Mathematical Statistician and Engineering Applications
ISSN: 2094-0343
2326-9865

A Short Review on Artifice Detection in Credit Cards using Machine learning and Data Science

Dewangini Tiwari[1], Meenakshi [1]

[1] Department of Mathematics, Chandigarh University, Mohali, Punjab

Email:- dewanginitiwari9981@gmail.com , chawlameenakshi7@gmail.com

Article Info
Page Number: 2950-2962
Publication Issue:
Vol. 71 No. 4 (2022)

Article History
Article Received: 25 March 2022
Revised: 30 April 2022
Accepted: 15 June 2022
Publication: 19 August 2022

Abstract
This paper surveys about the most dependable machine learning algorithms amongst the five-algorithm used i.e., decision tree, random forest, XGBoost, local outlier factor and isolation forest. The fundamental goal of paper is to procure 100% accuracy, while keeping the erroneous misrepresentation groupings to a minimum. Credit card Fraud Detection is an example of grouping. This paper also focusses on investigating and pre-handling informational collections as well as sending of different abnormality discovery algorithms, likewise, 'Local Outlier Factor' and 'Isolation Forest algorithm' on the 'Credit Card Transaction' information and also ML algorithm are applied on an informational collection of Visas cheats and the force of five ML algorithm is contrasted with recognize the fakes achieved utilizing credit cards. Later on every one of the classifiers/algorithms will be contrasted with know the most dependable algorithm among them.

Keywords- Credit card fraud detection, applications of machine learning, data science, isolation forest algorithm, local outlier factor, automated fraud detection, Decision Tree, XGBOOST algorithms, Random Forest.

I. INTRODUCTION

A variety of techniques are used to prevent the acquisition of money or property through false pretences, such as artifice detection and cleverness location in the financial sector. Many businesses, including those in banking and security, use false location. The use of stolen credit cards or the creation of checks are two examples of extortion in banking.

The fundamental instruments involved by analysts for 'credit card fraud detection' discovery incorporates, Machine Learning programs, Neural organizations, Clustering and classification strategies. Brain networks are dependent on the human mind, and they have successfully solved a wide range of problems in research and design thanks to their capacity for learning. For the purpose of detecting credit card fraud, Wang et al. [1] suggested a brain network-based whale computation. A cost-sensitive approach using brain organisation was put forth by Ghobadi and Rohani [2] for the identification of Mastercard fraud. Support Vector Machine was used by Gyamfi and Abdulai [3] to identify visa deception.

The methods used to divide the data into a few classes or groups include clustering and classification. Additionally, these procedures are quite beneficial for treating a variety of problems. To distinguish fakes in Mastercard exchanges, Kavitha and Suriakala [4] employed a meta classifier to a vast amount of data. Wang et al[5] .'s approach for charge card misrepresentation identification used dividing and grouping algorithms. Mishra and Ghorpade [6] used several arranging techniques for Visas misrepresentation identification while chipping

away at biassed information. A plan of action was put up for Visas misrepresentation identification by Alex G.C. de Sá et al. [7].

ML algorithms are ML calculations which are utilized in different disciplines to tackle issues essentially manages enormous measure of information. Numerous specialists applied ML, AI [8][9][10] [11] [12] and profound learning [13] [12] strategies to recognize fakes in charge cards. In any case, there is as yet a need to break down and apply the force of ML calculations to identify fakes in Visa exchanges. The regions wherein Machine learning algorithms are being used are as per the following:

Classification Classification discover a few determinations from a colossal measure of information. At the point when given a few information values from the information, the arrangement calculations endeavor to choose at least one results based on the info information. AI calculations are extremely valuable in characterization.

Regression - Regression is an administered learning strategy. Anticipating yield values from given input values is utilized. Foreseeing ceaseless data is for the most part utilized. Relapse strategies are AI procedures which are extremely valuable in forecast.

Clustering- It alludes to partitioning the issue space into bunches based on the likenesses between the information. The things in a single group are basically the same as one another. Things in various groups are different to one another in their properties. AI calculations are exceptionally helpful in grouping

The utilization of ML algorithms isn't restricted to these areas as it were. Indeed, even numerous analysts are chipping away at regions in which Machine Learning algorithms are relevant and will provide better outcomes.

ML methods are utilized in this paper on identification of cheats utilizing Mastercard exchanges. The following segment is examining the proposed work.

