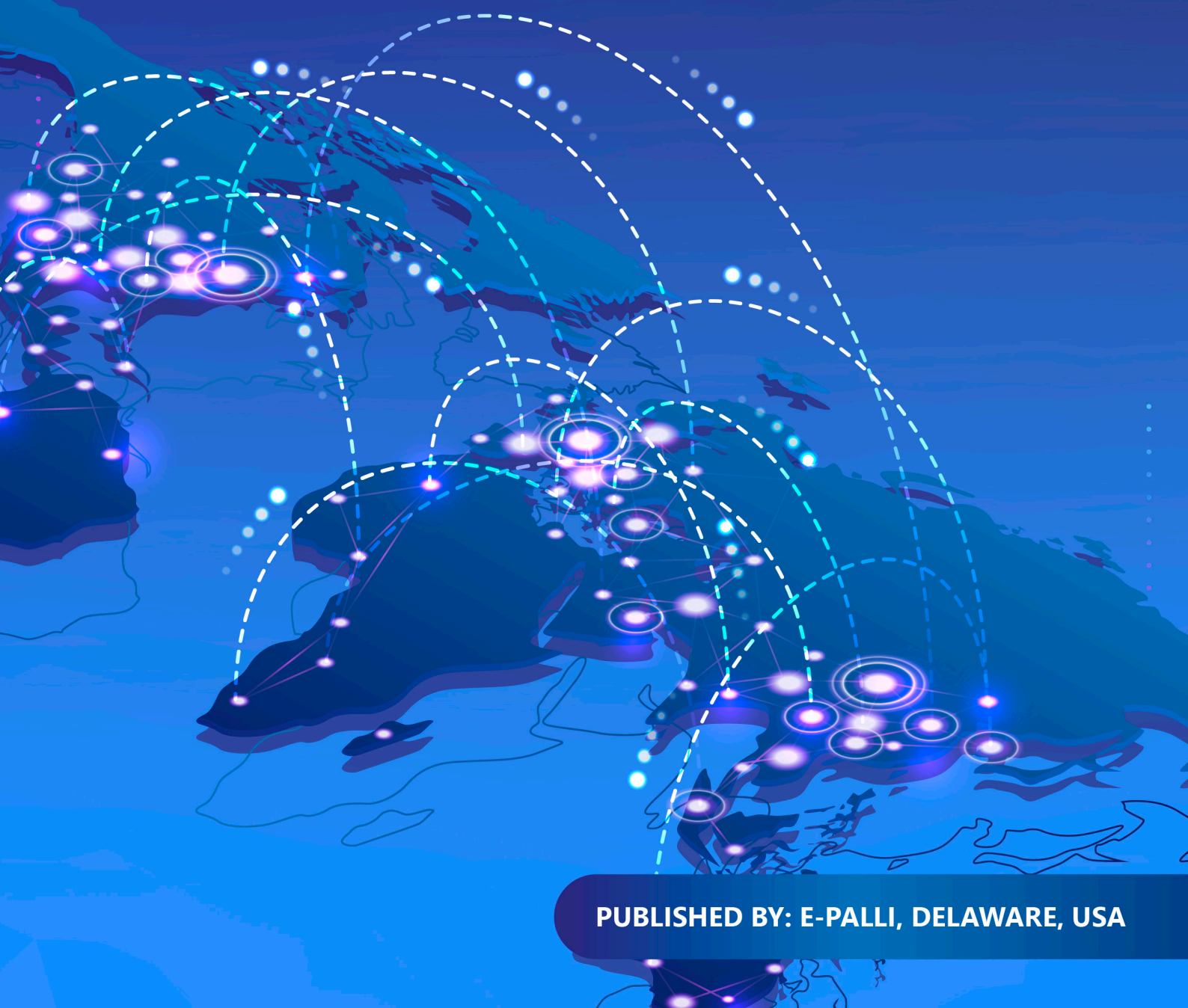
AMERICAN JOURNAL OF MULTIDISCIPLINARY RESEARCH AND INNOVATION (AJMRI)

ISSN: 2158-8155 (ONLINE), 2832-4854 (PRINT)

VOLUME 1 ISSUE 4 (2022)

Indexed in



PUBLISHED BY: E-PALLI, DELAWARE, USA

American Journal of Multidisciplinary Research and Innovation (AJMRI)

Volume 1 Issue 4, Year 2022
ISSN: 2158-8155 (Online), 2832-4854 (Print)
DOI: https://doi.org/10.54536/ajmri.v1i4.633
https://journals.e-palli.com/home/index.php/ajmri

Handling Class Imbalance in Credit Card Fraud using Various Sampling Techniques

Md.Alam Hossain[1*], Mst. Shimu Khatun[1], Rabiul Alam Bhuiyan[1], Md. Taslim[1]

Article Information

Received: September 22, 2022

Accepted: September 30, 2022

Published: October 02, 2022

Keywords

Credit Card Fraud, Class Imbalance, Under- Sampling, Over-Sampling, SMOTE, Adaptive Synthetic

ABSTRACT

Over the last few decades, credit card fraud (CCF) has been a severe problem for both cardholders and card providers. Credit card transactions are fast expanding as internet technology advances, significantly relying on the internet. With advanced technology and increased credit card usage, fraud rates are becoming a problem for the economy. However, the credit card dataset is highly imbalanced and skewed. Many classification techniques are used to classify fraud and non-fraud but in a certain condition, they may not generate the best results. Different types of sampling techniques such as under-over sampling, Synthetic Minority Oversampling, and Adaptive synthetic techniques have been used to overcome the class imbalance problem in the credit card dataset. Then, the sampled datasets are classified using different machine learning techniques like Decision Tree, Random Forest, K-Nearest Neighbors, Logistic Regression, and Naive Bayes. Recall, F1- score, accuracy, precision, and error rate used to evaluate the model performance. The Logistic Regression model achieved the highest result with 99.94% after under sampling techniques and Random Forest model achieved the highest result with 99.964% after over sampling techniques.

