**Credit Card Fraud Detection System using Machine Learning**

Ganesh A 1A, Dhivakar K 1B, Ganesh S 1C, Hammadh Ahmed A 1D

*DEPARTMENT OF INFORMATION TECHNOLOGY,*

*PANIMALAR ENGINEERING COLLEGE,CHENNAI , INDIA.*

A [ganesharum145@gmail.com](mailto:ganesharum145@gmail.com),

B [kdhivakar.75@gmail.com](mailto:kdhivakar.75@gmail.com),

C [sganeshmani20@gmail.com](mailto:sganeshmani20@gmail.com),

D [hamdh52001@gmail.com](mailto:hamdh52001@gmail.com),

**ABSTRACT :**

Technology is developing every day at a faster rate. As technology is developed, the usage of the Internet is also increasing among people all over the world. The rapid growth in the Electronic commerce industry has led to an impressive expansion within the usage of credit cards. Online transactions have increased their numbers and credit cards hold a huge share in it. Every day, millions of people do online transactions using credit cards. As the usage of credit cards increases day by day, credit card frauds are also increasing frequently which results in huge financial losses. To detect credit card fraud in transactions, machine learning is essential. For predicting these transactions banks make use of various machine learning methodologies, past data has been collected and new features are been used for enhancing the predictive power. We have explained the issue of credit card fraud in this paper. Fraudulent transactions can take many forms and fall under a variety of categories. This study examines four common types of fraud in real-world transactions. Each scam is dealt with by a series of machine learning models, with the optimal solution being chosen through an evaluation. This assessment provides a detailed guide to picking an effective algorithm based on the type of fraud, and it is illustrated with a suitable performance measure. Real-time credit card fraud detection is another important aspect of our project. To do so, we leverage predictive analytics powered by machine learning models and an API module to determine whether a transaction is legitimate or fraudulent. On an unbalanced dataset, we use boosting to implement various machine learning techniques such as logistic regression, naïve Bayes, and random forest with ensemble classifiers. The existing and proposed models for credit card fraud detection have been thoroughly reviewed, and a comparative evaluation of these strategies has been conducted. As a result, various classification models are applied to the data, and model performance is assessed using quantitative metrics like accuracy, precision, recall, f1 score, support, and confusion matrix. Our study's conclusion demonstrates how to train and assess the best classifier utilizing supervised techniques, which results in a better answer.

**KEYWORDS :**

Fraud detection, Credit card, Logistic regression, Decision tree, Random forest.

**INTRODUCTION :**

The most common payment method is a credit card. Identity theft and fraud are on the rise as the number of people who use credit cards grows around the world. Only the card information (card number, expiration date, security code, and so on) is necessary when purchasing a virtual card. The majority of these purchases are made over the phone or on the Internet. A person only has to know the card details to conduct fraud in these kinds of transactions. Credit cards are the most common means of payment for internet purchases. Credit card information should be kept confidential. Credit card information should not be released to protect customer privacy. At the moment of purchase, a manual signature, a PIN, or a card imprint are not required. In the majority of situations, the actual cardholder is unaware that his or her card information has been seen or taken by someone else. The simplest technique to identify this form of fraud is to examine each card's spending history and look for any deviations from the "normal" spending trends. The greatest strategy to lower the rate of successful credit card fraud is to identify fraud by examining the cardholder's existing data buy. Because the data sets aren't available, and the outcomes aren't made public, The available data sets, such as recorded data and user activity, should be used to detect fraud situations. Currently, fraud detection is accomplished using a variety of techniques, including data mining, statistics, and artificial intelligence.

**LITERATURE SURVEY :**

Many approaches have been proposed in previous studies to bring solutions to detect fraud, ranging from supervised approaches to unsupervised approaches to hybrid approaches; this necessitates a thorough understanding of the technologies involved in credit card fraud detection as well as a thorough understanding of the various types of credit card fraud. Fraud trends developed over time, creating new types of fraud, making it a hot topic among scholars. The rest of this section goes on individual machine learning algorithms, machine learning models, and fraud detection systems that have been employed in fraud detection. The issues that arose throughout the evaluation have been investigated to develop an effective machine learning model in the future.

