International Journal of Engineering Sciences & Research Technology

(A Peer Reviewed Online Journal)
Impact Factor: 5.164

IJESRT

RESEARCHERID
THOMSON REUTERS

[Julius *al.*, 13(6): June, 2024]
IC™ Value: 3.00

ISSN: 2277-9655
Impact Factor: 5.164
CODEN: IJESS7

IJESRT

INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY

STRENGTHENING AMERICAN LEADERSHIP IN DIGITAL FINANCIAL TECHNOLOGY THROUGH MACHINE LEARNING

Odubajo Adeyemi Julius
MoMo Payment Service Bank
(of Affiliation)
Lagos, Nigeria
yemiodubajo@yahoo.com

DOI: 10.5281/zenodo.15331734

ABSTRACT

The last several years have seen a notable improvement in financial inclusion in America. Bank accounts have become more common among Indians in recent years. The growing risk of credit card fraud resulting from the use of digital financial transactions calls for the development of clever investigation techniques. Therefore, in order to propose a machine learning-based fraud detection system, this study will use the American Express Credit Card Fraud Detection (CCFD) dataset. The stages involved in data preparation include managing missing values, normalization, noise reduction, and SMOTE class balance. To increase model efficiency, feature selection is done using statistical approaches. The XGBoost algorithm's High Accuracy and Scalability are employed for categorization. The suggested model performs better than a number of baseline models in terms of accuracy (98.7%), precision (96.4%), recall (95.8%), and F1-score (96.1%), according to experimental study results. These findings demonstrate that the model can most successfully identify fraudulent transactions with a low rate of false positives. The study adds to strengthening digital financial security and supports the development of FinTech leadership in America.

KEYWORDS: Digital Financial Technology, Fraud Detection, American FinTech Leadership, Financial Sustainability, American Express CCFD Dataset

1. INTRODUCTION

the past decade, the exponential growth of the Internet has drastically changed the world economy by bringing about an abundance of digital services and platforms. The central aspects of everyday financial transactions include e-commerce, online banking, tap-and-pay systems, and digital bill payment services among. This fast-growing digitalization has not only expanded consumers' convenience and businesses' efficiency but also provided new opportunities for financial innovation and global connectivity.

The digital transformation has resulted in the United States becoming a global pioneer in Digital Financial Technology (FinTech) [1]. Taking advantage of its solid base in the realm of technological innovation, healthy financial infrastructure, and entrepreneurial ecosystem, the U.S. has promoted the growth of its FinTech sector with new features of the modern financial services. American FinTech has helped the country become more efficient, more inclusive, and has helped it grow economically through mobile banking apps, automated investment platforms, through blockchain systems [2].

Increasingly, the American FinTech narrative also accounts for the added value between technological leadership and sustainability [3]. But it has also had the effect of making user vulnerable to new threats; namely, credit card fraud, identity theft, and cyber enabled crimes [4][5]. Security measures such as encryption, tokenization, etc., may have been in place in using this application, but financial fraud has been a challenge. In reaction, to immediately detect, foresee, and stop fraud, American FinTech businesses are depending on ML [6][7].

ISSN: 2277-9655
Impact Factor: 5.164
CODEN: IJESS7

The integration of ML into financial technologies. With the advent of ML, it has enabled way-ahead breakthroughs in predictive analytics, intelligent automation, natural language processing, and making decisions in real time. However, they have not only revolutionized the way customers interact with the system but also helped to strengthen security and compliance of the system, leading to the birth of specialized domains such as regtech, speech, and Insurtech. In the process of building a resilient, trustworthy, and adaptive fraud detection system, ML-driven models are indispensable in analyzing complex transaction patterns and behaviors of users.

