2021 2nd International Conference on Computing and Data Science (CDS) | 978-1-6654-0428-0/21/$31.00 ©2021 IEEE | DOI: 10.1109/CDS52072.2021.00026

A Comparative Study of Credit Card Fraud Detection Using the Combination of Machine Learning Techniques with Data Imbalance Solution

Faroque Ahmed
Research Associate, Research Wing
Bangladesh Institute of Governance and Management (BIGM)
Dhaka, Bangladesh
faroque.ahmed@bigm.edu.bd

Rittika Shamsuddin
Assistant Professor, Dept. of Computer Science
Oklahoma State University
Stillwater, OK, USA
r.shamsuddin@okstate.edu

***Abstract*—Due to the rapid spread of fraud and cybersecurity risks in digital economy, fraud detection stands as a prime issue of modern technology. However, the analysis of fraud cases is computationally difficult because, fraud cases conjure less than 0.2% of the transactions. Thus to figure out the best classification technique to use for fraud detection, this paper has conducted a thorough experimentation of Machine Learning (ML) techniques. It has implemented six ML techniques i.e. Logistic Regression (LR), Support Vector Machine (SVM), Naïve Bayes (NB), Random forest (RF), Decision Tree (DT), and K-nearest neighbour (KNN) classifiers to detect credit card fraud. The investigation used five type of datasets i.e. imbalanced data, Under Sampled (US) data, Over Sampled (OS) data, sampled data using Synthetic Minority Over Sampling Technique (SMOTE) and Adaptive Synthetic Sampling Method for Imbalanced Data (ADASYN). The best combination of these classification approaches is selected based on five performance evaluation criteria i.e. Accuracy, Area Under the Curve (AUC), Precision, Recall score and f1-score. After evaluation of the classifiers it has showed that among 30 different classification approaches, RF classifier with over sampling (OS) technique was found to be the best approach in terms of all the performance criteria. It showed 99.99 % accurate and precise results with 99.99 % AUC, f1-score and 100 % Recall rate. Our choosen approach has obtained the highest accuracy over other studies on the same dataset. The banking sector as well as other financial institutions might use this suggested machine learning based combination approach to minimize (debit/credit card) frauds.**

Keywords— Credit Card Fraud, Imbalance data solution, Machine Learning, ROC Curve, Classification.

I. Introduction

Fraud is as ancient as humankind itself and can take a boundless variety of diverse forms. Moreover, the improvement of new technologies provides supplementary ways in which criminals may commit fraud. When anyone acts in a false or deciteful way to get some benefit from a process is generally known as a criminal deception or "Fraud". In several ways Credit Card fraud can be conducted. By producing fake or forged cards, by lost or stolen cards, the original site can be cloned, by scanning or by thefting data, by phishing etc. [1]. Fraud hindrance deals in preventing online fraudulent activities. A successful transaction is identified as a fraudulent or genuine by fraud detection [2]. Annually billions of dollars are misplaced by the fraudulent activities of credit cards. According to the 10th annual report of online fraud, although 1.4 % is the amount of lost revenue among online payments during 2006-2008, due to increase in online sales the actual lost revenue percentage has gone up [3]. $4 billion is the estimated amount of loss due to online fraud in 2008 which was 11% higher than the previous year's loss of $3.6 billion [4]. During 2011 losses due to credit card was estimated around $5.07 per $100 which increases to $5.22 during 2012 [5]. 160 companies are investigated by [6] and they showed that offline fraud or physical fraud is 12 times lesser than online fraud. So detection of such fraud cases is important and is a critical need of recent time.

Classification problem is one of the important research themes in the arena of machine learning. In practical application there are a large number of imbalanced dataset. Remarkably, in the context of big data, data streams [7]and multi-class classifications recently the problem of imbalanced data has been painstaking. To reduce the imbalance ratio pre-processing methods are focusing on altering the original distributions. Commonly this is attained by either over- or under sampling. Further classy methods may also indirectly handle advanced data difficulty issues, such as the occurrence of blare or overlapping distributions. Classification algorithms are naturally additions of the prevailing learning approaches, which actually target at increasing the bias towards the minority class. The minority class, which is the abnormal transaction, is more important in case of fraud detection [8]. Handling the imbalanced data in classification problem, minority class sampling is a common technique. Accumulating the amount of smaller class observations is the main purpose of over sampling so that the original classification information can get better holding. Therefore in general, where there is greater demand for the classification accuracy, over sampling algorithm is chosen. On the other hand to balance the majority class with the minority class, under sampling may also be used. When the amount of collected data is sufficient under sampling is used. Cluster centroids and tomek links [9] are commonly used under sampling techniques both of which target potential overlapping features within the collected data sets to decrease the amount of majority class. In SMOTE instead of simply duplicating data from the minority class, it synthesizes new data from the minority class. synthetic data is generated by ADASYN algorithm, and its utmost benefits are not copying the same minority data, and producing more data for "harder to learn" datasets.

