International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Enhancing Security in E-commerce and E-payment Systems through the Implementation of Machine Learning for Credit Card Fraud Detection

Chetna Bisht[1] , Sachin Gaur[2] , Nidhi Mehra[3]

Submitted: 28/01/2024 **Revised**: 06/03/2024 **Accepted**: 14/03/2024

Abstract: The expeditious growth of electronic payment and e-commerce systems has brought unparalleled convenience to consumers globally. Yet, this digital transformation has concurrently fueled an increase in financial deception, notably under the guise of credit card fraud. It is paramount to identify and thwart these illicit activities to safeguard the trust and integrity of online transactions. In this paper, an innovative method is proposed that utilized the machine learning method in the realm of credit card fraud detection, with a specific emphasis on fortifying the security of e-commerce and electronic payment systems. In this scheme, we conducted extensive experiments using a dataset obtained from Kaggle in order to evaluate the effectiveness of our proposed technique. Different classifiers namely LR, DT and RF have been used in this proposed scheme. The RF model with oversampling yielded the highest accuracy, recall, precision, and F1 Score, which is 99.97%. Accordingly, it shows that RF classifiers are effective for oversampling. The outcomes of our assessment underscore its capability to notably bolster security within e-commerce and electronic payment systems. These findings emphasize the significance of harnessing machine learning methodologies in the continual effort to combat credit card fraud in the era of digital transactions.

***Keywords**: E-commerce, E-payment systems, Fraud detection, Machine learning.*

1. Introduction

In the age of digital transactions and e-commerce, the convenience of using credit cards has become an integral part of our daily lives. As the frequency of credit card usage has risen, so too has the unfortunate occurrence of credit card fraud [1]. Fraudulent activities, such as unauthorized transactions and identity theft, pose significant threats to both financial institutions and consumers alike. These vitriolic attempts not only result in substantial financial losses but also erode trust in electronic payment systems. Credit card fraud has two types: "inner card fraud" involving collusion between merchants and cardholders for cash fraud, and "external card fraud" with stolen or counterfeit cards used for unauthorized purchases and obtaining cash by buying easily convertible high-value items [2]. Though most activities are legal, the relatively small number of fraudulent ones can still result in substantial losses [3]. Machine learning methods have gained significant prominence in the sphere of credit card fraud detection owing to their capacity to rapidly scrutinize vast quantities of transaction data and pinpoint anomalies with exceptional accuracy. This fusion of technological prowess and financial vigilance constitutes a compelling avenue for upholding the integrity of credit card transactions [4]. In this research embarks on an exploration of credit card fraud detection by leveraging the capabilities of machine learning techniques. It delves deeply into the various methodologies, algorithms, and strategies employed to identify and counteract fraudulent activities, with the ultimate goal of bolstering the security and trustworthiness of electronic payment systems. By tackling the challenges, tracking advancements, and dissecting the subtleties within this constantly evolving field, this research endeavours to highlight the pivotal role that machine learning assumes in strengthening the foundations of financial security [5]. To traverse the fundamental principles underpinning credit card fraud detection, scrutinize the eclectic machine learning techniques employed, and critically assess their efficacy It also examines the limitations and ethical considerations associated with these systems, as well as exploring avenues for future research to further improve the accuracy and robustness of credit card fraud detection [6]. Through the utilization of data-driven knowledge and sophisticated algorithms, made a valuable contribution to the continuous strengthening of the underpinnings of electronic payment systems [7].

In light of the escalating challenge posed by credit card fraud, the convergence of cutting-edge technology and financial security has spawned a highly promising solution: machine learning [8]. In a realm where safeguarding financial integrity is of utmost importance, the synergy between machine learning and credit card fraud detection represents a profound opportunity [9]. In doing so, we can aspire to instil confidence and trust in both individuals and institutions, enabling secure digital transactions for all. Customer trust and satisfaction are essential to any

[1] *B.T.Kumaon Institute of Technology Dwarahat, India*
ORCID ID: 0009-0005-4000-7930
[2] *B.T.Kumaon Institute of Technology Dwarahat, India*
ORCID ID: 0000-0002-7638-3875
[3] *Graphic Era Hill University Dehradun, India*
ORCID ID: 0000-0003-3628-2356
Corresponding Author Email: ersgaur1234@gmail.com

company's success and sustainability. This precisely gave us the motivation to work on this research idea using machine Learning Techniques. The safety of consumer financial transactions must always be guaranteed. Customers will be more likely to continue using a financial institution's or retailer's services if they believe their transactions are secure. The remaining sections of this paper have been structured as a literature survey in Section II, which involves systematically reviewing and summarizing existing research on credit card fraud detection to inform and contextualize this research. Introduction to Classifiers in Section III, which provides a brief overview of the classifiers Proposed methods and techniques are described in Section IV. Experimental Results and Analysis are given in Section V, which provides a detailed analysis and outcome of our work. In Section VI, the compression study and then in Section VII conclusion and Future scope are given.

2. Literature Review

Despite the promising advances in machine learning-based credit card fraud detection for e-commerce, there is a critical research gap and a notable deficiency in comprehending and countering adversarial attacks uniquely crafted to exploit these systems. The research question that highlights this gap is as follows:

"To what extent are machine learning-driven credit card fraud detection systems in e-commerce susceptible to adversarial attacks, and what tactics can be devised to strengthen their resilience and security in the face of such threats?"

This research gap is vital due to evolving fraudster tactics exploiting machine learning model weaknesses, which are essential for long-term e-commerce fraud prevention and security. Various techniques, including neural networks, decision trees, logistic regression, and advanced methods like SVM and random forests, are applied in credit card fraud detection research.

➢ Traditional Fraud Detection Methods

In the initial stages of fraud detection, the predominant approaches are centered around rule-based systems and statistical methods. These earlier methodologies frequently fell short in terms of flexibility and precision compared to the capabilities afforded by machine learning, Plakandaras et.al.[9].

➢ Machine Learning in Credit Card Fraud Detection:

Numerous research investigations have delved into employing machine learning algorithms for the purpose of credit card fraud detection. Scholars have conducted experiments using a range of models with the aim of improving accuracy and effectiveness in this domain, Raj et al. [4].

➢ Imbalanced Datasets:

Several difficulties arise when credit card fraud is detected due to imbalanced datasets, Daniel et.al.[10].

- Bias: Due to models' inclination to favour the majority class, fraudulent transactions are not well detected.

High False Negatives: Models may fail to identify a large number of fraud situations, which could cost money.

- Evaluation Complexity: While accuracy, recall, and precision score
- Are more important metrics, they are more difficult to tune? Accuracy is not a trustworthy statistic.
- Costs of Data Collection: Model training is hampered by the limited and high cost of labelling fraud data.
- Model Complexity: Training time and model complexity may both rise with data balancing.
- Generalization: Models may have trouble making good generalizations if datasets are imbalanced.
- Concept Drift: Models could find it difficult to adjust to evolving fraud trends.
- Threshold Selection: Selecting a threshold for categorization can be difficult and have an impact on false positives and misses.

Addressing class imbalances in fraud datasets has been a recurring challenge. Sampling techniques can mitigate this problem.

➢ E-commerce and E-payment Specific Challenges:

Scholars have recognized the unique difficulties presented by e-commerce and electronic payment systems, which encompass a large number of transactions, a wide range of payment methods, and international transactions. In response to these challenges, they have tailored fraud detection solutions accordingly Wang et al. [11]. Credit card fraud detection distinguishes legitimate from fraudulent transactions based on spending behavior. Various techniques, including neural networks, genetic algorithms, and decision trees, have been applied to this field. Evgeniou et al. [12] and research has compared methods like logistic regression and naive bayes and assessed Bayesian models and neural networks. Some studies explore advanced data mining approaches like support vector machines and random forests, while others concentrate on neural networks and logistic regression for credit card fraud detection.