II. LITERATURE REVIEW

In order to obtain financial or personal gain, coercion is likely used in an illegal or criminal manner. An intentional show is an illegal practice that uses a regulation or method to achieve an unauthorized financial benefit. Various written works relating to irregularity or distortion acknowledged in this area have already been published and are available for use by the general audience. Data mining applications, robotized deception acknowledgment, and badly organized space are among the methods used in this field, according to a thorough evaluation performed by Clifton Phua and his collaborators. Suman, a research scholar with GJUS&T at Hisar HCE, offered approaches including supervised and unsupervised learning for the disclosure of visa coercion in a different publication. Despite how these methods and calculations made captivating improvements in a certain area, they forgot to offer an incredibly solid and unsurprising response for deception acknowledgment. Wen-Fang YU and Na Wang presented a specific area of investigation where they employed outlier mining, outlier recognisable proof mining, and distance total computations to categorically anticipate counterfeit trade in a mirroring examination of Visa trade enlightening file of one specific

business bank. Data mining's field of exemption mining is primarily utilised in the financial and web sectors. It is liable for detecting items that are removed from the regulatory framework, or trades that aren't real. The difference between the property's assessed value and its intended value was evaluated using the client's characteristics and taking into account the value of those characteristics. Impulsive methodologies, like creamer data mining/complex association plan estimation, for example, can identify illegal cases in a real card trade educational assortment by taking into account association revamping computation that licences for making depictions of the deviation of one case from a reference bundle have consistently proven capable on medium estimated online trade. Similar attempts have been made to advance from a completely novel perspective. Additionally, efforts have been taken to support the provided analysis affiliation in the event of deceptive transaction. If a misleading trade event were to occur, the supported system would be alarmed and information would be provided off denying the continuous trade. One of the processes that provided new insight in this area addressed deformation caused by a replacement bearing. It exhibited accurate in sorting out the underhanded trades and restricting the amount of misdirecting alerts. Notwithstanding the way that, it was gotten by plan issue together with factor misclassification costs.

The working of credit card fraud detection can be under stand through the figure mentioned below:

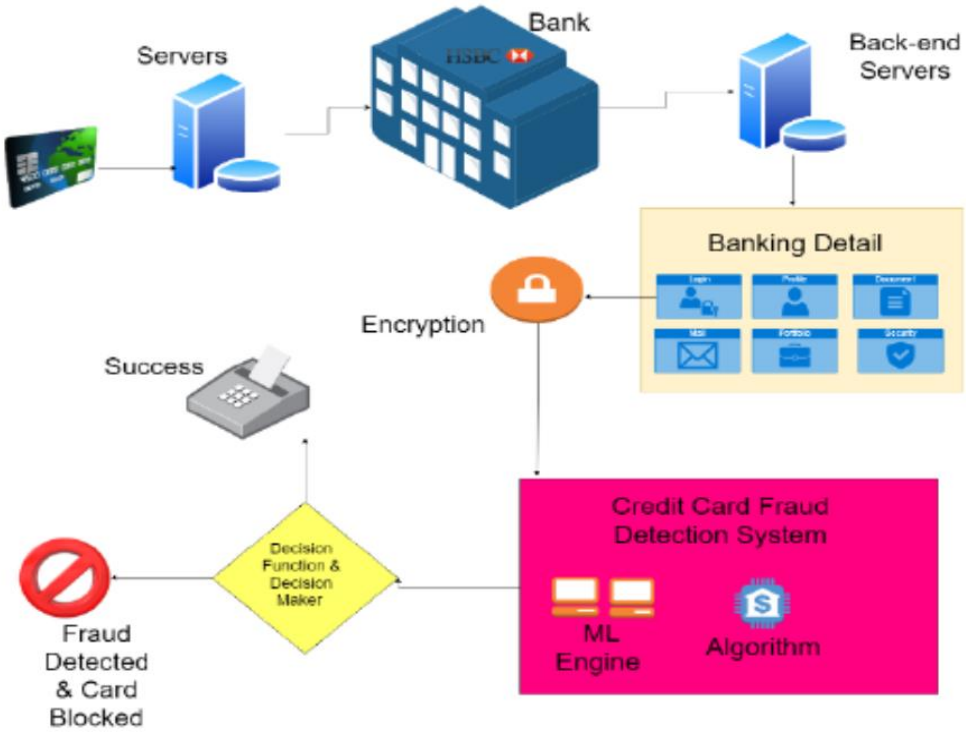
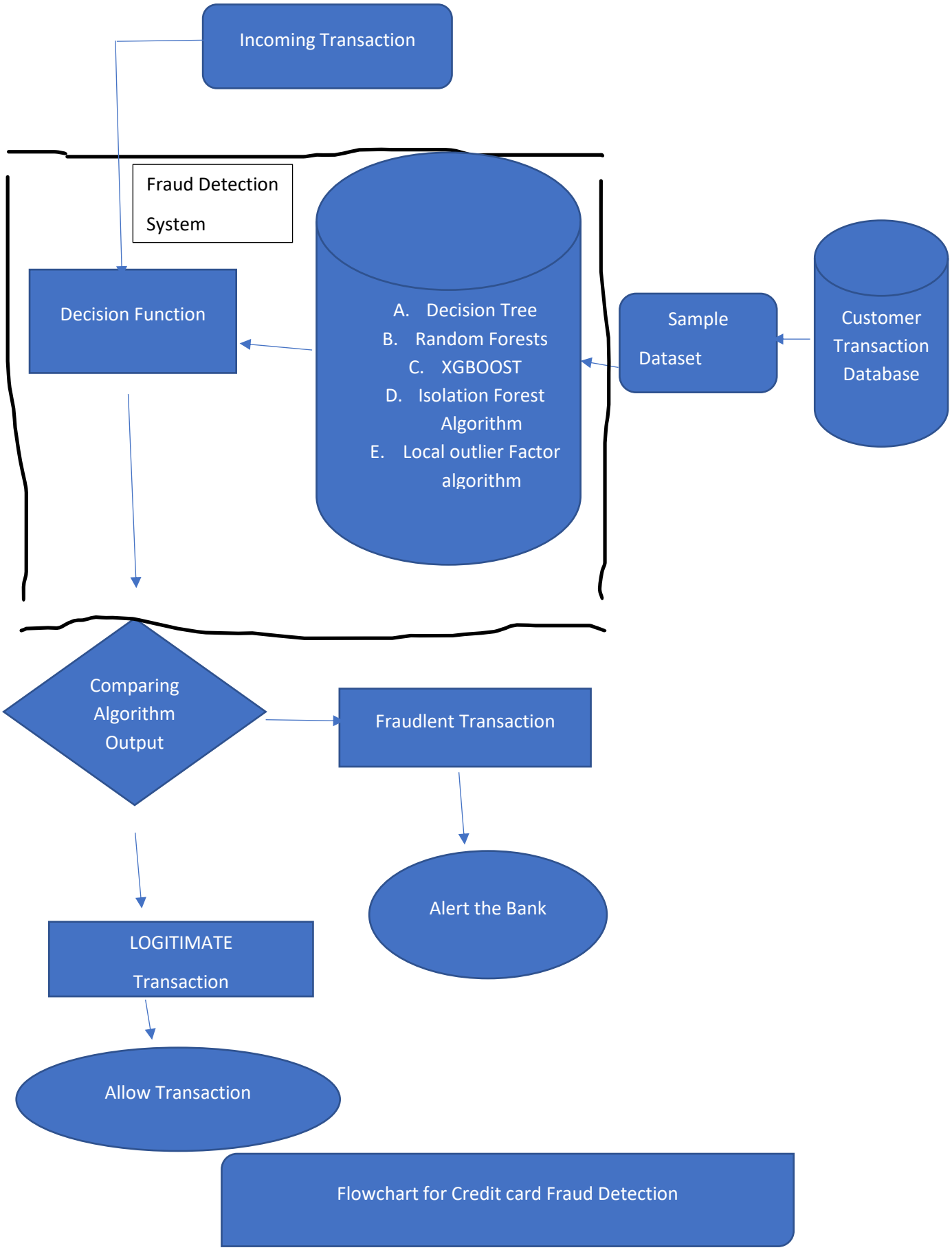


Figure1: The framework of the working of Credit Card Fraud Detection.

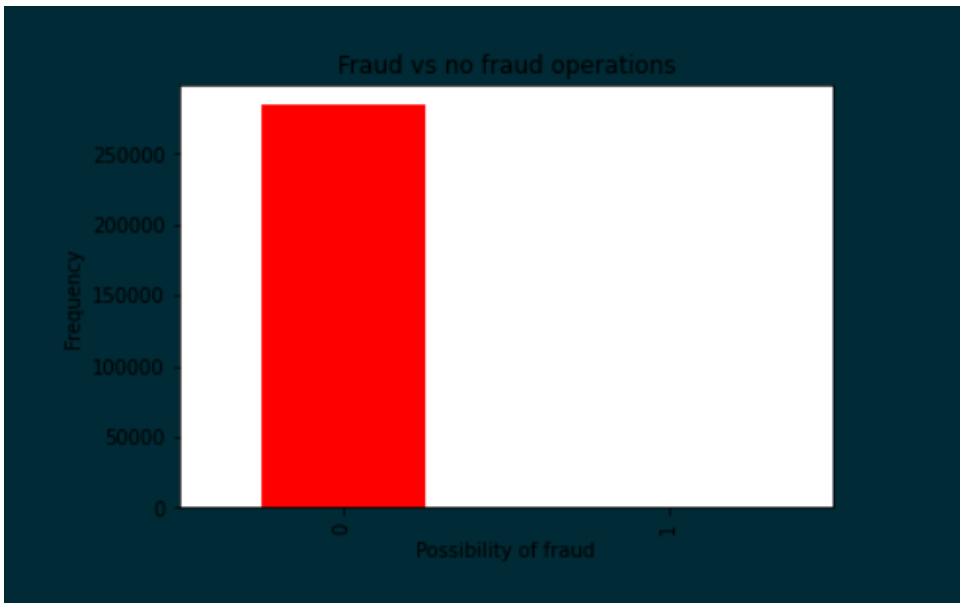
Figure 1 is taken from [15]

