https://doi.org/ 10.33472/AFJBS.6.5.2024. 6519-6532

African Journal of Biological Sciences

BLOCKCHAIN AND MACHINE LEARNING APPROACHES FOR CREDIT CARD FRAUD DETECTION

[1]Princy Usha M,[2]Suganthini C,[3]Jose P, [4]Nandhini I

[1]Assistant Professor,Department of Computer Science and Engineering,DMI Engineering College,Kanyakumari
[2]Assistant Professor(SR.Gr.I),School of Computer Science and Engineering,Vellore Institute of Technology,Vellore
[3]Associate Professor, Department of Computer Science and Engineering,School of Computing, Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology,Avadi,Chennai
[4]Assistant Professor, Department of Electronics and Communication Engineering,School of Electrical and Communication,
Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology,Avadi,Chennai

[1]princyusha@dmiengg.edu.in,[2]suganthini.c@vit.ac.in,[3]drjosep@veltech.edu.in,
[4]drnandhinii@veltech.edu.in

Article History
Volume 6, Issue 5, 2024
Received: 09 May 2024
Accepted: 17 May 2024
doi: ***10.33472/AFJBS.6.5.2024. 6533-6550***

Abstract:

A credit card is a simple way to make cashless purchases, both online and offline, and it is accepted almost everywhere. One of the most significant benefits of using a credit card rather than a debit card is that it allows you to borrow money to pay for your transactions. As well as the majority of online fraud occurs during a card or online transaction when a user attempts to buy something or move money. Now a days lots of technology introduced for secured money transaction, that's one of them blockchain technology. The blockchain has the potential to evolve into a distributed ledger, offering a revolutionary new form of trustworthy third-party authentication. Because of the long history of credit card systems, it is easier to understand and security has always been triggered by a process of delegating risk to third parties. Blockchain technology has the potential to avoid these types of losses from occurring in the first place. This study examines how Blockchain technology may be applied, how it might be made safe, and how it might be used to reduce the danger of credit card data being compromised. In addition this article defines and describes a system that may be developed to prevent theft of credit cards using existing technology, such as two-factor authentication, SMS random OTP (One Time Password) and biometrics tools.

Keywords: Blockchain, Creditcard, OTP, distributed ledger, SMS and biometrics

I. Introduction:

Credit card fraud is typically described because of the usage, either the use of physical card information or the cardholder, of any system or criminal activity. A little plastic card that was sent to the user as a payment mechanism may be the MasterCard. The misleading behaviors also Credit card fraud is generally defined as the use of any system or criminal conduct, whether it be the use of physical card information or the use of the cardholder, to commit a fraudulent transaction. The MasterCard is a little plastic card that was provided to the user as a means of making a financial transaction. In addition, whether the MasterCard is physical or virtual, the cardholder must present his or her card to a business representative to make a payment.

Technology in Blockchain is praised in technology as the "next great thing." Can I use it to prevent fraudulent purchases via credit card? You give details on your credit card. As soon as the transaction is verified, biometric evidence is required for the consumer to update the transaction history of the credit card, to which the Blockchain technology is used. This article will also analyze and evaluate the establishment of a new safe information system, a consumer protection mechanism, and a new technique for companies to take credit cards safely and securely. By analyzing Blockchain and other current technologies and learning from major consumer safety violations today we can develop a system that protects consumers and corporations from future safety violations.

Target magazines had been hacked in 2013, resulting in 40 million credit and debit card information. In addition to stolen credit and debit card data, the study also recorded additional 70 million customer names, addresses, telephone numbers, and e-mail addresses. Target reported in its May 2016 Securities and Exchange Commission report that data breach charges were $291 million and that the total costs were estimated to rise to $370 million[32]. Target made a net expenditure of $202 million on $291 million offset by insurance payments of $90 million. As part of the settlement, Target also agreed to pay an extra $18.5 million[26].

When the Home Depot website was hacked in 2014, another 40 million credit and debit cards and 50 million e-mail addresses were adopted. 143 million consumers' personal data, including phone numbers, names, residences, and dates of birth, were collected. The violation started in May 2017 and it is estimated that the attack led to up to five million credit card records. Although the financial effects of this violation are too early to be estimated, the other attacks detailed above demonstrate the enormous financial harm of maybe millions of dollars.

Is it not time to investigate alternative alternatives to avoid the financial consequences of these violations? A new way of processing the transaction of credit cards must be adopted to remove the possibility of credit card data being stolen and afterward sold on the open market. These expenditures do not cover replacement costs for robbing credit cards, reputation harm, investigative fees for transactions fraudulent, replacement costs for credit cards, and development costs for technology that prevents them from reoccurring. When taking all the expenses into account, finding a solution that may be efficient is meaningful and putting a stop to the practice of stealing credit card data. The cost of an assault alone for firms and customers should be a stimulus for changing the way we deal with credit card theft and fraud in the future. Both companies and consumers may profit from the usage of Blockchain technology to save money, save time, and genuinely respond to credit card theft. The use of Blockchain technology and the capacity to regulate the use of credit cards for customers are a great step in this direction.

1.1 Blockchain:

The global data base based on distributed ledger technologies, known as blockchain technology, safeguards cryptocurrencies like Bitcoin and ETHEREUM. BlockCoin or ETHEREUM offers the untrusted parties a way of reaching a shared digital historical accord (consensus) [28].

1.1.1 Transactions on the blockchain:

This section provides a high-level overview of blockchain transactions, as well as the application of cryptographic digital signatures and the signing procedure for the transactions in question. More sophisticated applications (also known as decentralized apps) may be built on the Ethereum platform, and code that implements arbitrary rules (also

known as smart contracts) can be used to directly govern digital assets (such as cryptocurrencies, coins, and smart properties) [28]. Blockchain technology allows transactions to be securely and irreversibly recorded in blocks that are then linked together to form a ledger [35] in a distributed ledger. It is worth noting that each new block comprises many transactions, each of which has been confirmed by digital signatures [26], and each of which has been recorded as a state in a distributed ledger that is shared among all participants in the blockchain network shows below figure 1.

1.2. Problem statement:

Nowadays, the majority of people use credit cards to purchase things that they desperately need but cannot afford at the present time. Credit cards are increasingly being utilized to fulfill consumer demands, and the amount of fraud connected with them is growing as a result. As a result, it is necessary to create a model that is well-fit and forecasts with more accuracy. Existing systems are carried out by taking into consideration machine learning techniques such as Support Vector Machine, Nave Bayes, k-Nearest Neighbor, and so on, and some of them utilized random datasets Artificial neural networks (ANNs) have been utilized for credit card fraud detection by a small number of companies.

1.3. Objectives:

The primary goal of this study is to identify fraudulent transactions in credit card transactions.. A comparison was made between supervised learning and deep learning, with the deep learning algorithm outperforming the supervised learning algorithm in terms of accuracy.