West et al. [13] proposed a method that categorizes financial fraud detection by algorithm and fraud type. It notes the prevalence of neural networks and logistic regression, and the success of credit card fraud detection. It emphasizes the need for innovation to address evolving fraud tactics, especially in under-researched areas like insurance fraud. A novel credit card fraud detection approach using a Convolutional Neural Network (CNN) and "trading entropy" is presented by Fu et al. [14]. It overcomes imbalanced data challenges and outperforms conventional methods. This research advances fraud detection in financial transactions. Awoyemi et al. [15] conducted a comparative analysis of various ML methods using the European cardholders' credit card fraud dataset. They adopted a hybrid sampling strategy to solve the dataset's imbalance and assessed KNN, LR and NB but this study did not explore feature selection techniques. Khare et al. [16] presented a credit card fraud detection technique, with a specific focus on highly imbalanced datasets. They assessed the performance of Decision Trees, SVM, Logistic Regression and Random Forest considering various metrics such as accuracy, sensitivity, specificity, and precision. Their findings showed that Logistic Regression achieved quite good accuracy, but they concluded that Random Forest was the most accurate algorithm for detecting fraud, while SVM faced challenges due to data imbalance, resulting in suboptimal performance in credit card fraud detection. Varmedja et al. [17] proffered a method for detecting credit card fraud using ML. They tested the Random Forest (RF), Naive Bayes (NB), and Multilayer Perceptron (MLP) algorithms on a dataset containing credit card transactions. To address the dataset's imbalance, they carried out the Synthetic Minority Oversampling Technique (SMOTE). The study's findings indicated a high level of fraud detection accuracy, but it additionally suggested more research be done to use feature selection strategies to boost the effectiveness of other ML methods.

A study examining various algorithms for credit card fraud detection, including Naïve Bayes, Logistic Regression, J48, and Adaboost is proposed by Naik et al. [18]. Naïve Bayes relied on Bayes' theorem for classification. Logistic Regression, typically used for classification tasks, is analogous to linear regression. J48 generated decision trees for classification, primarily focusing on constant and categorical variables. Ad boost, a widely used algorithm, aimed to enhance the performance of decision trees in binary classification. The research found that both Ad boost and Logistic Regression achieved the highest accuracy, with the choice between them often influenced by computational time. Ultimately, Adaboost was preferred due to its efficiency in detecting credit card fraud. Jain et al. [19] presented a study on credit card fraud detection techniques. They explored several algorithms, such as SVM, KNN, Decision Trees, ANN, Bayesian Networks, Hidden Markov Models and Fuzzy Logic systems. Their findings revealed that KNN, SVM and Decision Trees achieved moderate accuracy, while Logistic Regression and Fuzzy logic yielded lower accuracy. On the other hand, KNN, Neural Networks, Naive Bayes and Fuzzy systems demonstrated high detection rates. Logistic Regression, SVM, and Decision Trees also performed well but at a medium level. ANN and Naïve Bayesian Networks emerged as top performers across all parameters, despite being computationally expensive. However, they noted a drawback: these algorithms did not consistently produce the same results with different types of datasets. KNN and SVM excelled with small datasets, while logistic regression and fuzzy logic systems showed accuracy with raw, unsampled data.

Several ML methods, including Isolation Forest, Decision Tree and Logistic Regression were employed for credit card fraud detection using the European credit cardholder fraud dataset is suggested by Dornadula et al. [20]. For handling the dataset's imbalance, they used the SMOTE sampling technique. Sahayasakila et al. [21] introduced two crucial algorithmic techniques: Whale Optimization Techniques (WOA) and Synthetic Minority Oversampling Techniques (SMOTE). Their goal was to enhance convergence speed and address data imbalance in credit card fraud detection. They used SMOTE to balance the classes by generating synthetic transactions and employed WOA for optimization. This combination of techniques not only improved convergence speed but also enhanced the reliability and overall efficiency of the system. Khatri et al. [22], conducted a study evaluating Machine Learning techniques for credit card fraud detection. They considered Decision Tree (DT), k-Nearest Neighbor (KNN), Logistic Regression (LR), Random Forest (RF), and Naive Bayes (NB) algorithms. Their experiments used a dataset with a severe class imbalance, focusing on the precision metric. Precision scores were as follows: DT (85.11%), KNN (91.11%), LR (87.5%), RF (89.77%), and NB (6.52%). Ileberi et al. [23] suggested a method for credit card fraud detection using Machine Learning techniques. They applied Logistic Regression (LR), Decision Tree (DT), Support Vector Machine (SVM), and Random Forest (RF) algorithms to a dataset with a significant disparity between legitimate and illegal transactions. High accuracy scores were obtained, but the authors recommended that the performance of the classifiers could be further improved by using sophisticated pre-processing techniques. Seera et al. [24] developed an intelligent system for detecting payment card fraud using Genetic Algorithms (GA) for feature selection and aggregation. They implemented various ML algorithms to validate their approach, with GA-RF achieving 77.95% accuracy, GA-ANN reaching 81.82%, and GA-DT attaining 81.97%.

Prior research in credit card fraud detection has laid the foundation for the utilization of machine learning

techniques to combat financial fraud in the digital age Raghavan et al. [25].

This research builds upon this body of knowledge by introducing a novel approach and demonstrating its potential to significantly enhance the security of e-commerce and e-payment systems. These findings highlight the ongoing importance of utilizing machine learning in the ongoing battle against credit card fraud.

3. Classifiers

A machine learning classifier is a computational model that assigns input data points to predefined categories or classes based on learned patterns and features. Different classifiers are used in machine learning to solve various classification problems by leveraging their unique algorithms and capabilities, such as decision trees for interpretability, support vector machines for handling complex data, and deep neural networks for intricate pattern recognition Lim et al., [26] .The terms "classifiers," "algorithms," and "data mining techniques" all refer to the same class of techniques that are used in supervised machine learning systems Seify et al. [27].

Logistic Regression

The Logistic Regression (LR) classifier, often known as the Logit classifier, is a supervised machine learning technique commonly employed for binary classification tasks ,Velasco et al. [28]. A linear function is supplied to the logit function in LR, a unique kind of linear regression.

$$y = \alpha 0 + \alpha 1X1 + \alpha 2X2 + \cdots + \alpha nXn \quad (1)$$

Where, α0 is intercept and α1, α2,… are the slopes against independent variables X1-Xn. Logistic Regression is a versatile algorithm used for both regression and classification tasks, with a primary focus on binary classification. It predicts the probability of an instance belonging to a particular class and employs a threshold to make categorical predictions. Mathematically, Logistic Regression models the probability of a binary outcome (e.g., 0 or 1) using the logistic function. The logistic function maps the value of the input X to a probability between 0 and 1.

Functioning of Logistic Regression:

In a binary classification scenario, Logistic Regression operates Mehbodniya et al.,[29] as follows:

- Calculate the linear combination of input features and coefficients
- Apply the logistic function to obtain the probability:
- If the probability exceeds a threshold (usually 0.5), classify the instance as the positive class (e.g., 1); otherwise, classify it as the negative class (e.g., 0).

Mathematically, Logistic Regression models the probability of class membership and provides a versatile tool for classification tasks, taking into account the relationship between features and outcomes. Hussain et al. [30].