For the same the flowchart is:



Flowchart for Credit card Fraud Detection

We plot different graphs to check for inconsistencies in the dataset and to visually comprehend it:



The above bar graph shows the comparison of Fraudulent and non – Fraudulent operations and taken place according to the taken dataset [14]

III. PROPOSED WORK/METHODOLOGY

In this paper five ML calculations specifically Decision tree, Random Forest, XGBOOST, Isolation Forest Algorithm and nearby exception factor calculation are applied on a genuine informational collection having information of more than one needs charge cards. Initial three calculations were utilized in [16] and other two were utilized in [15]. The working of these AI calculations is as per the following:

A. Decision Tree (DT)

Decision TREE is a non-parametric directed machine learning technique. It has an already-set target variable that is typically used in issue arrangement. It is valuable for characterization and relapse both. It works downright and nonstop both for info and result factors

B. Random Forests (RF)

It is an exceptionally valuable ML algorithm. It is generally utilized in regions, for example, order, relapse examination and so on. At the preparation time RF calculation makes numerous choice trees. Random Forest is a directed learning strategy that requires test data to prepare a model. It creates arbitrary backwoods for the problem set and then uses them to observe the arrangement.

C. XGBoost

It is a notable ML algorithm which performs better much of the time. It depends on ANN. It turns out good for enormous measure of information that is accessible in tables.

D. Isolation Forest Algorithm (IFA)

Isolation Forest Algorithm (IFA) is a peculiarity discovery ML algorithm. It primarily identifies inconsistencies utilizing disconnection instead of displaying. The standard is that peculiarities are perceptions that are not many and unique, which ought to make them more straightforward to recognize. Woods utilizes outfit of disconnection Trees for the given information focuses to detach peculiarities.

E. Local Outlier Factor Algorithm (LOFA)

It is a solo peculiarity discovery ML technique which fundamentally processes the neighborhood thickness deviation of a given item as for its neighbors. It generally considers as exceptions the examples that have considerably lower thickness than their neighbors.

IV. RESULT AND ANALYSIS

Varied AI (ML) calculations are used to the informational index of more than one needs Mastercards. The ML calculations are executed utilizing python programming language and stage utilized is Jupyter Notebook Anaconda IDE. The outcome in this paper is checked on utilizing [15] and [16] with the assistance of another dataset [14].

The after effects of five AI calculations are as per the following:

A. Decision Tree

The confusion matrix of the predictions made using the Decision Tree machine learning technique is displayed in Figure 2. The Decision Tree method accurately predicted zeros in the final output 85272 times, while it predicted the zero wrongly 27 times, as can be shown in Figure 2. The Decision Tree algorithm predicts one correctly 112 times while getting one wrong 32 times.

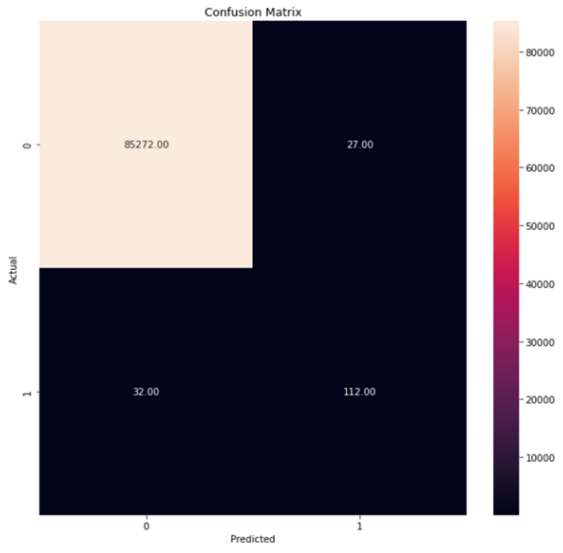


Figure 2: Confusion Matrix - Decision Tree Algorithm

	precision	recall	f1-score	support
0	1.00	1.00	1.00	85299
1	0.81	0.78	0.79	144
accuracy			1.00	85443
macro avg	0.90	0.89	0.90	85443
weighted avg	1.00	1.00	1.00	85443

Figure 3: Result - Decision tree Algorithm

The result when 70 % dataset is trained and 30% dataset is tested for Decision tree Algorithm.