The objective of this article is to provide a novel way in which blockchain transactions may be automated digitally signed, thereby reducing the user's burden to sign and verify each transaction manually. The suggested approach uses customized transaction data to ensure that digital signatures for blockchain-related transactions and other safety measures, including an anomaly detection mechanism, based on personalized transaction information are automatically and personalized. The designated transaction, which includes the shipment of crypto-monetary from address A (i.e. receiver), is automatically digitally signed in the sending site except in the case of a potentially anomalous transaction detection by including a proposed method into blockchain-dedicated software (for instance, a wallet) managing the digital signature process (for example, wallet). In that case, before the transaction can be completed a sender's manual approval is required to digitally sign a transaction. The proposed method's features also help avoid existing digital signatures which may be indirectly present in use and so serve as a barrier to the broad use of blockchain technology, which is now under progress. An innovative machine learning-based approach for automatic digital signature of blockchain transactions has been developed, which incorporates the following features:

- An abnormal transaction detection system that is tailored to each individual; and
- When compared to a centralized environment, a user's local environment is used to run and store the tailored data obtained from the anomaly detection model.

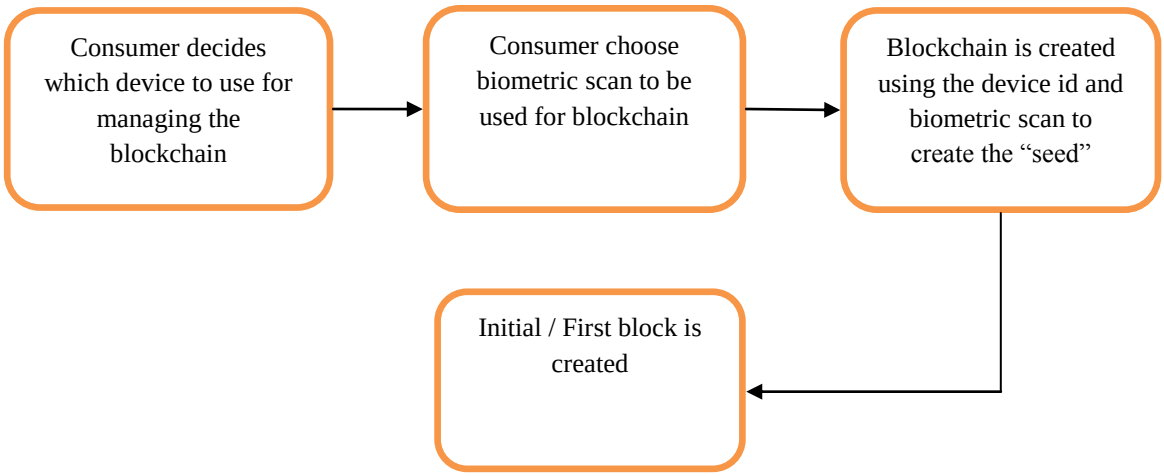

Figure 1: Creation of blockchain technology

II. Related Works:

This section contains descriptions of some of the relevant studies conducted by different researchers.Altab Althar Taha and Sareef Jameel Malbery said that advancements in e-commerce and communication technologies have made credit card use a more common method of payment, and that the amount of fraud associated with transactions is rising. They have utilized the improved light gradient boosting machine, in which Bayesian based hyper-parameter optimization is coupled with light gradient boosting machine parameter tuning to get the desired result (LightGBM). In this method, they utilized two sets of real-world public datasets, one including fraudulent transactions and the other containing non-fraudulent transactions. They found that their proposed method outperformed other strategies in terms of accuracy when compared to those other systems. The accuracy of the proposed system is 98.40 percent, the area under the receiver operating characteristic curve (AUC) is 92.88 percent, the precision is 97.34 percent, and the F1-score is 56.95 percent [1].

According to the findings of S. Makki, Z. Assaghir, Y. Taher, R. Haque, M. Hacid, and H. Zeineddin's study, credit card theft results in a significant financial loss. The majority of researchers have been working on this to develop novel approaches to eradicating this loss, and the majority of the techniques that are now accessible are expensive, time-consuming, and labor-incentive tasks. Many experimental investigations have been conducted by the authors, and they have discovered that an unbalanced categorization of the dataset is the primary cause of the incorrect findings. It is because of this unbalanced dataset that the model was unable to forecast accurately, resulting in a financial loss to the organization. As a result, they discovered that the LR, C5.0 decision tree method, SVM, and ANN are the best algorithms in terms of accuracy, AUCPR, and sensitiv- ity. For the purpose of training these models, they have made use of the balanced dataset [2].

In their paper, C. Jiang and colleagues presented a new method with several phases that included G. Liu, L. Zheng, and W. Luan. They begin by collecting the transactions made by the cardholder, then aggregate the transactions

based on behavioural patterns. After that, the dataset is categorized, the model is trained, and lastly the model is tested. If any aberrant behavior occurs, the system receives input on the abnormal behavior via the use of a feedback mechanism to alert it to the abnormal behavior. [3].Ishan Sohony, Rameshwar Pratap, and Ullas Nambiar suggested an ensemble learning method for credit card fraud detection since the ratio of fraudulent transactions to regular transactions is a little higher than it should be, according to their research. According to their findings, random forest is the best algorithm for providing greater accuracy, followed by neural networks for identifying fraud cases. They also conducted experiments using large-scale real-world credit card transactions. Random forest and neural networks [4] are used in ensemble learning, which is a combination of the two.

According to a study conducted by Phuong Hanh Tran, Kim Phuc Tran, Truong Thu Huong, Cédric Heuchenne, Phuong HienTran, and Thi Minh Huong Le, credit card theft has steadily risen over the last several years. Many techniques are used in conjunction with machine-learning algorithms to identify and prevent fraudulent transactions from taking place. Their research team developed two novel data-driven methods that make advantage of the most effective anomaly methodology for fraud detection in credit card transactions. Both kernel parameter selection and T 2 control chart [5] are methods for selecting kernel parameters.

The work of Imane Sadgali, Nawal Sael, and Faouzia Benabbou describes how banking transactions such as online transactions, credit card transactions, and mobile transactions, among other things, are gaining popularity because people are increasingly preferring digital and paperless methods of conducting business. Millions of transactions have been completed, each of which has been subjected to some kind of fraud. A large number of researchers have conducted analyses, created models, and produced models for detecting fraud using machine learning techniques. It was proposed that a comparison be made between the complete machine-learning process in order to determine which model is the most effective for fraud detection in card transactions [6].

Debachudamani Prusti and Santhnu Kumar Rath developed an application that used applied machine learning approaches such as the Decision Tree (DT), the k-nearest algorithm (kNN), the Extreme learning machine (ELM), the Mul- tilayer perceptron (MLP), and the support vector machine (SVM) to detect the accuracy in fraud identification. They received funding from the National Science Foundation. They presented a model that was a hybridization of the DT, SVM, and kNN methods, among others. In order to efficiently transmit data across various heterogeneous platforms, they utilized two web-based protocols, such as the simple object access protocol (SOAP) and the representational state transfer (REST). Based on the accuracy measure, they compared the outcomes of five machine learning algorithms. By 81.63 percent, SVM outperformed the other algorithms; however, the hybrid system suggested by them was more accurate, with an accuracy of 82.58 percent [7].

Mohamad Zamini and Golamali Montazar presented an unsupervised credit card fraud detection system based on autoencoders and clustering to identify fraudulent credit card transactions. They utilized three hidden layers and k means for clustering, and they evaluated their system on a European dataset, which performed well when compared to other existing systems [8].