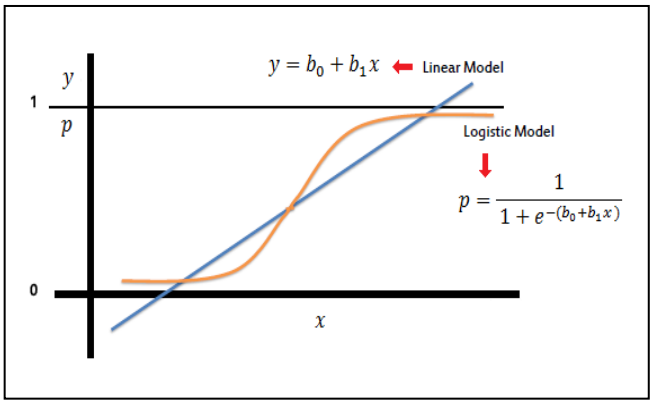

Fig 1. Graphical Representation of Logistic Regression

The sigmoid function is referred to as an activation function for logistic regression and is defined as:

$$F(x) = \frac{1}{1+e^{-(b0+b1x)}} \quad (2)$$

where,

- e = base of natural logarithms
- value = numerical value one wishes to transform
- f(x) value will range from 0 to 1. The probability f(x) is what determines a given class's prediction. A given class is predicted more accurately when q is closer to 1.

For categorical dependent variables, qualitative response models are suitable. Because the dependent variable in our study, "fraud," contains two categories (binary), logistic regression is a frequently used technique to handle situations like these. Bhattacharyya et al. [31].

Decision Trees

Decision Trees are a supervised learning methodology that offers a graphical representation of potential solutions based on specific conditions. Patel et al. [32]. This tree-like structure, serves as a classifier, primarily employed for solving classification problems. Decision Trees start at a root node, with internal nodes representing features within the dataset, branches denoting decision rules, and each leaf node signifying a potential outcome.

Mathematically, Decision trees attempt to divide the feature space into areas that match to several class labels. The root node initiates this process, and at each internal node, a decision rule is applied to determine the path to follow, based on the values of the features.

A class label or a result is represented by each leaf node. A Decision Tree asks a question at each node and divides the

dataset into sub-trees based on the answer. The algorithm's ability to solve both classification and regression problems makes it versatile.

Functioning of Decision Trees:

The operation of Decision Trees can be broken down into several steps:

- Starting with the root node (S), which encompasses the entire dataset.
- Identifying the best feature (Trait) within the dataset using an Attribute Selection Measure.
- When further splitting is not feasible, the node becomes a leaf node.
- The root node is divided into a decision node and a leaf node based on class labels.
- Nodes continue to subdivide into two leaves, ultimately providing classification outcomes.

Mathematically, the decision rule at each internal node can be represented as:

Decision Rule (Node):

Feature->Split Condition->Next Node

A "Decision Rule Node" in a decision tree is a point where a decision is made based on a specific feature and a split condition. The next node is determined by the outcome of that decision, leading to different branches in the tree. Decision trees are used in machine learning for making decisions or classifications based on input data features.

Random Forest

The strength of numerous decision trees is combined in Random Forest, a potent ensemble learning technique, to increase the precision of prediction. Rather than relying on a single decision tree, Random Forest creates and aggregates predictions from numerous decision trees. This ensemble approach significantly improves accuracy and reduces the risk of overfitting Lingjun et al. [33].

Mathematically, a random forest (RF) is defined as RF = {g(X, θk)}, where {θk } denotes independently distributed, identically distributed trees that vote on an input vector X, given a number of trees k. It is the prediction that has received the most votes.

Assume that there are N decision trees in the forest.

Random Forest (ensemble)= {Decision Tree1,Decision Tree 2,…,Decision Tree N}

Predictions from each decision tree are combined using majority voting:

Final Prediction=Majority Vote (Predictions from

Decision Trees)

This approach helps reduce variance and improve the model's overall learning capability.

Functioning of Random Forest:

The following steps can be used to describe how Random Forest works:

- Select a subset (K) randomly from the training set of data points.
- Using the chosen data points (subsets), construct decision trees.
- Specify how many (N) decision trees you want to build in the forest.
- Duplicate Steps 1 and 2 to generate multiple decision trees.
- Predict new data points by aggregating predictions from all decision trees using majority voting.

Mathematically, the ensemble of decision trees combines to form a robust predictor, ensuring accurate classification even with large datasets.

4. Proposed Methods and Techniques

In our proposed method for credit card fraud detection using machine learning involves a systematic approach to address the challenges associated with imbalanced datasets. Figure.2. illustrates the step-by-step procedure for identifying credit card fraud:

Algorithm of Proposed Method

This algorithm represents the strategy we suggested as a step-by-step solution.

Step 1: Add the dataset to Jupyter Notebook from the local storage.

Step 2: Change the data's format to that of data frames.

Step 3: Randomly sample the selected dataset.

Step 4: Decide how much data will be used for training and testing.

Step 5: Give 80% of the data for training and the remaining 20% for testing in step 5.

Step 6: Give the models the train dataset for use in training.

Step 7: Apply the algorithm to three different algorithms in step 7 and build a model for each.

Step 8: Making predictions for the test dataset for each method is step eight

Step 9: Use the accuracy_score, precision_score, recall_score, and f1_score to determine the accuracy of each method.

4a. Data Collection and Preprocessing

- Assemble an extensive dataset comprising credit card transactions, ideally covering a wide spectrum of both genuine and illegimtimate transactions.

- Execute data preprocessing tasks to address missing data and tackle imbalances, ensuring the dataset is well-suited for analysis using machine learning techniques.

4b. Data Analysis

Data analysis in ML involves extracting valuable insights from datasets through preprocessing, feature engineering, and statistical examination, enabling the development of accurate models.Pandas is a Python library used for working with data sets. It has functions for analyzing, cleaning, exploring, and manipulating data.

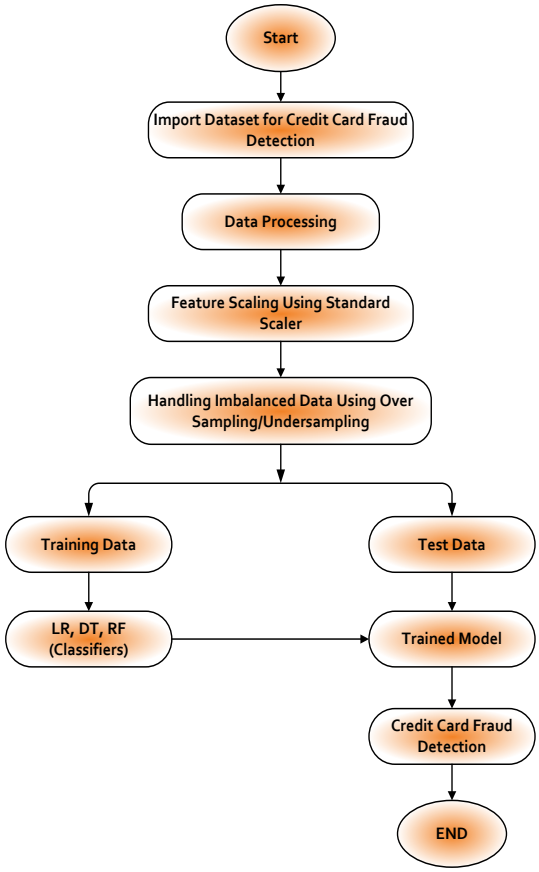

Fig 2. Block Diagram for the Proposed Model

4c. Data Sampling

A data set is considered class-imbalanced when there is a substantial disparity in the number of samples between the two classes, with one class having significantly more instances than the other, Khalilia et al.[34]. To address the problem of class imbalance in machine learning datasets, data sampling strategies such as under-sampling and over-sampling are used. Class imbalance happens when there are disproportionately fewer occurrences of one class than the other in a binary classification problem.

