IJIRSET
International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY

e-ISSN: 2319-8753 | p-ISSN: 2347-6710

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 12, Issue 10, October 2023

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.423

9940 572 462 | 6381 907 438 | ijirset@gmail.com | www.ijirset.com

|e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 10, October 2023 ||

| DOI:10.15680/IJIRSET.2023.1210100 |

Enhancing Security and Compliance in Fintech through Microservices: A Blockchain - Based Approach

Sumit Bhatnagar[1], Roshan Mahant[2]

Vice President, New Jersey, USA[1]

Senior Software Consultant, Texas, USA[2]

ABSTRACT: The fintech industry leverages technology to enhance and automate financial services, offering faster, more efficient, and accessible alternatives to traditional banking systems. Credit card fraud and data breaches are major problems in the rapidly expanding fintech industry and can result in huge losses for businesses as well as damage to their reputations and legal standing. In response to these problems, this study presents a new approach that utilizes blockchain technology: GBDT-APBT, which stands for Gradient Boosting Decision Tree-Based Anti-Fraud Protection. To combat credit card fraud, this methodology combines gradient boosting with blockchain technology, which guarantees transparency, independence, anonymity, and immutability while decentralizing the detection process. Training offline at the local blockchain node, the GBDT-APBT system uses each user's private data to improve weak classifiers, with anti-fraud transaction modeling being the approach it follows. The next step is to save the trained model to the cloud. Then, using voting processes, they can obtain the final consensus model. The GBDT-APBT method improves financial system security and surpasses current methods in detecting fraudulent transactions, according to experiments. This technology provides fintech companies with an exciting opportunity to tackle the increasing problem of credit card fraud in a decentralized and scalable way.

KEYWORDS: Blockchain Technology, Fintech, GBDT-APBT, Security

I. INTRODUCTION

The advent of internet banking has enabled banks to progressively open their doors to more people in the modern day. To facilitate the easy acquisition of goods and services in today's more competitive financial markets, electronic payment systems have grown indispensable. [1]. "Fintech" refers to technological advancements that automate or enhance financial services. By leveraging technology, fintech companies offer financial services that are more convenient, efficient, and economical compared to conventional banks [2]. A conceptual representation of the Fintech industry is shown in Figure 1 [3]. Companies in the financial technology sector provide consumers with the option of using cards instead of cash when making purchases. Among the many perks offered by credit cards is purchase protection, which protects buyers in the event that their items are damaged, lost, or stolen. Following the issuance of 1,023 million MasterCard cards during the first quarter of 2022, the subsequent quarters saw the issuance of 1,045 million, 1,061 million, and 1,034 million cards respectively. According to these statistics, the annual growth rate of MasterCard credit card issuance is showing no signs of slowing down, suggesting that card is gaining popularity among buyers. A total of 1.91 billion Visa and MasterCard credit, debit, and prepaid cards were in circulation as of December 31, 2022, according to The Nilson report. These numbers demonstrate how easy and commonplace card-based purchases have grown among consumers. Customers depend significantly on the ease that Fintech offers when it comes to bill payment, POS machine payments to merchants, and money transfers through digital wallets, credit/debit cards, Instant Bank Fund Transfer (IBFT), and Interbank Fund Transfer (IFT). There has been a dramatic increase in the counterfeiting of credit cards despite all these advantages for consumers. Because of the ease and rapidity with which large sums of money can be obtained using credit cards, they are a popular target for fraudsters [5]. Credit card fraud comes in many forms, including those that occur online, in physical stores, through applications, and even in counterfeit products. Credit card fraud can happen during online purchases, over the phone, or even when the cardholder isn't around. The use of stolen plastic cards in physical establishments is an example of offline counterfeiting. An extremely serious form of counterfeit is application counterfeit, which involves the fraudulent use of stolen or falsified personal information to obtain a credit card without any intention of paying it back. The fraudulent use of credit card information for online purchases is known as counterfeiting.

|e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 10, October 2023 ||

| DOI:10.15680/IJIRSET.2023.1210100 |

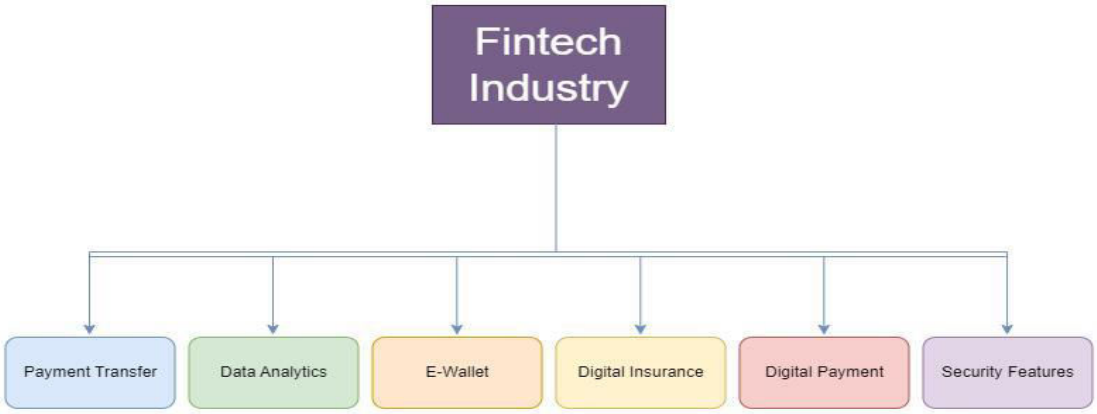

Figure 1. Fintech Industry

Phishing, skimming, and identity theft are just a few of the many techniques used by fraudsters, Nevertheless, their ultimate goal is the same: to steal money or information about individuals. Both clients and banks can lose a lot of money in these kinds of assaults. When customers' personal information is compromised, it can lead to the loss of their hard-earned money and damage the bank's reputation. Financial institutions should think about using AI-based systems that can detect and identify data breaches to reduce the likelihood of such incidents [6]. These days, ML-based models are widely used by banks to ward against this form of fraud. With the use of training data, ML models can identify fraudulent transactions. Banks are in a prime position to meet client demands thanks to their extensive branch networks. Customer information can be kept separate by each branch. For training purposes, it is necessary to have this information (dataset) in one central location. The potential of data breaches and the ensuing revelation of sensitive information about customers is now on the verge of becoming fully realized. To improve security and privacy while simultaneously reducing the risk of data breaches, it is possible to implement an additional layer of privacy and security through the use of FL. FL is a form of artificial intelligence that does not compromise the privacy of its users and has the ability to expedite the funding process through communication and collaboration [7]. FL does not anticipate data migration to a central repository, which protects data and reduces security threats. By combining AI and ML, fintech can improve efficiency, analyze data more effectively, assess risks, and provide individualized financial solutions. This will change the way the industry adapts and meets the demands of businesses and consumers.

Fintech with Blockchain - Based Federated Learning

Blockchains are used to construct a secure and unchangeable chain of data blocks and to record transactions that take place across a network of devices. This ensures that the blockchain is both transparent and unchangeable. An innovative method, blockchain-based FL integrates the trustworthiness and safety of blockchain technology with FL's advantages. There will be an unchangeable and certified record of all transactions and data transfers, and a huge number of different apps and entities can work on machine learning models without revealing their data thanks to this approach. In addition to enabling financial institutions to leverage machine learning for service improvement, these might aid in managing data privacy concerns and regulatory constraints [8].