A misrepresentation location framework with a non-overlapped risk based bagging ensemble (NRBE) model was developed by Akila and Srinivasulu reddy to cope with the unevenness dataset and to avoid being affected by the noise that was present in the transactions. The bagging model eliminates all of the irregularities in the dataset as well as the non-vital character of the data. The sacking model is reached out to by a group of students from a creativity and danger-based basis. The bag construction eliminates the problem caused by unbalanced data, and the Nave Bayes algorithm eliminates the problem caused by the noise generated during the trans- operations. It has been shown that by using NBRE, the suggested model outperformed the competition by 5 percent in BCR and BER as well as half the recall and 2x or 2.5x the cost of fraud detection. It has been determined that the NRBE model is the most effective for fraud detection and that it is the most suitable for business dynamic technology [9].With the use of Random forest methods, M. S. Kumar and his colleagues V. Soundarya and Kavitha and E. Aswini developed a model for the detection of credit card fraud in transactions using random forest techniques. The random forest algorithm (RFA) is a supervised machine learning method that utilizes a decision tree to classify credit card transactions, and then uses a confusion matrix to assess the performance of the algorithm. Using the suggested

method, an accuracy of 90 percent [10] may be achieved.Z. Li, G. Liu, S. Wang, S. Xuan, and C. Jiang presented a fraud detection system based on Kernel-based supervised hashing, which was implemented in a Java program (KSH). This KSH model is based on the concept of approximate closest neighbor. It is most appropriate for big datasets with a high number of dimensions. It is the first time that KSH has been utilized for prediction, and it performs much better than other current methods [11].Pawan Kumar and Fahad Iqbal conducted a study on all methods used to identify MasterCard fraud using machine-learning algorithms and evaluated the performance using metrics. The results were published in the journal Machine Learning and Applications. There has been a great deal of research done in this area. In their opinion, there is a pressing need for more efficient systems that function effectively in all situations [12].

III. Proposed Techniques:

revolutionary Blockchain technology to prevent the use of counterfeit credit cards. The operation would be regarded complete from the time a financial institution issues a consumer's credit card number. The consumer needs to make a Blockchain with the device he or she has chosen first before using the credit card. It can be anything from a laptop, smartphone, tablet, IPad, or any other device that has internet connection, depending on the item. This method is the same as what a client does when a credit card is obtained. For example, Alice will now have to activate the new credit card, which may be achieved by either dialing a telephone / mobile number or visiting a website. In the event of a later passcode or a transaction notification, Alice will also provide her cellphone number. After completing your Blockchain setting, customers can use the card as they are today. Once the transaction is started in a retail company or a financial institution a text message (such as a mobile phone) is delivered to the customer's device prior to the completion of the transaction. The cellphone can be protected with a biometric two-factor authentication method. The transaction is accepted by the consumer or denied. Upon acceptance, a new Blockchain entry is created which corresponds to this transaction and is then validated for accuracy, honesty and completed by previously recorded transactions and then uploaded to the Blockchain server. If the customer declines to participate, the transaction will be denied. The analytical model is used in order to determine if an incoming transaction is a legitimate transaction or not. For fraud detection, machine learning models such as logistic regression, decision tree, Naive Bayes, Artificial Neural Network, and Random forest are used in conjunction with each other. The model is based on a data set of credit card banking transactions. Five models for fraud detection classification are being used in this research, As shown figure 2.

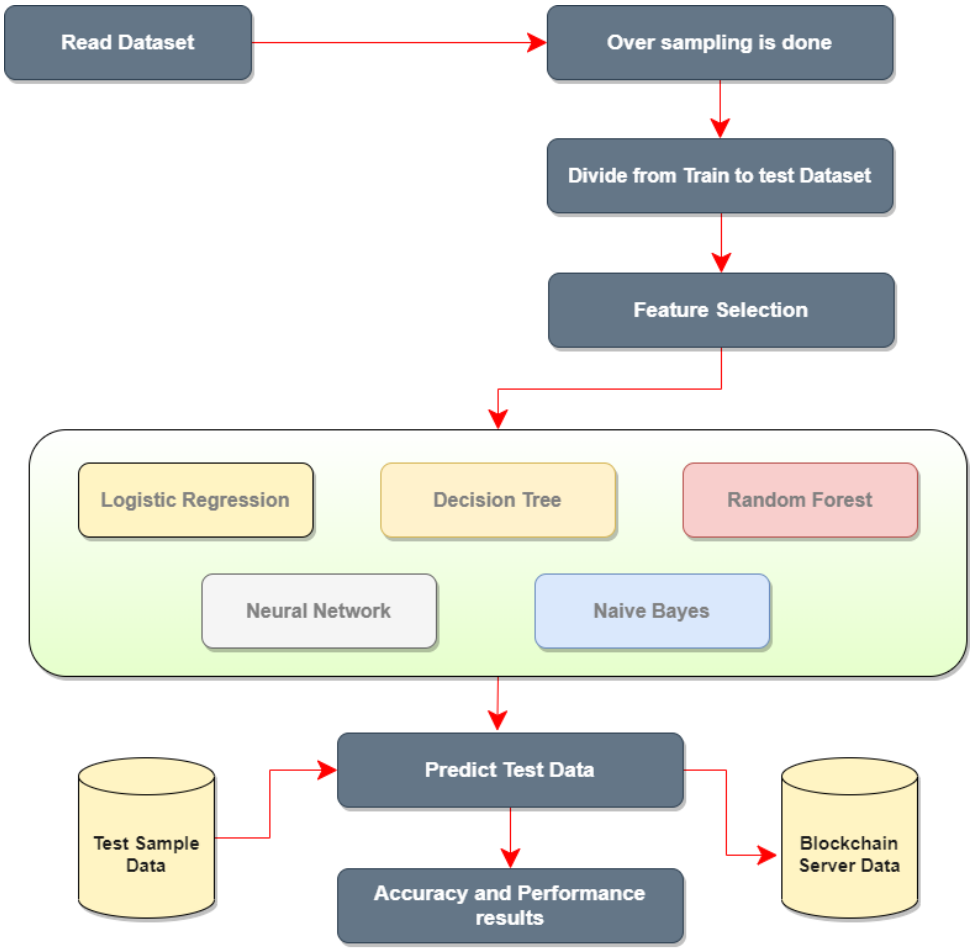

Figure 2: System Architecture

3.1 Machine Learning Method for Automated Digital Signing:

The mechanism for automatic and customized digital signature of blockchain with machine learning transactions is shown in this section, along with an example of how it works. In addition, we discuss the history of customized anomaly detection for blockchain transactions, as well as the personalized anomalous blockchain transaction detection system that we developed for this purpose. Furthermore, we will illustrate how our suggested method, as opposed to the existing manual approach, allows for an automatic evaluation and digital signature of transactions submitted by decentralized apps.

3.2 Logistic Regression:

Logistic regression is a supervised learning classification algorithm. The target or dependent variable is dichotomous, meaning there are only two classes. In other words, the dependent variable is binary, with data coded as 1 (yes) or 0 (no). A logistic regression model predicts P(Y=1) based on X. An easy to use ML algorithm for classification problems like spam detection, diabetes prediction, cancer detection, etc. We use Logistic Regression to classify fraud. Logistic regression uses a logistic curve to detect fraud. The univariate logistic curve formula,

$$p = \frac{e^{(c_0 + c_1 x_1)}}{1 + e^{(c_0 + c_1 x_1)}} \quad (1)$$

Because the logistic curve produces a number between 0 and 1, it may be understood as the likelihood of belonging to a certain class. The logarithmic function may be used to the logistic function in order to achieve the regression, as illustrated in the example below.

$$\log_e\left(\frac{p}{1-p}\right) \tag{2}$$

In this equation, p is the probability of a tuple being in class, and $1-p$ denotes the probability of a tuple not being in class. The model, on the other hand, selects values for the coefficients c_0 and c_1 that maximize the likelihood of an incoming transaction.