Credit card fraud datasets are heavily imbalanced, in real scenario, with the vast majority (98%) representing legal transactions and only a small fraction (2%) being fraudulent, Zareapoor et al.[35]. Due to the imbalance, the model may favor the dominant class while underperforming the minority class, which might result in biased model performance.

Class imbalance is seen in the bar plot below, where "Class 0" has more instances than "Class 1."

Class

0 275190

1 473

Name: count, dtype: int64

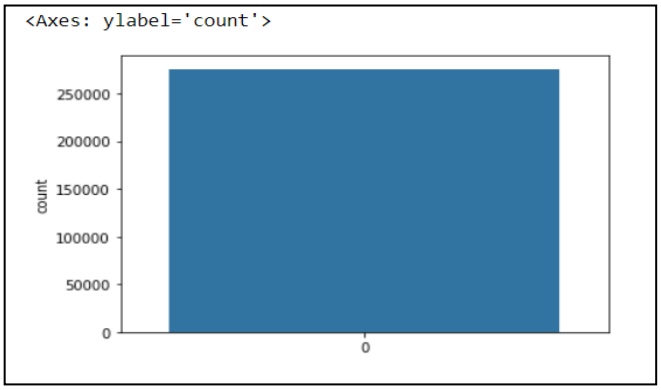

Fig 3. Class Imbalance

Figure.3. indicates class imbalance, with "Class 0" having more instances than "Class 1," highlighting the need for techniques like resampling, algorithm adjustments, or cost-sensitive learning in machine learning models. Now explore under-sampling and over-sampling in more detail next sub section.

4c.i. Undersampling

To balance the majority class with the minority class, undersampling requires reducing the number of instances in the majority class. To do this, a subset of instances from the majority class are chosen at random in order to equal the number of examples in the minority class. A more balanced dataset can be produced by undersampling, which also stops the model from favoring the majority class.

4c.ii. Oversampling

To balance the minority class with the majority class, oversampling entails adding more instances to the minority class. It is possible to accomplish this by copying already-existing instances or by creating synthetic instances by employing methods like Synthetic Minority Over-Sampling Technique (SMOTE). To help the model better understand the minority class, oversampling seeks to give it additional samples of this group.

- Given a data frame with N rows, random sampling will take out X randomly chosen rows, where X ≤ N. The sample () function in Python's pandas package allows for random sampling. There are two different ways to express how many samples need to be extracted:

The exact number of random rows to extract and the percentage of random rows to extract should be specified. Indicating the percentage is a number between 0 and 1.

4d. Data Visualization

Data graphs in machine learning are visual representations that simplify complex data using graphical elements like scatter plots and histograms. They help in pattern recognition, data exploration, and communication of insights. Data graphs are essential for understanding data, guiding modeling decisions, and conveying results effectively.

-Using matplotlib: Matplotlib is a data visualization library and 2-D plotting library of Python.

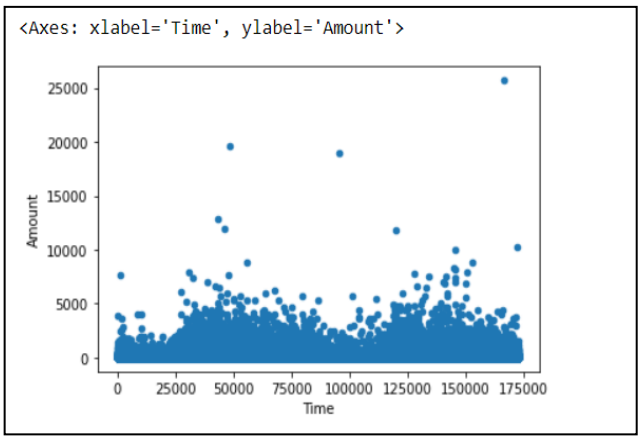

Fig 4 Amount Vs Time Plot

Figure.4. shows that most of the fraudlant transactions takes place below amount of 5000.

4e. Machine Learning Method

The train_test_split() method:The Sklearn train-test split is a vital step in machine learning model development. It involves dividing a dataset into two parts: a training set (used to train the model) and a testing set (used to evaluate the model's performance).

• **Train set:** A set of data used to fit the model is referred to as the training dataset. the dataset used to train the model. The model observes and absorbs this data.

• **Test set:** To accurately assess how well a final model fits, the test dataset—a subset of the training dataset—is used.

```
X_train,X_test,Y_train,Y_test=train_test_split(
X,Y,test_size=0.2,stratify=Y,random_state=42)
```

The line of code is used in Python for splitting a dataset into training and testing sets for machine learning. Here's an explanation of each part:

- 'X_train' and 'X_test' are variables that will contain, respectively, the feature data for the training and testing sets. The feature matrix or dataset that contains the input features for your machine learning model is typically represented by the letter "X."
- 'Y_train' and 'Y_test': These variables will contain the target or label data for the training and testing sets, respectively. The goal or output variable that your machine learning model is attempting to predict is often represented by the letter "Y."
- 'train_test_split()' is a function offered by the scikit-learn Python package, which is frequently used for machine learning applications. A dataset is divided into two subsets using this technique: one for training your model and the other for testing its effectiveness.
- 'X' and 'Y': These are your initial feature and target datasets that you want to divide into training and testing groups.
- 'test_size=0.2': This parameter indicates that you wish to give the testing set 20% of your data. In other words, you will train your machine learning model using 80% of the data.
- 'stratify=Y': This argument makes sure that the splitting procedure keeps the target variable 'Y"s class distribution the same in both the training and testing sets. It ensures that each class is fairly represented in both subsets, which is helpful when working with unbalanced datasets.
- `random_state=42`: This parameter sets a random seed for the random number generator used during the data splitting process. Setting a specific seed ensures that the data split is reproducible. In this case, `42` is an arbitrary value and can be any integer.

We can use the datasets 'X_train', 'X_test', 'Y_train', and 'Y_test' for training and testing our machine learning model after running this line of code. Our original data will make up 80% of the training data, while the remaining 20% will make up the testing data.

Algorithm for Computation of the Fitness Function

Input: X, Y, represent the input vector and the dependent variable respectively.

Output: Acc; the RF classifier's Accuracy

Step 1: Divide X and Y into X_train, X_test, Y_train, and Y_test

Step 2: Initiate the RF classifier, identified as rf.

Step 3: Fit rf via Y_train and X_train

Step 4: Use X_test to evaluate rf.

Step 5: Obtain the Y_pred3 predictions

Step 6: Utilizing Y_pred3 and Y_test, obtain the Acc.

Pseudocode for train_test_split function

```
import random

def train_test_split(data, test_size=0.2, random_state=42):

    if random_state:

        random.seed(random_state)

    num_test_samples = int(len(data) * test_size)

    random.shuffle(data)

    training_data = data[num_test_samples:]

    testing_data = data[:num_test_samples]

    return training_data, testing_data
```

Randomization helps ensure data impartiality. During training, the model learns patterns from the training set, while the testing set assesses its ability to generalize to new data. Evaluation metrics like accuracy measure the model's performance, guiding further refinement. A successful train-test split ensures a model's effectiveness in real-world predictions.

By incorporating these proposed methods and techniques into our research, we aim to provide valuable insights into the effectiveness of machine learning-based credit card fraud detection, contributing to the ongoing efforts to secure electronic payment systems and protect consumers from financial fraud.

Reducing the occurrence of credit card theft necessitates a comprehensive approach that combines both detection and prevention measures, Strelcenia et al. [36].

Traditional rule-based systems are budget-friendly with lower data and computing needs, but machine learning methods often deliver superior and cost-effective long-term results, Shah et al. [37].

