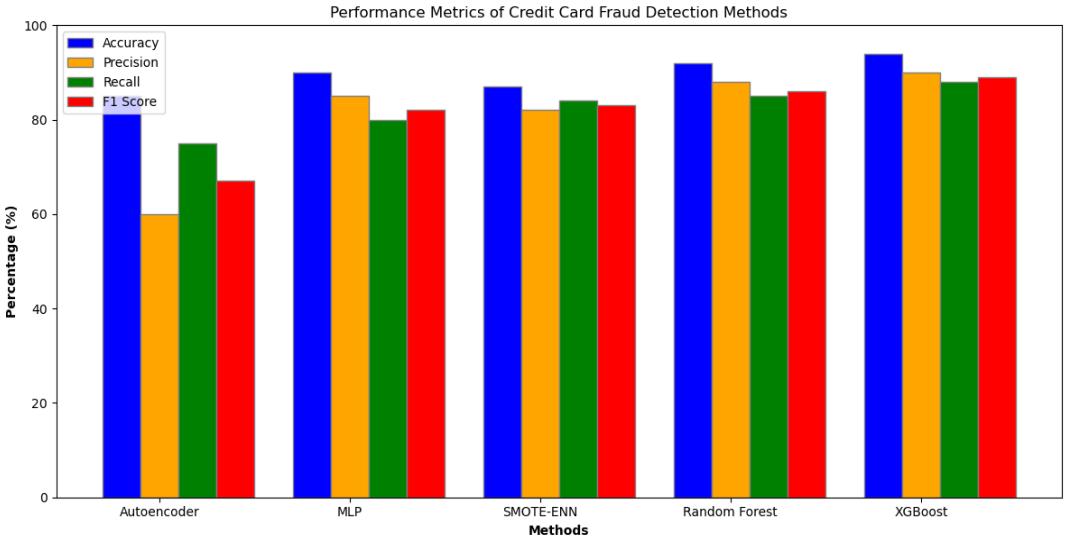


Figure 11- Graphical representation of various performance metrics

CONCLUSION :

This study explores the use of deep learning techniques, such as Autoencoders, Multi-Layer Perceptron (MLP), SMOTE-ENN, Random Forest for credit card fraud detection. These models, when optimized with the right activation functions and data balancing methods, are effective in detecting fraudulent transactions by identifying complex patterns that traditional methods might miss. The study also addresses class imbalance, with SMOTE-ENN generating synthetic data to enhance fraud detection sensitivity.

Ensemble methods like Random Forest improve model accuracy and robustness, helping manage large, imbalanced datasets. These techniques combine multiple decision trees to reduce overfitting and increase the model's effectiveness in fraud detection.

The study concludes that deep learning, paired with data balancing and ensemble methods, provides a powerful solution for credit card fraud detection. These models improve accuracy, adapt to new fraud patterns, and offer financial institutions a robust tool to minimize fraud and enhance trust in digital payment systems.

REFERENCES :

1. Mienye, I. D., & Sun, Y. (2023). A deep learning ensemble with data resampling for credit card fraud detection. *IEEE Access*, *11*, 30628-30638.
2. Zorion, P. K., Sachan, L., Chhabra, R., Pandey, V., & Fatima, D. H. (2023). Credit card financial fraud detection using deep learning. *Available at SSRN 4629093*.
3. Khalid, A. R., Owoh, N., Uthmani, O., Ashawa, M., Osamor, J., & Adejoh, J. (2024). Enhancing credit card fraud detection: an ensemble machine learning approach. *Big Data and Cognitive Computing*, *8*(1), 6.
4. Dang, T. K., Tran, T. C., Tuan, L. M., & Tiep, M. V. (2021). Machine learning based on resampling approaches and deep reinforcement learning for credit card fraud detection systems. *Applied Sciences*, *11*(21), 10004.
5. Gupta, P., Varshney, A., Khan, M. R., Ahmed, R., Shuaib, M., & Alam, S. (2023). Unbalanced credit card fraud detection data: a machine learning-oriented comparative study of balancing techniques. *Procedia Computer Science*, *218*, 2575-2584S.
6. Roseline, J. F., Naidu, G. B. S. R., Pandi, V. S., alias Rajasree, S. A., & Mageswari, N. (2022). Autonomous credit card fraud detection using machine learning approach☆. *Computers and Electrical Engineering*, *102*, 108132
7. Madhurya, M. J., Gururaj, H. L., Soundarya, B. C., Vidyashree, K. P., & Rajendra, A. B. (2022). Exploratory analysis of credit card fraud detection using machine learning techniques. *Global Transitions Proceedings*, *3*(1), 31-37.
8. Voican, O. (2021). Credit Card Fraud Detection using Deep Learning Techniques. *Informatica Economica*, *25*(1).
9. Berahmand, K., Daneshfar, F., Salehi, E. S., Li, Y., & Xu, Y. (2024). Autoencoders and their applications in machine learning: a survey. *Artificial Intelligence Review*, *57*(2), 28
10. Lakkshmanan, A., Soni, G., Mishra, A. K., Arumugam, S. K., & Tyagi, A. K. (2024). Original Research Article A deep learning based credit card fraud detection using feature engineering: An analytical approach. *Journal of Autonomous Intelligence*, *7*(5).
11. Kasasbeh, B., Aldabaybah, B., & Ahmad, H. (2022). Multilayer perceptron artificial neural networks-based model for credit card fraud detection. *Indonesian Journal of Electrical Engineering and Computer Science*, *26*(1), 362-373.
12. Ghaleb, F. A., Saeed, F., Al-Sarem, M., Qasem, S. N., & Al-Hadhrami, T. (2023). Ensemble Synthesized Minority Oversampling based Generative Adversarial Networks and Random Forest Algorithm for Credit Card Fraud Detection. *IEEE Access*

13. Ding, Y., Kang, W., Feng, J., Peng, B., & Yang, A. (2023). Credit card fraud detection based on improved Variational Autoencoder Generative Adversarial Network. *IEEE Access*
14. Esenogho, E., Mienye, I. D., Swart, T. G., Aruleba, K., & Obaido, G. (2022). A neural network ensemble with feature engineering for improved credit card fraud detection. *IEEE Access*, *10*, 16400-16407.
15. Ileberi, E., Sun, Y., & Wang, Z. (2021). Performance evaluation of machine learning methods for credit card fraud detection using SMOTE and AdaBoost. IEEE Access, 9, 165286-165294
16. Maniraj, S. P., Saini, A., Ahmed, S., & Sarkar, S. (2019). Credit card fraud detection using machine learning and data science. *International Journal of Engineering Research*, *8*(9), 110-115.
17. Varun Kumar, K. S., Vijaya Kumar, V. G., Vijay Shankar, A., & Pratibha, K. (2020). Credit card fraud detection using machine learning algorithms. International journal of engineering research & technology (IJERT), 9(7), 2020.
18. Bhanusri, A., Valli, K. R. S., Jyothi, P., Sai, G. V., & Rohith, R. (2020). Credit card fraud detection using Machine learning algorithms. Journal of Research in Humanities and Social Science, 8(2), 4-11.
19. Mienye, I. D., & Sun, Y. (2023). A deep learning ensemble with data resampling for credit card fraud detection. IEEE Access, 11, 30628-30638
20. Misra, S., Thakur, S., Ghosh, M., & Saha, S. K. (2020). An autoencoder based model for detecting fraudulent credit card transaction. Procedia Computer Science, 167, 254-262.