3.3 Decision Tree Algorithm:

It is a kind of supervised learning algorithm in the traditional sense. Building the decision tree by taking into account the entropy of the data set is accomplished via the usage of ID3 method. The entropy of a collection of data is used to determine the degree of uncertainty in the data. Each attribute's entropy value is used to determine the splitting criteria that will be used in the decision tree construction process. The entropy of a distinct state may be computed using the following equation:

$$H(p_1,p_2,\ldots\ldots,p_s) = \sum_{i=1}^{s}\left(pi\log\left(\frac{1}{p_i}\right)\right) \tag{3}$$

When the probabilities of the attributes in the dataset are $p_1,p_2,\ldots\ldots,p_s$ then the probability of the attribute is p_1.Entropy of each attribute in the dataset is computed, and the gain is obtained by subtracting the entropy of the whole dataset from the entropy of the splitting attribute, and the gain is expressed as The characteristic with the greatest benefit is chosen as the root node, and a decision tree is constructed in accordance with this selection. The ID3 determines the gain of a given split by using the following equation,

$$Gain(D,S) = H(D) - \sum_{i=1}^{s}(p(D_i)H(D_i)) \tag{4}$$

3.4 Artificial Neural Network

ANN is a deep learning concept that is implemented using the Keras programming language (in this case). Neurons are the building blocks of the ANN. The input neuron is the initial layer, also known as the input layer, and it contains information about each customer's transaction and amount. Weights, bias, and an activation function are all components of the hidden layer. We may create as many hidden layers as we like in order to fine-tune the performance. In this instance, we are using three layers. This is the last layer in the classification process, and it is where we receive our categorized output. The result will be either 1 or 0, with 1 indicating a fraud case and 0 indicating a regular case as shown figure 3. It is made up of three layers: an input layer, an output layer, and a hidden layer. The number of hidden layers depends on the issue we are trying to solve; there may be one or more hidden layers. The number of neurons in the input layer corresponds to the number of input attributes in the training dataset, which we will see later in this paper, and the number of neurons in the output layer is dependent on the type of problem you are trying to solve; for example, in the credit card fraud detection case, we have two outputs, one of which is fraud and the other of which is non-fraud, which are 0 and 1 respectively. The number of neurons in the input layer corresponds to the number of input attributes in the training dataset.

$$S = \sum_{2}^{x=0} 1x.Zx \tag{5}$$

The output of this summation function is then passed on to the activation function for further processing. The value of S is scaled in the appropriate range by the activation function. The most often used activation function is the

sigmoid activation function, which operates on a threshold; if the value of S exceeds the threshold value, the node passes the output signal.

3.5 Random Forest:

Random forest is a kind of supervised machine learning algorithm that is based on the concept of ensemble learning. It is a learning in which various types of algorithms are combined or the same method is repeated many times to create a more effective prediction model. When developing the trees, the Random Forest method introduces more arbitrariness into the model. Instead of searching for the primary component, it searches for the best piece from an irregular subset of highlights. This method outperforms single decision trees because it minimizes over-fitting by averaging the results. The algorithm can be used for both classification and relapse problems.

Bootstrapped samples are chosen at random from the standard training set with replacement to form the training set for each tree in the classification system. At each internal node, it chooses a subset of characteristics at random and computes the centers of various classes of data contained inside the current node's data structure. The centers of classes 0 and 1 are represented by the letters leftCenter and rightCenter, in that order. Following are several formulae that may be used to calculate the kth element of a center.

$$leftCenter[k] = \frac{1}{n}\sum_{i=1}^{n} x_{ik} I(y = 0) \tag{6}$$

$$rightCenter[k] = \frac{1}{n}\sum_{i=1}^{n} x_{ik} I(y = 1) \tag{7}$$

They are represented by the dictator functions I(y = 0) and I(y = 1). Currently, each record of the dataset is assigned to the appropriate class at the current node based on the Manhattan distance between the record and the center of the class hierarchy.

$$distance(center, record) = \sum_{i \in sub} |center[i] - record[i]| \tag{8}$$

Observe that sub denotes a subset of attributes randomly chosen from the set X whose size is the square root of the number of attributes in |X|. Each tree develops to its full potential without the need for trimming.

3.6 Naïve Bayes:

The Naive Bayes Classifier is a classification method that is based on the Bayes theorem and is used to classify data. The Nave Bayes Classifier is often considered to be superior than several other classification algorithms. As a result of the fact that the key characteristic of Naive Bayes is the very strong (naive) assumption of independence from each condition or event, Second, the model is straightforward and straightforward to construct. Third, the approach may be applied to big data sets without difficulty. The Bayes formula is used as the foundation for the Naive Bayes theorem, which is as follows:

$$P = \left(X_i = x_i : C = c_j\right) = \frac{1}{\sqrt{2\pi\sigma_{ij}}} \exp\left(-\frac{\left(x_i - \mu_j\right)^2}{2\sigma_{ij}^2}\right) \tag{9}$$

Whrere:

P : opportunity

Xi : the $i\ th$ attribute

xi : the $i\ th$ attribute value

C : class

Ci : the $i\ th$ sub class

μ : the mean of all attributes

σ : standard deviation

IV. EXPERIMENTAL RESULTS AND PERFORMANCE METRICS:

4.1 Dataset:

Using the dataset obtained from this website: www.kaggle.com, the suggested method may be tested and improved. The transactions performed by customers at a European bank over the year 2017–20 were utilized as the data set. The total number of individual transactions in the original dataset exceeds 30,000,000 in total. Each transaction record contains 62 attribute values, such as the transaction time, the transaction location, and the transaction amount. Each record is tagged with the words Fraud or Legal on it. According to the company's requirements, it is not allowed to provide specifics about the dataset's characteristics. There were about 82,000 transactions in the dataset that were classified as fraudulent, resulting in a fraud ratio of 0.27 percent and a dataset imbalance issue that should be taken into account.

4.2 Dataset Evaluation:

Dataset Evaluation is a metric. In addition to evaluating the final result using the confusion matrix, it is also necessary to quantify precision, recall, and accuracy as shown figure 4 to 6. It is divided into two categories: the actual class and the projected class. The following characteristics influence the confusion metrics:

True Positive: When both of the numbers are positive, this is referred to as 1.

True Negative : This is the situation in which both numbers are negative, in which case the result is 0.

False Positive: This is the situation in which the true class is 0 and the non-true class is 1.

False negative: occurs when the real class number is one and the non-true class number is 0.

Specifically, precision is defined as follows: Precision is true positive divided by actual result. Precision is defined as true positive divided by (true positive and false positive).