5. Experimental Result & Analysis

Machine learning models utilize diverse approaches for decision-making, like LR, DT, RF etc. Our analysis shows that The Random Forest classifier achieves the highest accuracy as shown in graph whereas the decision trees are interpretable but prone to overfitting; random forests, which combine multiple trees, offer better generalization and robustness, making them suitable for complex data with improved performance potential. This output has been obtained using seaborn library in jupyter notebook and the language used is Python.

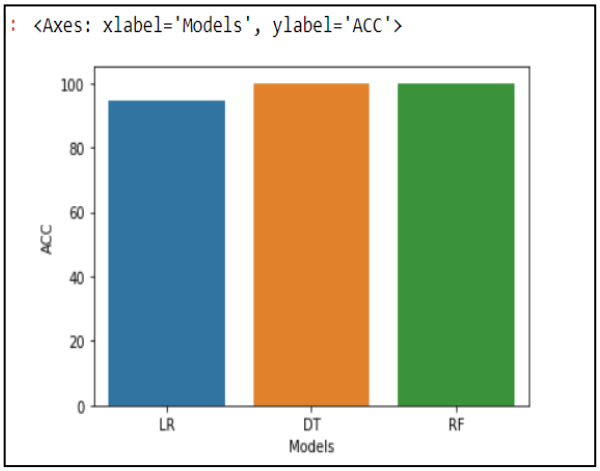

Fig 5. Classifiers Comparison

As shown in Figure 5, blue bar represents logistic regression, orange bar represents decision tree and green bar represents random forest classifier. The maximum accuracy is attained by RF classifier. ACC(Accuracy) is defined as the percentage of cases that are accurately classified. This classification performance measure is one of the most popular ones. Accuracy is the number of correctly predicted events. generalized or specific predictions for binary classification models.

A "density vs. class" graph visualizes the distribution of predicted and actual class labels, assisting in assessing the model's classification accuracy and any discrepancies in class predictions. Graphs like these are available for a range of classifiers, illustrating how each classifier's predictions compare to the actual class labels. Such graphs are given for different classifiers:

Under sampling

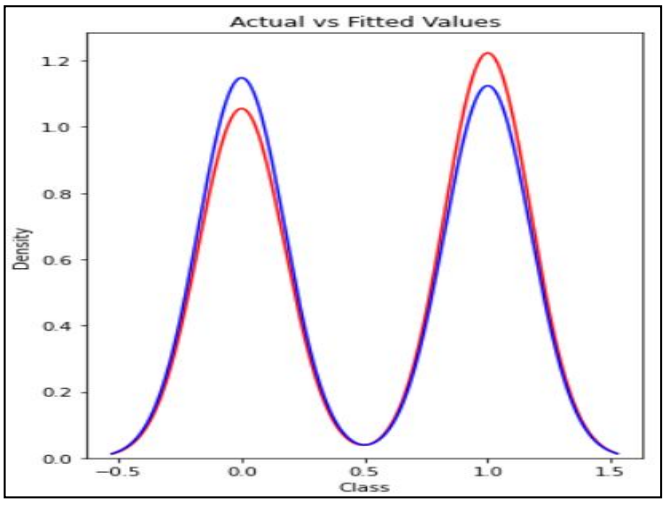

Fig 6 Undersampling Logistic Regression

In Figure.6. the fitted value is represented by blue line and the actual value is represented by the red line . It shows that the LR model doesn't provide a perfect fit as both the actual value and fitted value are not overlapping significantly. The maximum peak value attained is of density 1.2.

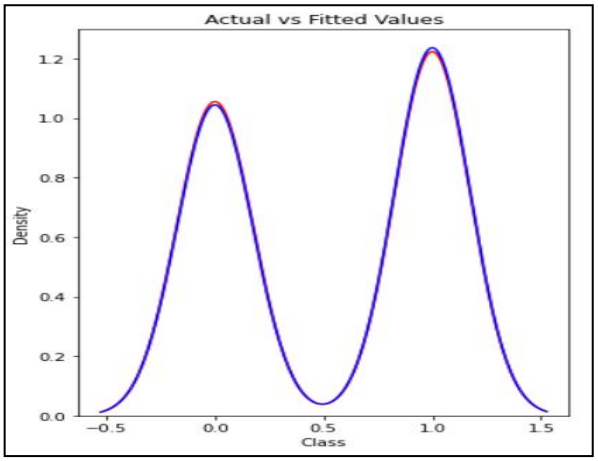

Fig 7 Undersampling Decision Tree

The fact that there is significant overlap between the fitted and real values in Figure.7. indicates that the DT model provides a commendable match. The maximum peak value attained is of density 1.2

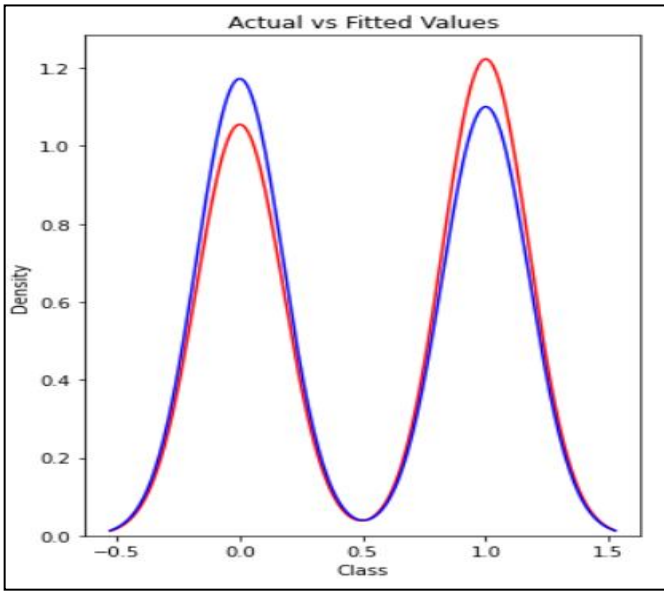

Fig 8 Undersampling Random Tree

The fact that the fitted value and real value do not greatly overlap in Figure.8. indicates that the RF model does not offer a perfect fit. The maximum peak value attained is of density 1.2.

Oversampling

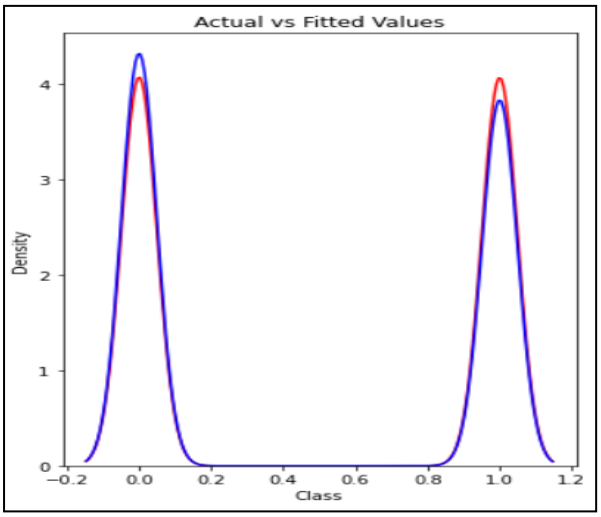

Fig 9 Oversampling Logistic Tree

A unsatisfactory match is offered by the LR model, indicated by the observation that the fitted and real values in Figure.9. do not appreciably overlap. The maximum peak value attained is above density 4

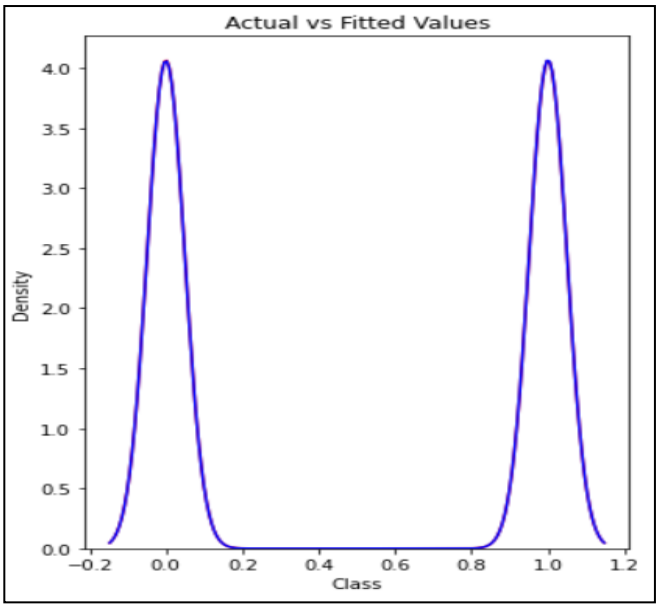

Fig 10 Oversampling Decision Tree

The DT model does not provide an ideal fit, as seen by the fact that the fitted value and real value in Figure.10 and do not substantially overlap. The maximum peak value attained is above density 4.