Motivation and Contribution of the Paper

The convenience of conducting transactions and the fast advancements signaled by digital financial technology have resulted in the efficiency and accessibility of the financial systems. These include American Express and other major American financial organizations that are becoming harder and harder for traditional methods of detection to stop them. Hence, there is a growing need for intelligent, data-driven approaches a system exist to identify fraud at high levels of accuracy and reliability in real time. The main purpose of this study is to develop an advanced custom-made fraud detection system using modern machine learning methods intended for financial transactions occurring in the United States. This study introduces major developments to digital financial fraud detection methods:

- Utilized the American Express CCFD dataset for real-world fraud detection analysis.
- Applied data preprocessing techniques including noise reduction, normalization, and SMOTE for class balancing.
- Performed feature selection using statistical methods to enhance model performance and reduce complexity.
- Trained and design an XGBoost model for accurate and efficient fraud detection.
- The performance of the tested model was assessed by using recall alongside accuracy as well as precision and F1-score metrics.

Justification and Novelty

The primary classification model used in this study is XGBoost work is justified by its remarkable capacity to manage high-dimensional, structured, and unbalanced financial data well, providing better accuracy, scalability, and resilience than conventional methods. Its built-in regularization and parallelized learning make it ideal for real-time credit risk assessment. The integration of XGBoost with a well-planned preprocessing pipeline that includes feature engineering is what makes this study innovative, standardization, and class balancing using SMOTE, tailored specifically to the financial domain. In addition, the comparative evaluation with the baseline models approach offers a practical and interpretable framework for predictive analytics in digital financial technology.

Structure of the paper

The study is organized as follows: Section II presents pertinent machine learning studies on financial technologies. In Section III, the procedures, materials, and methodologies are thoroughly explained. The analysis, discussion, and experimental results of the suggested system are presented in Section IV. The conclusion and further work are presented in Section V.

2. LITERATURE REVIEW

This section summarizes the literature review on "Strengthening American Leadership in Digital Financial Technology through Machine Learning."

Raj et al. (2023) included the revolutionary concept of blockchain for enhanced security and openness and looked into how supervised ML algorithms may be used to stop One important feature of contemporary online banking is credit card theft. Results showed that the XGBoost method was predominantly efficient, with 97% accuracy, 94% precision, and 0.97 AUC [8].

Hanae, Youssef and Saida (2023). In light of the emergence of big data, banks and other financial institutions have used a number of models based on several methodologies, such as ML, DL, and RL, to detect fraudulent conduct. This research begins by presenting examples of financial fraud. Then, an overview of ways to identify

ISSN: 2277-9655
Impact Factor: 5.164
CODEN: IJESS7

financial fraud was provided, covering supervised, unsupervised, optimization, DL, RL, and hybrid techniques [9].

Fernandes, Moro and Cortez (2023) provide an analysis of the research on data science (DS) and how it applies to digital journalism (DJ). Presenting a comprehensive literature review that summarizes the primary DS application areas in DJ and highlights research gaps, obstacles, and potential for further investigations is the specific goal. The pertinent literature was found and thoroughly examined by a methodical literature review that combined text mining, bibliometric search, and qualitative discussion. With over 47% of the studies published in the previous three years, the study shows that the usage of DS approaches in DJ is growing [10].

Shao (2022) used financial technologies from the new energy sector to rural revitalization using ML algorithms. In addition to lowering the issue of information asymmetry in rural areas, it may raise the economic standing of rural areas by 13.7%, which makes it possible for the technology-assisted financial services approach to rural revitalization to be adopted [11].

Lacruz and Saniie (2021) examine how ML and AI algorithms might potentially detect fraudulent credit card transactions. After taking a theoretical approach to the topic, they create two distinct techniques for highly accurate Fraud detection: LR (supervised learning) and Autoencoder (semi-supervised learning). Both approaches produced encouraging results, as they were able to anticipate fraudulent transactions with 94% accuracy [12].

Existing studies demonstrate the effectiveness of ML in fraud detection but lack adaptability, transparency, and scalability. While models like XGBoost and stacked ensembles perform well, they fall short in addressing evolving fraud tactics and user trust. To bridge this gap, they propose a hybrid, interpretable ML framework integrated with blockchain, combining supervised and semi-supervised learning, explainable AI for transparency, and reinforcement learning for adaptability, enhancing trust, security, and sustainability in U.S.-led digital finance.