In order to intensify the performance of fraud revealing rate and overcome the difficulty of the incorrect alarm rates several methods like Data Mining, Ensemble Learning methods, Bayesian Learning, Fuzzy Darwinian System,

978-1-6654-0428-0/21/$31.00 ©2021 IEEE
DOI 10.1109/CDS52072.2021.00026

Hidden Markow Model, Outlier Detection, Neural Networks, Support Vector Machines, Genetic Algorithm etc. have been used over times. In the detection and prevention of the fraudulent transactions in bank credit card fraud detection system (CCFDS), these approaches can be effectively used. Most of the researchers don't usually focus on the data imbalance even though imbalanced dataset can produce wrong/biased machine learning outcome/prediction. That is why we run thorough experimentation where we combine machine learning techniques with various class instance balancing techniques.

In the detection of fraudulent transactions, this paper implements a comparative study of six ML algorithms along with four data imbalance solution methods to discover the best combination of classifiers and data imbalance solution techniques. Machine learning classifiers namely: Logistic Regression (LR), Support Vector Machine (SVM), Naïve Bayes (NB), Random forest (RF), Decision Tree (DT), and K-nearest neighbour (KNN) classifiers are implemented on five types of datasets i.e. imbalanced data, Under Sampled (US) data, Over Sampled (OS) data, sampled data using (SMOTE) and sampled data using (ADASYN). The best combination of these classification approaches is selected based on five performance evaluation criteria i.e. Accuracy, Area Under the Curve (AUC), Precision, Recall score and f1-score.

Starting with Introduction in section 1, remaining paper progresses in this way, Section 2 narrates the pertinent literature pertaining to the topic, Section 3 details the methodological tactics adopted, the experimental results are shown in Section 4, and lastly Section 5 draws conclusions and provides policy implications/recommendations from this study.

II. Literature review

In case of credit card transactions the spontaneous detection of frauds has been prime focus to wide research. The discovery of new directions from data is typically focused by the literatures. Researchers proposed various methods of fraud detection which are by some means operational in fraud uncovering but due to security issues, accessibility of datasets is the real complicacy. Some popularly used methods are Fuzzy Darwinian System, Covering Algorithm, Data Mining, Meta Classifiers, SVM, Fusion of Dempster Shafer, Bayesian Learning, NN, Hidden Markov Model, Genetic Algorithm, Ensemble Learning, Machine Learning Algorithms etc.

A novel model which uses the Hidden Markov Model (HMM) is proposed by [2]. Random Forest, Support Vector Machine, and Logistic Regression on a real world dataset is compared by [10] using a wide variety of metrics and showed Random forests performed better than others.. The generation of false alarms is minimized while the mixture of clustering schemes with learning introduced. According to [11] in fraud detection outlier detection techniques and k-Nearest Neighbours (KNN) can also be proficient. They are verified suitable in increasing fraud detection rate and reducing false alarms. In experiment of [12] KNN algorithm also implemented well, where with other ordinary algorithms the authors tested and compared it. In the paper of [13] a contrast was made between deep learning methods and some traditional algorithms. Accuracy of approximately 80% is achieved by most of the tested methods. On a European cardholders dataset a mixture of certain classifiers like SVM, GB, RD, LR are used by [14], which returns high recall of over 91%. Simply when matching the dataset by under sampling improved accuracy and recall were achieved. While using European dataset [15] uses altogather following algorithms: LR, XGBoost (XGB), SVM, MLP, DT, GB, KNN,RF, NB and stacking classifier (a mixture of several machine learning classifiers). Due to a detailed data pre-processing, over 90% accuracy was accomplished by all of the algorithms. With accuracy and recall value of 95%, Stacking classifier was the most successful one. On the European dataset a Neural Network was implemented by [16]. Back propagation NN was enhanced with Whale algorithm. the Neural Network includes input (2), hidden (20) and output (2) layers. They attained excellent outcomes on 500 test samples: 97.83% recall and 96.40% accuracy due to optimization algorithm. Neural Network is used by [17], [18] in order to establish enhancement in consequences while ensemble techniques are used. Random Forest algorithm provides the top results according to [19] by means of different metrics, such as accuracy, recall, and precision.