INTRODUCTION

The use of credit cards has expanded considerably during the last decade, owing to the advent of e-commerce. The possibility of fraudulent transactions has also increased as a result of this. Credit card fraud occurs in around 0.1 percent of all card transactions, but it can result in significant financial losses because transactions can be relatively substantial. As a result, the overall credit card fraud dataset becomes highly skewed, with only a few examples of one class. In the previous credit card fraud detection, the impact of this extremely imbalanced situation was ignored (Mallidi, *et al.*, 2021), (Ali, *et al.*, 2015). When a credit card was first used to make a transaction in the early 1970s, with a slide machine it was processed manually which imprinted the credit card number on a multi-part receipt. The merchant received the actual document while the clients received the carbon copy. Fraud is described as illegal deception to obtain monetary benefit. Credit card transactions have surged due to the high reliance on internet technologies. Credit card providers are dealing with a severe challenge. In 2004, credit card transactions in the United States resulted in a total 800 million dollars loss cause of fraud. In the same year, the United Kingdom suffered a loss of 425 million pounds due to credit card theft (750 million U.S. dollars). Inner card fraud and exterior card fraud are the two types of credit card fraud. The goal of inner card fraud is to defraud people of their money. It is usually the result of conspiracy between merchants and cardholders. External card fraud is defined as the consumption of stolen cards, phony or counterfeit credit cards, or the use of cards to obtain cash in disguised forms.

Different types of fraud can be follows as:

1. Application Fraud: When a well-known fraudster gains access to an authorized user's application by exploiting sensitive user information such as the user's email address, user name, and password.

2. Card Not Present (CNP): This occurs when the fraudster knows the credit card's expiration date and account number.

3. Lost/Stolen Card: When a credit card is misplaced by the account holder, this happens. Any scammer who wants to make a payment with it can get their hands on it. Making a payment, on the other hand, is difficult since it requires a PIN that is unknown to the fraudsters

4. Electronic Card: The card information on the magnetic stripe is skimmed by this electronic card. It saves the data, and the fraudster can access it at any time.

5. Phishing: It is a type of social media victim in which the customer's personal information, login passwords, and payment card details are stolen.

6. Account Takeover: This is the most typical type of deception. The fraudster will have access to the card- holder's account information as well as all relevant documents.

Various techniques are used to detect fraud such as Machine Learning, Deep Learning, and Data mining approachs. Credit card dataset is highly skewed and imbalanced where Under- sampling, Over-sampling, SMOTE, and AdaSyn techniques are used for getting balanced data. Sampling techniques implementation is very difficult for skewed data. These types dataset don't get easily from the providers of the dataset. Credit card detection has a variety of challenges, including the fact that fraudulent behavior profiles are dynamic, making fraudulent transactions appear to be lawful. Credit card transaction databases are hard to come by and are often skewed (unbalanced). Models with the best feature

[1] Dept.of Computer Science and Engineering Jashore University of Science and Technology Jashore, Bangladesh
[*] Corresponding author's email: alam@just.edu.bd

(variable) selection; suitable metric for assessing approach success on skewed data. Many techniques Support vector machine, Na¨ıve Bayes, K-means clustering, Decision tree,Genetic algorithm, and Logistic regression have been applied to detect credit card fraud. The performance of Artificial neural networks,Convolution neural network, Multilayer Perceptrons, Random forest, and K-nearest neighbor is also evaluated on credit card fraud. The main goal of this study is to find out the best sampling techniques and evaluate the different classifier performances, also abilities to classify fraudulent and non-fraudulent transactions.

The rest of this paper is organized as follows: Section II gives Related works. Section III describes the experimental setup approach including the four sampling techniques and five classifier methods. Section IV describes the performance evaluation. Section V describes the results and comparison. Lastly, section VI displays future work and concludes this paper.

LITERATURE REVIEW

Multiple Machine Learning Techniques (Asha, *et al.*, 2021) like Support Vector Machine (SVM), K-Nearest Neighbor, and Deep learning techniques like Artificial Neural Networks have been used to detect fraud and non-fraud transactions. They have collected a dataset total of 284807 from Kaggle where 492 or 0.1792 percent are fraud cases. The dataset is highly skewed and imbalanced. The proposed system shows that ANN gives better accuracy than SVM and KNN. ANN pro- duces 99.92% accuracy, 81.15% precision, and 76.19% recall. ANN is the best method among them for credit card fraud detection. In paper (Awoyemi, *et al.*, 2017), research showed how to get a better result after using the hybrid sampling techniques in an imbalanced dataset. In this paper, they have used Over- sampling and Under-sampling techniques to get a balanced dataset and get an equal number of fraud & non-fraud trans- actions. They differentiate the results between before using the sampling techniques and after the sampling techniques. K- nearest neighbor is outperformed across the evaluation metrics and got better results for precision and specificity (that is 1.0). The KNN performs better than other classifiers. The NB, KNN, and logistic regression classifiers are obtained 97.92%, 97.69%, and 54.86% accuracy respectively. In paper (Dejan, *et al.*,2019) used 4 different machine learning algorithms in their research on the balanced datasets. SMOTE sampling techniques have been used to balance the dataset.