By enabling financial firms to utilize blockchain technology, FL improves the precision and efficacy of financial models [9]. Financial institutions are able to communicate data and information more efficiently and accurately through blockchain-based FL. Financial institutions may build more precise models to better detect fraud, predict market movements, and enhance risk management by combining data from several sources.

Contribution

They show how smart contracts in blockchain-based FL are crucial for protecting privacy when training ML models on many decentralized peripheral devices as a whole, which is the main contribution of our paper. Here are the key points from this paper:

1. An improved fraud detection model, continuous learning, and data privacy are all features of the proposed federated learning system based on blockchain technology.
2. Establishing a privacy-preserving smart contract to govern the training and sharing of ML models.

|e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 10, October 2023 ||

| DOI:10.15680/IJIRSET.2023.1210100 |

II. LITERATURE SURVEY

An outline of recent developments in financial technology, with an emphasis on measures taken to protect financial systems from fraud. They go over the recent groundbreaking studies and strategies that have been used to beef up Fintech security. Additionally explore how to securely enable collaborative model training while protecting sensitive financial data via FL, a privacy-preserving approach. The emphasis here is on the noteworthy strides taken to safeguard Fintech systems and forestall counterfeiting. In 2017, Google initially proposed the idea of FL as a crucial push to aid data scientists in their job. This idea was conceived with the primary intention of providing assistance to data scientists and motivating them to pursue initiatives that are driven by data. Fl developed as a novel methodology that solved the difficulties of data privacy and security. It did this by encouraging cooperative model preparation while simultaneously decentralizing and restricting sensitive data. Because of this innovative design, which caused a disruption in the industry, data scientists were able to make use of the combined expertise and experiences of a variety of data sources without jeopardizing the confidentiality of individual data [10]. The necessity to safeguard consumer information is on the rise alongside the introduction of cutting-edge technological solutions. Applying a protective layer of FL across different ML models is one effective strategy. This makes it possible to more reliably and safely anticipate fraudulent transactions. The authors of [11] look into the potential of FL in open banking to safeguard data privacy and security and thwart Fintech fraud.

Without revealing any raw consumer data, it emphasizes FL's collaborative character and its potential advantages in enhancing and strengthening fraud detection. An overview of the development and influence of Fintech on the financial sector is given in the authors' study in [12]. It explains what Fintech is, how it works, the pros and pitfalls, and why regulations are needed to protect customers and keep the system stable. Digital payments, robo-advisors, blockchain, peer-to-peer finance, and mobile banking apps were among the other significant Fintech concepts and uses discussed by the author. Despite the improvement of development methodologies, software projects that are large, complicated, poorly specified, and involve unfamiliar aspects, are still vulnerable to large, unanticipated problems [13]. By enabling cooperative model training without disclosing raw data, the study highlights methods FL safeguards privacy. It went on to discuss how blockchain technology helps with privacy-preserving decentralized data protection. The authors of [14] focus primarily on personally identifiable information (PII) while discussing the significance of data protection in the financial technology sector. When it comes to cooperative, model-based training, data security and privacy protection, and the extraction of important insights from confidential financial data, they discuss and investigate the benefits of FL.

Blockchain Technology Based Client-Server Model

They have implemented a client-server paradigm using blockchain technology in the suggested framework. In this configuration, the server node acts as a central authority that controls and monitors various learning models. The client node, meanwhile, hosts a plethora of financial apps. Integrating numerous trained models without data exchange is one of the server's tasks. Problems with communication security and efficiency are among the concerns they bring to light, along with the need for robust encryption mechanisms and trust frameworks. To combat credit card theft, offer FL for theft Detection (FFD). With FFD, financial institutions can apply database data scattered across several locations to train fraud detection models. Considering that these locally calculated model updates are integrated to establish a shared Fraud Detection System (FDS), financial institutions are able to reap the benefits of a collective model without compromising the confidentiality of cardholder information during the process. The difficulties of detecting fraud in the Fintech industry using blockchain technology were discussed in [16]. Delays, scalability challenges, increased energy use, operational inefficiencies, and expenses are some of the potential problems covered in the article. Another recognized difficulty with ML is the intricate structure of training data, which raises privacy problems during data collecting. Identifying fraudulent charges on credit cards and how well ML algorithms distinguish between real and fake purchases are the main points of [17]. Using the Credit Card Fraud Detection dataset, they demonstrate how ML may help with a major problem: protecting customers' sensitive financial information from being stolen or lost.

|e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 10, October 2023 ||

| DOI:10.15680/IJIRSET.2023.1210100 |

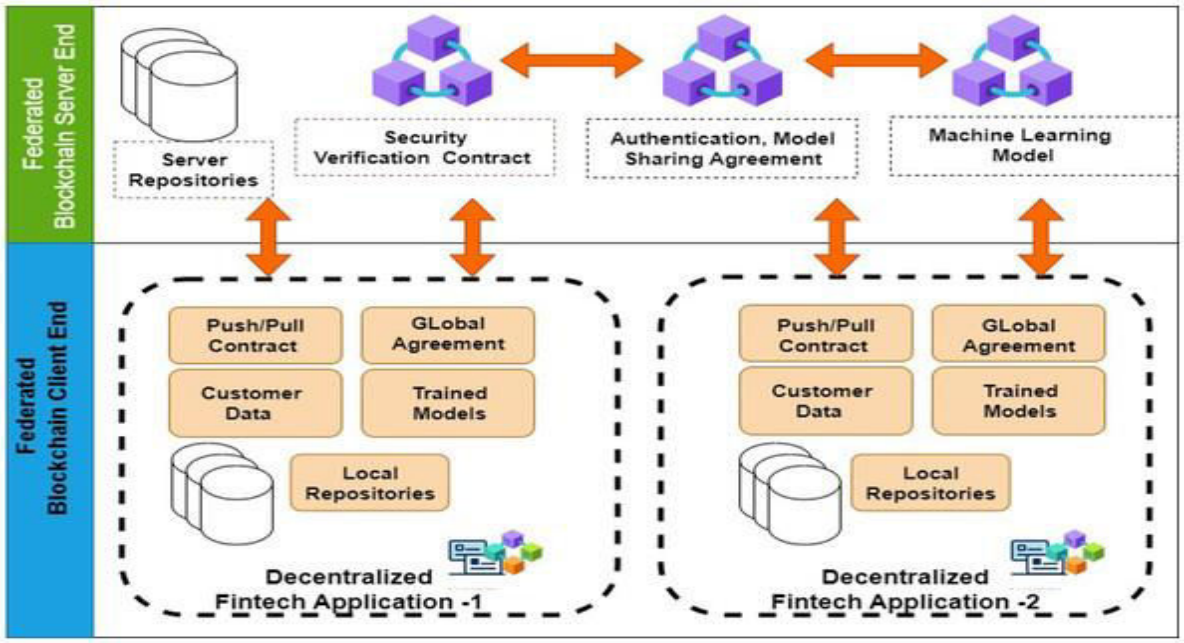

Figure 2 blockchain fintech architecture

III. METHODOLOGY

Every one of these parts is essential to the research process, and the technique as a whole rest on them. The following main components make up the methodology of this study:
1. Collecting datasets and initially analyzing them for patterns.
2. The application of machine learning models in the financial technology sector.

Dataset Collection & Exploratory Data Analysis
To safeguard customer privacy, we utilized a synthetic credit card transaction dataset from The Sparkov data generator was used to generate the Kaggle data, which was then categorized as either fraudulent or non-fraudulent. Over 1.8 million occurrences and over 20 categories are included in this dataset. These attributes include both numerical and categorical characteristics.