**Comparative Analysis of Various Credit Card Fraud Detection Techniques:**

Any harmful behaviour aimed at causing financial loss to the other party is considered fraud. As the usage of digital money or plastic money grows, so does the fraud linked with it, especially in underdeveloped nations. Credit card fraud has cost customers and banks billions of dollars throughout the world. Fraudsters are always attempting to come up with new ways and tactics to commit fraud, even though there are several measures in place to prevent it. To combat these scams, we need a robust fraud detection system that not only identifies the fraud but also detects it before it occurs and in a precise manner. We also need to make our systems capable of learning from previous frauds and adapting to new fraud tactics in the future. We have discussed the notion of credit card fraud and the numerous forms of fraud in this article. Support Vector Machine (SVM), Artificial Neural Networks (ANN), Bayesian Network, K- Nearest Neighbor (KNN), Hidden Markov Model, Fuzzy Logic Based System, and Decision Trees are some of the approaches available for a fraud detection system.

**METHODOLOGY :**

**TYPES OF ALGORITHMS USED :**

**DATA SET :**

A credit card fraud detection dataset was utilized in this work, which may be acquired from Kaggle. This dataset covers transactions done by European cardholders in September 2013 over two days. There are 31 numerical characteristics in the dataset. Because some of the input variables contain financial information, the PCA transformation of these variables was used to ensure that the data remained anonymous. Three of the listed characteristics were not altered. The feature "Time" displays the time between the first and subsequent transactions in the dataset. The amount of credit card transactions is displayed in the "Amount" feature. The label is represented by the feature "Class," which accepts just two values: 1 in the event of a fraud transaction and 0 otherwise.

**SAMPLING :**

Further, the data set is minimized to 560 transactions. Where 228 fraud and 332 normal transactions.

**DIVIDE THE DATASET :**

The dataset is split into two sections: training data and test data. Seventy per cent of the data set is being trained, while the remaining thirty per cent is being tested. We're employing supervised machine learning methods in this case. Naive Bayes, Logistic Regression, and Random Forest with Boosting Technique are the algorithms.

**NAÏVE BAYES :**

Bayes theorem: Bayes theorem finds the probability of an event occurring given the probability of another event that has been already occurred. Naïve Bayes algorithm is easy and fast. This algorithm needs less training data and highly scalable P (A/B) = (P (B/A) P (A)) / P (B) Where P (A) – Priority of A P (B) – Priority of BP (A/B) – Posteriori priority of B

**LOGISTIC REGRESSION :**

Logistic regression is used to predict binary values in a set of independent variables (1 / 0, Yes / No, True / False). Dummy variables are used to represent binary/categorical values. When the resultant variable is categorical, the log of chances is utilized as the dependent variable in logistic regression. It also predicts the likelihood of occurrence of an event by fitting data to a logistic function.

**RANDOM FOREST :**

The random forest is a supervised learning technique that combines numerous decision trees into a single "forest" at random. To enhance accuracy, the idea is to use a set of decision models rather than a single learning model. The main difference between this method and the usual decision tree technique is that the root nodes have randomly generated splitting nodes.

**BOOSTING TECHNIQUE :**

AdaBoost (machine learning algorithm) is a machine learning algorithm. Developed mostly for binary categorization. Each occurrence in the training dataset is weighted in AdaBoost. The starting weight is set to:

(1/n) = weight (xi)

Where xi – the first training instance

n – the total number of training instances

**ALGORITHMS FOR FINDING THE BEST ALGORITHM :**

Step 1: Import the dataset

Step 2: Convert the data into data frames format

Step3: Do random oversampling using the ROSE package

Step4: Decide the amount of data for training data and testing data

Step5: Give 70% data for training and the remaining data for testing.