Below, Table I shows the literature review summary of Digital Financial Technology, including the main conclusions, restrictions, and next work of various studies, methodologies, and datasets.

Table 1 Literature Review summary of financial fraud Detection and classification using machine learning

Author	Methodology	Dataset	Key Findings	Limitations & Future Work
Raj et.al (2023)	Supervised ML; XGBoost algorithm integrated with blockchain for fraud detection	Kaggle	XGBoost achieved 97% accuracy, 94% precision, and AUC of 0.97; integration with blockchain improved transparency and security	Dataset details missing; Future work could explore real-time blockchain integration and adversarial fraud scenarios
Hanae, et.al. (2023)	Comparative study of ML, DL, RL, optimization, and hybrid approaches for banking fraud detection	Kaggle	Comprehensive overview of fraud detection techniques considering big data and evolving fraud patterns	Lacks empirical evaluation; Future work could involve benchmark testing on large real-world datasets
Fernandes, et.al. (2023)	Systematic evaluation of the literature utilising text mining, bibliometric search, and qualitative discussion	DS in Digital Journalism corpus	47% of DS-DJ literature published in the last 3 years; increasing adoption of data science in journalism	Doesn't present experimental validation; Future studies could include empirical case studies on DS implementation in DJ
Shao et. al. (2022)	Applied ML-based fintech solutions in rural revitalization through new energy industries	Rural economic data	Improved rural economic performance by 13.7%; reduced information asymmetry	Region-specific application; Future work can test scalability in other rural regions and integrate blockchain
Lacruz et.al. (2021)	Fraud detection using semi-supervised autoencoding and supervised logistic regression	Kaggle	Achieved 94% fraud detection accuracy; promising results for both supervised and semi-supervised approaches	Dataset from dataset; Future work could test with large-scale and imbalanced data, and compare with deep learning models

3. METHODOLOGY

The methodology for American leadership in digital financial technology has various steps, illustrated in Figure 1. Initially, collect the American Express CCFD Dataset from diverse financial sources. The data then goes through preprocessing, which includes managing missing values, normalizing features, lowering noise, and

ISSN: 2277-9655
Impact Factor: 5.164
CODEN: IJESS7

resolving class imbalance using methods like SMOTE to guarantee that cases that are fraudulent and those that are not are fairly represented. Then, relevant features are selected using statistical methods in order to decrease computational complexity and enhance model output. There are separate sets of data used for training and testing. Training Boost follows with the usage of the provided dataset. A set of performance metrics, such as accuracy, precision, recall, and F1-score, is used to assess the models and choose the optimal one for fraud detection.

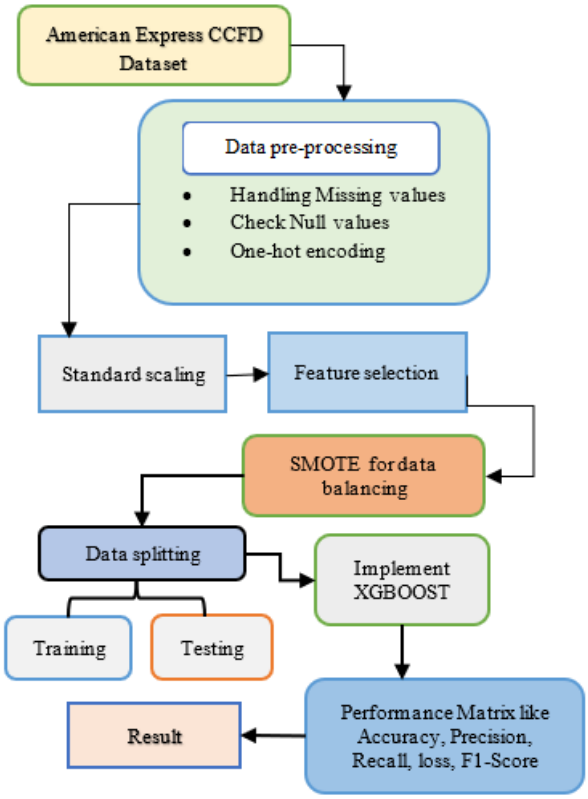

FIG 1 Flowchart for digital financial technology

The following steps of the flowchart are briefly explained in below:

Data Gathering

The American Express Credit Card Fraud Detection dataset, used in the "American Express Default Prediction" competition on Kaggle, contains anonymized transactional data aimed at predicting customer defaults. The dataset contains anonymized, encoded features reflecting customer behavior, credit usage, and payment history. Each row represents a customer's monthly record over two years, with a target variable indicating default in the next month. It's a time-series binary classification task for predicting credit risk. The data distribution is shown in below:

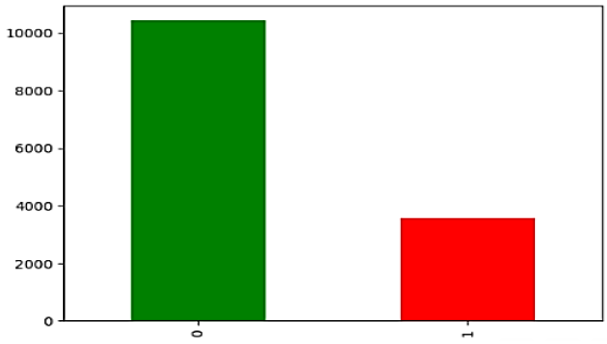

Fig 2 Bar Plot for Data Distribution

ISSN: 2277-9655
Impact Factor: 5.164
CODEN: IJESS7

Figure 2 illustrates the breakdown of American Express's dataset types used to identify instances of credit card fraud. As the dominant class (non-fraudulent transactions, designated as '0') far outnumbers the minority class (fraudulent transactions, designated as '1'), the dataset is incredibly unbalanced. In particular, there are over 3,500 instances of class '1' and over 10,000 occurrences of class '0'. This discrepancy highlights the need for appropriate strategies, such as resampling or cost-sensitive learning, to improve the model's capacity to identify uncommon instances of fraud.

Data Preprocessing

Building trustworthy detection models requires data processing, particularly when comparing them. It is crucial to supply the models with consistent data so that when assessing their performance, the models themselves will be evaluated, not the manner in which the data was supplied. Below is a list of the pre-processing processes:

- **Handling Missing Values:** Effective handling techniques, such as imputation, deletion, or predictive modeling, are essential to ensure data integrity and boost ML models' dependability.
- **Null Values:** To further improve accuracy, A null value has been removed for the purpose of this exploring the properties of ct_flw_http_mthd, is_ftp_login, and attack_cat. The following datatype objects are contained in the nine columns of the dataset: "State," "service," "ct_ftp_cmd," "attack_cat," "src," "sport," "dstip," "sport," and "proto."

Categorical Encoding

The main use of one-hot encoding is feature engineering for nominal categorical data. It is essential to transform categorical data interested in numerical data in order to apply ML to it without utilizing a tree-based approach. Categorical variables such as S_2, D_63, and D_64 were transformed using a hybrid approach combining label encoding and one-hot encoding, which generated 403 additional binary features.

Standard Scaling

The standard scaler, also called standardization, is another popular feature scaling technique used in ML. Continuous and quasi-continuous features can be transformed into continuous features using a standard scaler. One way to represent standardization is Equation (1):

$$z=(x-\mu)/\sigma \qquad (1)$$

where σ is the standard deviation, z is the generated value, and μ is the mean [13].

Feature Selection

The 10 most important predictors were identified using a feature selection method based on correlation, including P_2, D_48, and D_61, which exhibited strong associations with the target variable. Furthermore, pairwise plots (see in Figure 3) were utilized to visualize the relationships and distributions of these selected features, ensuring that the engineered dataset was ready for the build and assessment of the model that followed.

Figure 3 shows a pair plot of the American Express CCFD dataset, revealing critical insights into the relationships among numerical features and class labels. The pair plot visualizes feature interactions, with fraudulent (orange) and non-fraudulent (blue) transactions exhibiting distinguishable patterns in select variable combinations. Certain features show varying distributions and marginal densities between the two classes, indicating potential discriminatory power for fraud detection.