Training Accuracy: 0.9998846331333641

Testing Accuracy: 0.9993094811745842

B. Random Forest

The confusion matrix of the projection outcomes produced by the application of the Random Forest ML algorithm is shown in Figure 4. It accurately predicted zeros in the final output 85291 times, and it predicted the zero wrong 8 times, as can be shown in Figure 4. The Random Forest predicts one's actions 111 times accurately and 33 times erroneously.

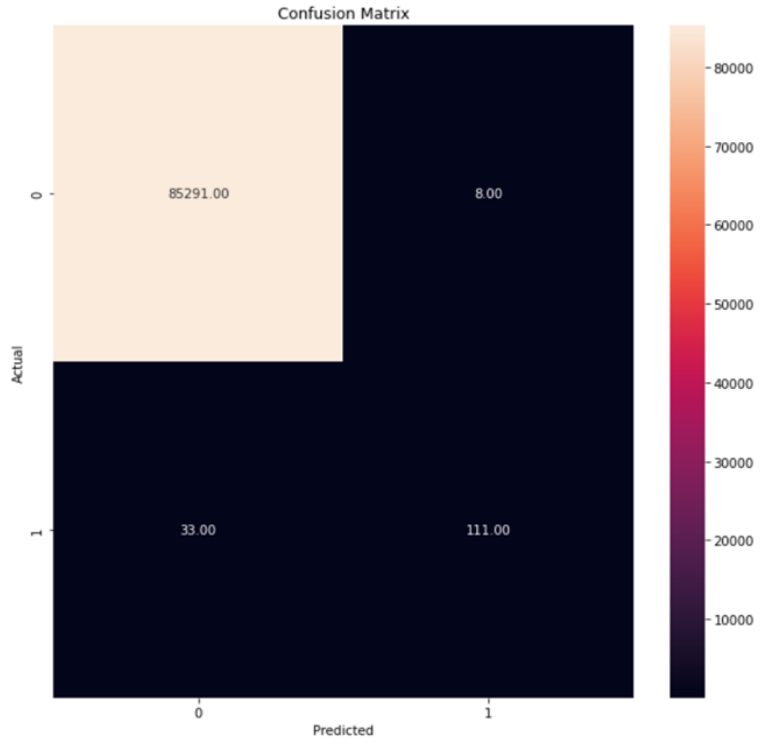


Figure 4: Confusion Matrix for Random Forest Algorithm

	precision	recall	f1-score	support
0	1.00	1.00	1.00	85299
1	0.93	0.77	0.84	144
accuracy			1.00	85443
macro avg	0.97	0.89	0.92	85443
weighted avg	1.00	1.00	1.00	85443

Figure 5 : Result for the Decision tree Algorithm

The result when 70 % dataset is trained and 30% dataset is tested for Decision tree Algorithm.

Training Accuracy: 0.9999147288377039

Testing Accuracy: 0.9995201479348805

C. XGBOOST

The confusion matrix of the predicted using the XGBOOST ML algorithm is displayed in Figure 6. The XGBOOST algorithm properly predicts zeroes in the end output 85290 times, and 9 times it predicts the zero inaccurately, according to figure 6. One is accurately predicted by the XGBOOST algorithm 115 times, whereas one is wrongly predicted 29 times.