Step6: Assign train dataset to the models

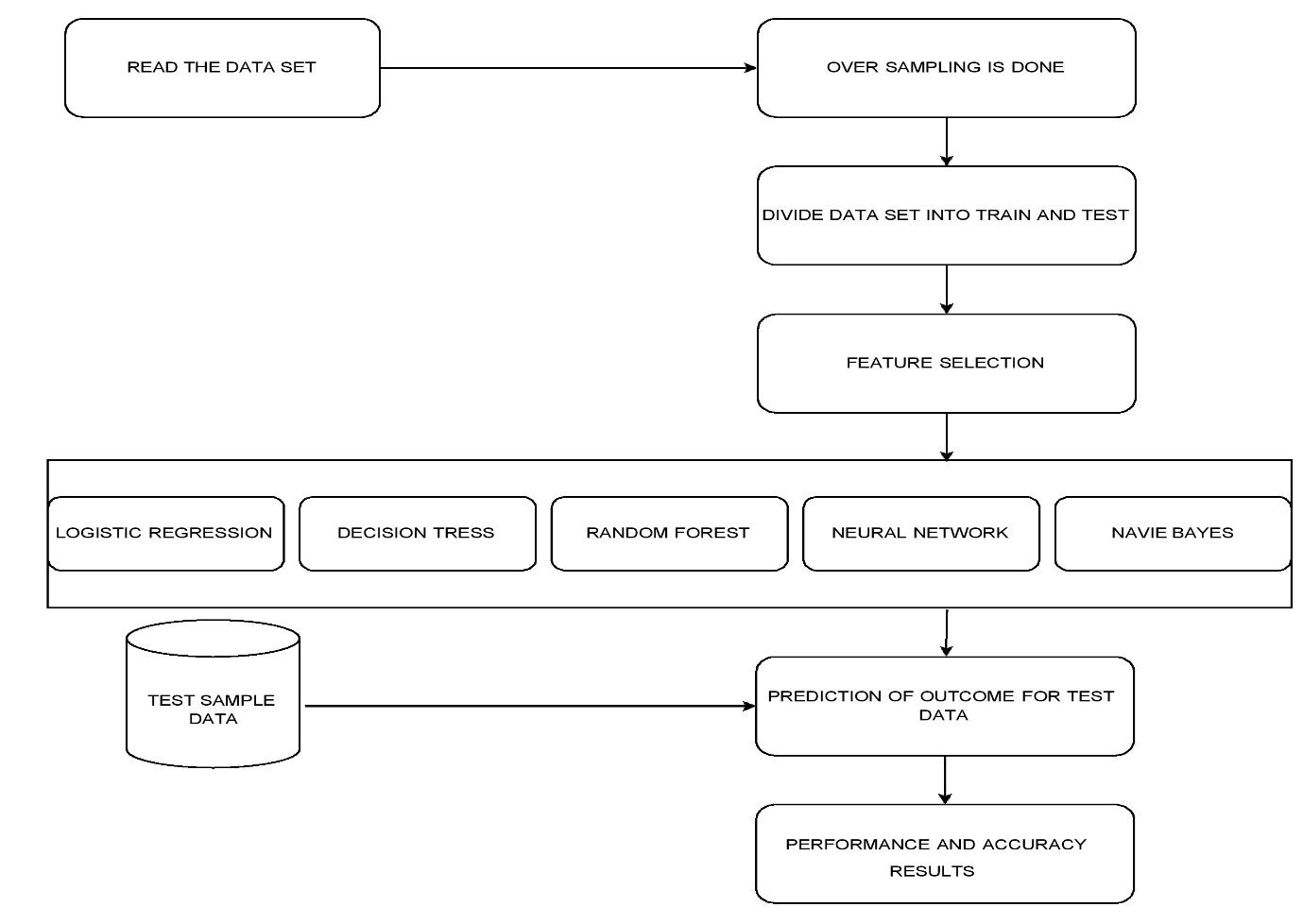
Step7: Choose the algorithm among 3 different algorithms and create the model Step8: Make predictions for the test dataset for each algorithm

Step9: Calculate accuracy for each algorithm

Step10:Apply confusion matrix for each variable

Step11:Compare the algorithms for all the variables and find out the best algorithm.