Data Collection
The initial step involved identifying reliable data sources, specifically a dataset from Fintech that includes credit card details. This dataset is divided into a training file with approximately 1.3 million instances and a test file with around 0.5 million instances.

Data Pipeline
Finding and comprehending the data source is the first step in the data pipeline. Data preparation, which includes cleaning and processing the data, follows. They then enter the model planning phase, when we establish goals, choose ML algorithms, and create the model's framework. The chosen algorithms are trained using the prepared data during the model-building process. The models are then deployed and integrated into real-world applications. Lastly, they bring stakeholders up-to-date on our findings by analyzing them.

Tools and Approach
Python is used for scripting, Pandas is used for data processing, Scikit-learn is used for machine learning models, and Seaborn is used for visualization. These are just some of the technologies that everyone use to efficiently handle data. To solve the issues that arise throughout the data collection process, a methodical and iterative strategy is utilized. The implementation of federated learning is accomplished through the utilization of Java Spring Boot, which includes both client modules and a federated server.

|e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 10, October 2023 ||

| DOI:10.15680/IJIRSET.2023.1210100 |

Machine Leaning Models

Machine learning relies on models, which are algorithms designed to learn from data and make judgments or predictions without human intervention [18; 26].

GBDT (Gradient Boosting Decision Trees): A method for building models in an ensemble whereby subsequent models fix mistakes caused by earlier ones. It works well for tasks involving classification and regression.

Bayes (Naive Bayes): A probabilistic classifier that bases its predictions on Bayes' theorem and assumes that the predictors are independent of one another. Detection of spam and text classification are two common applications of this technology.

DT (Decision Trees): A classification and regression model that resembles a tree. It uses feature values to partition data into branches, with decisions being made at each node.

SVM (Support Vector Machine): A supervised learning technique that locates the hyperplane in a high-dimensional space that most effectively differentiates between the many classes inside the space. It is useful for classification jobs that are linear as well as those that are non-linear.

KNN (K-Nearest Neighbors): A basic instance-based learning technique that uses the feature space's most populous class for categorization.

XGBoost (Extreme Gradient Boosting): Implementation of gradient boosting that has been tuned, making it both faster and more efficient, especially when used to big datasets. It includes regularization to reduce overfitting.

RF (Random Forest): A method that combines the predictions of numerous decision trees that are constructed using an ensemble approach. Precision is enhanced, and overfitting is prevented as a result.

DNN (Deep Neural Network): A neural network with multiple layers that can model complex patterns in data. It's versatile and can be applied to various tasks such as image recognition, speech recognition, and more.

GBDT-APBT (Gradient Boosting Decision Trees with Adaptive Pruning Based on Trees): A variation of GBDT that incorporates pruning techniques to enhance model performance by removing less significant splits.

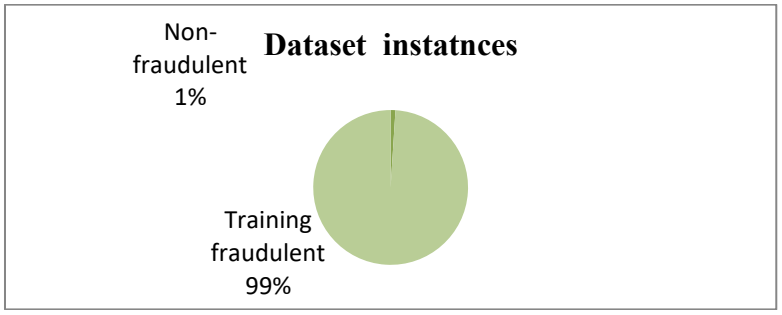

Figure 3 Splitting of Dataset

Data pre-processing

To guarantee high-quality data for further analysis, the data pre-processing stage seeks to clean the dataset by removing noise, outliers, and discrepancies. Standardization, normalization, and dealing with missing data are all part of this process. The Ordinal Encoder is used to convert ordinal integers from categorical data like merchant, location, and transaction date. Assigning a distinct integer to each category makes better use of memory, decreases dimensionality, keeps interpretability, and retains ordinal relationships than one-hot encoding. For the sake of keeping the training of the model away from features with larger magnitudes, MinMaxScaler is used to scale numerical features such that their values are contained within a specified range of 0 to 1. In order to solve the class imbalance that exists within the dataset, they employ the NearMiss under-sampling method to reduce the number of instances that belong to the majority class, which consists of non-fraudulent transactions. This is done in order to bring the distribution of the minority class, which consists of fraudulent transactions, into equilibrium. They opt for the NearMiss-1 variation because it ensures that a representative subset of the majority class is preserved for the purpose of conducting an

|e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 10, October 2023 ||

| DOI:10.15680/IJIRSET.2023.1210100 |

efficient analysis. To achieve this, they only use samples from the majority group that closely resemble the minority group.

Exploratory Analytics

EDA is used to investigate the underlying structure and relationships of the 405 data set after the data is cleaned and pre-processed. Features and models can be better informed by the insights it gives regarding data trends, correlations, and 406 patterns. A fraudulent transaction is recorded if it is determined that the transaction is not authentic. Figure 3 presents a comparison of the distribution of the target variable (fraud) with respect to the 408 dataset both before and after the implementation of under-sampling. According to the findings of a statistical study, the incidence of fraud is 410 times lower among males and 409 times higher among females. These findings are displayed in Figure 4, which describes the findings. In Figure 5, the frequency of fraudulent use of credit cards is displayed across a number of employment 411 categories. The data shown in the chart indicates that certain professions have a higher probability of being victims of forgery involving 412 cards than others. When compared to workers in other industries, those who are employed in the service and retail sectors, for example, have a risk that is 413 percent higher of being victims of an identity theft involving their credit card.