They have found Random Forest classifier shows better results from other classifiers. The accuracy, precision, and recall are 99.96%, 81.63%, and 96.938% respectively. (Samidha, *et al.*, 2020) have specified different supervised machine learning techniques like Decision tree, Logistic regression, KNN, random forest, and Naive Bayes to differentiate between fraudulent and legitimate transactions. Precision, sensitivity, and time used for the evaluation performance. They have evaluated performance based on the threshold value between 0.5(de- fault value) & 0.4. When the threshold value is 0.4, the proposed system gives the best output. The Decision tree gives better results according to the parameters like precision, sensitivity, and time. (Mohammed, *et al.*,2020), have discussed how Machine learning and artificial neural networks use to identify the potential fraudsters who referred to their previous mistakes and fraudster's details. They also work with the same dataset.

They evaluated their performance based on an imbalanced dataset distributed into the training and testing set (10:90). The research applied Naïve Bayes, Random Forest, Logistic regression, Multiple linear regression, and Neural networks classifiers. To get better results, they have suggested additional techniques like under-sampling, over-sampling, and cost-sensitive loss functions applied on an imbalanced dataset to get the balanced dataset solving the problem. The paper (Deepti, rt al., 2018), used KNN, Logistic regression, Decision tree, and Naive Bayes classifiers on the balanced dataset getting the better result. They used different sampling techniques like oversampling and undersampling to convert an imbalanced dataset into a balanced one. The KNN shows better results from all the algorithms based on evaluation metrics.

The paper (Manoj, *et al.*,2021) have discussed supervised algorithms like LR, KNN, DT and un-supervised learning such as DBSCAN and K-Means. Getting better results, they have also used advanced algorithms like multi-layer perceptron and few of the ensemble techniques like RF, Gradient Boost and XGBoost. This paper showed an imbalanced dataset result and result after balancing. The RF gives satisfactory result among 10 algorithms depends on accuracy, precision, recall and F1-score. This study mainly focus the value of accuracy and F1-score in both cases.

The paper (Dilip, rt al., 2017) applied different oversampling techniques (SMOTE, SMOTE TL, SMOTE ENN, SAFE SMOTE and ROS) to handle an imbalanced problem and also applied ensemble classifier (Adaboost, Bagging) and cost sensitive (C4.5, CSVM) evaluating the performances using recall, G-mean, specificity, and Area under ROC. They observed the SMOTE ENN sampling technique detects the fraud better than other sampling techniques and TL taken better place on the underampling technique. In the paper (Hordri, *et al.*, 2018) they have collected total of 284,807 transactions made in 2013 by European cardholders. Their dataset is highly imbalance containing only 492 fraud transactions. They have used three widely used methods RUS, ROS and SMOTE for sampling.

They have used Naïve Bayes, Linear Regression, Random Forest and Multilayer Perception for classification. And among all four classification techniques Random Forest showed a robust performance in three sampling methods. Random Forest shows much higher accuracy than the Naïve Bayes, Linear Regression and Multilayer Perception for resampling methods. And it was also found that ROS

gave convincing results compared to SMOTE. They also showed that ROS gives much better result than the SMOTE. In the paper (Zhenchuan, *et al.*, 2021), they have proposed a hybrid method to handle the problem of class imbalance with overlap based on a divided and conquer idea. Firstly, they have divided the main dataset into two parts the overlapping subset and non-overlapping subset which is performed by k-Nearest Neighbor and its variations.

After that a model is trained on the minority samples for excluding both a few outliers of minority class and a majority of the majority class from the main dataset. And thus, the overlapping class has a very low imbalance ratio. For achieving better result, they also proposed Dynamic Weighted Entropy (DWE) to determine its quality. They also consider hyper-parameters to achieve greater results. In the paper (Alam, *et al.*, 2020), they mentioned that imbalanced data is crucial to enhance the performance of the model because imbalanced data provides some important insights. It is very difficult to properly train the model on that dataset. So various resampling techniques are used to balance the dataset.

They also used data normalization. But first they started with GBDT model, then they compare the results with traditional machine learning models. The obtained 88.7 percent accuracy by using GBDT method which is much higher than other traditional method on a Taiwan client's credit dataset. They also deployed their method on the web to assist the different stakeholders. In this paper (Luthra, *et al.*, 2019), they have discussed about class imbalance correction techniques such as RUS, ROS, some hybrid methods and some traditional methods. They also minimize the class imbalance issue that affects the performance of the methods. They performed RUS, ROS, SMOTE, Tomak, ENN, SMOTE + Tomak, SMOTE + ENN on Logistic Regression, Decision Trees, Naïve Bayes and Random Forest to check which combination gives best result. Among all the methods the hybrid methods tend to give much higher results.

METHODOLOGY

Dataset

The European cardholders' dataset had collected over two days in September 2013 from the ULB group which can be downloaded from the Kaggle (Awoyemi, *et al.*, 2017), (Varmedja, *et al.*, 2019), (Samidha, *et al.*, 2020), (Azhan, *et al.*, 2020), and (Zareapoor, *et al.*, 2017), Only 0.172 percent of the transactions were fraud cases (positive class), whereas the remainder were negative class. The highly skewed dataset has 27 characteristics, all of which are numeric values transformed using Principal Component Analysis. Two features time and amount are used as" non-class attributes" that don't change in this example, another one is the target class used as a class attribute. The time feature displays the amount of time in seconds that has passed from the initial and the current transaction. The amount feature shows the transaction amount. The label identifies the feature" class" which has two values (0, 1) where 1 chooses fraud and 0 chooses non-fraud cases.

