Improving predictions on Imbalanced credit card transaction data by using Hybrid Sampling and Ensemble methods

Sai Krishna B[1], Ateeq Ur Rahman[2], Imtiyaz Khan[3]

[1]*PG Scholar,Department of Information Technology, Shadan College of Engineering and Technology, Hyderabad, Telangana, India - 500086*

saikrishnagodfather@gmail.com

[2]*Professor, Department of Computer Science and Engineering, Shadan College of Engineering and Technology,Hyderabad, Telangana, India – 500086. Email:* mail_to_ateeq@yahoo.com

[3]*Professor, Department of Information Technology, Shadan College of Engineering and Technology,Hyderabad, Telangana, India - 500086*

imtiyaz.khan.7@gmail.com

Abstract: credit card fraud is a big problem now that more and more people are buying things online. It costs a lot of money for both individuals and businesses. it is hard to find fraud since the datasets are very unbalanced, with only a small number of illicit transactions compared to genuine ones. fixing this mismatch is very important for making fraud detection systems that work well and give accurate results. To solve this problem, this study looks into advanced hybrid undersampling and oversampling methods that can help find fake transactions while still keeping excellent performance across a range of assessment criteria. To balance the dataset and make it easier to work with, "we used sampling methods including SMOTE, B-SMOTE, ADASYN, SMOTE-Tomek, SMOTE-EEN, and Hybrid BIRCH Borderline SMOTE. The voting Classifier, which combines Boosted decision trees and ExtraTree, always did better than the other models in terms of accuracy, precision, recall, and F1-score across all sampling methods". This shows that it is strong and works well with credit card fraud datasets that are not balanced. The results show that using a mix of hybrid sampling approaches and ensemble learning could make fraud detection systems a lot better.

"Index Terms - *Borderline SMOTE, class imbalance, credit card, fraud detection, sampling techniques, Tomek links".*

1. INTRODUCTION

The fast growth of e-commerce has changed the way people shop by letting them buy things online with credit cards or mobile wallets. credit cards are now the most common way to pay for things online, which has led to a huge increase in the number of transactions that happen every day. Criminals have also found ways to steal credit card information that are more complicated because of this convenience. As a result, credit card theft is now a major problem for businesses, causing them to lose a lot of money

and personal property. Cases of fraud are on the increase hence companies are working on devising new methods of preventing frauds and detecting them, securing the data of the consumers as well as gaining their trust in their services.

In order to create an effective credit card fraud detection model, what you have to do is view transactions in terms of their qualities, features, and values. through the patterns in the past information, these models attempt to determine whether the new transactional information is genuine or not. The inherent issue with regards to the field, however, is that credit card records are highly skewed, with actual purchases vastly exceeding fraudulent ones. This is a great issue to classification models because they may attain a large overall accuracy by predicting the majority classes primarily, and they do not determine fraudulent transactions. Therefore, correcting the problem of class imbalance is a significant element of developing effective fraud detection systems [1].

The inequality of classes in datasets has attracted a lot of attention over the past years because it leaves a significant impact on the manner in which individuals learn and how items can be categorized. when a certain group of classes is extremely tiny like in the situation of fake transaction it becomes challenging to identify special patterns and bizarre behaviors. This problem is aggravated by the fact that the minority magnificence cases are few in numbers and they occur infrequently and that makes them even difficult to classify in the right way [1]. The data balancing strategies have become one significant method to combat this issue.

The three popular methods of striking a data balance are facts-level strategies, algorithm-level techniques, and hybrid approaches [2]. All the methods of data-level include oversampling, undersampling and hybrid sampling. The most common one is oversampling. Oversampling minimizes the influence of the majority group through normalization of the dataset that is going to be classified. This allows models to devote their attention to the minority class [3]. on the other hand, those methods at the level of algorithms attempt to address the vulnerability to class imbalance through the modification of the mechanism of operation of the classification algorithm internally [3].

The latest technology to detect credit card fraud has allowed these techniques to be mixed and used as a hybrid system providing a more accurate and reliable result. Hybrid strategies make the dataset balanced as well as improves the functioning of the algorithm, which is the prospective solution to the persistent issue of locating fraud in non-balanced datasets [3].

2. RELATED WORK

The issue of credit card fraud detection in the existing digital economy is a huge problem because it endangers businesses and individuals, as well. There has been an even greater need to design powerful and strong fraud detection models, as the number of people purchasing items online using credit cards keeps on increasing. There have been some studies using various data sampling strategies, classification algorithms and combinations of the two that have attempted at solving the issue.

Mahesh et al. [5] conducted a comparison of various mechanisms of gathering and categorizing data in order to identify false credit card transactions. They illustrated in their work the great significance of balancing datasets that are not balanced to ensure the models become useful. They demonstrated that balanced datasets attain superior results of classification by experimenting with some sampling methods, such as, “synthetic Minority Oversampling

technique (SMOTE) and undersampling". The guiding things of their experiment is just how essential ensuring that sampling tactics and classification designs are aligned to present the most fruitful outcomes.

A DL classification model proposed by Rtayli [6] is effective to work with non-smoothly distributed datasets. The paper employed high-level designs of neural networks to handle the issues that arise through imbalanced data. To make attributes more valuable in differentiating between classes and possible data augmentation strategies in making the minority class more representative, feature engineering and strategies of data augmentation were deployed as the model. Accuracy and recall was improved by this practice, particularly in the identification of fraudulent transactions. By pairing this method with effective data pretreatment practices, the findings in this research indicate that deep learning practices have an extremely high potential of finding fraud.