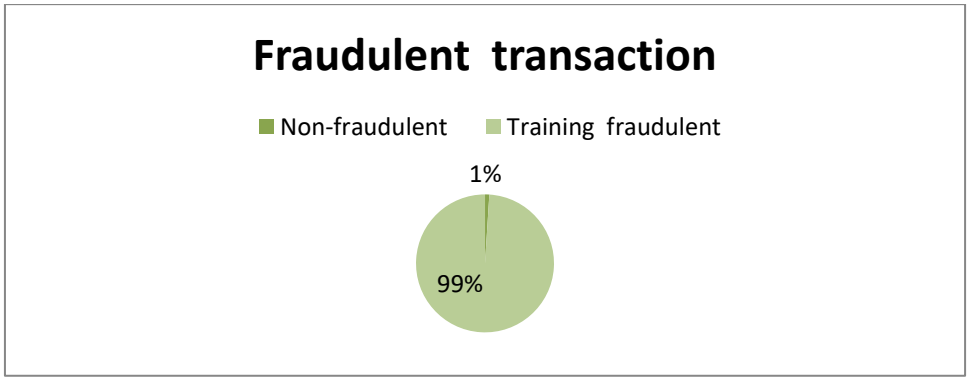

Figure 4 after sampling

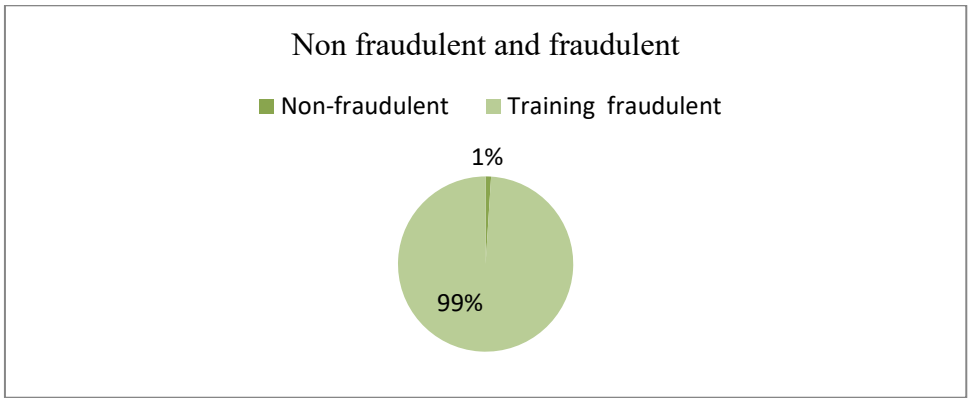

Figure 5 Non fraudulent and fraudulent

|e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 10, October 2023 ||

| DOI:10.15680/IJIRSET.2023.1210100 |

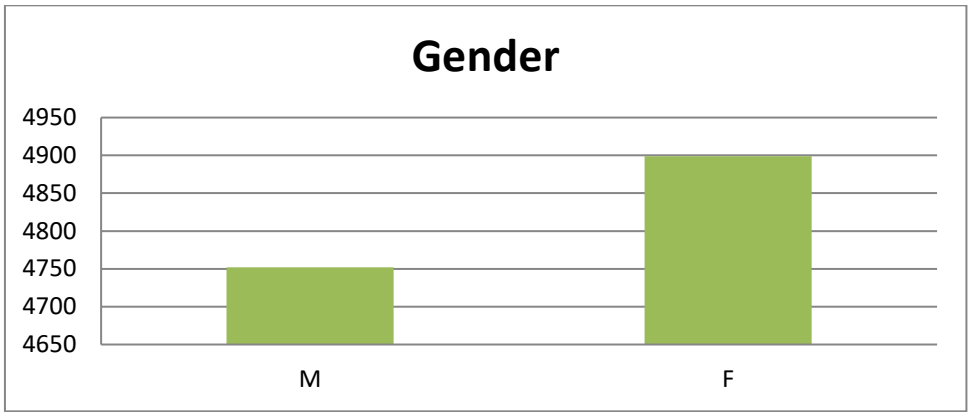

Figure 6. Gender Vs Fraud

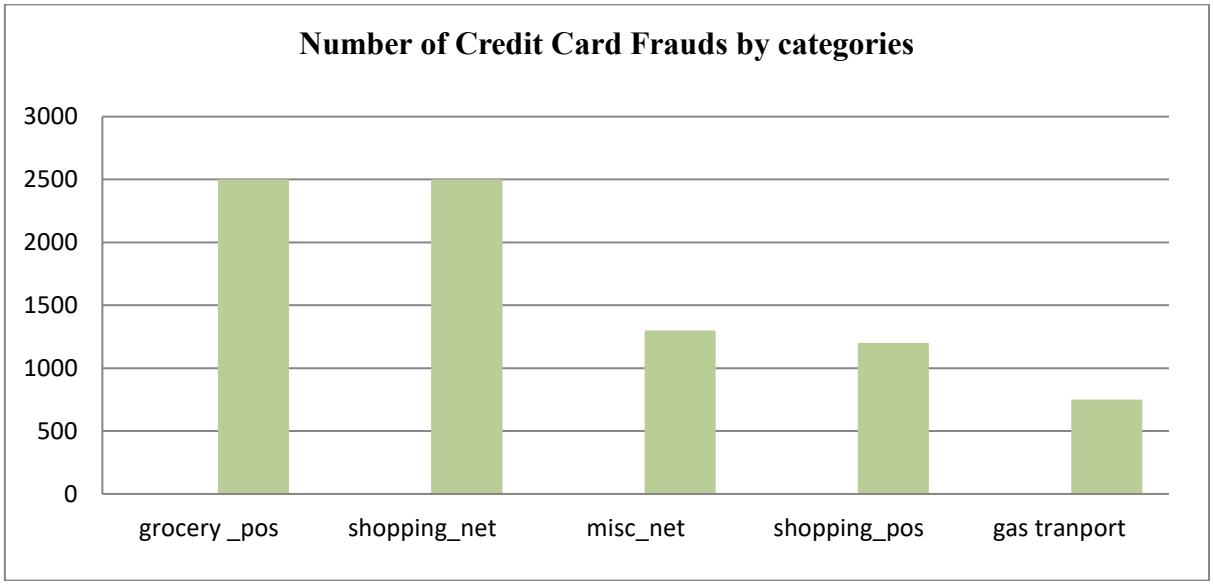

Figure **7** Credit Card Frauds by categories

IV. PROPOSED FRAMEWORK

In the proposed GBDT-APBT (Gradient Boosting Decision Tree-Based Anti-Fraud Protection with Blockchain Technology) model, microservices can significantly enhance its functionality and efficiency by enabling a modular architecture that allows for independent development, deployment, and scaling of system components such as data collection, model training, and fraud detection. This approach facilitates real-time data processing, ensuring quick identification of fraudulent transactions while also accommodating increased transaction loads through horizontal scalability. The decentralized nature of microservices aligns with blockchain technology, allowing each node to operate independently and contribute to fraud detection without a central authority, thus enhancing security and privacy. Furthermore, microservices support seamless integration with other fintech services, enabling collaboration and the utilization of external data sources. In order to keep client data secure, the decentralized federated server and clients update and synchronize their models over APIs. Client nodes, like banks, are protected against cyber-attacks that could compromise critical consumer data because this method only shares models and not raw data sets. Decentralized FL is a good solution to the problem of data leakage since it enables customers to train and run machine learning models on their own private data sets for training purposes. Within the framework of this architecture, they made use of GBDT-APBT, which is an acronym that stands for Gradient Boosting Decision Tree-Based Anti-Fraud Protection. This method guarantees that the suggested framework does not divulge any sensitive information. The federated server, which is the hub of the organization, only receives the trained models. Customers can request and incorporate updated models whenever needed because the federated server gathers and aggregates data about all the customers' trained models. Within this collaborative and secure framework, the privacy of customers is safeguarded, which enables the

|e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 10, October 2023 ||

| DOI:10.15680/IJIRSET.2023.1210100 |

sharing and improvement of models in an effective manner. Next, we'll go over the server and client nodes that house the smart contracts and the processes that execute them.

All updates to the global model are managed and controlled by the federated server that is based on the blockchain. In order to keep the trained models from blockchain-based federated clients consistent and up-to-date, the federated server needs to retrieve and synchronize the data from the client nodes. When a client submits its trained model to the server, the server ensures data privacy and uses security validation smart contracts to verify the client's identity.

A mechanism to prevent fraud and safeguard user information throughout the business. Users' data is no longer uploaded to a central database due to the employ blockchain technology to construct a distributed network architecture that breaks the old model of centralized databases. Most of this chapter is devoted to studying the model that banks use to detect credit card fraud. To address issues where there is a mismatch between the distribution of non-fraud transaction data and fraud transaction data in the sample and where the efficiency of individual classifiers is lower than that of combined ones? To help banks better detect fraudulent transactions, we provide a blockchain-based gradient-boosting decision tree expert classification model that can decide whether a given sample represents a fraudulent transaction and, if so, how to improve the model's classification effect. Figure 8 displays the overall model's structure. Computer program the establishment of a binary classification model with strong generalizability is typically required for anti-fraud transactions in order to ascertain whether or not they constitute fraud. It is a first for us to combine blockchain technology with the gradient-boosting decision tree. By putting model parameters on the blockchain, training consensus models no longer requires uploading user private data to the cloud, which is a huge plus. The local blockchain node finishes training each user's private data offline, and then the parameters of the model are sent straight to the cloud consensus model. The elimination of the prospect of user privacy data leaking and the logical necessity to threaten the model to verify it are both addressed physically.