4.3 Classification and Tranactions:

A new method for identifying fraud accounts on the Ethereum network is proposed in this paper. Ethereum is a blockchain that offers several major advantages over Bitcoin in terms of security. These enhancements make it easier to create and execute contracts (also known as smart contracts) in the future. These contracts provide a means for a wide range of parties to enter into complicated agreements that are fully executable and can be validated via the usage of the protocol that underpins them. In the first step, we automatically collected information on accounts and transactions that were already accessible. After that, we used the raw data to construct explanatory variables. They represent aggregates and statistics that have been calculated across large quantities of data and over time. In the next step, we evaluated five different classifiers are Decision Tree, Random Forest, Artificail Neural Network,

Naïve bayes, Logistic regression and compared their findings in the context of potential applications for each. They may be highly dependent on various use cases, which may place a higher value on accuracy than on recall, or the opposite, depending on the situation. The following achieved as the contribution of this work:

presented a new method for detecting fraudulent accounts on the Ethereum blockchain that is readily transferrable to other blockchains, such as the Bitcoin blockchain.For the purpose of classifying accounts into the "fraudulent" or "not fraudulent" categories, we performed a comprehensive study of five alternative machine learning algorithms.The sensitivity analysis was performed in order to determine how much we rely on specific explanatory factors. As a result of this test, we will be able to address the possible issue of a look-ahead bias that may or may not exist in the data that have collected.

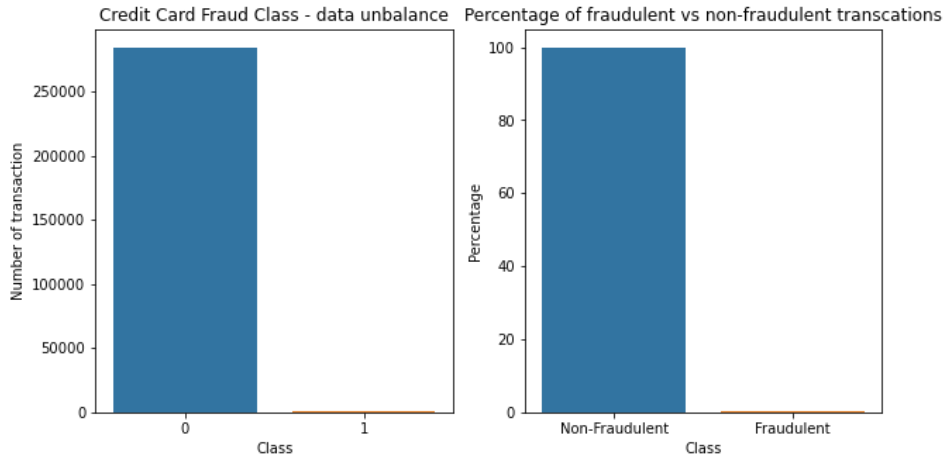

Figure 3**:** Fraud and Non Fraud transaction

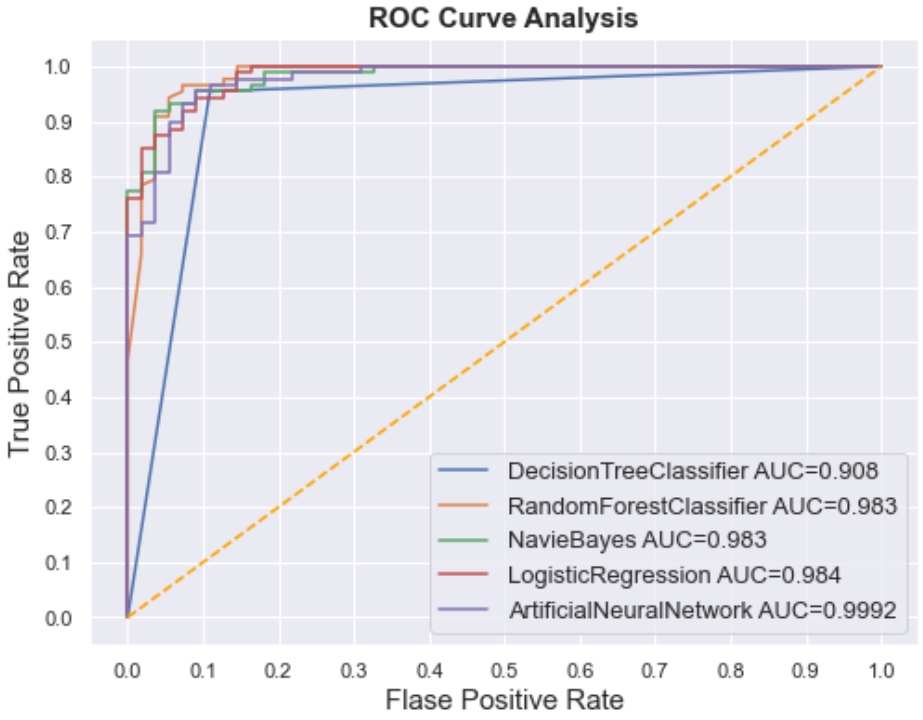

Figure 4**:** ROC Curve Analysis

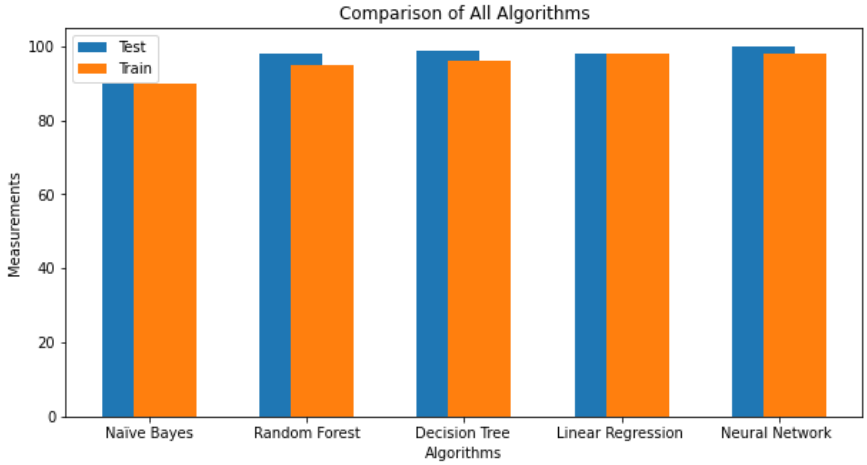

Figure 5**:** Comparison of ML Algorithms Train/Test set

4.4 Classification Results:

For evaluatting the effectiveness of the framework in identifying charge card extortion, precision, error rate (sensitivity), and specificity are used. Three artificial intelligence computations are implied in this article to differentiate between extortion and fraud in Mastercard settings. 70 percent of the dataset is used for arranging, and the remaining 30 percent is used for testing and evaluation to validate the computations. To demonstrate how various factors such as exactness, blunder rate, affectability, and explicitness may be used to test different factors for three calculations, Table 3 illustrates how they are used to test different factors for three computations. When calculating computed relapse, the exactness results for the choice tree and arbitrary timberland classifier are 99.2 percent, 98.4 percent, 98.3 percent, and 90.8 percent, respectively. As shown by the table 1, the random woodlands methods outperform the calculated relapse and choice tree processes in terms of effectiveness.