Akinwamide [7] observed the ease in which ML techniques can be used to predict when a transaction is false. It analyzed a publicly published credit card fraud detection dataset and "considered a large number of classification algorithms, including the decision trees, Random forest and support Vector Machines". To obtain good detection rates, Akinwamide emphasized the issue of selecting an appropriate approach towards correcting the imbalance in the classes. It was also found in the research that ensemble methods, particularly combined with the analog of sampling strategies, can aid to recognize fraud in a better manner.

Li and Xie [8] have devised an incongruous structure of classifying credit card fraud, which is founded on the classification of behaviours. They clustered transaction behaviors to form balanced subsets of data, which they used to train models of classification. The procedure simplified the detection of fraudulent transactions because it reduced the possibility of having most of the majority party winning. This paper demonstrated the relevance of employing domain-specific clustering approaches to rectify the class imbalance problem in order to ensure the ease of understanding the models of fraud detection.

The neural network ensemble architecture invented by Esenogho et al. [9] together with feature engineering were used to facilitate the ease of detecting credit card fraud. They applied a combination of multiple neural network models to discover various styles in transaction data. feature engineering played a great role in extracting the useful features of raw data and this enhanced the performance of the model. The research indicated that ensemble techniques perform much better than individual models in combination with powerful feature selection algorithms.

Ullastres and Latifi [11] considered the way of utilizing DL algorithms to identify credit card fraud. As their master work, they examined the effectiveness of ensemble methods such as bagging and boosting. Ensemble models were more accurate and resistant than single classifiers because they employed two or more basic classifiers. The study revealed that the ensemble methodology can be employed in the real-life processes of detecting frauds where a vast amount of resources and flexibility is needed.

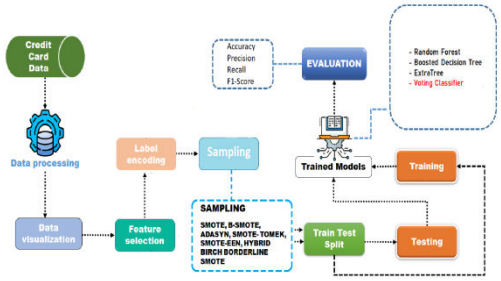
For addressing the issues of imbalanced classification in detection of credit card fraud, Zhu et al. [12] developed "the Noisy-sample-removed Undersampling Scheme (NUS)" before training classifiers, the plan involved the removal of data points that were either noisy or at the onset of

majority class. The approach improved the overall dataset and reduced the possibility of overfitting. The research found that well thought targeted undersampling techniques can cause classification models to perform better.

In attempting to arrive at credit card fraud, Mondal et al. [15] examined various methods of handling skewed data. Their research investigated not only the traditional over-sampling approaches such as SMOTE but also emerging hybrid approaches which were combinations of the undersampling and oversampling. In the study, it was pointed out that the hybrid solutions were the most effective as it employed the best aspects of both approaches. It was also emphasized by the authors that multiple statistics should be used, which would include precision, recall, and F1- score, so as to obtain a complete image of the effectiveness of a model work.

3. MATERIALS AND METHODS

By combining advanced sampling methods with strong ML algorithms, the suggested approach solves the problem of finding credit card fraud in datasets that are very unbalanced. To make the dataset more even and make it easier to find fake transactions, other sampling methods are used, "such as SMOTE, B-SMOTE, ADASYN, SMOTE-Tomek, SMOTE-EEN, and Hybrid BIRCH Borderline SMOTE". these methods work well to reduce the difference between real and fake transactions, which makes the model work better. "We use 4 ML methods to look at the resampled datasets: Random forest, Boosted decision Tree, ExtraTree, and a voting Classifier that combines Boosted decision Tree and ExtraTree". The voting Classifier uses the best parts of each of its parts to make predictions more accurate. This hybrid method combines good sampling techniques with ensemble learning models. Its goal is to offer a solution for real-world credit card fraud detection problems that is scalable, efficient, and reliable. The system is built to do well on all of the important evaluation metrics.



"Fig.1 Proposed Architecture"

This picture shows a ML pipeline that can find credit card fraud. It starts with raw credit card data, which is then preprocessed by being visualized, labeled, and split into training and testing sets. feature selection is used to find the most critical attributes. "After that, the data is put into different ML models, such as ExtraTree, voting Classifier, and Random forest". The training data is used to train these models, and the testing data is used to test them using metrics like "accuracy, precision, recall, and F1-score. also, to deal with data that isn't balanced, oversampling methods like SMOTE, B-SMOTE, ADASYN, SMOTE-TOMEK, SMOTE-EEN, and HYBRID BIRCH BORDERLINE SMOTE are used".

i) "Dataset Collection:"

The PaySim simulator created the dataset used for fraud detection [13]. It creates fake credit card transaction data based on trends seen in the real world. The dataset is based on aggregated financial logs from mobile money services. It simulates normal transaction behaviors while adding fake ones to test fraud detection methods. It has more than six million records and 11 attributes: "step (transaction

time), type (transaction type), amount, nameOrig (origin account), oldbalanceOrg and newbalanceOrig (account balances before and after the transaction), nameDest (destination account), oldbalanceDest and newbalanceDest (destination account balances), isFraud (fraud indicator), and isFlaggedFraud (illegal transaction flag)." The dataset lets us look at fraudulent transactions in great detail, making it a great place to train fraud detection systems.