Dataset Division

The credit card dataset is divided into two halves, one for training and the other for testing. The training and testing ratio is 80:20. Then the training set is fed into resampling techniques such as ROS, RUS, SMOTE, and Adasyn to balance the dataset. In this study, a 50:50 ratio is carried out in all the sampling techniques. It refers to the same number of occurrences of fraud and non-fraud transactions.

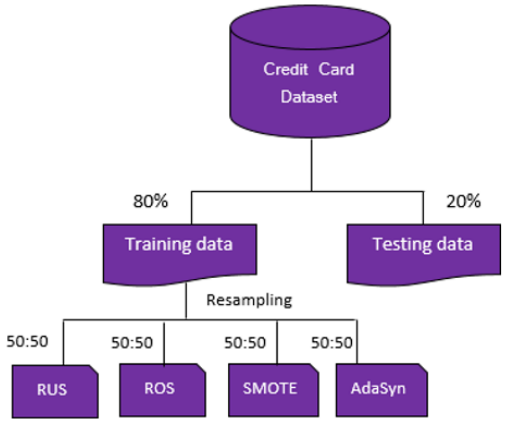


Figure 1: Division of credit card fraud dataset

Preprocessing

On the dataset, several experiments have been conducted. There are a total of 284,807 extremely imbalanced credit card transactions in this dataset where 492 fraudulent transactions. It only contains continuous input variables that are transformed by PCA. As a result, this study employs a total of 30 input features. Time and Amount features have been normalized us- ing a standard scaler. This dataset is highly skewed. To remove the skewness, power transformation has been used. Removing the highly class imbalance problem in the credit card dataset, different types of sampling techniques such as under-sampling, over-sampling, SMOTE, AdaSyn have been used to balance the data. Naive Bayes, k-Nearest Neighbor, Decision tree, Random Forest, and Logistic Regression classifier techniques are employed in the research. The paper has been used hyper- parameter tuning to find the best classifier parameter. Then the classifier model has been trained on preprocessed training data and evaluated the performance of each classifier using real test data.

Sampling Techniques

Random Under-Sampling: The Under-sampling technique balances the dataset by removing random instances from the majority class before performing the classification technique. The Under-Sampling has a simple premise and is much faster than SMOTE. The drawback of this technique is that it can eliminate vital information from the majority class, which in some situations may not be acceptable. Before using under-sampling in the

Figure 2: Distribution of Target Variables After Under-sampling

training dataset have 227,451 non-fraudulent and 394 fraud transactions. After applying RUS techniques, the majority class converted into 394 equally shown in fig 2. Random Over-Sampling: Random over-sampling technique duplicates the random instances of the minority class, using this technique which result get be a reason over-fitting dataset. It is the process of randomly picking and replacing instances from the minority class then added to the training dataset. Replacement uses to select examples from the training dataset at random. It means that minority class examples would be selected firstly, then adding new highly balanced trained dataset more times. After using the ROS technique in the training dataset, the minority fraud class converted into 227,451 equally to non-fraud transactions shown in Fig.3. 3) Synthetic

Figure 3: Distribution of Target Variables After Oversampling

Minority Oversampling Technique (SMOTE): SMOTE is an over-sampling technique that generates random instances selected from the minority class. To generate instances the interpolation method is used between the selected point and its closest instances. Considered every minority class and generated new minority class along line segment is joined with k-nearest neighbors. Based on the required over-sampling percentage, synthetic instances are generated. After applying the SMOTE technique, the distribution of the target variable changed equally shown in Fig.4 4)

Figure 4: Distribution of Target Variables After SMOTE

ADASYN(Adaptive Synthetic): Adaptive Synthetic algorithm generates the synthetic data. It Creates minority data samples based on their distributions and generates more data for balancing the dataset.

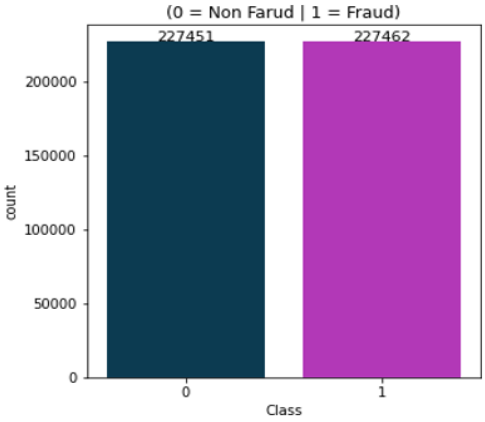


Figure 5: Distribution of Target Variables After AdaSyn

Machine Learning Algorithms

1) Na¨ıve Bayes Classifier: Na¨ıve Bayes classifier is a simpler Bayesian classifier that is built on the Bayes Theorem approach. It works on basis of the conditional independence. It has two attributes ”non-class”, and ”class” independent of each other. Here p(cj | fl) is called the posterior probability when non-class values are known. It will determine the probability belongs to cj ∈ C. P(cj) called the prior probability when fl belongs to cj ∈ C having no prior knowledge about cj . The Bayesian classifier's non-class attribute is uncommitted to the other non-class attributes and considers only one cj non-class value whenp(cj | fl) is calculated.The equations (1) and (2) of the binary classes (fraud and non fraud) taken from paper

$$p(c_j|f_l) = \frac{p(f_l|c_j) * p(c_j)}{p(f_l)} \quad (1)$$

$$p(f_l|c_j) = \prod_{j=1}^{n} p(f_l|c_j) l = 1, \ldots, n; j = 1, 2 \quad (2)$$

Logistic Regression

Logistic regression is a method for developing machine learning models with a dichotomous (binary) dependent variable. Dependent and independent variables describe by using this technique. Nominal, ordinal, or interval variables use as independent variables. Its basis on the probabilistic concept used to solve classification problems. Its similarity with the Linear Regression model and its complex cost function is called 'Sigmoid function' can be denoted by S shape curve converted real value to between 0 and 1. If probability> 0.50 value rounded off to 1, probability <0.5 value rounded off to 0. The sigmoid function (σ) and the input (y) to the sigmoid function are shown in (3) and (4) and these two equations have been taken from paper (Awoyemi, *et al.*, 2017). 3)