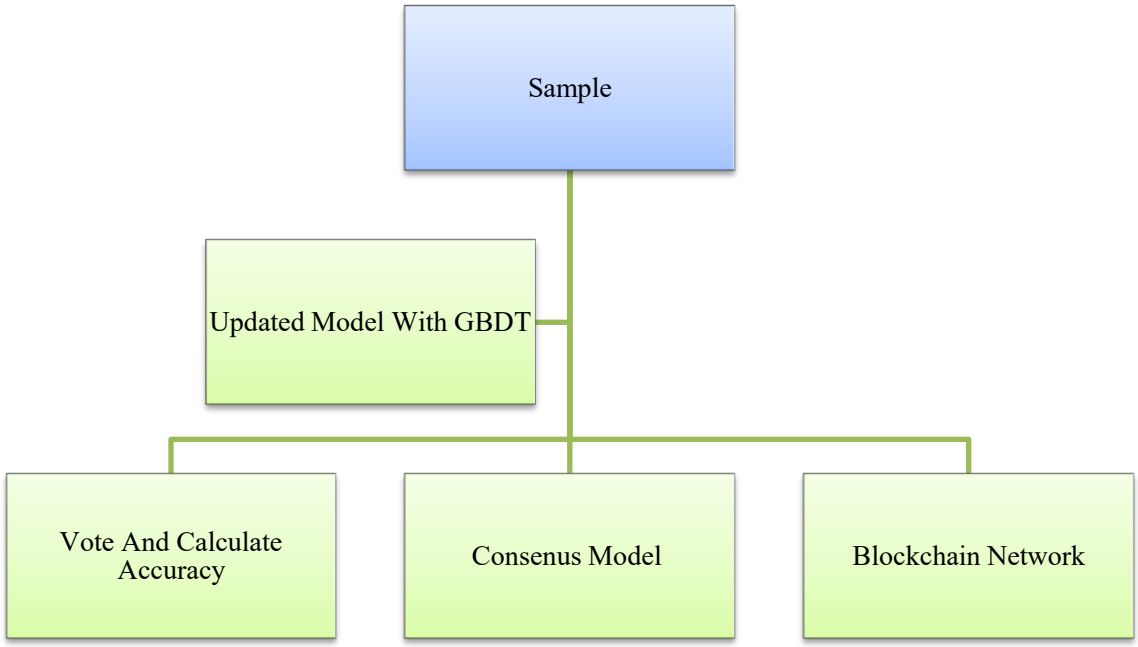

Figure 8 system structure

Figure 8 an overview of GBDT-APBT, followed by a consensus model that determines whether it is a fraud transaction. A model for preventing fraudulent transactions as a result of weak classifiers, represented as

$$\mathbf{f(x) = W_0 + \sum_{m=1}^{m} w_m \phi_m(x)}$$

|e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 10, October 2023 ||

| DOI:10.15680/IJIRSET.2023.1210100 |

In this case, w represents the weight and φ stands for the collection of weak classifiers. One way to look at the equation is as a product of linear functions. By combining the decentralized nature of blockchain with the residual removal of gradient-boosting decision trees, fraudulent transactions can be identified in real-time. They train models on a blockchain node-by-node basis, with each node training its own subset of the dataset before voting on a final model to form a consensus. As is the case in the real world, one can avoid uploading aggregated privacy data to the cloud by training each user's data locally and then selecting the best model parameters to upload to the consensus model. Rebuilding the anti-fraud transaction model in the direction of gradient residual reduction will get rid of the residual. The goal of each subsequent model construction in gradient boost—a technique that differs greatly from classic boost with respect to the weight of right and wrong samples is to reduce the residual of the previous model along the gradient direction. Since p is the collection of parameters and F(x, P) is the classification function, we may extend the addition function to the following equation:

$$F(X; \{\beta_{m,}a_m\}_1^M) = \sum_{m=1}^{m} \beta_m h_{(x;a_m)} \quad \text{Eq.1}$$

In this case, a = a1, a2,..., am stands for the classifier's partition variable, and h(x; a) is the base function, which is a singularly parameterized function for the input variable x. By optimizing the loss function, they are able to solve Equation (2) for the parameters {βm, am}.

$$\{\beta_{m,}a_m\}_1^M = \arg^{\min}_{\{\beta'_{m,}a'_m\}_1^m \; i=1} \sum_{I=1}^{N} L\left(y_{i,} \sum_{i=1}^{m} \beta'_m h((x; a_m)\right) \text{Eq.2}$$

The regression tree model, specifically the fraud transaction model, is subjected to the basis function, or GBDT. So, this is the how the fraud prevention model with addition looks:

$$h\,(x_{;}\{b_j, R_j\}_1^j) = \sum_{j=1}^{j} b_j\, 1(x \epsilon R_j) \quad \text{Eq.3}$$

Where γ represents the loss function:

$$\gamma_{jm} = \beta_m b_{jm} \quad \text{Eq.4}$$

In the given function, H represents the level of confidence and b is the value of the tree model's node. Algorithm.1 displays the GBDT-APBT algorithm. Inputted with a constant beginning value, the algorithm iteratively trains the decision tree to produce the categorization model.

Algorithm 1

Input:

- The loss function γ is minimized by setting an initial constant value.
- The sets of transaction data (x0, y0), the number of iterations (T), and the pairs of values (x1, y1) through (xN, yN)

Output: Transaction information model of the classification tree

Initialize

$f0(x)=\arg\min_{f0}\frac{1}{N}\sum_{I=1}^{N} L(y_i, \gamma)$

For t=1 to T:

For i=1,2,…,N

Calculate the negative gradient:

$$r_{it} = -\frac{\partial L(yi, ft-1(xi))}{\partial f(xi)}$$

End For.

Fit a classification tree to the targets r_{it} resulting in terminal regions Rjt for J =1,2,…,Jt.

For j=1,2,…,Jt

Update:

$\gamma_{jt}=\text{argmin}\sum_{x\in R_{jt}} L(y_i, ft-1(x_I)+\gamma)$

End For.

Update the model:

$f_{t(x)} = f_{t-1}(x)+\sum_{j=1}^{jt} \gamma_{jt} I(x \in R_{jt})$

|e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 10, October 2023 ||

| DOI:10.15680/IJIRSET.2023.1210100 |

End For.

Return $\hat{f}(x)=f_T(x)$.

In order to avoid fraud while using the GBDT classifier, it is required to estimate the constant value that minimizes the loss function. This function is a single-root tree. The model must be initialized before this can be done. To determine the residual estimate, Algorithm 1 takes the negative gradient of the loss function in the current model in line 5, of the first algorithm. Finding the residual estimation requires this action. Once that is accomplished, the area of the leaf node in Algorithm 1 should closely match the approximate residual value; next, the approach employs linear search to minimize the loss function. Utilizing all of the data samples is the final stage in the process of constructing a hierarchical classification tree model.

In order to classify data related to anti-fraud transactions, this article adjusts the GBDT tree. The bank issues the training request for the model when the network is initiated. Once the user gives their consent to join the training network, every client will have their first model, f0, shared. After receiving new data from the user (xi, yi), the model is updated accordingly. With y being the lable, x being the input, and f (x) being the output of the trained model, the client uses Equation (7) to determine the minimal loss function of GBTD locally.