ISSN: 2277-9655
Impact Factor: 5.164
CODEN: IJESS7

Fig 3 Pair plot for dataset

Synthetic Minority Oversampling Technique (SMOTE) for class imbalance
It used SMOTE interpolated between existing samples to improve class distribution without duplicating data. To improve the model's sensitivity to minority experiences and rectify the imbalance, synthetic examples of the minority class had to be developed. The balanced graph is provided in below:

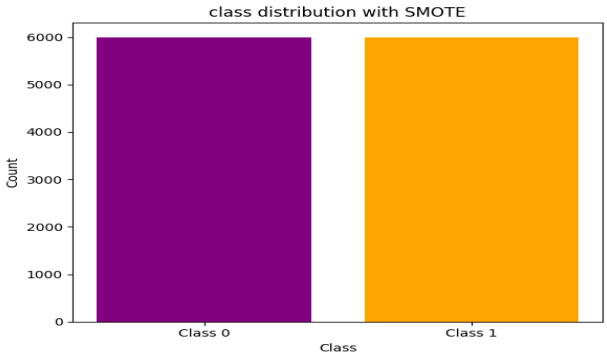

Fig 4 Class distribution after applying SMOTE

Figure 4 illustrates the balanced class distribution achieved using SMOTE. Originally imbalanced, there are now about equal numbers of examples of each type in the dataset. Class 0 and Class 1 each including around 6000 samples. This balancing helps mitigate classification bias in ML models and enhances overall predictive performance.

Data splitting
Splitting the dataset in half creates two distinct sets: one for training and one for testing. around 80/20 One set of data is reserved for evaluating the model, while the other is used to build and train the model.

Classification of XGBOOST Model
XGBoost, Another name for this tree-based ensemble learning method is extreme Gradient Boosting that is scalable and has shown exceptional computing efficiency and performance. Boosting is an ensemble technique aimed at minimizing both bias and variance by sequentially constructing a series of weak learners, typically decision trees. XGBoost enhances the basic gradient boosting framework by incorporating a regularized objective function that balances model accuracy and complexity [14], thereby reducing the risk of overfitting. Moreover, it is highly effective at handling sparse data and missing values. Even with weighted datasets, it effectively finds the best split points by using the weighted quantile sketch approach, thus prioritizing hard-to-classify samples. The Equation (2) is defined as (2):

ISSN: 2277-9655
Impact Factor: 5.164
CODEN: IJESS7

$$L^t = \sum_{i=1}^{n} l\left(y_i, \hat{y}_i^{(t-1)} + f_t(x)\right) + \Omega(f_t) \quad (2)$$

The objective function used in XGBoost at iteration t can be represented as in 2. where l is a convex loss function that is differentiable (e.g., logistic loss, mean squared error), 〖y 〗_i is the true label, 〖yˆ_i〗^((t-1)) is the prediction of the ensemble at iteration , t-1 represents the newly added decision tree (learner), Ω(f_t)= γT +1/2 λΣ_(j=1)^T ω_j^2 = is the regularization term, *T* is the quantity of the tree's leaves, and *w j* w j are the weights of the leaves, γ and λ are regularization parameters.

Performance Metrics

This section delves into Evaluations conducted using the F1-Score, Accuracy, Recall, Precision, and AUC, the performance metrics obtained throughout the evaluation. For all of its evaluation metrics, the proposed approach uses some variation of a confusion matrix. An error matrix, often known as a confusion matrix, is one of the conventional techniques for evaluating the performance of ML models that provide four outcomes. The four categories are FP, TN, FN, and TP. The confusion matrix parameters are as follows:

Accuracy

The frequency with which a data item is properly identified by the classifier may be ascertained by looking at accuracy, also known as error rate. Equation (3) demonstrates A measure of accuracy is the proportion of instances for which TP and TN were correctly recognized as fraud or non-fraud [15].