The above literatures confirm that Neural Networks along with machine learning methods are popularly used in the improvement of accuracy, and scalability in this type of research. Very few of the previous studies focusses extensively on imbalance characteristic of dataset and the way out from this problem. In order to put emphasize on the data imbalance problem in classification and to find a way out, this paper uses four popular sampling methods i.e. US, OS, SMOTE and ADASYN in data pre-processing. Six popular classifiers i.e. LR, SVM, NB, RF, DT, and KNN are used to pick the best classification combination for fraud detection of credit cards. On this dataset using Neural Network optimized with Whale algorithm [16] got 96.40% accuracy and 97.83% recall where we get 99% accuracy and almost 100% recall rate by the combination approach of Random Forest (RF) classifier with over sampled (OS) dataset. This is where this paper adds value over the previous literatures using a comparative approach among the combination of four sampling techniques and six machine learning classifiers in the detection of credit card frauds.

III. Methodology

In this article, the ML classifiers namely: LR, SVM, NB, RF, DT, and KNN are implemented with Python. There are several phases of generating and handling the classifiers which comprise; assembly of data, pre-processing of data, training of processes, testing of processes and investigation of classifiers are involved. In the pre-processing phase of data, different sampling techniques were used for getting balanced dataset. The class imbalance solution was addressed by four sample selection methods i.e. US, OS, SMOTE and ADASYN. Primarily all the classifiers were implemented on imbalanced dataset. Then consecutively all classifiers were implemented on those sampled datasets for comparison. In total 5 (Datasets) * 6 (Classifiers) = 30 models were performed to effectively classify fraudulent transactions of credit cards.

A. Dataset

The dataset is collected from the Kaggle [20]. From the Master Card dealings of European cardholders on Sept 2013 the data set was created [21]. The transactions of only 2 days were documented that volumes to 284,807 records. Only 0.172% of the transactions information was the positive category (fraud cases). Applying PCA on the features 28 principal components were constructed i.e. V1, V2,…,V28. The features comprise credit history, earlier months bills, credit boundary, masculinity, status of prevailing account, nuptial status, earlier months payments, wage assignments, savings account, persistence, volume of credit, possessions, employment status, age in months, housing, time, amount, class etc. According to the descriptive measeurments the dataset was very disturbed and was biased in the direction of the negative class. Due to confidentiality issues the background details of the features were out of sight and cannot be exposed. In the dataset information about the seconds passed amongst every exchange and the primary exchange was defined by the time variable. The exchange amount was stored in the 'Amount' feature. The 'class' feature is utilized to characterize the transactions as fraud or non-fraud. It symbolizes the fraud transactions with value 1 and genuine ones with value 0.

B. Machine Learning Classifiers

Logistic Regression (LR) is known as one of the utmost common classification algorithms. When the dependent variable is binary or categorical and predictors are both continuous, and categorical LR can easily describe it. Nature of the dependent variable may be binary. Whether something will happen or not is predicted based on some predictors. The likelihood of going to each class is anticipated for a particular set of regressors.

Support vector machines (SVMs) are found to be very fruitful in a wide range of classification tasks. It works in a high-dimensional feature space. Two important properties they own make them strong — margin optimization and kernel representation. By the use of a kernel function SVMs, mapps to a high-dimensional feature space and learns the classification task in that space without any supplementary computational difficulty are attained.

Based on analogy learning K-nearest neighbour (KNN) is a popular classification algorithm. The classifier examines the pattern space for the K-nearest neighbours when given a new unknown sample, that are nearby to the new sample. It is a supervised method that assigns the class to the new pattern as that of the secret pattern class. In order to define closeness the algorithm uses distance.

Decision Tree (DT) is another widely used method for prediction and classification. A tree encompasses of some interior knobs which represent a test on an characteristic, an outcome of that test is denoted by each branches and separately greenery knob (terminal node) grasps a class marker. A dataset is repeatedly partitioned via any breadth first greedy approach or depth first greedy approach and stoped when all the features have been allocated to a specific class. The best divider will be the one in which the subgroups do not itersect i.e. they are noticeably separate to a maximum extent.