Algorithm 2: Updating of GBDT-APBT

Input:

- Mean quantity of user nodes
- The following sets of transaction data are available: (x0,y0), (x1,y1),..., (xN,yN)
- Time iteration number T

1. **Initialize** the GBDT tree model $f_0$with an initial constant value to minimize the loss function.
    a. **For** each transaction n=0,1,2,…,N
    b. **For** each user node m=1,2,…, M:
        i. Update model parameters using Algorithm 1.
        ii. Update f_n to f_{n+1}{n+1}fn+1 with the sample (xn,yn)for user node m.
    c. **End For**
2. **For** each user node m=1,2,…,Mm = 1, 2, \ldots, Mm=1,2,…,M:
    a. Calculate the 0-1 loss function
    b. $L(y_n,f_m(x_n))$
    c. **End For**.
3. **For** each user node m=1,2,…,M
    a. Determine whether to vote for model f_m based on the results of the 0-1 loss function.
    b. **End For**.
4. **For** each user node m=1,2,…,M
    a. Calculate the accuracy $P(f_{nm})$
    b. **End For**.
5. Calculate max P(fm) and select the corresponding model fm as the consensus model.
6. The model parameters are uploaded to the blockchain network after being packed and signed by node m.
7. After the next model update, node m will leave.
8. End For

Return The completed model fn One can see the model in Figure 4. The procedure for updating the model is illustrated in Algorithm 2. Using Algorithm 1 as a point of reference, every user node updates its model for each loop. This is further illustrated in Figure 5.

$$L(y, f(x)) = \log(1 + \exp(-yf(x))) \quad \text{Eq.5}$$

Presumably, one of the nodes in the network can update the most recent model ft−1 at time t. The timestamp t is used to write the learned model parameters ft into the block. Nodes a, b, c, d, e, etc. in the network are authorized to verify the model parameters in the prior block according to the update method they developed. The values of ft(a), ft(b), ft(c), ft(d), ft(e), etc. can be derived from the data of individual nodes. The data pool is a collection of model parameters that each node packages and uploads. Nodes also vote on each other's models. Nodes determine which nodes to support by using the 0-1 loss function, as shown in Equation (8), where y is the label, x is the input, and ft(b)(x) is the output of node b's trained model at time t.

|e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 10, October 2023 ||

| DOI:10.15680/IJIRSET.2023.1210100 |

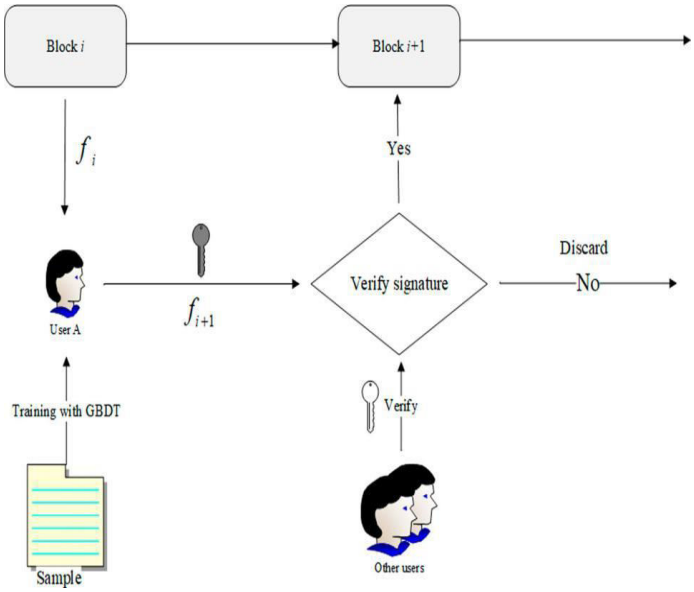

Figure 9 Transfer process.

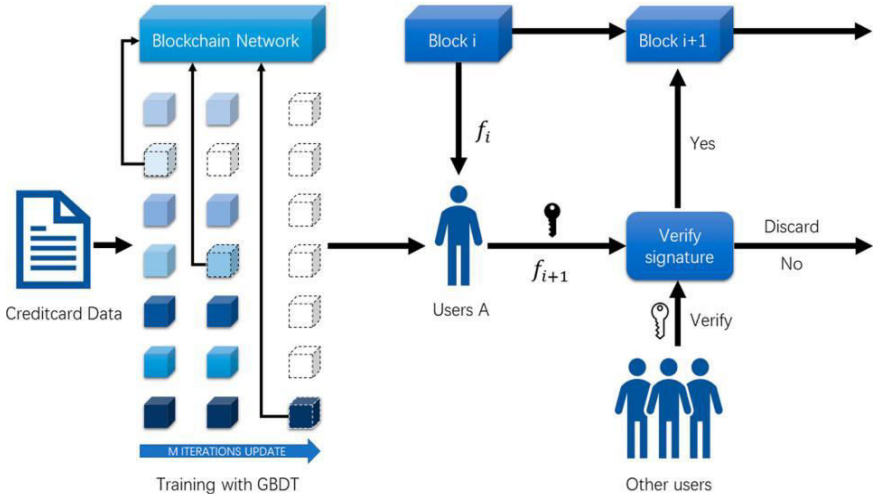

Figure 10 Credit card fraud detection frameworks

V. EXPERIMENT AND ANALYSIS

A system for detecting credit card fraud was put into place using a combination of gradient-boosting decision trees and decentralized blockchain. This method can be used to process data and train detection models in offline, imbalanced datasets, as well as in landed circumstances where user privacy is a concern and data cannot be transferred to the cloud. When learning the skewed data set, they compared the GBDT-APBT (anti-fraud prevention based on blockchain technology with gradient boosting decision tree) algorithm which updates the GBDT model on the blockchain to other important machine learning techniques. Decision trees, XGBoost (extreme gradient boosting), Random Forest, CNN, and DNN are among the other machine learning algorithms. Bayes and decision trees are also included.

|e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 10, October 2023 ||

| DOI:10.15680/IJIRSET.2023.1210100 |

Table 1 performance of predictions rate

Model	TN	FP	FN	TP	Prediction Rate
GBDT	85,069	78	10	138	93.24
Bayes	85,027	120	48	100	67.57
DT	78,960	6187	49	99	66.89
SVM	85,055	92	25	123	83.11
KNN	76,684	8463	52	96	64.86
XGBoost	85,279	16	37	111	75.00
RF	85,300	7	26	110	80.88
CNN	85,278	24	30	111	78.72
DNN	85,278	18	31	116	78.91

Table: 2 Performance Metrics for Machine Learning Models

Model	AUC-ROC (%)	AUC-PR (%)
GBDT	98.65	76.59
Bayes	96.11	68.39
Decision Tree (DT)	96.20	71.20
Support Vector Machine (SVM)	95.30	48.70
K-Nearest Neighbors (KNN)	96.28	63.10
XGBoost	98.20	87.10
Random Forest (RF)	97.23	80.25
Convolutional Neural Network (CNN)	98.23	95.23
Deep Neural Network (DNN)	91.25	92.36
GBDT with Adaptive Boosting (GBDT-APBT) (Proposed)	90.23	92.25