Table 1: Accuracy Comparison Table

Algorithms	Accuracy	False Positive Rate	Recall	Precision	F1-Score
Random Forest	0.9831	0.224490	0.775510	0.950000	0.853933
Decision Tree	0.9086	0.755102	0.792857	0.773519	0.244898
Artificial Neural Network	0.9992	0.217687	0.761900	0.811512	0.812721
Naive Bayes	0.9835	0.102041	0.897959	0.060746	0.113793
Logistic Regression	0.9842	0.228687	0.742313	0.7555484	0.823832

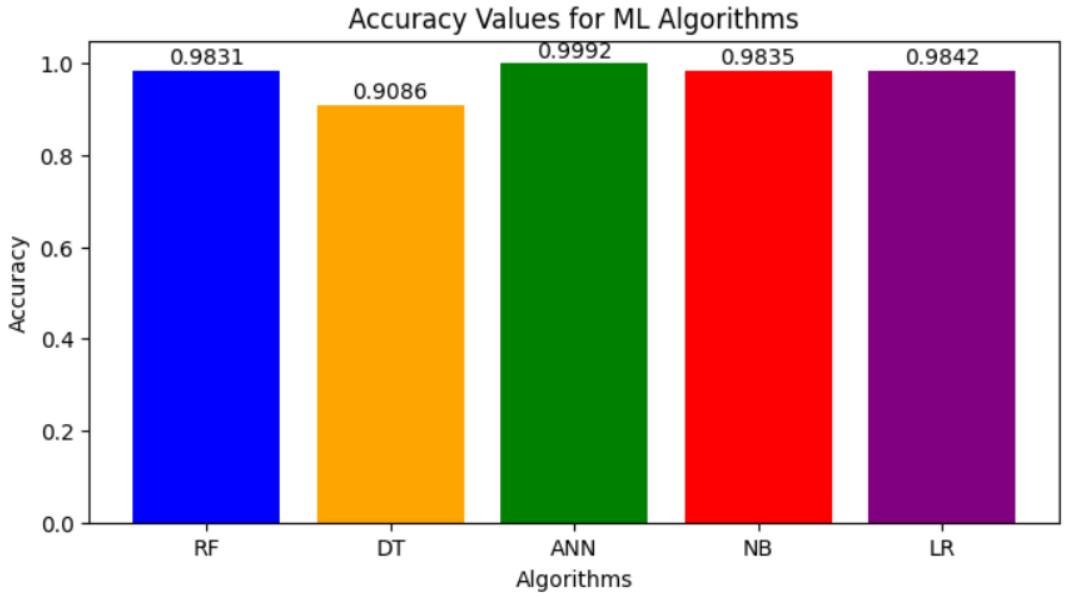

Figure 6**:** Accuracy Values for Machine Learning Algorithms

Figure 6. shows The Artificial Neural Network (ANN) achieved the highest accuracy (0.9992), while the Decision Tree had the lowest accuracy (0.9086). The other algorithms (Logistic Regression, Naive Bayes, and Random Forest) showed comparable accuracy, slightly below that of ANN, ranging from 0.9831 to 0.9842.

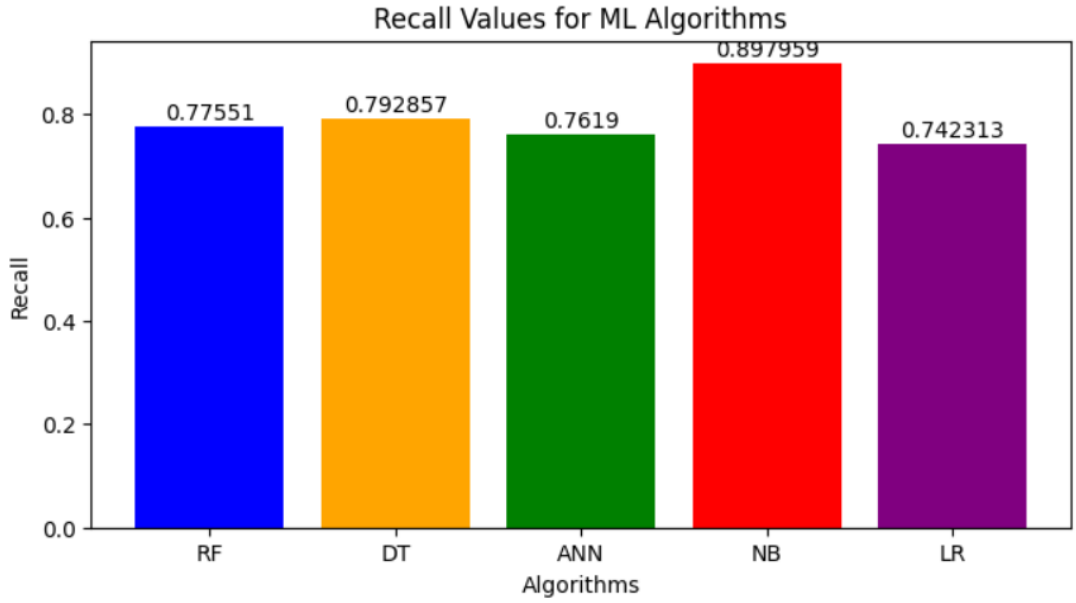

Figure 7**:** Recall Values for Machine Learning Algorithms

Figure 7. shows Naive Bayes achieved the highest recall (0.897959), indicating it is the best at identifying relevant instances among the algorithms tested. Logistic Regression had the lowest recall (0.742313). The other algorithms (Decision Tree, Random Forest, and ANN) have recall values ranging from 0.761900 to 0.792857.

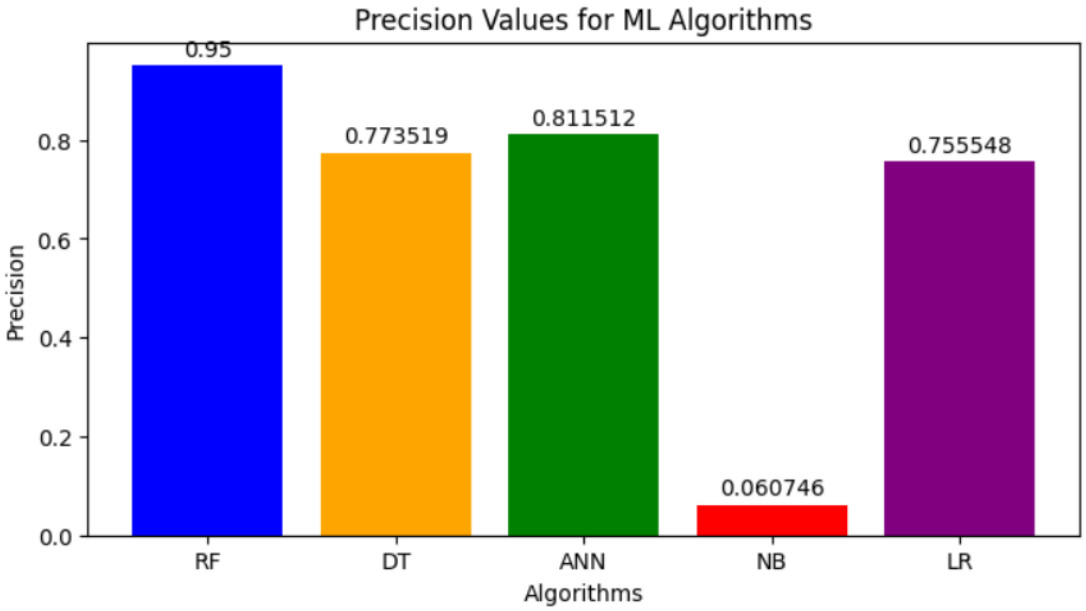

Figure 8**:** Precision Values for Machine Learning Algorithms

Figure 8. shows Random Forest achieved the highest precision (0.950000), indicating it has the lowest rate of false positives among the algorithms tested. Naive Bayes had the lowest precision (0.060746), suggesting it has a high rate of false positives. The other algorithms (ANN, Decision Tree, and Logistic Regression) have precision values ranging from 0.755548 to 0.811512.