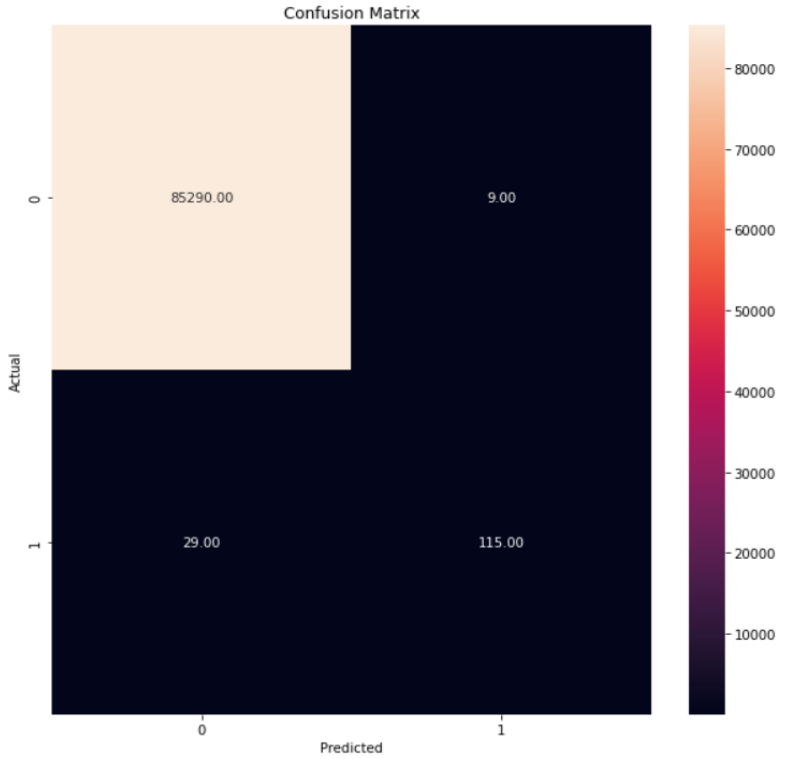


Figure 6: Confusion Matrix for XGBOOST Algorithm

	precision	recall	f1-score	support
0	1.00	1.00	1.00	85299
1	0.93	0.80	0.86	144
accuracy			1.00	85443
macro avg	0.96	0.90	0.93	85443
weighted avg	1.00	1.00	1.00	85443

Figure 7: Result for the Decision tree Algorithm

The result when 70 % dataset is trained and 30% dataset is tested for Decision tree Algorithm.

Training Accuracy: 0.9997441865131117

Testing Accuracy: 0.9995552590615966

D. Isolation Forest Algorithm (IFA)

The confusion matrix of the predicted using the IFA ML algorithm is displayed in Figure 8. The IFA method successfully anticipates zeroes in the end output 81925 times, and it predicts the zero wrongly 3374 times, as per figure 8. IFA's system predicts people 121 times accurately and 23 times erroneously.

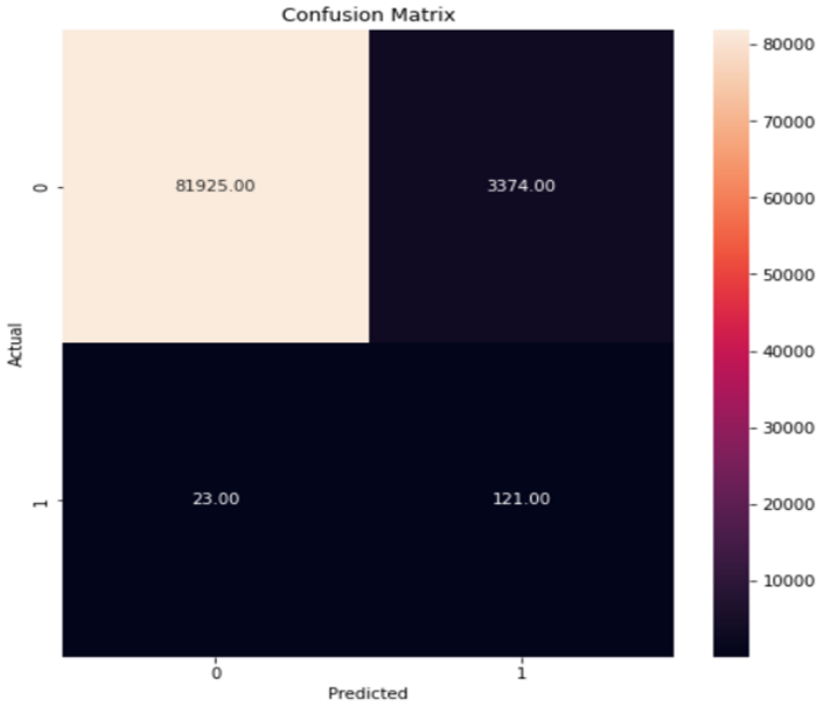


Figure 8: Confusion Matrix - Isolation forest Algorithm

	precision	recall	f1-score	support
0	1.00	0.96	0.98	85299
1	0.03	0.84	0.07	144
accuracy			0.96	85443
macro avg	0.52	0.90	0.52	85443
weighted avg	1.00	0.96	0.98	85443

Figure 9: Result - Isolation forest Algorithm

The result when 70 % dataset is trained and 30% dataset is tested for Decision tree Algorithm.