$$\sigma(y) = \frac{1}{1+\ell^{-y}} \tag{3}$$

$$y = w_0 z_0 + w_1 z_1 + \cdots + w_n z_n \tag{4}$$

K-Nearest Neighbor

The K-Nearest Neighbor is a supervised classifiers that any data scientist should be familiar with. It mainly used for classification tasks. It uses similarity to classify a new data point. The KNN algorithms choose a number k cause the data point of the closest neighbor will reseparated. In this paper has been set the value of k is 5, which value will search for the 5 closest Neighbors to that data point. Euclidean, Manhattan, and Minkowski distance functions can be used for calculating the input and current data points. The Euclidean and the Manhattan distances are used for continuous variables while the Minkowski distance suited for statistical variables. The KNN classifier have been used the Euclidean distance measure where distances are stored in increasing order and k items with lowest distances are selected.

Random Forest

Random Forest is another type of supervised machine learning technique, used for solving classification and regression problems. It generates multiple decision trees at training time for getting more accurate and consistent predictions based on majority voting. Using hyper-parameter tuning this work found out that Random Forest performs considerably better when 50 trees are trained in parallel. Entropy criteria and Bootstrap aggregation are used for getting more accurate results.

Decision Tree

Decision Tree algorithm is a supervised learning algorithm where the root node splits into sub node continuously. Root node can be identified by calculating the entropy and information gain from the trained dataset and sub node also. Another entity is leaf node which is the final outcome that doesn't separate any sub node. At first, it begins with the root node. On each iteration, it calculates Entropy and Information gain of this attribute. It Selects the smallest Entropy or Largest Information gain. Then the selected attribute is split up the root node for producing a sub node. It continuously recurs on each subset, considered only those attributes which never been selected before. In this paper performed gini criteria and the maximum depth is set to 5. The following formula is used to compute the gini impurity

$$GiniIndex = 1 - \sum_i p_i^2 \tag{5}$$

Where pi is the probability of class i

Performance Evaluation

In this thesis have been proposed five approaches named as Naïve Bayes, k-Nearest Neighbor, Logistic Regression, Random Forest, and Decision Tree. The research has trained the model by using the prepared data set which is made available in Kaggle. Many simulations are run to verify the results, and these methods solve the flaws that prior techniques had. Basis of an accuracy, recall, precision, F1-score, and false positive rate, the NB, KNN, RF, DT, and LR classifiers are evaluated. False positive are actually negative cases but classified as positive cases. The confusion matrix, which describes the difference between the dataset's ground truth and the model prediction, is a traditional approach of evaluating machine learning classifiers (see table I). In a perfect model, all positive examples would be predicted to be positive, and all negative examples would be predicted to be negative. The confusion matrix can be used to compute a variety of metrics to compare the performance of classifiers: Recall

Table 1: Confusion Matrix of Credit Card Dataset

	Predicted as Non-Fraud	Predicted as Fraud
actual non-fraud	True negative	False positive
actual fraud	False negative	True positive

$$Accuracy = \frac{TruePos + TrueNeg}{TruePos + FalsePos + TrueNeg + FalseNeg} \tag{6}$$

$$Sensitivity = \frac{TruePos}{TruePos + FalseNeg} \tag{7}$$

$$Precision = \frac{TruePos}{TruePos + FalsePos} \tag{8}$$

$$F1score = 2 * \frac{Precision * Recall}{Precision + Recall} \tag{9}$$

$$FPR = \frac{FalsePos}{FalsePos + TrueNeg} \tag{10}$$

and Precision give the positive (fraud) cases classification. Accuracy is the Proportion of well classified examples among all the testing examples. Where Precision and recall are diametrically opposed it's called F1-score.

RESULTS AND DISCUSSION

In this study used Five classifier models; RF, DT, LR, KNN, and NB. This paper has been used different types of balancing techniques such as under-sampling, oversampling, Synthetic Minority Oversampling, and

Adaptive synthetic techniques. sampling techniques, where the best classifier is Random Forest. The RF accuracy, recall, and F1-score are 99.95%, 78.57%, and 87.00%. Here the precision score is not enough to predict the fraud and the error rate will be high.

Table II displays the comparison results before using sampling techniques, where the best classifier is Random Forest. The RF accuracy, recall, and F1-score are 99.95%, 78.57%, and 87.00%. Here the precision score is not enough to predict the fraud and the error rate will be high.

Table 2: Before Using Sampling Techniques

Classifiers	**Metrics**			
	Accuracy	**Precision**	**Recall**	**F1-score**
LR	99.85%	82.14%	23.46%	36.50%
DT	99.90%	74.43%	67.34%	70.71%
KNN	99.95%	93.90%	78.57%	85.55%
RF	99.95%	97.46%	78.57%	87.00%
NB	97.75%	05.91%	80.61%	11.01%

Table 3: After Applying Under-Sampling Techniques

Classifiers	**Metrics**				
	Accuracy	**Precision**	**Recall**	**F1-score**	**FPR**
LR	99.940%	84.78%	79.59%	82.10%	0.024%
DT	98.400%	08.30%	82.65%	15.09%	01.57%
KNN	97.740%	06.51%	90.81%	12.15%	02.24%
RF	98.609%	10.20%	90.81%	18.35%	01.37%
NB	96.938%	04.63%	85.71%	08.78%	03.04%

Table III provides the information of different classifiers after using the under-sampling technique. Here, Random Forest and K-Nearest Neighbor performed well. The Recall is equal between KNN and RF respectively and The FPR is and 1.37%. But the LR has a higher precision rate 2.24% and 1.37%. But the LR has a higher precision rate compared to other classification methods.