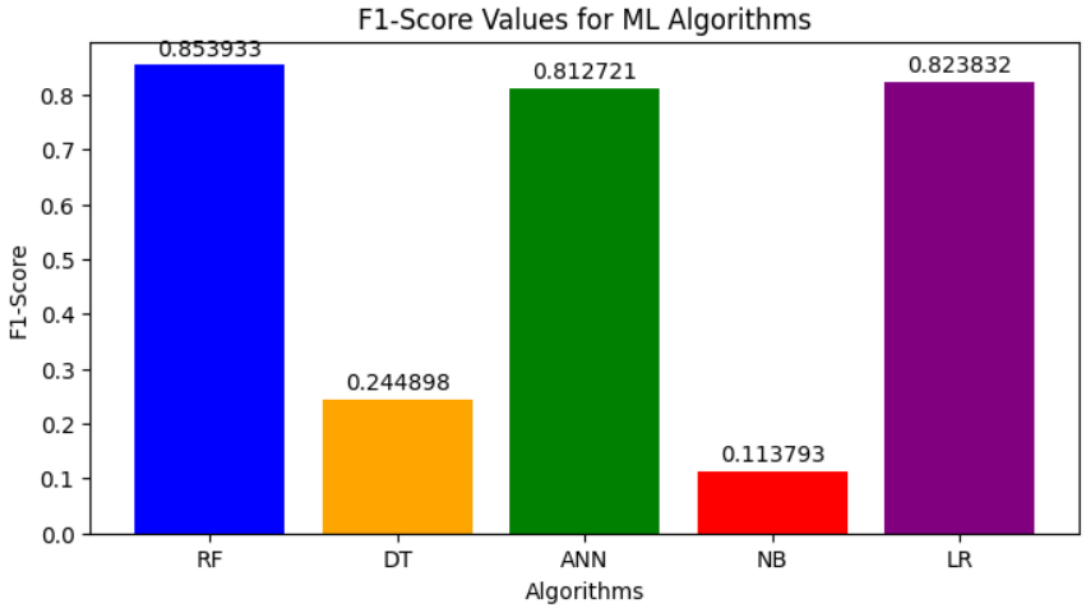

Figure 9. F1-Score Values for Machine Learning Algorithms

Figure 9. shows Random Forest achieved the highest F1-Score (0.853933), indicating a good balance between precision and recall. Logistic Regression (0.823832) and ANN (0.812721) also performed well with high F1-Scores.

Decision Tree (0.244898) and Naive Bayes (0.113793) had significantly lower F1-Scores, indicating less balanced performance between precision and recall for these algorithms.

V. Conclusion:

To provide a customized and automated way to sign blockchain transactions, which term personalized and automated signature. New usage within the blockchain technology sector of artificial intelligence, especially machine learning approaches, is utilized to automate the signature process while safeguarding users against the digital signing of possibly fraudulent blockchain transactions. The business and financial sector can fight fraud by moving from one of hiding to another of prevention and by using the technology of Blockchain. The distributed Blockchain technology leader feature can ensure that all parties associated with Blockchain customers have legitimate and permitted transactions. Blockchain technology is becoming increasingly popular. When it comes to credit card fraud, utilizing technology that has already been created and incorporating Blockchain technology can completely transform the game.

VI. Future Scope:

During the creation and testing of the approach provided, some intriguing issues were asked and these might be further studied. If an anomalous transaction is signed by a user, the answer to that prompt might be used to determine if the transaction is anomalous or not. Other approaches of uncontrolled anomaly and novelty detection, which include statistical to nature-inspired meta-heuristics, optimized type-based artificial neural networks and other techniques for modest data volumes, can also be employed. In conclusion, a comparison may be conducted between the suggested technique and the original procedure, providing further insight into the usability aspect of such an approach.

VII. REFERENCES:

[1] A.A. Taha, S.J. Malebary, An intelligent approach to credit card fraud detection using an optimized light gradient boosting machine, IEEE Access 8 (2020) 25579–25587, doi: 10.1109/ACCESS.2020.2971354.

[2] S. Makki, Z. Assaghir, Y. Taher, R. Haque, M. Hacid, H. Zeineddine, An experimental study with imbalanced classification approaches for credit card fraud detection, IEEE Access 7 (2019) 93010–93022, doi: 10.1109/ACCESS.2019.2927266.

[3] C. Jiang, J. Song, G. Liu, L. Zheng, W. Luan, Credit card fraud detection: a novel approach using aggregation strategy and feedback mechanism, IEEE Internet Things J. 5 (5) (Oct. 2018) 3637–3647, doi: 10.1109/JIOT.2018.2816007.

[4] I. Sohony, R. Pratap, U. Nambiar, Ensemble learning for credit card fraud detection, in: Proceedings of the ACM India Joint International Conference on Data Science and Management of Data Association for Computing Machinery, New York, NY, USA, 2018, pp. 289–294, doi: 10.1145/3152494.3156815.

[5] P.H. Tran, K.P. Tran, T.T. Huong, C. Heuchenne, P. HienTran, T.M.H. Le, Real time data-driven approaches for credit card fraud detection, in: Proceedings of the 2018 International Conference on E-Business and Application. Association for Computing Machinery, New York, NY, USA, 2018, pp. 6–9, doi: 10.1145/3194188.3194196.

[6] I. Sadgali, N. Sael, F. Benabbou, Fraud detection in credit card transaction using neural networks, in: Proceedings of the 4th International Conference on Smart City Applications (SCA '19). Association for Computing Machinery, New York, NY, USA, 2019, pp. 1–4, doi: 10.1145/3368756.3369082 . Article 95.

[7] D. Prusti, S.K. Rath, Web service based credit card fraud detection by applying ma- chine learning techniques, in: Proceedings of the TENCON 2019 - 2019 IEEE Re- gion 10 Conference (TENCON), Kochi, India, 2019, pp. 492–497, doi: 10.1109/TEN- CON.2019.8929372.

[8] M. Zamini, G. Montazer, Credit card fraud detection using autoencoders based clus- tering, in: Proceedings of the 9th International Symposium on Telecommunications (IST), Tehran, Iran, 2018, pp. 486–491, doi: 10.1109/ISTEL.2018.8661129.

[9] S. Akila, U.S. Reddy, Credit card fraud detection using non-overlapped risk based bagging ensemble (NRBE), in: Proceedings of the IEEE International Conference on Computational Intelligence and Computing Research (ICCIC), Coimbatore, 2017, pp. 1–4, doi: 10.1109/ICCIC.2017.8524418.

[10] M.S. Kumar, V. Soundarya, S. Kavitha, E.S. Keerthika, E. Aswini, Credit card fraud detection using random forest algorithm, in: Proceedings of the 3rd International Conference on Computing and Communications Technologies (ICCCT), Chennai, In- dia, 2019, pp. 149–153, doi: 10.1109/ICCCT2.2019.8824930 .