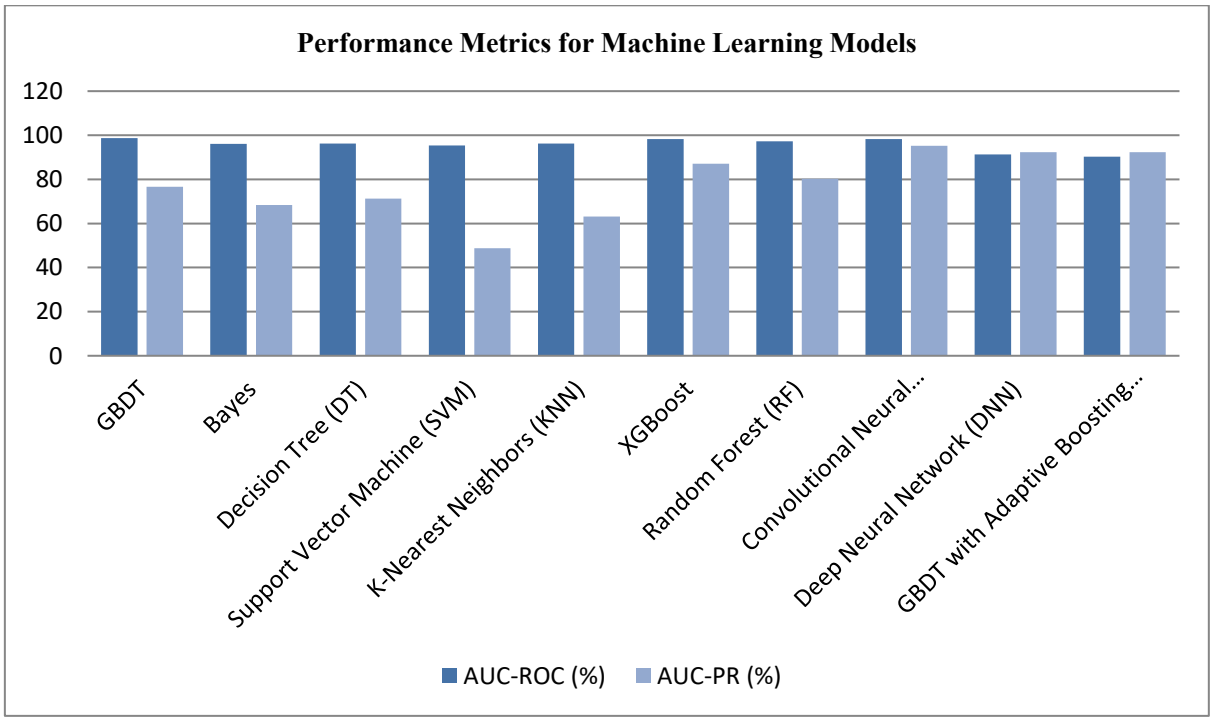

Figure 11. Performance Metrics for Machine Learning Model

|e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 10, October 2023 ||

| DOI:10.15680/IJIRSET.2023.1210100 |

The likelihood of this method correctly predicting negative samples is higher than that of the methods that came before it. They conducted comparison experiments with the Heart-Failure Prediction Dataset and the Default of Credit Card Clients Dataset, both of which were obtained from KAGGLE. For the experiments, employed a 5-fold cross-validation method. Credit card customers in Taiwan who paid their bills in full between April 2005 and September 2005 can be found in the Default of Credit Card Clients Dataset, which also includes demographic information, payment histories, credit scores, and more. The dataset contains thirty thousand samples, all of which are similarly split into two groups, with each group having twenty-five characteristics. On the other hand, the 0.3 data set was used as the test set, while the 0.7 data set was used as the training set. Customers who made timely payments were automatically assigned the label 1, while all other customers were assigned the label 0.

Table 3. performance of credit-card client's dataset.

Model	Precision (%)	Recall (%)	F1-score (%)
GB-DT	86.50	84.00	85.25
Bayes	87.80	85.90	86.84
Decision Tree (DT)	84.00	79.50	81.65
SVM	82.00	85.00	83.50
KNN	80.00	74.00	77.00
XGBoost	81.50	87.00	84.20
Random Forest (RF)	85.50	88.50	86.95
GBDT-APBT(Proposed)	89.00	90.00	89.50

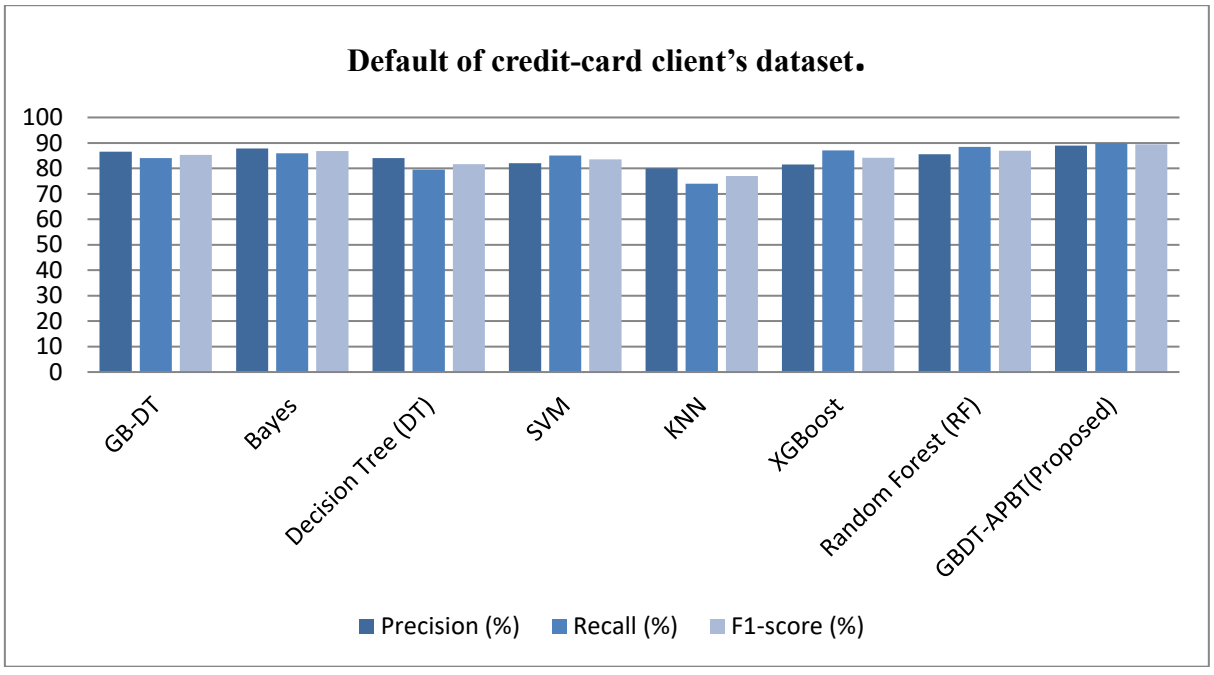

Figure 12. Performance of credit-card client's dataset.

Tables 4 and 5 present performance metrics for machine learning models applied to heart failure prediction and credit card default prediction datasets, respectively. In Table 4, the models exhibit strong performance, with the GBDT-APBT achieving the highest scores of 92.00% in precision and 90.00% in F1-score, indicates its effectiveness in accurately predicting heart failure. Other notable models include Gradient Boosting Decision Trees (GBDT) and K-Nearest Neighbors (KNN), which score 85.50% and 86.20% in precision, respectively. In Table 5, the credit card default prediction results show that GBDT-APBT leads with an impressive precision of 89.00% and a recall of 90.00%, demonstrating its superiority in identifying potential defaults. Meanwhile, the Random Forest model also performs well with a precision of 85.50% and a recall of 88.50%. Overall, both tables highlight the effectiveness of various machine learning techniques in predicting critical outcomes in healthcare and finance, showcasing their potential for real-world applications.

|e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 10, October 2023 ||

| DOI:10.15680/IJIRSET.2023.1210100 |

VI. CONCLUSION

This paper highlights the transformative potential of a microservices architecture supported by blockchain technology in enhancing security and compliance within the fintech sector. By harnessing the decentralized nature of blockchain, fintech organizations can achieve superior data integrity, transparency, and security, effectively mitigating the risks associated with fraud and data breaches. The modular design of microservices facilitates rapid development and scaling, enabling firms to swiftly adapt to changing regulatory requirements and security challenges. Furthermore, this architecture supports improved compliance through automated auditing and real-time monitoring capabilities, fostering a robust and trustworthy financial ecosystem. Additionally, we address the critical issue of asymmetrical data distribution in credit card transactions, particularly in the realm of fraud detection. Our proposed GBDT-APBT classification method effectively integrates the strengths of blockchain technology with gradient-boosting decision trees, allowing for real-time model updates without extensive preprocessing. This innovation not only maintains data privacy but also creates a collaborative environment for sharing insights among participants. The use of cryptographic tools further enhances security against potential attacks, as evidenced by our experimental results validating the model's efficacy in fraud identification.

REFERENCES

1. Ammour, N., Bashmal, L., Bazi, Y., Al Rahhal, M. M., & Zuair, M. (2018). Asymmetric adaptation of deep features for cross-domain classification in remote sensing imagery. IEEE Geoscience and Remote Sensing Letters, 15(4), 597–601.
2. Balagolla, E., Fernando, W., Rathnayake, R., Wijesekera, M., Senarathne, A., & Abeywardhana, K. (2021). Credit Card Fraud Prevention Using Blockchain. 2021 6th International Conference for Convergence in Technology (I2CT). Institute of Electrical and Electronic Engineers.
3. Boot, P., Volk, A., & de Haas, W. B. (2016). Evaluating the role of repeated patterns in folk song classification and compression. Journal of New Music Research, 45(3), 223–238.
4. Bottou, L., Curtis, F. E., & Nocedal, J. (2018). Optimization methods for large-scale machine learning. Siam Review, 60(2), 223–311.
5. Castrillón-Santana, M., De Marsico, M., Nappi, M., & Riccio, D. (2017). MEG: Texture operators for multi-expert gender classification. Computer Vision and Image Understanding, 156(3), 4–18.
6. Ding, Y., Pu, H., Liang, Y., & Wang, H. (2019). Blockchain technology in the registration and protection of digital copyright. In International Conference on Applications and Techniques in Cyber Security and Intelligence (pp. 608–616).
7. Eom, C., Kaizoji, T., Kang, S. H., & Pichl, L. (2019). Bitcoin and investor sentiment: Statistical charac-teristics and predictability. Physica A: Statistical Mechanics and Its Applications, 514(1), 511–521.
8. Fernández, A., Chawla, N. V., & Herrera, F. (2017). An insight into imbalanced big data classification: outcomes and challenges. Complex and Intelligent Systems, 3(2), 105–120.
9. Granik, M., & Mesyura, V. (2017). Fake news detection using naive Bayes classifier. In 2017 IEEE First Ukraine Conference on Electrical and Computer Engineering (UKRCON) (pp. 900–903). IEEE, Institute of Electrical and Electronic Engineers.
10. Gupta, S., Malhotra, V., & Singh, S. N. (2020). Securing IoT-driven remote healthcare data through blockchain. In Advances in data and information sciences (pp. 47–56).
11. Hitaj, B., Ateniese, G., & Perez-Cruz, F. (2017). Deep models under the GAN: Information leakage from collaborative deep learning. In Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security (pp. 603–618). ACM.
12. Kasa, N., Dahbura, A., Ravoori, C., & Adams, S. (2019). Improving credit card fraud detection by profiling and clustering accounts. In 2019 Systems and Information Engineering Design Symposium (SIEDS) (pp. 1–6). Institute of Electrical and Electronic Engineers.
13. (Salibindla & Solutions, 2018) alibindla, J., & Solutions, K. (2018). Microservices API Security. International Journal of Engineering Research And, 7(01).
14. Lebichot, B., Borgne, Y. A. L., He-Guelton, L., Oblé, F., & Bontempi, G. (2019). Deep-learning domain adaptation techniques for credit cards fraud detection. In Recent advances in big data and deep learning (pp. 78–88). Springer.
15. Li, P., & Wang, E. A. (2016). High-performance personalized heartbeat classification model for long-term ECG signal. IEEE Transactions on Biomedical Engineering, 64(1), 78–86.
16. Liu, J., Li, B., Chen, L., Hou, M., Xiang, F., & Wang, P. (2018). A data storage method based on blockchain for decentralization DNS. In 2018 IEEE Third International Conference on Data Science in Cyberspace (DSC) (pp. 189–196). IEEE, Institute of Electrical and Electronic Engineers.

|e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 10, October 2023 ||

| DOI:10.15680/IJIRSET.2023.1210100 |

17. Mahapatra, S., Sinha, D., & Kumar Das, A. (2022). A secure health management framework with anti-fraud healthcare insurance using blockchain. In Pattern recognition and data analysis with applications (pp. 491–505). Springer.
18. Maillo, J., Ramírez, S., Triguero, I., & Herrera, F. (2017). kNN-IS: An iterative spark-based design of the k-nearest neighbors classifier for big data. Knowledge-Based Systems, 117(2), 3–15.
19. Pan, Z., Wang, Y., & Ku, W. (2017). A new k-harmonic nearest neighbor classifier based on the multi-local means. Expert Systems with Applications, 67(1), 115–125.
20. Pang, G., Cao, L., Chen, L., & Liu, H. (2018). Learning representations of ultrahigh-dimensional data for random distance-based outlier detection. In Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 2041–2050). ACM.
21. Pang, G., & Shen, C. (2019). Deep anomaly detection with deviation networks. In Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 353–362). ACM.
22. Priscilla, C. V., & Prabha, D. P. (2020). Influence of optimizing XGBoost to handle class imbalance in credit card fraud detection. In 2020 Third International Conference on Smart Systems and Inventive Technology (ICSSIT) (pp. 1309–1315). Institute of Electrical and Electronic Engineers.
23. 2021 Rani, P., Shokeen, J., Agarwal, A., Bhatghare, A., Majithia, A., & Malhotra, J. (3). Credit card fraud detection using blockchain and simulated annealing K-means algorithm. In International Confer-ence on Innovative Computing and Communications (pp. 51–59). Springer.
24. Ruff, L., Vandermeulen, R., Goernitz, N., Deecke, L., Siddiqui, S. A., Binder, A., Kloft, M., & Kloft, M. (2018). Deep one-class classification. In Proceedings of the 35th International Conference on Machine Learning (pp. 4393–4402). PMLR.
25. Saberi, S., Kouhizadeh, M., Sarkis, J., & Shen, L. (2019). Blockchain technology and its relation-ships to sustainable supply chain management. International Journal of Production Research, 57(7), 2117–2135.
26. Tanha, J., van Someren, M., & Afsarmanesh, H. (2017). Semi-supervised self-training for decision tree classifiers. International Journal of Machine Learning and Cybernetics, 8(1), 355–370.

INNO SPACE
SJIF Scientific Journal Impact Factor

Impact Factor: 8.423

doi cross ref

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

9940 572 462 6381 907 438 ijirset@gmail.com

www.ijirset.com

Scan to save the contact details