Table IV provides the information of different classifiers after using the Over-sampling technique. Here, RF generates the best result compared to other classification methods. It has higher accuracy (99.964%), precision rate (98.75%), recall (80.61%), and F1-score (88.76%). The error rate (0.0017%) is too low. During this table comparison with RF, KNN also performed well but KNN takes more time than RF.

Table 4: After Applying Over-Sampling Techniques

Classifiers	**Metrics**				
	Accuracy	**Precision**	**Recall**	**F1-score**	**FPR**
LR	98.546%	9.80%	90.81%	17.69%	1.44%
DT	96.209%	03.76%	85.71%	07.21%	03.77%
KNN	99.915%	70.83%	86.73%	77.98%	0.06%
RF	99.964%	98.75%	80.61%	88.76%	0.001%
NB	97.349%	05.37%	86.73%	10.11%	02.63%

Table 5: After Applying Over-Sampling Techniques

Classifiers	**Metrics**				
	Accuracy	**Precision**	**Recall**	**F1-score**	**FPR**
LR	98.439%	9.18%	90.81%	0.68%	1.54%
DT	96.209%	03.80%	86.73%	07.29%	03.77%
KNN	99.808%	46.99%	87.75%	61.20%	0.17%
RF	99.965%	91.30%	85.71%	88.421%	0.014%
NB	97.394%	05.46%	86.73%	1.027%	02.58%

Table V provides the information of different classifiers after using the SMOTE technique. Here, RF performed well compared to other classifiers. In SMOTE, RF has higher accuracy (99.965%), precision (91.30%), recall (85.71%), F1-score (88.421%) and lower FPR (0.014%).

Table VI displays the Adasyn sampling techniques. Here,

Table 6: After Applying Adaptive Synthetic Techniques

Classifiers	Metrics				
	Accuracy	**Precision**	**Recall**	**F1-score**	**FPR**
LR	92.491%	02.10%	93.87%	04.12%	07.51%
DT	87.802%	01.30%	93.87%	02.57%	12.20%
KNN	99.808%	46.96%	86.73%	60.93%	0.16%
RF	99.952%	89.88%	81.63%	85.56%	0.015%
NB	95.884%	03.67%	90.81%	07.05%	04.10%

RF has higher accuracy (99.952%), precision (89.88%), recall (81.63%), and F1-score (85.56%). It has also lower False positive rate than other four algorithms. In the overall table and bar graph analysis, SMOTE generates higher Accuracy in the RF classifier compared to other techniques shown in Fig.6. In comparison to other sampling techniques, Oversampling creates the highest precision rate in the RF classifier in Fig. 7. The Recall rate of LR and DT is higher than other classifiers in Fig.8 but their precision rate is comparatively low. Not only a higher recall rate is not considered in fraud detection but also a higher precision rate also considered. F1 score measures the harmonic mean of precision and recall rate. In fig.9 SMOTE generates a higher f1 score in the RF classifier. The RF in SMOTE creates the lowest false-positive rate among the four sampling techniques shown in Fig.10. In summary, Random Forest always performed well than other classifiers in the SMOTE sampling technique.