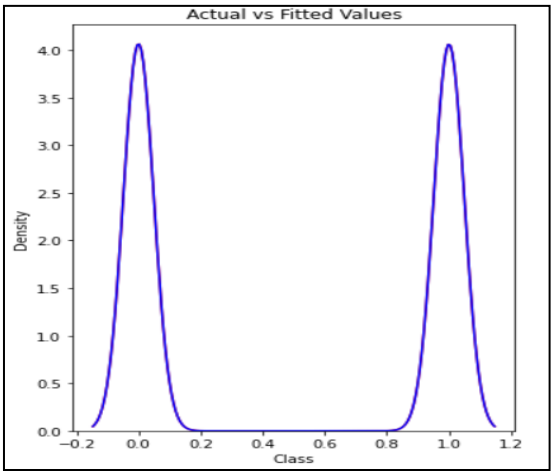

Fig 11 Oversampling Random Tree

The fitted and real values in Figure.11 show a considerable overlap, which suggests that the RF model offers a good match with oversampling. The maximum peak value attained is above density 4.

The following graph shows that the Random Forest classifier achieves the highest accuracy as their predicted values are more close to actual values.

In this section, we also present the results of our experiments in credit card fraud detection using various machine learning algorithms. Our evaluation focused on several key performance metrics, including accuracy, precision, recall and F1-score. We employed a representative dataset of credit card transactions.

The validation set was used to improve the models after they had been trained using the training set. With the use of the evaluation metrics below, we evaluated their performance.

Table.1. Undersampling Result of LR, DT, RF

Model	Accuracy	Recall	Precision	F1 Score
LR	94.73%	91.17%	98.93%	94.89%
DT	87.89%	92.15%	86.23%	89.09%
RF	93.68%	90.19%	97.87%	93.87%

• **Accuracy:** Accuracy is the proportion of instances that are correctly classified. One of the most used classification performance measures is this one. Number of accurately foreseen events = accuracy Predictions made overall or for binary classification models. The accuracy is described as follows:

$$\boldsymbol{Accuracy} = \frac{\boldsymbol{Tp} + \boldsymbol{Tn}}{\boldsymbol{Tp} + \boldsymbol{Tn} + \boldsymbol{Fp} + \boldsymbol{Fn}}$$

Where,

TP (True Positives): Occurrences that were accurately identified as positive (for example, recognizing real fraud situations).

TN (True Negatives): Situations that were accurately identified as negative (for example, recognizing legal transactions).

False positives (FPs) are instances that are mistakenly identified as positive when they are actually negative (for example, falsely identifying a valid transaction as fraudulent).

False Negatives, or FNs, are events that are wrongly foreseen as positive when they are actually negative (for example, failing to recognize a fraudulent transaction).

• **Precision and Recall**

Precision is the proportion of instances that are fraudulently categorized as positive that are genuinely positive instances.

$$\boldsymbol{Precision} = \frac{\boldsymbol{Tp}}{\boldsymbol{Tp} + \boldsymbol{Fp}}$$

• **Recall** is a metric that measures the proportion of accurate positive predictions among all possible positive predictions. Recall gives an indicator of missed positive predictions, unlike precision, which only comments on the accurate positive predictions out of all positive predictions. The number of true positives divided by the sum of true positives and false negatives is used to determine recall.

$$\boldsymbol{Recall} = \frac{\boldsymbol{Tp}}{\boldsymbol{Tp} + \boldsymbol{Fn}}$$

F1 score: The weighted average of Precision and Recall is the F1 score. Therefore, both false

positives and false negatives are considered while calculating this score (Dheepa et al., 2013).

$$\boldsymbol{F1\ Score} = \frac{2*(\boldsymbol{Recall} * \boldsymbol{Precision})}{\boldsymbol{Recall}+\boldsymbol{Precision}}$$

accuracy score:0.99

precision score:0.99

recall score:1.0

F1 score:0.99

Once anomalies are found, it can be notified to authorities. Testing and comparing these algorithms can gauge their accuracy and precision [39].

The Table .1. shows that the highest accuracy, precision and F1 Score attained are 94.73%, 98.93%, 94.89% respectively

using LR model while DT model gives the highest recall score of 92.15% hence it can be affirmed that DT classifiers works efficiently for undersampling.

Table 2. Oversampling Result of LR, DT, RF

Model	Accuracy	Recall	Precision	F1 Score
LR	94.46%	91.48%	97.28%	94.29%
DT	99.80%	99.89%	99.70%	99.80%
RF	99.97%	100%	99.98%	99.99%

The Table .2. shows that the highest accuracy, recall, precision and F1 Score attained are 99.97%, 100%, 99.98% and 99.99% respectively using RF model using oversampling. hence it can be affirmed that RF classifiers works efficiently for oversampling

6. Comparative Study

In this section we compare our proposed method with previous presented scheme in terms of accuracy and find out that our presented method is rich interma of accuracy. Awoyemi et al.[15] The research focused on the European cardholders credit card fraud dataset, addressing the imbalance by employing a hybrid sampling strategy. The study evaluated the performance of K-Nearest Neighbor (KNN), Logistic Regression (LR), and Naive Bayes (NB) machine learning methods. Notably, the investigation did not delve into the exploration of feature selection techniques. Our model optimizes performance by engaging in data pre-processing and sampling techniques, indicating a focus on refining the input data for better model outcomes. [16] focused on credit card fraud detection in highly imbalanced datasets. They compared Decision Trees, SVM, Logistic Regression, and Random Forest, finding that Logistic Regression had good accuracy, but Random Forest was deemed the most accurate for fraud detection. SVM faced challenges due to data imbalance, resulting in suboptimal performance. Our model fine-tunes data through preprocessing, sampling, and utilizes the Random Forest classifier to enhance accuracy in predictions. Alenzi et al. [40] proposed a machine learning model to mitigate credit card fraud detection using logistic regression whereas our model works on random forest classifier which is for efficient for imbalanced dataset.

Bhanusri et al. [41] proposed a model that do not efficiently deals with imbalanced data whereas our model uses oversampling to overcome this problem. Khatri et al.[22] compared machine learning algorithms for credit card fraud detection, including Decision Tree, k-Nearest Neighbor, Logistic Regression, Random Forest, and Naive Bayes. The study focused on a dataset with class imbalance, emphasizing precision. Precision scores were DT (85.11%), KNN (91.11%), LR (87.5%), RF (89.77%), and NB (6.52%). Ileberi et al.[23] utilized Logistic Regression, Decision Tree, Support Vector Machine, and Random Forest for credit card fraud detection using a dataset with a significant class imbalance. The algorithms achieved high accuracy, but the authors recommended the application of sophisticated pre-processing techniques to further improve classifier performance. Our model performs data pre-processing and sampling to enhance the model performance. Seera et al.[23] designed a sophisticated system to identify payment card fraud. Their approach incorporated Genetic Algorithms (GA) for feature selection and aggregation. The researchers applied several Machine Learning algorithms to assess their methodology. The results showed that the combination of Genetic Algorithms with Random Forest (GA-RF) achieved an accuracy of 77.95%, Genetic Algorithms with Artificial Neural Network (GA-ANN) reached 81.82%, and Genetic Algorithms with Decision Tree (GA-DT) attained an accuracy of 81.97%. our model employs data pre-processing, sampling techniques, and utilizes a Random Forest (RF) classifier to achieve improved accuracy in its predictions.