	step	type	amount	nameOrig	oldbalanceOrg	newbalanceOrig	nameDest
0	1	PAYMENT	9839.64	C1231006815	170136.0	160296.36	M1979787155
1	1	PAYMENT	1864.28	C1666544295	21249.0	19384.72	M2044282225
2	1	TRANSFER	181.00	C1305486145	181.0	0.00	C553264065
3	1	CASH_OUT	181.00	C840083671	181.0	0.00	C38997010
4	1	PAYMENT	11668.14	C2048537720	41554.0	29885.86	M1230701703

"Fig.2 Dataset Collection Table"

ii) "Pre-Processing:"

Data processing, visualization, label encoding, feature selection, and using sample techniques are all important elements in the pre-processing phase. these steps are very important for getting the dataset ready so that the fraud detection model can work better and more accurately.

a) "Data Processing:" the first step in processing data is to import the dataset into a pandas DataFrame, which is a good way to organize data for analysis and manipulation. during this step, columns that aren't needed for fraud detection, like those that aren't needed for the fraud detection procedure, are removed to make the dataset smaller. This helps cut down on noise and makes sure the model pays attention to important features. also, missing or null values can be dealt with by either filling them in or getting rid of them, which keeps the data consistent and ready for further analysis and modeling.

b) "Data Visualization:" In order to understand where the data is distributed and how various components of the data relate with each other, you must visualize it. there are many types of plots that you can create with Seaborn and Matplotlib such as box plots, histograms, and heatmaps. these plots and graphs allow easier visualization of patterns, trends and potential outliers in the data. a heatmap can be used as a graph to reveal how many fake and real transactions there are, a heatmap can be used to illustrate a relationship between diverse aspects. This move assists you in executing wise decisions whenever you are selecting features and formulating a model.

c) "Label Encoding:" Label encoding It involves converting the categorical variables into numbers, which allows the ML algorithms to work on it. This project makes use of the scikit-learn library LabelEncoder class which is applied to transform categorical features, such as transaction type or name, into numeric variables. this is also relevant because most ML models require numeric data to work on because it enables them to perform calculations. There is label encoding that ensures that the categorical variables are treated appropriately and that no info is lost in the process of transformation.

d) "Feature Selection:" An important process in identifying the most important variables to detect fraud is in deciding the right features that are to be adopted. This method involves the examination of the usefulness and importance of some features in predicting fraudulent transactions. it can be used to drop features that are of no use or are too correlated to each other in terms of correlation analysis and univariate selection. This assists the model to be able to generalize to new information.

e) "Sampling:" To fix the class imbalance in the dataset, sampling methods are used. To balance the data, methods like "SMOTE (synthetic Minority Over-sampling technique) and its derivatives, such as B-SMOTE and ADASYN", make fake examples of the minority "class (fraudulent transactions). SMOTE-Tomek and SMOTE-EEN" go this a step further by getting rid of samples that are noisy or on the edge, which makes the model more generalizable. "The Hybrid BIRCH Borderline SMOTE method uses both clustering and over-sampling" to make sure that fake instances are made in the most useful parts of the feature space. these methods make guarantee that the training dataset is more balanced and trustworthy.

iii) Training & Testing:

We use the train_test_split function from the scikit-learn module to divide the data into training and testing sets. "The data set is usually split into a training set (80%) and a testing set (20%). However", the split ratio can be changed to fit the objectives of the project. The training set is what you use to teach the ML models, and the testing set is what you use to see how well the models work on data they haven't seen before. The model can generalize better and we can better judge how well it finds fake transactions if we make sure the split is right. The random_state argument makes guarantee that the results can be repeated.

iv) Algorithms:

"Random Forest" combines several decision trees to make classification more accurate. "It uses a number of sampling methods, including as SMOTE [4], B-SMOTE, ADASYN, SMOTE-Tomek, SMOTE-EEN, and Hybrid BIRCH Borderline SMOTE, to fix class" imbalance and make it easier to find fraud by correctly telling the difference between real and fake transactions.

By integrating "weak learners, **Boosted decision Tree** [10]" makes predictions more accurate. It uses sampling methods "including SMOTE, B-SMOTE, ADASYN, SMOTE-Tomek, SMOTE-EEN, and Hybrid BIRCH Borderline SMOTE to make fake data". This makes the model more sensitive and helps find fraud more easily while lowering the number of false negatives.

"**ExtraTree [14]"** speeds up training and makes models more diverse by constructing several trees with random splits. It effectively fixes class imbalance by using sampling "methods including SMOTE, B-SMOTE, ADASYN, SMOTE-Tomek, SMOTE-EEN, and Hybrid BIRCH Borderline SMOTE". This makes it easier to find fraud in a variety of transaction situations.

The "**voting Classifier"** combines predictions from "the Boosted decision Tree and ExtraTree to make the overall accuracy better". It improves the identification of fake transactions while making sure it is strong against false positives and negatives by using sampling "methods including SMOTE, B-SMOTE, ADASYN, SMOTE-Tomek, SMOTE-EEN, and Hybrid BIRCH Borderline SMOTE."

4. RESULTS & DISCUSSION

"Accuracy:" A test is accurate if it can correctly tell the difference between sick and healthy people. To figure out how accurate a test is, we need find the ratio of true positives to true negatives in all the cases that were tested. this can be said in math as:

$$Accuracy = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{FP} + \mathrm{TN} + \mathrm{FN}} (1)$$

"Precision:" Precision looks at the percentage of accurately labeled instances or samples among those that were labeled as positives. So, the formula for figuring out the precision is:

$$Precision = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}} (2)$$

"Recall:" In ML, recall is a measure of how well a model can find all the relevant examples of a certain class. it is the ratio of accurately predicted positive observations to the total number of real positives. This tells you how well a model captures all occurrences of a certain class.

$$Recall = \frac{\text{TP}}{\text{TP} + \text{FN}} (3)$$

"F1-Score: The F1 score is a way to check how accurate a ML model is". It takes the precision and recall scores of a model and combines them. The accuracy statistic counts how many times a model produced a valid prediction on the whole dataset.

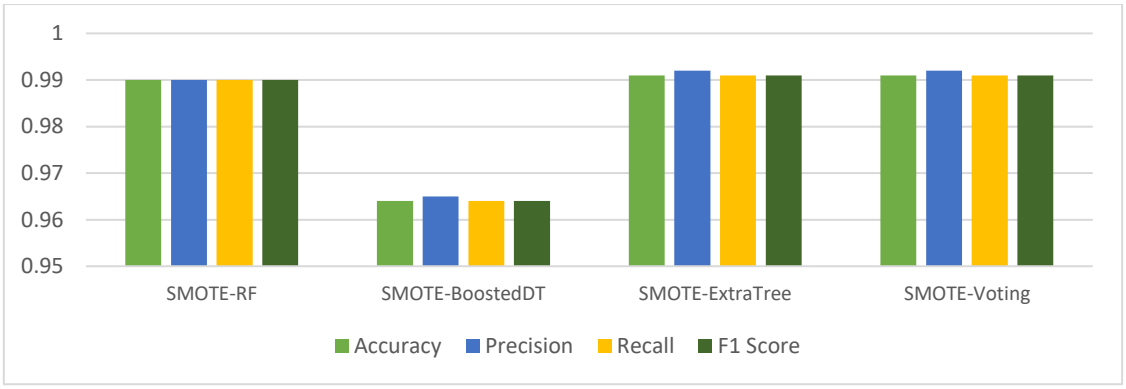
$$F1\ Score = 2 * \frac{Recall\ \text{X}\ Precision}{Recall + Precision} * 100 (4)$$

The "voting Classifier (Boosted DT + ExtraTree) had the best accuracy and performance of all the sampling methods: SMOTE, B-SMOTE, ADASYN, SMOTE-Tomek, SMOTE-EEN, and Hybrid BIRCH Borderline SMOTE". It always did better than competing algorithms on all criteria, "such as accuracy, precision, recall, and the F1 score".

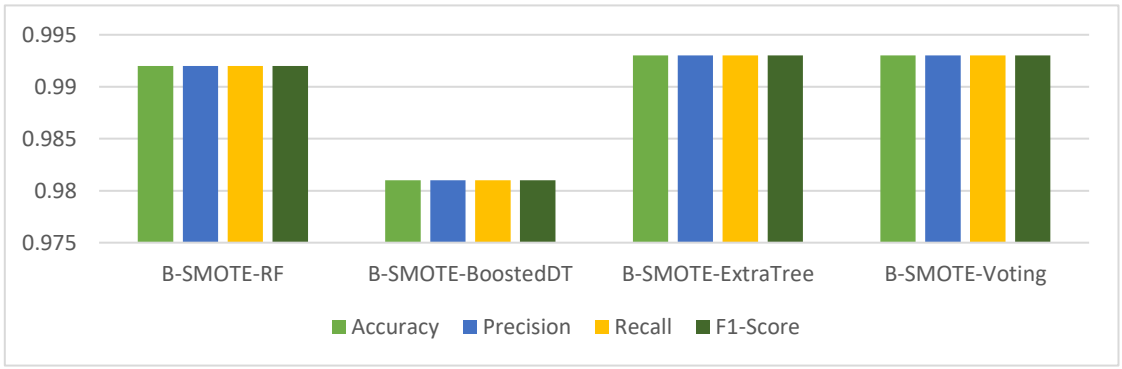
"Table.1 Performance Evaluation Metrics"