E. Local Outlier factor Algorithm (LOFA)

The confusion matrix of the predictions made using the LOFA ML algorithm is displayed in Figure 10. Figure 10 shows that the LOFA algorithm accurately predicts zeroes in the final output 85291 times while doing so wrongly 8 additional times. 33 times out of the 111 predictions made by the LOFA algorithm are inaccurate.

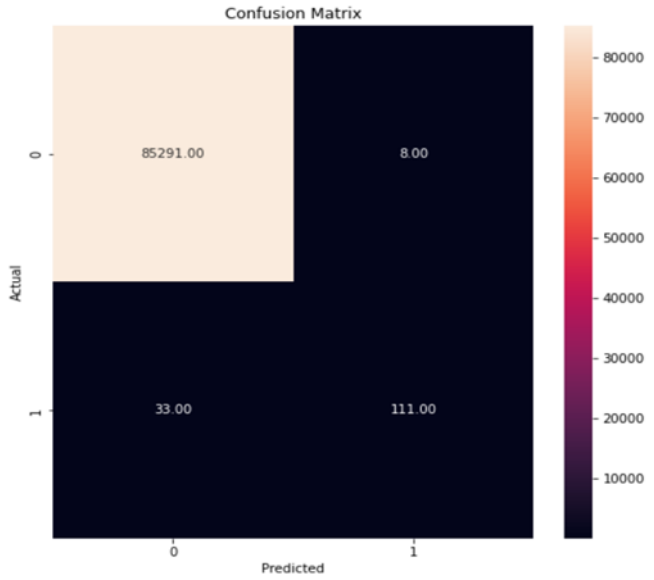


Figure10: Confusion Matrix for Isolation Forest Algorithm

	precision	recall	f1-score	support
0	1.00	0.97	0.98	85299
1	0.01	0.26	0.03	144
accuracy			0.97	85443
macro avg	0.51	0.62	0.50	85443
weighted avg	1.00	0.97	0.98	85443

Figure 11: Result for Isolation forest Algorithm

The result when 70 % dataset is trained and 30% dataset is tested for Decision tree Algorithm.

The prediction accuracy of the ML algorithms decision tree, random forest, XGBOOST, isolation forest method, and local outlier factor is represented in Table 1. Decision trees have a 99.993 prediction accuracy. The random forest has a prediction accuracy of 99.52 percent. The XGBOOST forecast accuracy stands at 99.955. Its prediction accuracy is 99.743 for the isolation forest algorithm. The prediction accuracy for the algorithm using local outlier factors is 99.659.

TABLE I. ACCURACY- ML ALGORITHMS FOR CREDIT CARD FRAUD DETECTION

Sr. No.	ML Algorithm	Predicted Accuracy of ML Algorithm
1.	Decision Tree	99.930
2.	Random Forest	99.952
3.	XGBoost	99.955
4.	Isolation Forest Algorithm	99.743
5.	Local Outlier Factor	99.659

V. CONCLUSION

In this review paper, ML Algorithms are utilized for credit card fraud detection. The force of ML is utilized to distinguish Visas cheats and the presentation of various machine it is contrasted with learn calculations. Five ML algorithms i.e., Random Forest, Decision Tree, Isolation woodland Algorithm, XGBoost, and Local Outlier Factor Algorithm are applied on an informational collection have the information of 284807 Visas. In which 70% (199364) of the dataset was prepared for the model and 30 % (85443) of the information was tried to obtain the outcome.

From the above outcome and investigation, we can obviously see that the XGBoost algorithm that is 99.955% is giving more precision in the model, secondly Random Forest with 99.952 % accuracy, thirdly Decision Tree with 99.930% precision, at fourth Isolation Forest Algorithm with 99.743 % exactness and at the last nearby outlier factor calculations with minimal exactness among five calculations utilized i.e., 99.659 %.

VI. FUTURE ENHANCEMENTS

We ultimately came up with a framework that can, with enough opportunity and information, go very close to that objective. Alas, we were unable to link the goal of 100% accuracy in misrepresentation identification. Like every such endeavour, there is room for improvement in this one as well. In order to increase the accuracy of the final product, the real concept of this endeavour considers a number of computations to be organized together as components and their results to be combined. The option of adding more calculations to this model allows for further improvement. Nevertheless, these computations should produce results that are similar

to those of the others. The modules can easily be added as written in the code after that criterion is met. As a result, the assignment has an amazing degree of measured quality and flexibility. In the dataset, there is more room for growth. As has already been shown, as the dataset size is increased, calculation accuracy increases. As a result, more data will undoubtedly improve the model's ability to spot frauds and reduce the number of fabricated advantages. Anyhow, the real banks must provide governmental assistance in this.