Table 3 Comparison Between Presented and Previous Proposed model

SR.No.	Model	Classifier	Accuracy
1.	Awoyemi et al. [15]	KNN,LR,NB	58.83%
2.	Khare et al., [16]	DT,LR,SVM, RF	98.60%
3.	Alenzi et al.,[40]	LR	97.2%
4.	Bhanusri et al.,[41]	DT, KNN	87.18%
5.	Khatri et al.,[22]	LR,RF,KNN, RF	89.77%
6.	Ileberi et al.,[24]	LR,DT,SVM, RF	98.60%
7.	Seera et al. [23]	GA-RF,GA-DT	77.95%
8.	Our	LR, DT, RF	99.97%

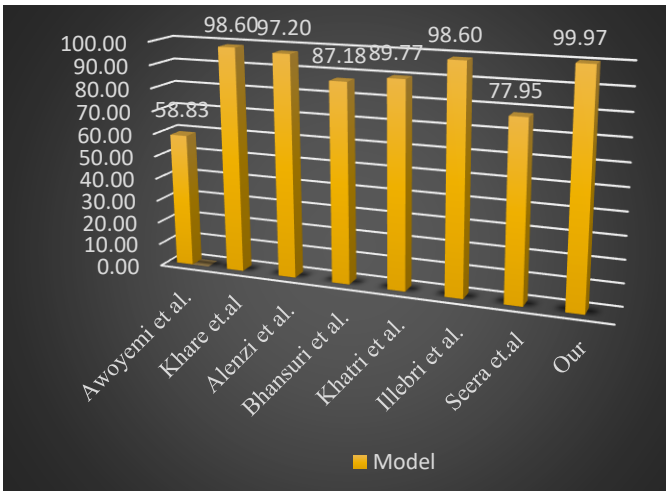

Fig 12 Comparative Chart between our and previous proposed model

Table.3 shows that the accuracy obtained by presented model is higher than the previous presented models. The visual representation of our presented and previous proposed model are depicted in Figure. 12. Which shows that the presented model has more accuracy than the previous scheme analysis.

7.Conclusion

This research highlights the significance of machine learning in enhancing credit card fraud detection systems. We employed three distinct classifiers: LR, DT, and RF. By employing an oversampled RF model, the maximum accuracy, recall, precision, and F1 Score were reached, with respective values of 99.97%, 99.98%, 99.98%, and 99.99%. hence it can be said that RF classifiers are effective when used for oversampling. Additionally, our suggested model's output outperforms previous models that were offered.

Through a comparative analysis of different machine learning algorithms, the study reveals that the choice of algorithm is crucial, with some algorithms outperforming others in specific metrics. The research emphasizes the importance of selecting the right algorithm based on the context and goals of the fraud detection system. It acknowledges the limitations of the study and suggests future research directions, such as using larger datasets and more advanced techniques. Overall, the conclusion underscores the evolving nature of fraud detection and the role of machine learning in making credit card transactions more secure.

Future Scope

The future of credit card fraud detection is bright, adapting to complex fraud tactics and growing digital transactions. Random Forest with Boosting excels but can't identify transaction identities. To advance the project, explore innovative research directions and emphasize ongoing system monitoring and adaptation to combat evolving fraud tactics.

Conflicts of interest

The authors declare no conflicts of interest.

REFERENCES

[1] Y. Kou, C.T. Lu, S. Sinvongwattana, and Y.P. Huang, "Survey of Fraud Detection Techniques," in Proceedings of the IEEE International Conference on Networking, Sensing & Control, Taipei, Taiwan, March 21-23, 2004, vol. 2, pp. 749-754, doi: 10.1109/ICNSC.2004.1297040

[2] Shen, R. Tong, and Y. Deng, "Application of Classification Models on Credit Card Fraud Detection," in International Conference on Service Systems and Service Management, Chengdu, China, pp. 1-4, 2007, doi: 10.1109/ICSSSM.2007.4280163.

[3] M.H. Ozcelik, E. Duman, M. Işik, and T. Çevik, "Improving a credit card fraud detection system using genetic algorithm," in International Conference on Networking and Information Technology, Manila, Philippines, pp. 436-440, 2010, doi: 10.1109/ICNIT.2010.5508478

[4] S.B.E. Raj and A.A. Portia, "Analysis on credit card fraud detection methods," in International Conference on Computer, Communication and Electrical Technology (ICCCET), Tirunelveli, India, 2011, pp. 152-156, doi: 10.1109/ICCCET.2011.5762457.

[5] Phua, V. Lee, K. Smith, and R. Gayler, "A Comprehensive Survey of Data Mining-based Fraud Detection Research," https://www.researchgate.net/publication/236852878.

[6] Abdallah, M.A. Maarof, and A. Zainal, "Fraud Detection System: A survey," Journal of Network and Computer Applications, vol. 68, 2016, doi:10.1016/j.jnca.2016.04.007.

[7] S.M. Lim and C.P. Kumar, "An intelligent payment card fraud detection system," Ann Oper Res, 2021 https://doi.org/10.1007/s10479-021-04149-2

[8] Ali et al., "Financial Fraud Detection Based on Machine Learning: A Systematic Literature Review," Appl. Sci. (MDPI), vol. 12, pp. 9637, 2022, [Online]. Available: https://doi.org/10.3390/app12199637

[9] V. Plakandaras et al., "Credit Card Fraud Detection with Automated Machine Learning Systems," Applied Artificial Intelligence, vol. 36, no. 1, pp. 2086-354, 2022, doi: 10.1080/08839514.2022.2086354

[10] S. Daniel, A. Olumide, and O. Oluwabunmi, "E-payment Challenges: The Genesis and Remedies to the Problem," Journal of Scientific Research and Reports, vol. 28, no. 5, pp. 14-19, 2022, doi:10.9734/jsrr/2022/v28i530518.

[11] Wang, S., Dai, Y., Shen, J.,Xuan, J. "Research on expansion and classification of imbalanced data based on SMOTE algorithm". *Scientific Report Springer Nature* **11**,2021, 24039 https://doi.org/10.1038/s41598-021-03430-5.

[12] Evgeniou, T., Pontil, M."Support Vector Machines: Theory and Applications," *Lecture Notes in Computer Science book series (LNAI,volume 2049)*,2001, Springer-Verlag Berlin Heidelberg.

[13] West, J., Bhattacharya, M. "Intelligent financial fraud detection: A Comprehensive Review," *Computers & Security*,2015, http://dx.doi.org/doi:10.1016/j.cose.2015.09.005

[14] JFu, K., Cheng, D., Tu, Y., Zhang, L. "Credit Card Fraud Detection Using Convolutional Neural Networks." In: Hirose, A., Ozawa, S., Doya, K., Ikeda, K., Lee, M., Liu, D. (eds) Neural Information Processing. ICONIP 2016. Lecture Notes in Computer Science(), vol 9949. Springer Cham. https://doi.org/10.1007/978-3-319-46675-0_53

[15] Awoyemi, J. O., Adetunmbi, A.O., Oluwadare,S. A."Credit card fraud detection using machine learning techniques: A comparative analysis," *2017 International Conference on Computing Networking and Informatics (ICCNI)*, Lagos, Nigeria, 2017, pp. 1-9, doi: 10.1109/ICCNI.2017.8123782

[16] Khare, N., Sait, S.Y. "Credit Card Fraud Detection Using Machine Learning Models and Collating Machine Learning Models." *International Journal of Pure and Applied Mathematics* 118(20). 825-838., 2018 url: http://www.ijpam.eu Special Issue.

[17] Varmedja, D., Mirjana, K., Srdjan, S., Marko, A., Andras, A. "Credit Card Fraud Detection - Machine Learning methods" *18th International Symposium INFOTEH-JAHORIN*A (INFOTEH)1-5. ,2019, 10.1109/INFOTEH.2019.8717766. DOI:10.1109/INFOTEH.2019.8717766..

[18] Naik, H., Kanikar, P."Credit card Fraud Detection based on Machine Learning Algorithms". *International Journal of Computer Applications Foundation of Computer Science. 182(44),2019, DOI:10.5120/ijca201991852*.

[19] Jain, Y., Tiwari, N., Dubey, S., Jain, S. "A comparative analysis of various credit card fraud detection techniques". *International Journal of Recent Technology and Engineering.* 7(5) S2, 7. 402-407. ,2019,Retrieval Number: ES2073017519/19©BEIESP.

[20] Dornadula, V.N., Geetha, S."Credit Card Fraud Detection using Machine Learning Algorithms", *INTERNATIONAL CONFERENCE ON RECENT TRENDS IN ADVANCED COMPUTING*, ICRTAC Procedia Computer ScienceVolume 165, 2019, Pages 631-641 https://doi.org/10.1016/j.procs.2020.01.057

[21] Sahayasakila, V. D., Sikhakolli, A., Yasaswi, V. "Credit Card Fraud Detection System Using Smote Technique and Whale Optimization Algorithm." *International Journal of Engineering and Advanced Technology (IJEAT).* 8(5). ,2019,Retrieval Number D6468048419/19©BEIESP

[22] Khatri, S., Arora, A., Agrawal, A. P. "Supervised Machine Learning Algorithms for Credit Card Fraud Detection: A Comparison," *10th International Conference on Cloud Computing, Data Science & Engineering (Confluence)*, Noida, India, 2020, pp. 680-683, 2020, doi: 10.1109/Confluence47617.2020.9057851

[23] Seera, M., Lim, C.P., Kumar, A."An intelligent payment card fraud detection system." *Ann Oper Res* ,2021, https://doi.org/10.1007/s10479-021-04149-2.

[24] Ileberi, E., Sun, Y., Wang, Z. "A machine learning based credit card fraud detection using the GA algorithm for feature selection. " Journal of Big Data 9, 24,2022, https://doi.org/10.1186/s40537-022-00573-8

[25] Raghavan, P., Gayar, N. E. "Fraud Detection using Machine Learning and Deep Learning," *International Conference on Computational Intelligence and Knowledge Economy (ICCIKE)*, Dubai, United Arab Emirates, pp. 334-339, 2019,doi: 10.1109/ICCIKE47802.2019.9004231

[26] Lim, K.H.,Lee, L.H., Sim, Y.W."A Review of Machine Learning Algorithms for Fraud Detection in Credit Card Transaction," *IJCSNS International Journal of Computer Science and Network Security*, Vol.21 No.9, September 2021. https://doi.org/10.22937/IJCSNS.2021.21.9.4

[27] Seify, M.,Sepehri, M., Hosseinian-far, A., Darvish, A. "Fraud Detection in Supply Chain with Machine Learning," *IFAC-PapersOnLine*, Volume 55, Issue 10, 2022, Pages 406-41, https://doi.org/10.1016/j.ifacol.2022.09.427

[28] Velasco, A.R., Cortes, P., Munuzuri, J., Onieva, L."Prediction of pipe failures in water supply networks using logistic regression and support vector classification," *Reliability Engineering and System Safety*, *2019*,doi: https://doi.org/10.1016/j.ress.2019.106754.

[29] Mehbodniya, A., Alam, I., Pande , S., 2 Neware, R.,Rane, K.P., Shabaz , M.,Madhavan, M.V. "Financial Fraud Detection in Healthcare Using Machine Learning and Deep Learning Techniques,"

Security and Communication Networks. 2021, 1-8. 10.1155/2021/9293877.

[30] Hussain, M.N., Reddy, M.S.C. "Fraud Detection of Credit Card Using Logistic Regression" ,2022,Available at SSRN: https://ssrn.com/abstract=4135514 or http://dx.doi.org/10.2139/ssrn.4135514.

[31] Bhattacharyya, S.,Jha, S., Tharakunnel, K., Westland J. C. "Data mining for credit card fraud: A comparative study," *Decision Support Systems*, Volume 50, Issue 3, February, Pages 602-613

[32] H.H. Patel and P. Prajapati, "Study and Analysis of Decision Tree Based Classification Algorithms," *International Journal of Computer Sciences and Engineering*, vol. 6, no. 10, 2018. DOI: 10.26438/ijcse/v6i10.7478

[33] He Lingjun, Richard A. Levine, Juanjuan Fan, Joshua Beemer, and Jeanne Stronach, "Random Forest as a Predictive Analytics Alternative to Regression in Institutional Research," *Practical Assessment, Research, and Evaluation*, vol. 23, Article 1, 2019. DOI: 10.7275/1wpr-m024.

[34] M. Khalilia, S. Chakraborty, and M. Popescu, "Predicting disease risks from highly imbalanced data using random forest," *BMC Med Inform DecisMak.*, vol. 11, p. 51, 2011. DOI: 10.1186/1472-6947-11-51.

[35] M. Zareapoora and P. Shamsolmoali, "Application of Credit Card Fraud Detection: Based on Bagging Ensemble Classifier," in *International Conference on Intelligent Computing, Communication & Convergence*, Procedia Computer Science, vol. 48, pp. 679–686, 2015. DOI: 10.1016/j.procs.2015.04.201.

[36] E. Strelcenia and S. Prakoonwi, "Improving Classification Performance in Credit Card Fraud Detection by Using New Data Augmentation," *AI 2023*, vol. 4, pp. 172–198, 2023. DOI: 10.3390/ai4010008

[37] Shah and Y. Makwana, "Credit Card Fraud Detection," 2023.

[38] V. Dheepa and R. Dhanapal, "Hybrid Approach for Improvising Credit Card Fraud Detection Based on Collective Animal Behaviour and SVM," in *Security in Computing and Communications*, vol. 377, S.M. Thampi, P.K. Atrey, CI. Fan, and G.M. Perez, Eds. Springer, Berlin, Heidelberg, 2013. DOI: 10.1007/978-3-642-40576-1_29.

[39] S.P. Maniraj, A. Saini, S. Ahmed, and S.D. Sarkar, "Credit Card Fraud Detection using Machine Learning and Data Science," *INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY (IJERT)*, vol. 08, no. 09, September 2019. DOI: 10.17577/IJERTV8IS090031

[40] H.Z. Alenzi and N.O. Aljehane, "Fraud Detection in Credit Cards using Logistic Regression," *International Journal of Advanced Computer Science and Applications*, vol. 11, no. 12, 2020. DOI: 10.14569/ijacsa.2020.0111265

[41] Bhanusri, K.R.S. Valli, P. SaiV.G Jyothi, and R.R.S. Subash, "Credit card fraud detection using Machine learning algorithms," Quest Journals Journal of Research in Humanities and Social Science, vol. 8, no. 2, pp. 04-11, 2020. Available: www.questjournals.org.