ML Model	Accuracy	Precision	Recall	F1 Score
SMOTE-RF	0.990	0.990	0.990	0.990
SMOTE-BoostedDT	0.964	0.965	0.964	0.964
SMOTE-ExtraTree	0.991	0.992	0.991	0.991
SMOTE-Voting	0.991	0.992	0.991	0.991
B-SMOTE-RF	0.992	0.992	0.992	0.992
B-SMOTE-BoostedDT	0.981	0.981	0.981	0.981
B-SMOTE-ExtraTree	0.993	0.993	0.993	0.993
B-SMOTE-Voting	0.993	0.993	0.993	0.993
Adasyn-RF	0.990	0.990	0.990	0.990
Adasyn-BoostedDT	0.956	0.959	0.956	0.956
Adasyn-ExtraTree	0.989	0.990	0.989	0.989
Adasyn-Voting	0.989	0.990	0.989	0.989
SMOTETomek-RF	0.990	0.990	0.990	0.990
SMOTETomek-BoostedDT	0.967	0.968	0.967	0.967
SMOTETomek-ExtraTree	0.992	0.992	0.992	0.992
SMOTETomek-Voting	0.992	0.992	0.992	0.992
SMOTEEEN-RF	0.996	0.996	0.996	0.996
SMOTEEEN-BoostedDT	0.978	0.978	0.978	0.978
SMOTEEEN-ExtraTree	0.998	0.998	0.998	0.998
SMOTEEEN-Voting	0.997	0.997	0.997	0.997
Hybrid-RF	0.993	0.993	0.993	0.993
Hybrid-BoostedDT	0.975	0.976	0.975	0.975
Hybrid-ExtraTree	0.995	0.995	0.995	0.995
Hybrid-Voting	1.000	1.000	1.000	1.000

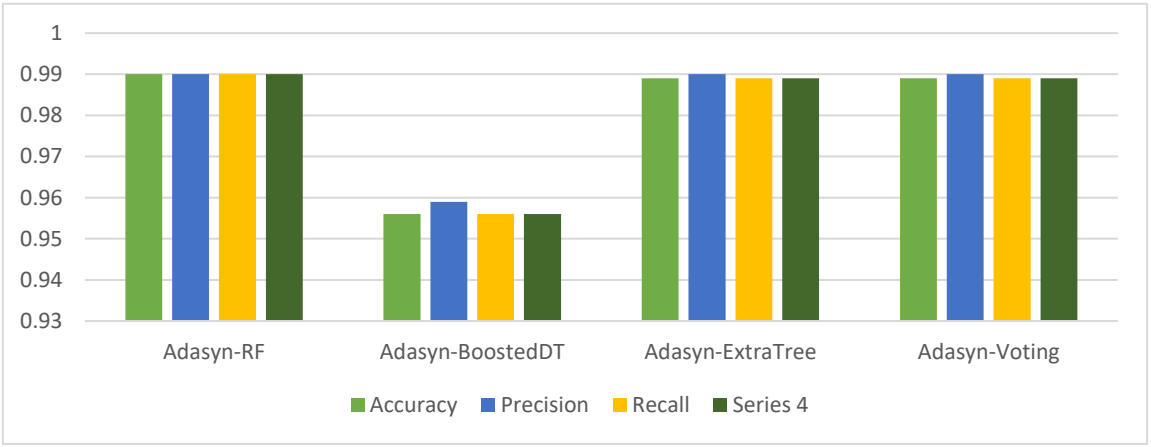
"Graph.1 Comparison Graphs – SMOTE Sampling"



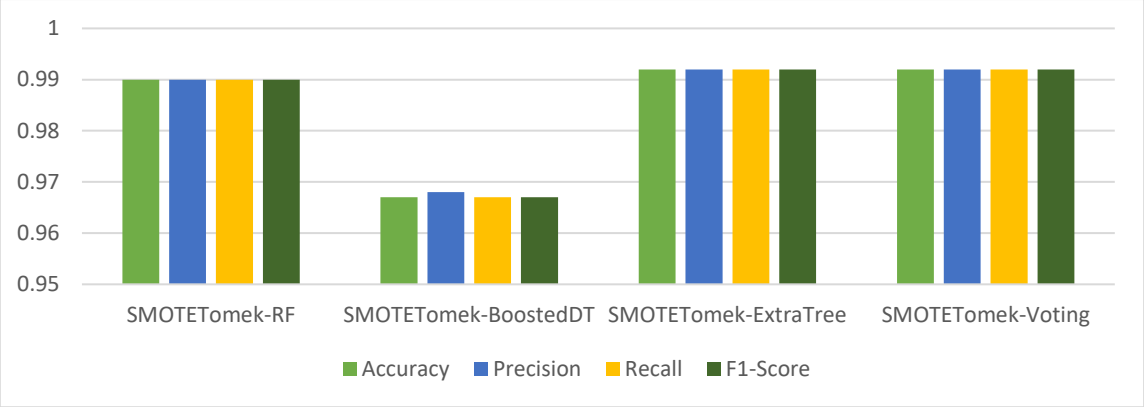
"Graph.2 Comparison Graphs – B-SMOTE Sampling"



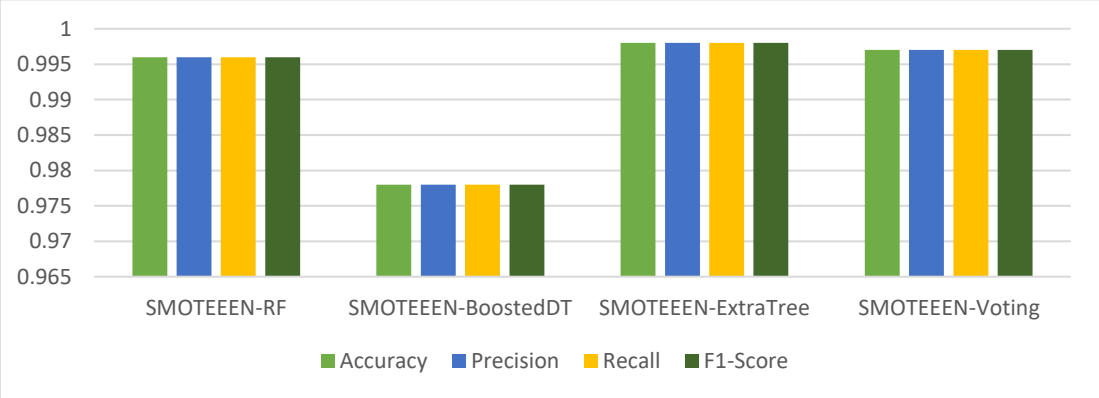
"Graph.3 Comparison Graphs - Adasyn Sampling"



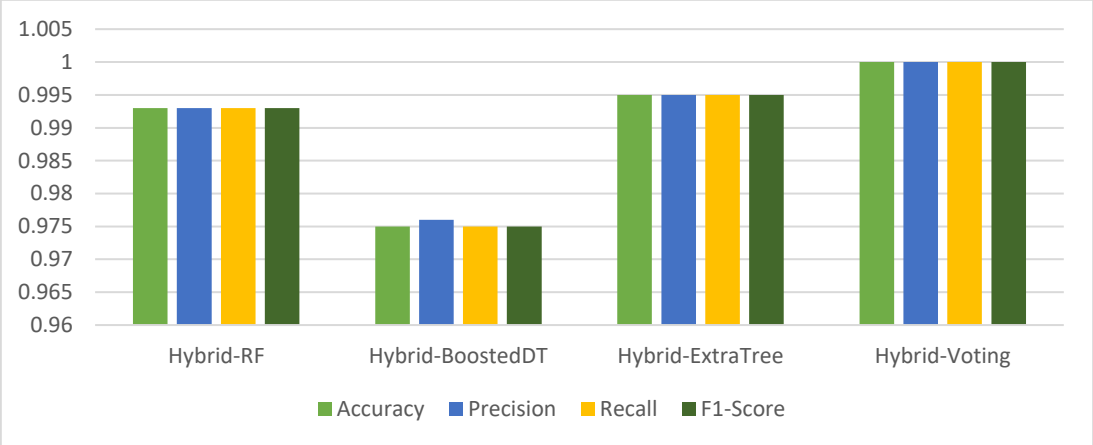
"Graph.4 Comparison Graphs - SMOTE- Tomek Sampling"



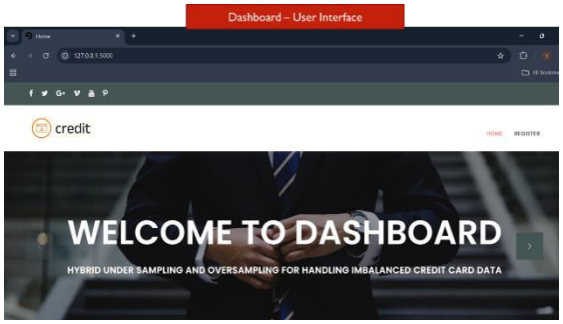
"Graph.5 Comparison Graphs - SMOTE-EEN Sampling"



"Graph.6 Comparison Graphs - Hybrid BIRCH Borderline Smote Sampling"

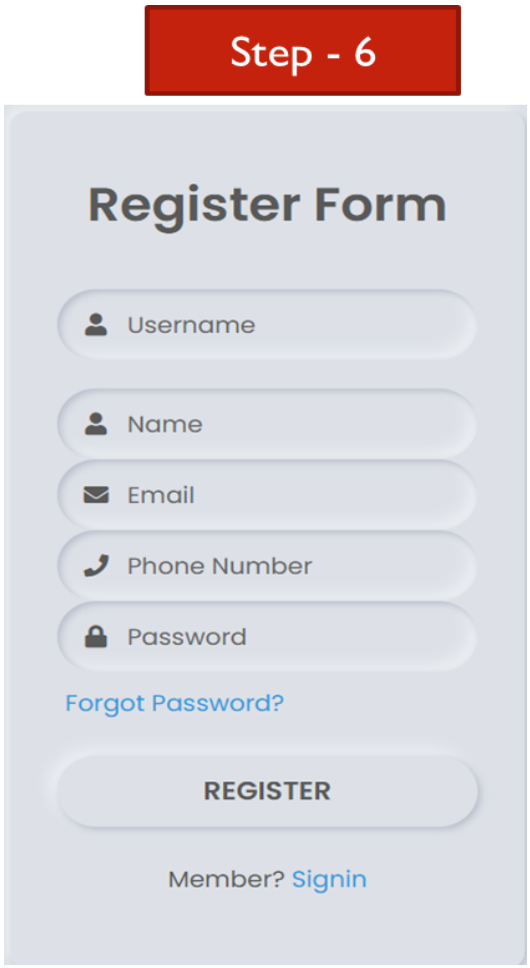


Graphs 1, 2, 3, 4, 5, and 6-show "accuracy in light green, precision in blue, recall in light yellow, and F1-score in green". The voting Classifier does better than the other algorithms on all criteria, with the highest values when compared to the other models. The graph above shows these details in a way that is easy to see.