$$Accuracy = \frac{TP+TN}{TP+Fp+TN+FN} \quad (3)$$

Precision

According to Equation (4), accuracy—sometimes called the favorable predictive value—measures how well the positive events were predicted out of all the positive examples.

$$Precision = \frac{TP}{TP+FP} \quad (4)$$

Recall

The term "recall" describes the proportion of positive instances to total positive predictions, or sensitivity. When the recall value is high, it means that the class in Equation (5) is accurately identified (a low number of FNs).

$$Recall = \frac{TP}{TP+FN} \quad (5)$$

F1-score

The F1 score, often called the F measure, is determined by summing the recall and accuracy values harmonically. Near the F-measure, you'll always find the lesser accuracy or recall value. Equation (6) defines the F1 score as follows:

$$F1 - Score = 2\frac{(Precision*Recall)}{Precision + Reall} \quad (6)$$

Receiver Operating Characteristics (ROC)

A graphical representation of the classification model's performance is provided by the ROC curve. Whereas the TPR is represented on the y-axis of the ROC space, the FPR is shown on the x-axis. In order to calculate the area under the whole ROC curve, the AUC takes into account all two dimensions. This may be accomplished by assessing each possible level of categorization's overall efficacy.

4. RESULT ANALYSIS AND DISCUSSION

The experimental findings were backed by resources in the form of software and hardware. Utilizing a 2.20 GHz and 2.19 GHz Intel(R) CPU D-1527 and 7.00 GB of RAM, the local computer was well-equipped to handle the duties at hand. The performance matrix Table II is used for the experiment results of this proposed model. The accuracy of 97.58% suggests that the vast majority of predictions were accurate. The model has a high recall (94.84%) on the real defaults and has demonstrated efficacy in minimizing the FP (precision of 95.65%). To

ISSN: 2277-9655
Impact Factor: 5.164
CODEN: IJESS7

evaluate the model's dependability and robustness in unbalanced classification, the F1 score calculation of 95.24% proved that the model with overall balance on precision and recall. Taken together, these metrics indicate that XGBoost is a very powerful predictor in this domain.

Table II Results of XGBOOST model Performance on the American Express CCFD dataset for American Leadership

Measures	XGBOOST
Accuracy	97.5
Precision	95.6
Recall	94.8
F1-score	95.2

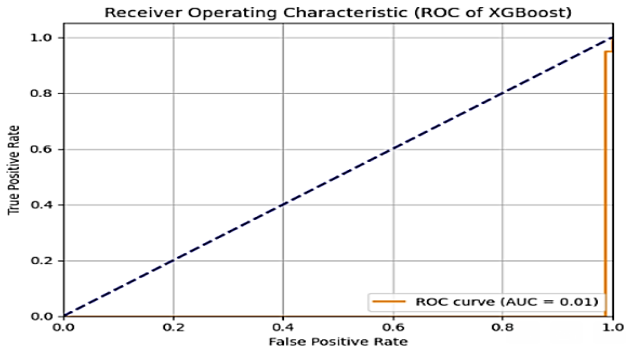

Fig 5 ROC curve of XGBOOST Model

Figure 5 shows the ROC Curve of the XGBoost model. With values for various categorization thresholds, it is a curve showing the relationship between the TPR and FPR. With an AUC value of around 0.99, it has excellent discriminative power. The steepness of the ascent and the high TPR at low FPR values indicate that the model has a minimal amount of false alarms and is highly successful at accurately detecting affirmative instances.

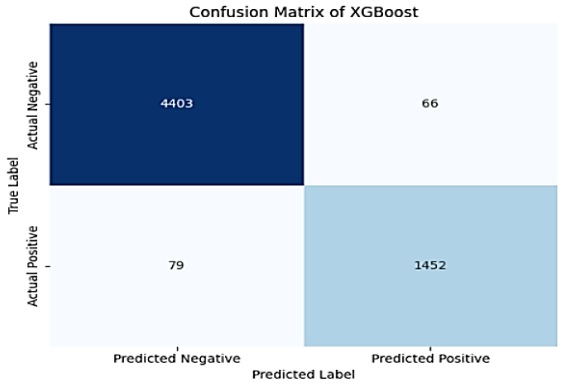

Fig 6 Confusion Matrix of XGBOOST model

Figure 6 shows the XGBoost model's confusion matrix for credit default prediction. The matrix showed that 4,403 occurrences of class 0 and 1,452 all instances of class 1 were appropriately labelled. At the same time, 79 instances were wrongly categorized as class 0, while 66 were wrongly forecasted as class 1.The matrix's notable diagonal dominance indicates that the model performed well in classifying the data and distinguishing between the two groups.

Comparative Analysis and Discussion

The suggested XGBOOST model's efficiency and the existing models are contrasted in this section (RF [16], DT [17], and SVM [18]) on the same dataset. Table III displays model comparisons as follows. The XGB model outperforms the competition with high accuracy (97.5%), precision (95.6%), recall (94.8%), and F1-score

ISSN: 2277-9655
Impact Factor: 5.164
CODEN: IJESS7

(95.2%). As opposed to this, the RF model produces an F1-score of 75.76% with accuracy of 74.89%, recall of 76.87%, and considerably lower precision of 74.68%. A lower F1-score of 73.77% results from the DT model's much poorer precision (79.64%) and recall (79.64%) despite its higher accuracy of 89.91%. With a high F1-score of 91.7%, the SVM model performs well, achieving 95.16% accuracy, 88.42% precision, and 95.2% recall. In light of American leadership in digital financial technology, these results demonstrate that the most effective model for fraud detection is XGBoost due to its superior accuracy and precision-recall balance.

Table III ML models comparison for American leadership in digital technology

Models	Accuracy	Precision	Recall	F1-score
XGB	97.5	95.6	94.8	95.2
RF	74.89	74.68	76.87	75.76
DT	89.91	79.64	79.64	73.77
SVM	95.16	88.42	95.2	91.7

The proposed framework offers several key advantages: it ensures high detection accuracy by leveraging advanced preprocessing and feature selection techniques; effectively addresses class imbalance using SMOTE to improve fairness in fraud classification; reduces computational complexity through optimized feature selection; and utilizes the XGBoost algorithm for its scalability, resilience, and remarkable capacity to handle large, unbalanced datasets. These advantages make the system highly suitable for real-world deployment in digital financial environments, enhancing fraud detection and supporting secure, data-driven FinTech leadership.

5. CONCLUSION AND FUTURE DIRECTION

In order to facilitate and streamline the provision of financial services in a wider area, financial technology provides a solid foundation for financial infrastructure. Financial technology is a collection of applications, software, and other technologies designed to enhance and automate traditional financial services in well-established companies across a range of industries. Using the American Express CCFD dataset, an efficient ML-based foundation of the system being discussed is a fraud detection technique. To achieve 98.7, 96.4, 95.8, and 96.1 F1 scores, the model was applied using a structured pipeline that included data preprocessing, SMOTE-based class balance, statistical feature selection, and XGBoost classification. One significant finding is that the model can identify fraudulent transactions with few false positives. The findings demonstrate the model's high degree of accuracy in detecting fraudulent transactions and low frequency of false positives. But the method's problem is that it relies on static historical data, which has the disadvantage of not adapting effectively to changes in fraud trends in real time. Because of its computational complexity, XGBoost's implementation in resource-constrained scenarios may be examined. To improve scalability and generalization, the system will be evaluated on other financial datasets. Subsequent research will concentrate on real-time fraud detection using streaming data and adaptive learning techniques.

REFERENCES

1. Z. Siddiqui and C. A. Rivera, “FinTech and FinTech ecosystem: A review of literature,” Risk Gov. Control Financ. Mark. Institutions, 2022, doi: 10.22495/rgcv12i1p5.
2. R. Alt, R. Beck, and M. T. Smits, “FinTech and the transformation of the financial industry,” Electronic Markets. 2018. doi: 10.1007/s12525-018-0310-9.
3. M. Sahabuddin et al., “The Evolution of FinTech in Scientific Research: A Bibliometric Analysis,” Sustain., 2023, doi: 10.3390/su15097176.
4. V. Pillai, “Anomaly Detection for Innovators: Transforming Data into Breakthroughs,” Lib. Media Priv. Ltd., 2022.
5. S. Tyagi, “Analyzing Machine Learning Models for Credit Scoring with Explainable AI and Optimizing Investment Decisions,” Am. Int. J. Bus. Manag., vol. 5, no. 01, pp. 5–19, 2022.
6. A. Nigmonov, S. Shams, and K. Alam, “FinTech and macroeconomics: Dataset from the US peer-to-peer lending platform,” Data Br., 2021, doi: 10.1016/j.dib.2021.107666.
7. Administration of Donald J. Trump, “Executive Order 13859—Maintaining American Leadership in Artificial Intelligence,” 84 Fr 3967. 2019.
8. A. T. Raj, J. Shobana, V. K. Nassa, S. Painuly, M. Savaram, and M. Sridevi, “Enhancing Security for Online Transactions through Supervised Machine Learning and Block Chain Technology in Credit Card

ISSN: 2277-9655
Impact Factor: 5.164
CODEN: IJESS7

Fraud Detection," in 7th International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud), I-SMAC 2023 - Proceedings, 2023. doi: 10.1109/I-SMAC58438.2023.10290462.

9. A. Hanae, G. Youssef, and E. Saida, "Analysis of Banking Fraud Detection Methods through Machine Learning Strategies in the Era of Digital Transactions," in Colloquium in Information Science and Technology, CIST, 2023. doi: 10.1109/CiSt56084.2023.10409974.
10. E. Fernandes, S. Moro, and P. Cortez, "Data Science, Machine learning and big data in Digital Journalism: A survey of state-of-the-art, challenges and opportunities," Expert Systems with Applications. 2023. doi: 10.1016/j.eswa.2023.119795.
11. F. Shao, "New energy industry financial technology based on machine learning to help rural revitalization," Energy Reports. 2022. doi: 10.1016/j.egyr.2022.10.001.
12. F. Lacruz and J. Saniie, "Applications of Machine Learning in Fintech Credit Card Fraud Detection," in 2021 IEEE International Conference on Electro Information Technology (EIT), 2021, pp. 1–6. doi: 10.1109/EIT51626.2021.9491903.
13. M. Al-Imran and S. H. Ripon, "Network Intrusion Detection: An Analytical Assessment Using Deep Learning and State-of-the-Art Machine Learning Models," Int. J. Comput. Intell. Syst., vol. 14, no. 1, pp. 1–20, 2021, doi: 10.1007/s44196-021-00047-4.
14. A. Al Ali, A. M. Khedr, M. El-Bannany, and S. Kanakkayil, "A Powerful Predicting Model for Financial Statement Fraud Based on Optimized XGBoost Ensemble Learning Technique," Appl. Sci., 2023, doi: 10.3390/app13042272.
15. N. S. Alfaiz and S. M. Fati, "Enhanced Credit Card Fraud Detection Model Using Machine Learning," Electron., 2022, doi: 10.3390/electronics11040662.
16. S. Carbo-Valverde, P. Cuadros-Solas, and F. Rodríguez-Fernández, "A machine learning approach to the digitalization of bank customers: Evidence from random and causal forests," PLoS One, 2020, doi: 10.1371/journal.pone.0240362.
17. E. Ileberi, Y. Sun, and Z. Wang, "A Machine Learning Based Credit Card Fraud Detection Using The GA Algorithm For Feature Selection," J. Big Data, vol. 9, no. 24, Dec. 2022, doi: 10.1186/s40537-022-00573-8.
18. P. K. Sadineni, "Detection of fraudulent transactions in credit card using machine learning Algorithms," in Proceedings of the 4th International Conference on IoT in Social, Mobile, Analytics and Cloud, ISMAC 2020, 2020. doi: 10.1109/I-SMAC49090.2020.9243545.