[11] Z. Li, G. Liu, S. Wang, S. Xuan, C. Jiang, Credit card fraud detection via kernel- based supervised hashing, in: Proceedings of the IEEE SmartWorld, Ubiquitous In- telligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City In- novation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI), Guangzhou, 2018, pp. 1249–1254, doi: 10.1109/SmartWorld.2018.00217.

[12] P. Kumar, F. Iqbal, Credit card fraud identification using machine learning ap- proaches, in: Proceedings of the 1st International Conference on Innovations in Infor- mation and Communication Technology (ICIICT), CHENNAI, India, 2019, pp. 1–4, doi: 10.1109/ICIICT1.2019.8741490 .

[13] "HOBA: A novel feature engineering technique for credit card fraud detection with a deep learning architecture," by Xinwei Zhang, Yaoci Han, Wei Xu, and Qili Wang. 2019 is the year of information sciences.

[14] Gregory Vaughan, Gregory Vaughan, Gregory Vaughan, Gregory Vaughan "Efficient big data model selection for fraud detection applications." The International Journal of Forecasting published an article in 2018.

[15] Véronique Van Vlasselaer, Cristián Bravo, Olivier Caelen, Tina Eliassi-Rad, Leman Akoglu, Monique Snoeck, and Bart Baesens, "APATE: A novel method for automatic credit card transaction fraud detection using network-based extensions," Decision Support Systems 75, 2015, pp. 38-48.

[16] "Champion-challenger study for credit card fraud detection: Hybrid ensemble and deep learning." Kim, E., Lee, J., Shin, H., Yang, H., Cho, S., Nam, S. K.,... & Kim, J. I. 214-224 in Expert Systems with Applications, vol. 128, no. 2, 2019.

[17] Siddhartha Bhattacharyya, Sanjeev Jha, Kurian Tharakunnel, and J. Christopher Westland. [17]. Bhattacharyya, Siddhartha, Sanjeev Jha, Kurian Tharakunnel, and J. Christopher Westland. "A comparative analysis of data mining for credit card fraud." 602-613, Decision Support Systems 50, no. 3, 2011.

[18] "Credit card fraud detection: a practical modeling and a novel learning technique," by Andrea Dal Pozzolo, Giacomo Boracchi, Olivier Caelen, Cesare Alippi, and Gianluca Bontempi. IEEE Transactions on Neural Networks and Learning Systems, vol. 29, no. 8, pp. 3784-3797,

[19] IEEE Transactions on Neural Networks and Learning Systems, vol. 29, no. 8, IEEE Transactions on Neural "Hybrid methods for detecting credit card fraud," by Yiit Kültür and Mehmet Ufuk alayan. e12191 in Expert Systems 34, no. 2 (2017).

[20]. Carcillo, Fabrizio, Yann-Al Le Borgne, Olivier Caelen, Yacine Kessaci, Frédéric Oblé, and Gianluca Bontempi. "Detecting credit card fraud using a combination of unsupervised and supervised learning." 2019 is the year of information sciences.

[21]. Rayana, Shebuti, Wen Zhong, and Leman Akoglu. "A bias-variance perspective on sequential ensemble learning for outlier detection." pp. 1167-1172 in IEEE 16th International Conference on Data Mining (ICDM), 2016. 2016 IEEE.

[22]. Veeramachaneni, Kalyan, Ignacio Arnaldo, Vamsi Korrapati, Constantinos Bassias, and Ke Li. "AI 2: teaching a big data machine to protect," says the author. IEEE International Conference on Big Data Security on Cloud (BigDataSecurity), on Intelligent Data and Security (IDS), pp. 49-54. IEEE, 2016.

[23] Yvan Lucas, Pierre-Edouard Portier, Léa Laporte, Liyun He-Guelton, Olivier Caelen, Michael Granitzer, and Sylvie Calabretto are among the authors of "Using multi-perspective HMMs for automatic feature engineering for credit card fraud detection." 393-402 in Future Generation Computer Systems, vol. 102, 2020.

[24] "Managing a pool of rules for credit card fraud detection through a Game Theory-based strategy," by Gabriele Gianini, Leopold Ghemmogne Fossi, Corrado Mio, Olivier Caelen, Lionel Brunie, and Ernesto Damiani. 549-561 in Future Generation Computer Systems, vol. 102, 2020.

[25] "Cost-sensitive Risk Induced Bayesian Inference Bagging (RIBIB) for credit card fraud detection," by S. Akila and U. Srinivasulu Reddy. pp. 247-25 in Journal of Computational Science, vol. 27, no. 1, 2018.

[26] R. Abrams [26] (2017). "Pay for Security Breach Settlement $18.5 million to 40 States." The Times of New York. Business day. Day of business. Retracted: https://www.nytimes.com/2017/05/23/business/settlement-settlement.html.

[27] F. Augustine's (2017). "They're not science fiction things anywhere any more than password-free cellphones." Business Insider. Business Insider. The smartphone/biometrics-are the-stuff-of-science-fiction-2017-12 no longer-informed at https://www.businessinsider.com

[28] CB Research Letters INSIGHTS. "What is the technology of Blockchain?" (2018). The technology https://www.cbinsights.com/research/what-is-blockchain.

[29] R. Choo and L. Zhu, K. Gai and L. (2018). "Cloud Datacenter blockchain-activated reengineering." IEEE Cloud Computing, vol. 5, pp. 6, nov. 2018 pages. 21-25. DOI (2018.06418116): 01109/MCC.

[30] Kurtz, I. Myzyri, M. Wheeler, S. Rizvi, Gualdoni. (2017). "Secure Transaction Algorithm: Securing the use of a two-factor authentication online transaction." CAS, Chicago, Illinois, USA, 30 Octobern - 1 November 2017, 93-99. Procedia Computer Science, 114 (Complex Adaptive Systems Conference with Topic: Cyber-Physical Systems Engineering; doi:10.1016/j.procs.2017.09.016

[31] Secretary of State. Kerner (2018). "The 5M Cardholders, Saks, Lord & Taylor Hit by Data Breach." Eweek.com. Recovered from:https://www.eweek.com/safety/saks-lord-taylor-hit-by-data-breach impacting/(language)/eng-user/cardholder

[32] A. We read about it (2016). "Researchers are fighting to identify real costs of data violations." Eweek.com. Get from: https://www.eweek.com/security/investigator-to-determine true-cost-of-data violations.

[33] S. Muppidi Muppidi (2017). "Companies need to maintain safe users via more than two-factor authentication." Digital Articles, 2-4, Harvard Business Review.

[34] S. Kess and Primoff (2017). "Data Breach in Equifax." CPA Newspaper, 87(12), 14-17.

[35] Qiu, K. Gai, Thuraisingham, Thuraisingham, Tao, and H Thuraisingham (2016). "Proactive user-centric and secure data scheme employing semanthe attribute access controls for financial industry mobile clouds." Computer Systems Future Generation, Volume 80, March 2018, pages 421-429.

[36] L. Zhu, Y. Wu and K. Gai and K. Choo, respectively (2019). "Controllable and reliable cloud data management based on blockchain." Computer Systems for the next generation, Volume 91, February 2019. Overlay panel connections Author LiehuangZhuaYuluWuaKekeGaia Kim Kwang RaymondChoob

[37] K. Gai, Y. Wu, L. Zhu, M. Qiu and M. Shen (2019), Blockchain in Smart Grid, IEEE Transactions on Industrial Computer Sciences, "Privacy-preserving Energy Trading." DOI:109-10-2019-2893433