According to Bayes' theorem the Naïve Bayes (NB) classifier is a method with an supposition of individuality among predictors. It assumes that the existence of a specific feature in a class is dissimilar to the existence of any other feature. It is mainly useful for very large data sets and easy to build.

One of the most popular methods is the Random Forest (RF) classifier which consists of a large number of specific decision trees that function togather. A class prediction is predicted by each specific tree in the random forest and the class with the maximum votes elected as the classifiers final prediction. It is superior to a single decision tree since it decreases the over-fitting by averaging the results.

IV. Results and Discussion

Only 0.172% of the transactions information was the positive category (fraud cases). According to the dataset description all the features but for time and amount passed through a PCA conversion (Dimensionality Reduction technique). By seeing the spreading (Fig.1.) we can sense about the skeweness of the dataset.

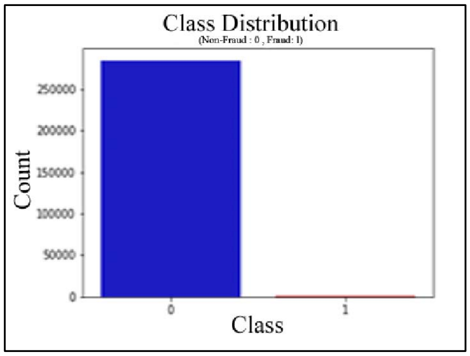

Fig. 1. Class spreading

We train six types of classifier and choose which one is more operational in recognition of fraud transactions. In order to meetup the class imbalance problem we use four methods i.e. US, OS, SMOTE and ADASYN. As the dataset is highly imbalanced, we shouldn't use only the accuracy score as a metric because it will be usually high and misleading, so additionally we focus on AUC, precision, recall score and f1-score in order to assess the enactment of different approches.

A. Data Pre-processing

We don't use the full data for creating the model. Some data is randomly selected and kept aside for checking how good the model is. This is known as "Testing Data" and the remaining data is called "Training Data" on which the model is built. Naturally the original dataset partitioned into training (70%) and testing (30%) dataset.

Where the classes are not characterized equally is generally known as imbalanced data. Classifiers are likely to guess everything as the majority class when someone use this imbalanced dataset. This is the prime difficlty of learning from extremely imbalanced datasets. In order to regulate the class distribution of a data set usually OS and US techniques are used.

During training the models random over sampling can result in overfitting in some cases due to over replication of elements from the smaller class.

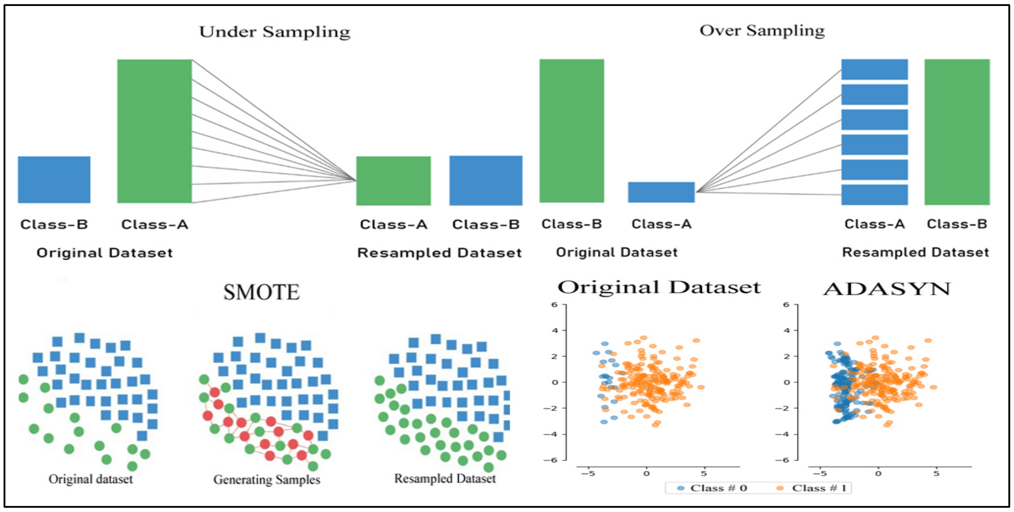

Fig. 2. Sampling Techniques for Imbalance Datase

On the other hand random under sampling erases instances from the bulky class which can result in losing vital information. From Fig. 2. we can see that instead of simply duplicating data from the minority class, SMOTE synthesizes new data from the minority class.

ADASYN (Adaptive Synthetic) is an procedure that also produces synthetic data, but its utmost benefits are not replicating the same minority data, and producing additional data for "harder to learn" datasets. ADASYN and SMOTE are distinguishable in a sense that the prior uses a density distribution, to spontaneously adopt the number of synthetic samples that must be produced for each minority sample by adaptively altering the loads of the different minority samples to compensate for the skewed distributions as a criterion. For each original minority sample the latter produces the equivalent number of synthetic samples.

B. *Model performance*

At first on the imbalanced dataset all the 6 classifiers were used for classification purpose. Secondly, using the under sampled dataset all the classifiers were trained and then validated on the test dataset. Then those classifiers were applied on the whole dataset. Thirdly, using over sampled dataset classification was carried out by all the considered classifiers. Fourthly, using SMOTE dataset all the classifiers were used for classification. At last using ADASYN dataset all classifiers were used for classification. In total 5 (Datasets) * 6 (Classifiers) = 30 models were performed to classify fraudulent transactions of credit cards.

To get a comparative view among those classification approaches following ROC curves (Fig. 3) were obtained. From the ROC curve it is clear that in all the cases except for Naive Bayes (NB), classification performance on imbalanced dataset is poorer than on other four sampled dataset. Moreover classifiers are performing better on over sampled, SMOTE and ADASYN dataset.

For the evaluation of the overall performance of classifiers depending on Accuracy, Recall score, Precision, AUC, and f1-score following performance Table I is generated. The ranking is done among 30 different models based on the aggregate value of all five criteria.

For the evaluation of the overall performance of classifiers depending on Accuracy, Recall score, Precision, AUC, and f1-score following performance Table I is generated. The ranking is done among 30 different models based on the aggregate value of all five criteria.

From Table I we find that Random Forest (RF) with over sampled dataset provides most accurate as well as precise result. All of the scores for Random Forest (RF) with over sampling (OS) technique and the Random Forest (RF) with SMOTE technique are doing well on this dataset. High true positive rate as well as low false-positive rate are obtained according to our desire. On a European cardholders dataset a mixture of certain classifiers like SVM, GB, RD, LR were used by [14], which returns high recall of over 91%. Simply when matching the dataset by under sampling improved accuracy and recall were achieved. While using European dataset [15] uses altogather following algorithms: LR, XGBoost (XGB), SVM, MLP, DT, GB, KNN,RF, NB and stacking classifier (a mixture of several machine learning classifiers). Due to a detailed data pre-processing, over 90% accuracy was accomplished by all of the algorithms. With accuracy and recall value of 95%, Stacking classifier was the most successful one. Random Forest algorithm provides the top results according to [19] by means of different metrics, such as accuracy, recall, and precision. Our comparative result is also identical with the results of these papers. We were able to precisely classify fraudulent transactions of credit cards using a Random Forest model with over sampling technique. We, therefore, choose the Random Forest (RF) classifier with over sampling (OS) technique as the superior combination model, which attained recall score of 100%.

V. Conclusion and Policy Implications

Using dataset of European cardholders of September 2013, the paper compares the performance of six machine learning classifiers namely: LR, SVM, NB, RF, DT, and KNN in prediction of credit card fraud transactions. The paper extensively investigats the solution of the problem for imbalanced dataset by using four sampling techniques. Among 30 different approaches of classification, all classifiers on imbalanced dataset performs poorer than others. NB classifier's performance using all the sampling techniques is unsatisfactory. LR performs better than NB classifier. KNN,