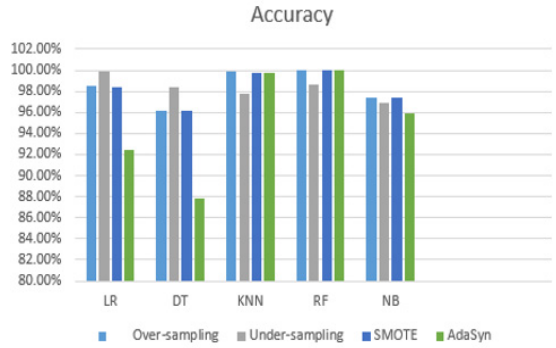


Figure 6: Comparison chart among four sampling techniques according to accuracy

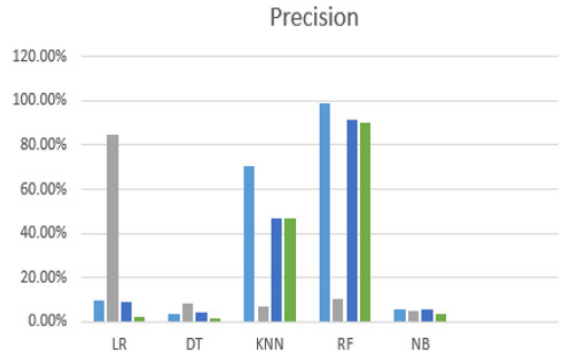


Figure 7: Comparison chart among four sampling techniques according to precision

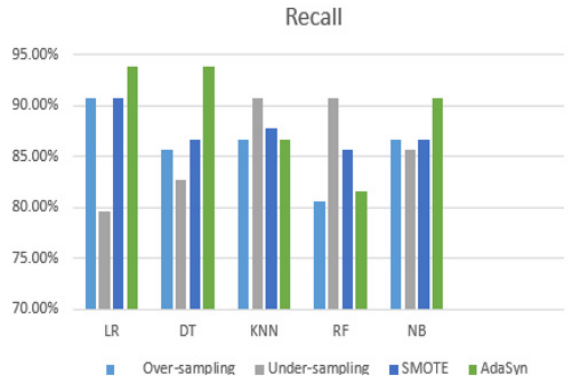


Figure 8: Comparison chart among four sampling techniques according to recall

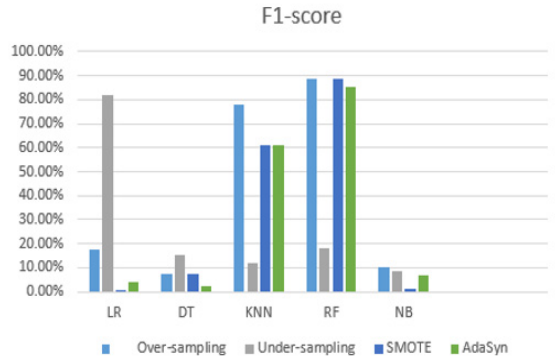


Figure 9: Comparison chart among four sampling techniques according to F1- score

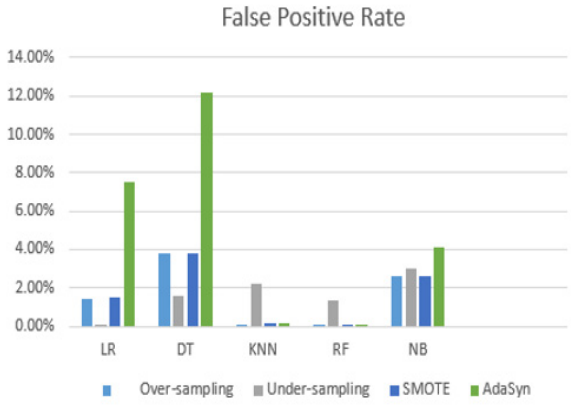


Figure 10: Comparison chart among four sampling techniques according to False positive rate

Table 7: Comparison Between Previous And Proposed Work

Performance	Previous method	Proposed method
Best Technique	SMOTE	SMOTE
Best Classifier	Random Forest	Random Forest
Accuracy	99.96%	99.97%
Recall	81.63%	85.71%
Precision	96.38%	91.30%
F1-score	88.40%	88.42%

Table 7 shows that the comparison between proposed and previous methods (Varmedja, *et al.*, 2019) performance. The main focus of this table is the recall rate. A high recall rate means how accurately a model can detect frauds. In the proposed method, acquired a higher recall rate (85.71%) than the previous method (81.63%). In cost-sensitive issues, a higher recall rate leads to a lower loss of money. From figure-11, see the comparison

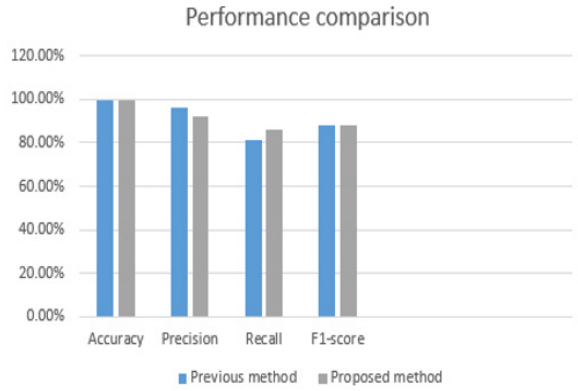


Figure 11: Performance comparison between proposed method and previous method in paper

between proposed method and the previous method. Here, the accuracy, recall, and f1-score of the random forest classifier are higher than the previous method.

CONCLUSIONS

Due to highly imbalanced and skewed data, real-time credit card fraud detection is a difficult task. The main goal of this study is to find out the best sampling techniques and evaluate the different classifier performances and abilities to classify fraudulent and non-fraudulent transactions. It's worth noting that the Random Forest dominated all five classifiers in four sampling techniques. Out of four sampling techniques, SMOTE performed well.In SMOTE, Random Forest generates a higher recall rate as well as a comparatively well precision rate.

To improve results, further study should be done on different SMOTE techniques. Meta-classifier and meta-learning tech- niques can effectively manage such unbalanced data. In this context, several sampling approaches can be investigated. A hybrid technique can be utilized to increase the suggested model's accuracy.

REFERENCES

Asha, R. B., & KR, S. K. (2021). Credit card fraud detection using artificial neural network. *Global Transitions Proceedings, 2*(1), 35-41. https://doi.org/10.1016/j.gltp.2021.01.006

Awoyemi, J. O., Adetunmbi, A. O., & Oluwadare, S. A. (2017, October). Credit card fraud detection using machine learning techniques: A comparative analysis. *In 2017 international conference on computing networking and informatics (ICCNI)*, 1-9.

Ali, A., Shamsuddin, S. M., & Ralescu, A. L. (2013). Classification with class imbalance problem. *Int. J. Advance Soft Compu. Appl, 5(*3).

Azhan, M., & Meraj, S. (2020, December). Credit card fraud detection using machine learning and deep learning techniques. *In 2020 3rd International Conference on Intelligent Sustainable Systems (ICISS),* 514-518.

Alam, T. M., Shaukat, K., Hameed, I. A., Luo, S., Sarwar, M. U., Shabbir, S., ... & Khushi, M. (2020). An investigation of credit card default prediction in the imbalanced datasets. *IEEE Access, 8,* 201173-201198

Bansal, A., & Garg, H. (2021, April). An Efficient Techniques for Fraudulent detection in Credit Card Dataset: A Comprehensive study. *In IOP Conference Series: Materials Science and Engineering, IOP Publishin, 1116*(1), 012181).

Dighe, D., Patil, S., & Kokate, S. (2018, August). Detection of credit card fraud transactions using machine learning algorithms and neural networks: A comparative study. *In 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA),* 1-6.