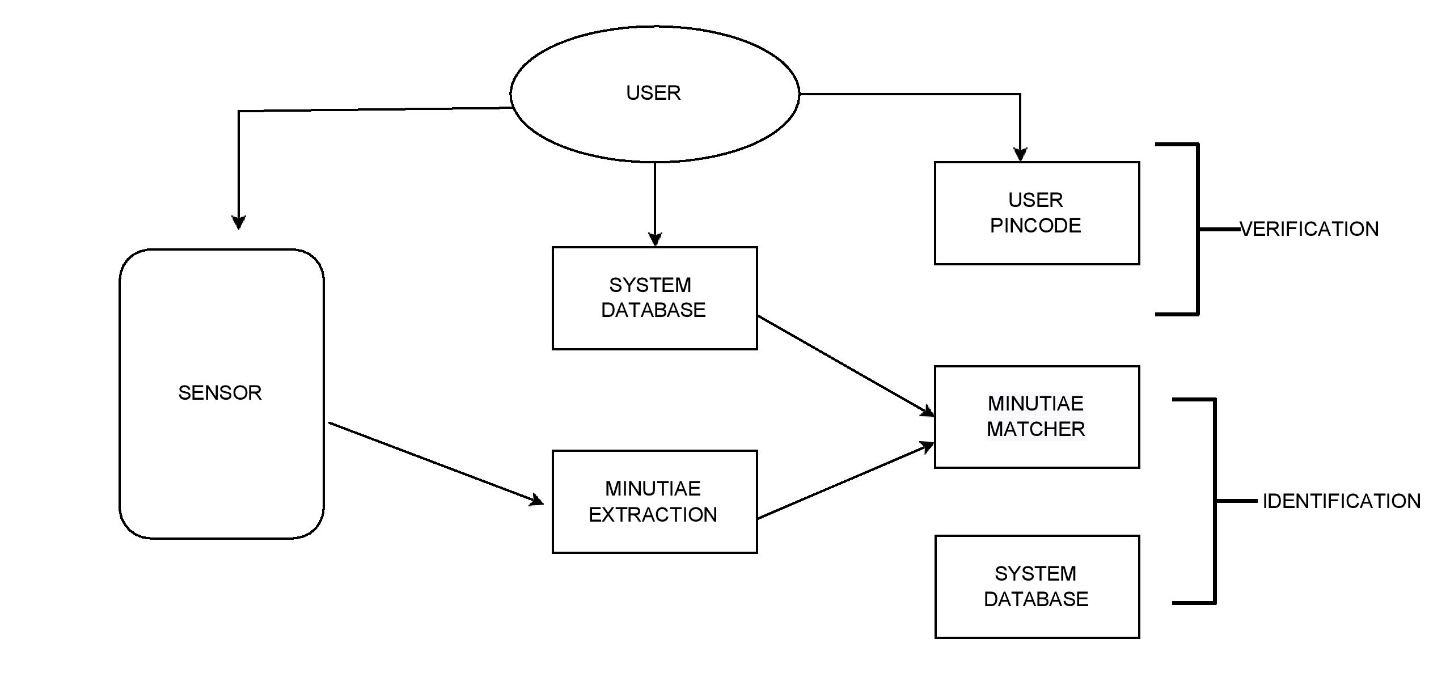
**FLOWCHART :**

****

**FRAUD DETECTION USING FINGERPRINT AND FACE RECOGNITION :**

* One of the new features we are going to add to the credit card fraud detection system is to detect fraud by fingerprint and face recognition.
* First, the credit card holder’s details should be verified by the bank.
* The fingerprint and photo of the credit card holder should be scanned and verified with the identity card of the person by the respective banks.
* The fingerprint and photo of the credit card holder should be printed on the credit card.
* During the usage of credit cards, the fingerprint and face of the credit card holder should be verified.
* If it is not verified, they cannot use the credit card.
* This will reduce credit card fraud.
* If the cardholder wishes to let his/her card be used by some other person whom he trust, the cardholder needs to give access to the respective persons.
* The access can be given the identifying the face of the respective user.
* When a user uses credit for an online transaction, the face of the user is captured by a camera, and a photo is sent to the credit card holder’s mobile number.
* The credit card holder can give access by verifying the photo of the respective user and then the payment will be processed.
* A special camera and a fingerprint sensor should be inserted in all ATMs.
* If the cardholder wishes can let access to the person whom he/her trust can add the fingerprints to the card.
* A fingerprint sensor and a camera should be inserted in the payment terminal also.

**FLOWCHART :**

****

**CONCLUSION :**

In this paper, we concentrated on the utilization of AI like Naïve Bayes, Logistic relapse, Arbitrary timberland with supporting and shows that it demonstrates exact in deducting false exchange and limiting the number of misleading alarms. Regulated learning calculations are novel in this writing as far as application space. If these algorithms are applied to the bank credit card misrepresentation identification framework, the likelihood of extortion exchanges can be anticipated not long after charge card exchanges. Also, a progression against extortion systems can be embraced to keep banks from incredible misfortunes and lessen gambles. The target of the review was taken uniquely in contrast to the regular order issues in that we had a variable misclassification cost. Precision, recall.f1-score, support, and precision are utilized to assess the presentation for the proposed framework. By looking at every one of the three strategies, we tracked down that arbitrary woods classifier with supporting method is better compared to the calculated relapse and credulous Bayes techniques

**REFERENCES :**

[1]. Fabiana Fournier, Ivocarriera, Inna skarbovsky, The Uncertain Case of Credit Card Fraud Detection, The 9th ACM International Conference On DistributedEvenBaseddd Systems(DEBS15) 2015.

[2]. Yashvi Jain, Namrata Tiwari, ShripriyaDubey, Sarika Jain, A Comparative Analysis of Various Credit Card Fraud Detection Techniques, Blue Eyes Intelligence Engineering And Sciences Publications 2019

[3]. Dinesh L. Talekar, K. P. Adhiya, Credit Card Fraud Detection System-A Survey, International journal of modern engineering research(IJMER) 2014.

[4]. SamanehSorournejad, Zahra Zojaji, Reza Ebrahimi Atani, Amir Hassan Monadjemi, A Survey of credit card fraud detection techniques: Data and techniques oriented perspective.

[5]. Lakshmi S V S S, Selvani Deepthi Kavila, Machine learning for credit card fraud detection system, International Journal Of Applied Engineering Research ISSN 2018