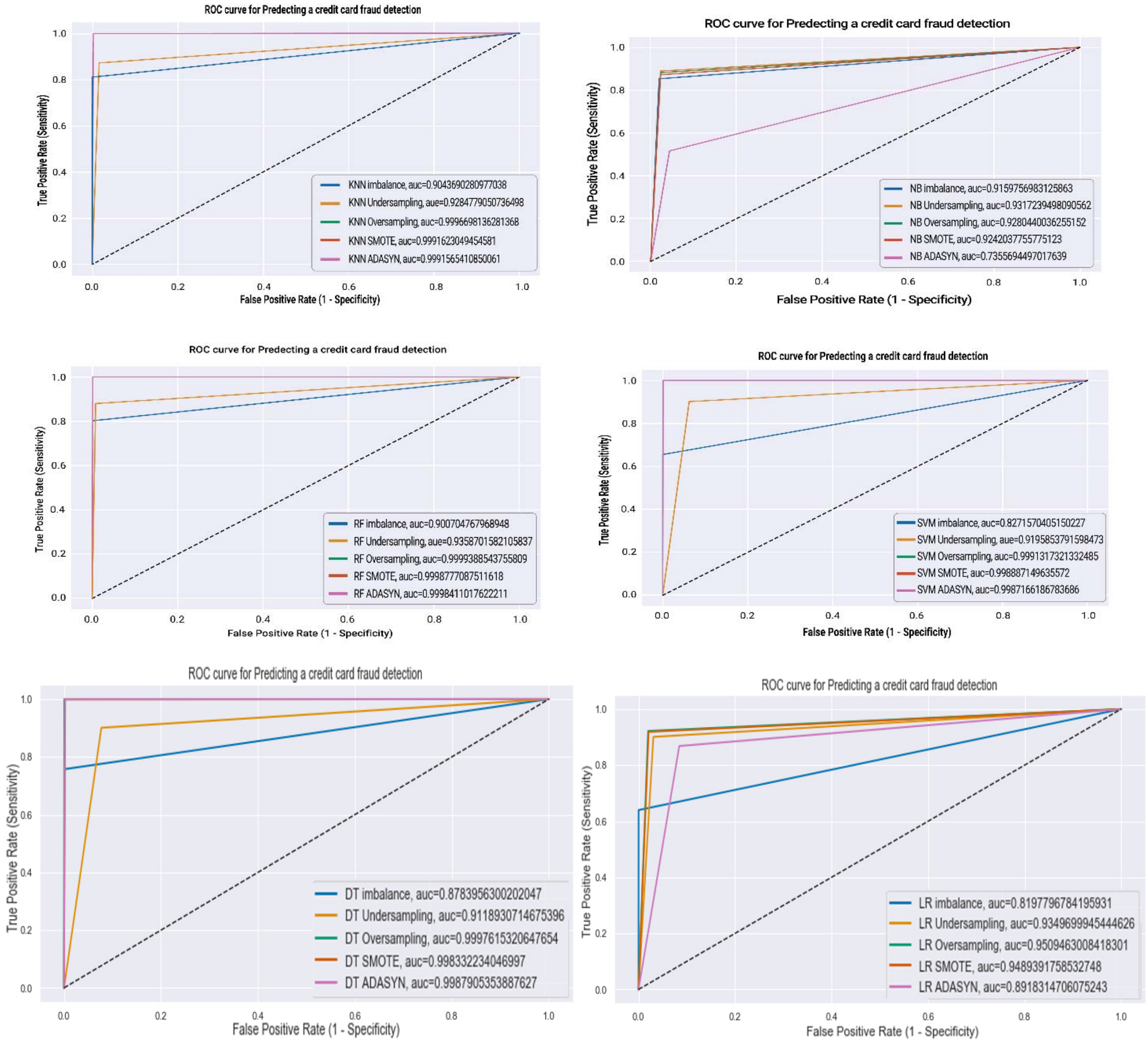

Fig. 3. Performance of different classification approaches by ROC Curve

DT and SVM classifier perform more or less in the same manner based on the employed performance criteria. In the ROC graph (Fig. 3), the AUC scores for Random Forest with over sampling technique is pretty high. If we go forward along the curve, more true positives will be captured with the generation of more false positives. Which is actually more fraudulent transactions are being captured, but also more normal transactions are being flaged as fraudulent. So RF with over sampling technique (OS) is our final model, since it provide highest Recall score of almost 100% on both train and test datasets.

The RF is a much more powerful algorithm than the DT, because it basically is an ensemble of DTs. Where the DT or SVMs overfit, the RF usually performs relatively well, because internally many DTs seeing all a different set of features are casting a vote for the final result.

RF over sampling can be further improved using a recently developed over sampling approach. Reference [22] proposes a fuzzy representativeness difference-based oversampling technique, using affinity propagation and the chromosome theory of inheritance (FRDOAC). The fuzzy representativeness difference (FRD) is adopted as a new imbalance metric, which focuses on the importance of samples rather than the number.

In Bangladesh a continuous growth of the (debit/credit card) transactions over the period of time has recorded. But all over the world as well as in Bangladesh (debit/credit card) fraud has turn into a severe problem. In this circumstance banks and other non banking financial institutions are trying to render sufficient measures of control to evade (debit/credit card) frauds. So the financial institutions as well as banking sector all over the world as well as Bangladesh should use this proposed combination of machine learning based technology to minimize (debit/credit card) frauds.