Ema RR, Akhi Khatun M, Hossain A, Akhond MR, Hossain N, Arafat MY.(2022). Protein Secondary Structure Prediction using Hybrid Recurrent Neural Networks.

Hordri, N. F., Yuhaniz, S. S., Azmi, N. F. M., & Shamsuddin, S. M. (2018). Handling class imbalance in credit card fraud using resampling methods. *Int. J. Adv. Comput. Sci. Appl, 9*(11), 390-396.

Harshit Lamba,(2022). Credit Card Fraud Detection In Real Time.

Hordri, N. F., Yuhaniz, S. S., Azmi, N. F. M., & Shamsuddin, S. M. (2018). Handling class imbalance in credit card fraud using resampling methods. *Int. J. Adv. Comput. Sci. Appl, 9*(11), 390-396.

Itoo, F., & Singh, S. (2021). Comparison and analysis of logistic regression, Naïve Bayes and KNN machine learning algorithms for credit card fraud detection. *International Journal of Information Technology, 13*(4), 1503-1511. https://doi.org/10.1007/s41870-020-00430-y

Islam T, Rizan RU, Tusher YA, Shafiuzzaman M, Hossain MA, Galib S. Nitrogen fertilizer recommendation for paddies through automating the Leaf Color Chart (LCC). *International Journal of Advanced Computer Science and Applications, 11*(8).

Khatri, S., Arora, A., & Agrawal, A. P. (2020, January). Supervised machine learning algorithms for credit card fraud detection: a comparison. In *2020 10th International Conference on Cloud Computing, Data Science & Engineering (Confluence)*, 680-683.

Kundu M, Nashiry MA, Dipongkor AK, Sumi SS, Hossain MA.(2014). An optimized machine learning approach for predicting Parkinson's disease. *Int. J. Mod. Educ. Comput. Sci. (IJMECS), 13*(4), 68-74.

Kaggle.com. (2019). *Credit Card Fraud Detection*. Retrieved 10 Jan. 2019. https://www.kaggle.com/mlg-ulb/creditcardfraud

Luthra, R., Nath, G., & Chellani, R. (2019). A Review on Class Imbalanced Correction Techniques: A Case of Credit Card Default Prediction on A Highly Imbalanced Dataset. *Praxis Business School.*

Li, Z., Huang, M., Liu, G., & Jiang, C. (2021). A hybrid method with dynamic weighted entropy for handling the problem of class imbalance with overlap in credit card fraud detection. *Expert Systems with Applications, 175*, 114750. https://doi.org/10.1016/j.eswa.2021.114750

Mˆınaˇstireanu, E.A., and Mes¸ni¸taˇ, G. (2020). Methods of Handling Unbalanced Datasets in Credit Card Fraud Detection. BRAIN. *Broad Research in Artificial Intelligence and Neuroscience, 11*(1), 131-143. https://doi.org/10.18662/brain/11.1/19

Mohammed, R. A., Wong, K. W., Shiratuddin, M. F., & Wang, X. (2018, August). Scalable machine learning techniques for highly imbalanced credit card fraud detection: a comparative study. I*n Pacific Rim International Conference on Artificial Intelligence* (pp. 237-246). Springer, Cham.

Mallidi, M. K. R., & Zagabathuni, Y. (2021). Analysis of Credit Card Fraud Detection using Machine Learning models on balanced and imbalanced datasets. *International Journal of Emerging Trends in Engineering Research, 9*(7). https://doi.org/10.30534/ijeter/2021/02972021

Najadat, H., Altiti, O., Aqouleh, A. A., & Younes, M. (2020, April). Credit card fraud detection based on machine and deep learning. *In 2020 11th International Conference on Information and Communication Systems (ICICS)*, 204-208.

Ramentol, E., Caballero, Y., Bello, R., & Herrera, F. (2012). SMOTE-RSB*: a hybrid preprocessing approach based on oversampling and undersampling for high imbalanced data-sets using SMOTE and rough sets theory. *Knowledge and information systems, 33*(2), 245-265.

Sisodia, D. S., Reddy, N. K., & Bhandari, S. (2017, September). Performance evaluation of class balancing techniques for credit card fraud detection. *In 2017 IEEE International Conference on Power, Control, Signals and Instrumentation Engineering (ICPCSI)*, 2747-2752.

Sohony, I., Pratap, R., & Nambiar, U. (2018, January). Ensemble learning for credit card fraud detection. In Proceedings of the ACM India Joint *International Conference on Data Science and Management of Data* (pp. 289-294). https://doi.org/10.1145/3152494.3156815

Saheed, Y. K., Hambali, M. A., Arowolo, M. O., & Olasupo, Y. A. (2020, November). Application of GA feature selection on Naive Bayes, random forest and SVM for credit card fraud detection. *In 2020 international conference on decision aid sciences and application (DASA)*, 1091-1097.

Shen, A., Tong, R., & Deng, Y. (2007, June). Application of classification models on credit card fraud detection. *In 2007 International conference on service systems and service management*, 1-4.

Varmedja, D., Karanovic, M., Sladojevic, S., Arsenovic, M., & Anderla, A. (2019, March). Credit card fraud detection-machine learning methods. *In 2019 18th International Symposium*. 1-5.

Vipsita S, Shee BK, Rath SK. An efficient technique for protein classification using feature extraction by artificial neural networks. In *2010 Annual IEEE India Conference (INDICON)*. 1-5.

Zareapoor, M., & Yang, J. (2017). A novel strategy for mining highly imbalanced data in credit card transactions. *Intelligent Automation & Soft Computing,* 1-7. https://doi.org/10.1080/10798587.2017.1321228