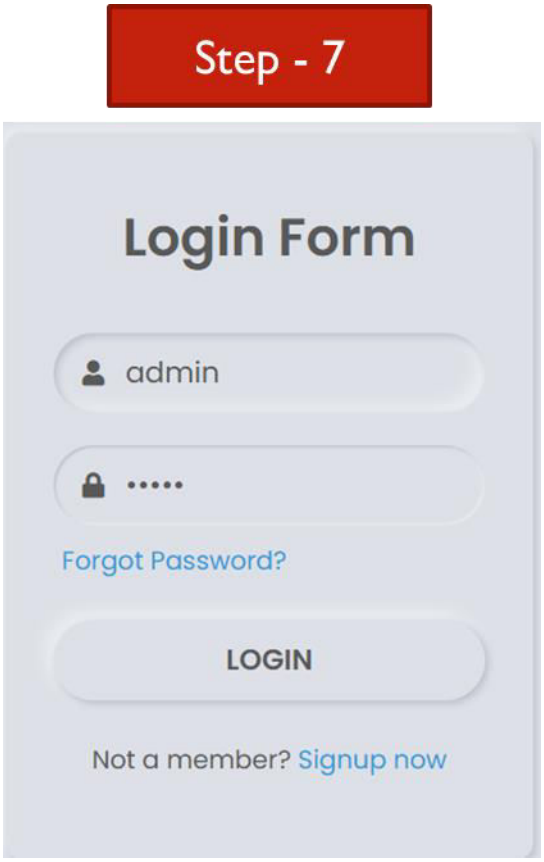
"Fig. 3 Dash Board"

There is a web page called "Dashboard - user Interface" in Fig. 3. It includes a "credit" logo and a welcome message that says, "Hybrid under Sampling and Oversampling for handling Imbalanced credit Card data."



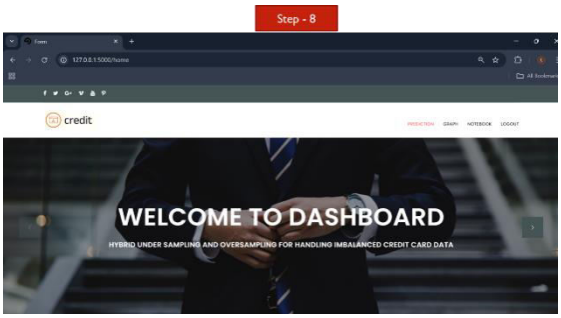
"Fig. 4 Registration page"

The Fig. 4 depicts a form for registering users. You need to give it a username, name, email address, phone number, and password. there is also a "register" button and a link that says "Forgot Password?"



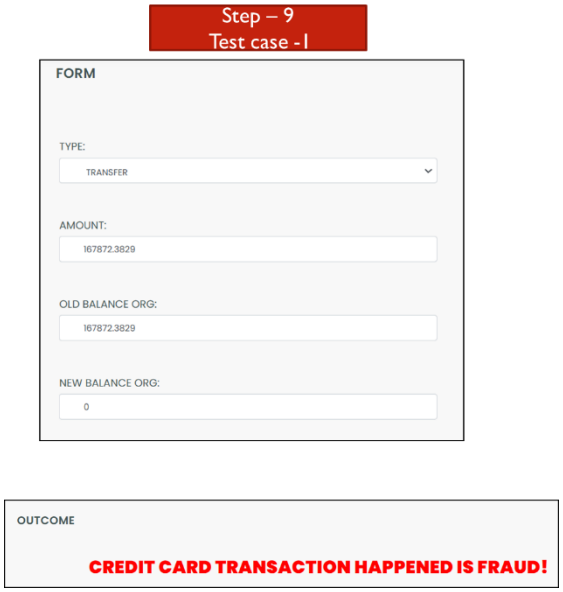
"Fig. 5 Login Page"

The Fig. 5 illustrates a form for logging in. The word "admin" is already in the username field. It also features a "LOGIN" button, a "Forgot Password?" link, and a password field. there is also a "Signup now" button for new users.



"Fig. 6 Home page"

the main page of a web app that deals with credit card data is shown in Fig. 6. The title is "Welcome to Dashboard" and the tagline is "Hybrid under Sampling and Oversampling for handling Imbalanced credit Card data."



"Fig. 7 Test case – 1"

The Fig. 7 provides a way to find fraud in credit card transactions. It gathers information on things like the type of transaction, the amount, and the balances. The form guesses that the transaction is "FRAUD" once you enter the details.

"Fig. 8 Test case – 2"

figure 8 provides a way to find fraud in credit card transactions. It keeps track of things like the type of transaction, the amount, and the balances. after you enter the information, the form says that the transaction is "not FRAUD."

5. CONCLUSION

This study shows that using advanced sampling methods with strong ML models can help find credit card theft in datasets that aren't balanced. "The voting Classifier, which combines Boosted decision Tree and ExtraTree", always performed better than the other algorithms tested, no matter what sampling approach was used. "The voting Classifier got 99.1% correct with SMOTE, which shows that it may manage the dataset's natural imbalance". using B-SMOTE sampling made it much better, "bringing its performance up to 99.3%. ADASYN sampling also kept a high accuracy of 98.9%, and SMOTE-Tomek sampling had an accuracy of 99.2%. The voting Classifier had one of its greatest accuracies at 99.7% with SMOTE-EEN sampling". lastly, the voting Classifier reached a flawless "accuracy of

100% thanks to the Hybrid BIRCH Borderline SMOTE sampling method". these results show that the voting Classifier is a strong and reliable tool for finding fake transactions, making it a great way to find credit card fraud.

the next step in this research is to look at more advanced DL methods, such neural networks, to make fraud detection even more correct. adding the ability to monitor things in real time can also provide notifications right away for transactions that seem suspicious. adding more types of transactions to the dataset and using more advanced sampling methods could potentially make the model work better. also, working with banks and other financial organizations might make it easier to use this system in real life, which would help the fight against credit card fraud continue.

REFERENCES

[1] H. Shamsudin, U. K. Yusof, A. Jayalakshmi, and M. N. A. Khalid, ''Combining oversampling and undersampling techniques for imbalanced classification: A comparative study using credit card fraudulent transaction dataset,'' in Proc. IEEE 16th Int. Conf. Control Autom. (ICCA), Oct. 2020, pp. 803–808, doi: 10.1109/ICCA51439.2020.9264517.

[2] W. W. Soh and R. Yusuf, ''Predicting credit card fraud on a imbalanced data,'' Int. J. Data Sci. Adv. Anal., vol. 1, no. 1, pp. 12–17, Apr. 2019. [Online]. Available: http://ijdsaa.com/index.php/welcome/article/view/3

[3] P. Kaur and A. Gosain, ''Comparing the behavior of oversampling and undersampling approach of class imbalance learning by combining class imbalance problem with noise,'' in Advances in Intelligent Systems and Computing. Singapore: Springer, 2017, pp. 23–30, doi: 10.1007/978-981-10-6602-3_3.

[4] R. Qaddoura and M. M. Biltawi, ''Improving fraud detection in an imbalanced class distribution using different oversampling techniques,'' in Proc. Int. Eng. Conf. Electr., Energy, Artif. Intell. (EICEEAI), Nov. 2022, pp. 1–5, doi: 10.1109/EICEEAI56378.2022.10050500.

[5] K. Praveen Mahesh, S. Ashar Afrouz, and A. Shaju Areeckal, ''Detection of fraudulent credit card transactions: A comparative analysis of data sampling and classification techniques,'' in Proc. J. Phys., Conf., Jan. 2022, vol. 2161, no. 1, Art. no. 012072, doi: 10.1088/1742-6596/2161/1/012072.

[6] N. Rtayli, ''An efficient deep learning classification model for predicting credit card fraud on skewed data,'' J. Inf. Secur. Cybercrimes Res., vol. 5, no. 1, pp. 57–71, Jun. 2022, doi: 10.26735/tlyg7256.

[7] S. O. Akinwamide, ''Prediction of fraudulent or genuine transactions on credit card fraud detection dataset using machine learning techniques,'' Int. J. Res. Appl. Sci. Eng. Technol., vol. 10, no. 6, pp. 5061–5071, Jun. 2022, doi: 10.22214/ijraset.2022.44962.

[8] Q. Li and Y. Xie, ''A behavior-cluster based imbalanced classification method for credit card fraud detection,'' in Proc. 2nd Int. Conf. Data Sci. Inf. Technol. New York, NY, USA: ACM, Jul. 2019, pp. 134–139, doi: 10.1145/3352411.3352433.

[9] E. Esenogho, I. D. Mienye, T. G. Swart, K. Aruleba, and G. Obaido, ''A neural network ensemble with feature engineering for improved credit card fraud detection,'' IEEE Access, vol. 10, pp. 16400–16407, 2022, doi: 10.1109/ACCESS.2022.3148298.

[10] X. Yi, Y. Xu, Q. Hu, S. Krishnamoorthy, W. Li, and Z. Tang, ''ASNSMOTE: A synthetic minority

oversampling method with adaptive qualified synthesizer selection,'' Complex Intell. Syst., vol. 8, no. 3, pp. 2247–2272, Jun. 2022, doi: 10.1007/s40747-021-00638-w.

[11] E. F. Ullastres and M. Latifi, ''Credit card fraud detection using ensemble learning algorithms MSc research project MSc data analytics,'' M.S. thesis, Nat. College Ireland, Dublin, Ireland, May 2022.

[12] H. Zhu, M. Zhou, G. Liu, Y. Xie, S. Liu, and C. Guo, ''NUS: Noisy-sample-removed undersampling scheme for imbalanced classification and application to credit card fraud detection,'' IEEE Trans. Computat. Social Syst., pp. 1–12, Mar. 2023, doi: 10.1109/TCSS.2023.3243925.

[13] E. A. Lopez-Rojas, A. Elmir, and S. Axelsson, ''PaySim: A financial mobile money simulator for fraud detection,'' in Proc. 28th Eur. Modeling Simulation Symp. (EMSS), Sep. 2016, pp. 249–255.

[14] A. A. Arfeen and B. M. A. Khan, ''Empirical analysis of machine learning algorithms on detection of fraudulent electronic fund transfer transactions,'' IETE J. Res., pp. 1–13, Mar. 2022, doi: 10.1080/03772063.2022.2048700.

[15] I. A. Mondal, Md. E. Haque, A.-M. Hassan, and S. Shatabda, ''Handling imbalanced data for credit card fraud detection,'' in Proc. 24th Int. Conf. Comput. Inf. Technol. (ICCIT), Dec. 2021, pp. 1–6, doi: 10.1109/ICCIT54785.2021.9689866.