TABLE I. COMPARATIVE PERFORMANCE OF DIFFERENT CLASSIFICATION APPROACHES

Ranking of Models	**Models Combinations**	**Accuracy**	**AUC**	**Precision Score**	**Recall Score**	**f1-Score**
1	**RF Over sampling**	99.993%	99.993%	99.986%	100%	99.993%
2	**RF SMOTE**	99.986%	99.986%	99.973%	100%	99.986%
3	**RF ADASYN**	99.986%	99.985%	99.973%	99.998%	99.986%
4	**DT Over sampling**	99.977%	99.977%	99.954%	100%	99.977%
5	**KNN Oversampling**	99.967%	99.967%	99.934%	100%	99.967%
6	**KNN SMOTE**	99.916%	99.916%	99.833%	100%	99.916%
7	**KNN ADASYN**	99.915%	99.915%	99.832%	100%	99.915%
8	**SVM Oversampling**	99.913%	99.913%	99.827%	100%	99.913%
9	**SVM SMOTE**	99.888%	99.888%	99.778%	100%	99.889%
10	**DT ADASYN**	99.879%	99.879%	99.819%	99.939%	99.879%
11	**SVM ADASYN**	99.871%	99.871%	99.756%	99.987%	99.872%
12	**DT SMOTE**	99.833%	99.833%	99.766%	99.901%	99.833%
13	**LR Over sampling**	95.091%	95.094%	97.897%	92.175%	94.951%
14	**LR SMOTE**	94.889%	94.893%	97.783%	91.876%	94.738%
15	**RF Under sampling**	93.357%	93.587%	99.201%	87.943%	93.233%
16	**LR Under sampling**	93.357%	93.497%	96.946%	90.071%	93.382%
17	**NB Under sampling**	92.988%	93.172%	97.656%	88.652%	92.936%
18	**KNN Under sampling**	92.619%	92.847%	98.401%	87.234%	92.481%
19	**NB Over sampling**	92.797%	92.804%	97.374%	87.988%	92.443%
20	**NB SMOTE**	92.412%	92.421%	97.366%	87.207%	92.007%
21	**SVM Under sampling**	91.881%	91.958%	94.074%	90.071%	92.029%
22	**NB imbalance**	97.880%	91.597%	06.321%	85.294%	11.771%
23	**DT Under sampling**	90.405%	90.391%	90.781%	90.781%	90.781%
24	**KNN imbalance**	99.959%	90.436%	94.017%	80.882%	86.956%
25	**RF imbalance**	99.957%	90.068%	93.162%	80.147%	86.166%
26	**LR ADASYN**	89.181%	89.183%	91.157%	86.807%	88.928%
27	**DT imbalance**	99.917%	88.581%	73.943%	77.205%	75.539%
28	**SVM imbalance**	99.933%	82.715%	91.752%	65.441%	76.394%
29	**LR imbalance**	99.925%	81.978%	87.878%	63.971%	74.042%
30	**NB ADASYN**	73.531%	73.556%	92.066%	51.568%	66.108%

REFERENCES

[1] Y. Jain, N. Tiwari, S. Dubey, and S. Jain, "A comparative analysis of various credit card fraud detection techniques," *Int. J. Recent Technol. Eng.*, vol. 7, no. 5, pp. 402–407, 2019.

[2] A. Srivastava, A. Kundu, S. Sural, and A. Majumdar, "Credit card fraud detection using hidden Markov model," *IEEE Trans. dependable Secur. Comput.*, vol. 5, no. 1, pp. 37–48, 2008.

[3] CyberSource, "Online fraud report: online payment, fraud trends, merchant practices, and bench marks," 2009. [Online]. Available: http://www.cybersource.com.

[4] H. Leggatt, "CyberSource: 2008 online fraud to reach $4 billion," 2008.[Online].Available:http://www.bizreport.com/2008/12/cybersource_2008_online_fraud_to_reach_4_billion.html.

[5] A. Rolfe, "Card Fraud Report 2015," 2015. [Online]. Available: https://www.paymentscardsandmobile.com/wp-content/uploads/2015/03/PCM_Alaric_Fraud-Report_2015-1.pdf.

[6] T. K. Behera and S. Panigrahi, "Credit Card Fraud Detection: A Hybrid Approach Using Fuzzy Clustering & Neural Network," *Proc. - 2015 2nd IEEE Int. Conf. Adv. Comput. Commun. Eng. ICACCE 2015*, pp. 494–499, 2015, doi: 10.1109/ICACCE.2015.33.

[7] T. R. Hoens, R. Polikar, and N. V Chawla, "Learning from streaming data with concept drift and imbalance: an overview," *Prog. Artif. Intell.*, vol. 1, no. 1, pp. 89–101, 2012.

[8] Y. Sahin, S. Bulkan, and E. Duman, "A cost-sensitive decision tree approach for fraud detection," *Expert Syst. Appl.*, vol. 40, no. 15, pp. 5916–5923, 2013.

[9] M. Zeng, B. Zou, F. Wei, X. Liu, and L. Wang, "Effective prediction of three common diseases by combining SMOTE with Tomek links technique for imbalanced medical data," in *2016 IEEE International Conference of Online Analysis and Computing Science (ICOACS)*, May 2016, pp. 225–228, doi: 10.1109/ICOACS.2016.7563084.

[10] S. Bhattacharyya, S. Jha, K. Tharakunnel, and J. C. Westland, "Data mining for credit card fraud: A comparative study," *Decis. Support Syst.*, vol. 50, no. 3, pp. 602–613, 2011, doi: 10.1016/j.dss.2010.08.008.

[11] N. Malini and M. Pushpa, "Analysis on credit card fraud identification techniques based on KNN and outlier detection," in *2017 Third International Conference on Advances in Electrical, Electronics, Information, Communication and Bio-Informatics (AEEICB)*, 2017, pp. 255–258.

[12] J. O. Awoyemi, A. O. Adetunmbi, and S. A. Oluwadare, "Credit card fraud detection using machine learning techniques: A comparative analysis," in *2017 International Conference on Computing Networking and Informatics (ICCNI)*, 2017, pp. 1–9.

[13] Z. Kazemi and H. Zarrabi, "Using deep networks for fraud detection in the credit card transactions," in *2017 IEEE 4th International Conference on Knowledge-Based Engineering and Innovation (KBEI)*, 2017, pp. 630–633.

[14] A. Mishra and C. Ghorpade, "Credit Card Fraud Detection on the Skewed Data Using Various Classification and Ensemble Techniques," in *2018 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS)*, 2018, pp. 1–5.

[15] S. Dhankhad, E. Mohammed, and B. Far, "Supervised machine learning algorithms for credit card fraudulent transaction detection: a comparative study," in *2018 IEEE International Conference on Information Reuse and Integration (IRI)*, 2018, pp. 122–125.

[16] C. Wang, Y. Wang, Z. Ye, L. Yan, W. Cai, and S. Pan, "Credit card fraud detection based on whale algorithm optimized BP neural network," in *2018 13th International Conference on Computer Science & Education (ICCSE)*, 2018, pp. 1–4.

[17] N. Kalaiselvi, S. Rajalakshmi, J. Padmavathi, and J. B. Karthiga, "Credit card fraud detection using learning to rank approach," in *2018 International Conference on Computation of Power, Energy, Information and Communication (ICCPEIC)*, 2018, pp. 191–196.

[18] F. Ghobadi and M. Rohani, "Cost sensitive modeling of credit card fraud using neural network strategy," in *2016 2nd International Conference of Signal Processing and Intelligent Systems (ICSPIS)*, 2016, pp. 1–5.

[19] V. N. Dornadula and S. Geetha, "Credit Card Fraud Detection using Machine Learning Algorithms," *Procedia Comput. Sci.*, vol. 165, no. March, pp. 631–641, 2019, doi: 10.1016/j.procs.2020.01.057.

[20] Kaggle.com., "Credit Card Fraud Detection.," 2020. https://www.kaggle.com/mlg-ulb/creditcardfraud (accessed May 04, 2020).

[21] A. D. Pozzolo, O. Caelen, R. A. Johnson, and G. Bontempi, "Calibrating Probability with Undersampling for Unbalanced

Classification," in *2015 IEEE Symposium Series on Computational Intelligence*, Dec. 2015, pp. 159–166, doi: 10.1109/SSCI.2015.33.

[22] R. Ren, Y. Yang, and L. Sun, "Oversampling technique based on fuzzy representativeness difference for classifying imbalanced data," *Appl. Intell.*, vol. 50, no. 8, pp. 2465–2487, Aug. 2020, doi: 10.1007/s10489-020-01644-0.