VII. REFERENCES

1) 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," 2018 13th International Conference on Computer Science & Education (ICCSE), Colombo, 2018, pp. 1-4. doi: 10.1109/ICCSE.2018.8468855
2) F. Ghobadi and M. Rohani, "Cost sensitive modeling of credit card fraud using neural network strategy," 2016 2nd International Conference of Signal Processing and Intelligent Systems (ICSPIS), Tehran, 2016, pp. 1-5. doi: 10.1109/ICSPIS.2016.7869880
3) N. K. Gyamfi and J. Abdulai, "Bank Fraud Detection Using Support Vector Machine," 2018 IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), Vancouver, BC, 2018, pp. 37-41. doi: 10.1109/IEMCON.2018.8614994
4) M. Kavitha and M. Suriakala, "Real time credit card fraud detection on huge imbalanced data using meta-classifiers," 2017 International Conference on Inventive Computing and Informatics (ICICI), Coimbatore, 2017, pp. 881-887. doi: 10.1109/ICICI.2017.8365263
5) H. Wang, P. Zhu, X. Zou and S. Qin, "An Ensemble Learning Framework for Credit Card Fraud Detection Based on Training Set Partitioning and Clustering," 2018 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI), Guangzhou, 2018, pp. 94-98. doi: 10.1109/SmartWorld.2018.00051
6) A. Mishra and C. Ghorpade, "Credit Card Fraud Detection on the Skewed Data Using Various Classification and Ensemble Techniques," 2018 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS), Bhopal, 2018, pp. 1-5. doi: 10.1109/SCEECS.2018.8546939
7) Alex G.C. de Sá, Adriano C.M. Pereira, Gisele L. Pappa, A customized classification algorithm for credit card fraud detection, Engineering Applications of Artificial Intelligence, Volume 72, 2018, Pages 21-29, ISSN 0952-1976, https://doi.org/10.1016/j.engappai.2018.03.011.
8) M. Kavitha and M. Suriakala, "Hybrid Multi-Level Credit Card Fraud Detection System by Bagging Multiple Boosted Trees (BMBT)," 2017 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC), Coimbatore, 2017, pp. 1-5. doi: 10.1109/ICCIC.2017.8524161
9) M. F. Zeager, A. Sridhar, N. Fogal, S. Adams, D. E. Brown and P. A. Beling, "Adversarial learning in credit card fraud detection," 2017 Systems and Information Engineering Design

Symposium (SIEDS), Charlottesville, VA, 2017, pp. 112-116. doi: 10.1109/SIEDS.2017.7937699

10) Ugo Fiore, Alfredo De Santis, Francesca Perla, Paolo Zanetti, Francesco Palmieri, Using generative adversarial networks for improving classification effectiveness in credit card fraud detection, Information Sciences, Volume 479, 2019, Pages 448-455, ISSN 0020-0255, https://doi.org/10.1016/j.ins.2017.12.030.
11) Fabrizio Carcillo, Yann-Aël Le Borgne, Olivier Caelen, Yacine Kessaci, Frédéric Oblé, Gianluca Bontempi, Combining unsupervised and supervised learning in credit card fraud detection, Information Sciences, 2019, ISSN 0020-0255, https://doi.org/10.1016/j.ins.2019.05.042.
12) G. Rushin, C. Stancil, M. Sun, S. Adams and P. Beling, "Horse race analysis in credit card fraud—deep learning, logistic regression, and Gradient Boosted Tree," 2017 Systems and Information Engineering Design Symposium (SIEDS), Charlottesville, VA, 2017, pp. 117-121. doi: 10.1109/SIEDS.2017.7937700
13) Z. Kazemi and H. Zarrabi, "Using deep networks for fraud detection in the credit card transactions," 2017 IEEE 4th International Conference on Knowledge-Based Engineering and Innovation (KBEI), Tehran, 2017, pp. 0630-0633. doi: 10.1109/KBEI.2017.8324876
14) https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud
15) S P Maniraj, Aditya Saini, Swarna Deep Sarkar Shadab Ahmed " Credit card Fraud Detection Using Machine Learning and Datascience," 2019 International Journal of Engineering Research & Technology (IJERT) Vol. 8 Issue 09, September-2019. http://www.ijert.org IJERTV8IS090031
16) Anuj Kumar, Vinod Jain, Mayank Agrawal "Performance Analysis of Machine Learning Algorithms in Credit Card Fraud Detection," Amity University, Noida, India, June 4-5, 2020, 8th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO).