Improved Credit card fraud detection by optimizing feature weighting with Ensembled of ANN and SVM

Aaushi Sharma[1], Neha Bathla[2]
Yamuna Group of Institutions Engineering and Technology, Yamuna Nagar, Haryana

Abstract-
Payment cards like credit and debit cards are absolute basics in international economic transactions. As these payment strategies are popularizing, the data related to these is being accessed by heaps additional individuals than needed increasing the chance of a potential. The employment of credit cards is daily in modern society. Fraud may be a million-dollar business, and it's rising per annum. Fraud presents a significant value to our economy worldwide. Trendy techniques supported data processing, Machine learning, Sequence Alignment, mathematical logic, Genetic Programming, computing, etc., are introduced for detective work Mastercard fraudulent transactions. This paper shows, however, data processing techniques will be combined with success to get high fraud coverage combined with a coffee or high warning rate.

Keywords- credit,cnn,ANN,smote,fraud,detection

I. INTRODUCTION

At the terminal stage, transactions are tested for validity or not, as is shown in figure1.2. Such critical requirements are validated and transactions are filtered at the terminal stage, such as appropriate cash, a valid PIN (personal identification numbers), etc. The predictive process then defines each appropriate transaction and ultimately categorises the transactions as genuine or fraudulent. Every fraudulent warning was investigated and feedback on the forecasting model was given to improve the efficiency of the model [30]. This article discusses the predictive model only.

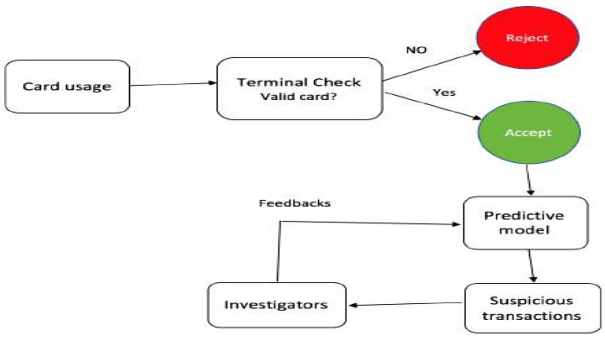

Figure 1.1: Fraud detection process

1.3 Challenges in fraud detection

It is not as simple as it looks to create a fraudulent discovery method. The practitioner must determine which approach to learning to use, which algorithms to use, which features to use, and how to manage class inequality (fraudulent cases are very sparse compared to legitimate cases) [30]. The main problem in the fraud detection system is not just class imbalance. Another issue in the classification task was the combination of legitimate and false classes because of insufficient knowledge about operational records,[52] and the majority of machine learning algorithms are less successful within these scenarios [48]. A model of fraud detection forecasts the existence of the class (true or fraudulent) in a specific scenario and sends the investigators the notice of the most suspicious transaction. In order to enhance its presentation, researchers then carry out further analysis and respond to the fraud detection scheme. This method can, however, be an overall process for the investigators, because of which the investigators only verify few transactions on time. In this case, the predictive model has just a few reviews, usually resulting in a less accurate model [30]. Finally, considering the very rare public disclosure of consumer data by financial institutions due to confidentiality concerns, the discovery of the true financial databases would be a challenge. It is one of the major obstacles to fraudulent recognition testing.

Problems of Credit Card Fraudulent finding

Fraudulent prevention scheme is prune to many problems as well as challenges mentioned bellow. A successful strategy for detecting fraud would have the potential to overcome such challenges and produce the highest results.

- ***Unbalanced results:*** Results on credit card fraud identification is unbalanced in design. This indicates that low numbers of all purchases by credit card are fraudulent. Which causes quite complex as well as inaccurate identification of fraudulent transactions?
- ***Lack of adaptability:*** Classification algorithms continue to face the difficulty of identifying different forms of natural or deceptive patterns. The managed as well as unmonitored fraud identification methods, however, are ineffective in identifying emerging forms in regular as well as fraud behaviour.
- ***Specific value of misclassification:*** Various misclassification failures have different meaning in fraudulent identification activities. Misclassification of a legitimate activity as theft is not as dangerous as

legitimate fraud identification. Since the recognition problem during the first case can be found in more investigations .

- ***Overlapping data:*** Many transactions might be deemed fake, although they are completely legal (false positive) as well as vice versa, fraud transactions may often seem genuine (false negative). A main obstacle for fraud identification programs is also maintaining a small incidence of false positives and false negatives .
- ***Fraud identification costs:*** The program will take into consideration both the identified criminal behavior costs and the avoidance costs. For eg, by preventing a dishonest transaction of a few dollars no revenue is received .

II. RELATED WORK

Saad M. Darwish [1] Suggested a smart two-level card fraudulent identification system through extremely unnecessary databases, focused on the semiconductor fusion of k-means as well as artificial bee colony algorithm (ABC) to get better recognition accuracy as well as speeding up convergence. ABC as a second category stage carries out a kind of neighborhood quest in conjunction with the worldwide search to manage the failure of the k-means classifier to find the real cluster if the similar data is entered in a different format it may yield specific clusters. In fact, the classifier k-means will be encircled by the maximum central, because it is prone to the original state. The recommended framework filters the attributes of the dataset that used a built-in rule engine to determine if the purchase is legitimate or illegitimate depending on several criteria of consumer activity (profile) such as regional areas, frequency of usage as well as book balance. Investigation demonstrates that perhaps the new model will advance the precision of detection in opposition to the probability of irregular transactions as well as have increasing consistency relation to conventional approaches.

Altyeb Altaher Taha et al. [2] proposed Smart method for identifying credit card theft utilizing an Automated Light Gradient Boost System. A Bayesian-based hyperparameter optimization algorithm is smartly implemented into the planned approach to tuning the parameters of a light gradient enhancing system (LightGBM). In order to show the efficacy of our experimental OLightGBM in identifying credit card fraud, tests were conducted utilizing two real-world government card transaction data sets composed of fraud as well as legal purchases. Focused on a compare with other methods utilizing the two data sets, the suggested solution exceeded the other solutions and obtained the best accuracy value (98.40 percent), accuracy (92.88 percent), accuracy (97.34 percent), as well as F1 (56.95 percent).

E. Saraswathiet al. [3] elucidate Artificial Neural Network (ANN) has the capacity to act like a human mind if appropriately equipped. We have also introduced SOM for reason of exactness. In this article the author addressed the network efficiency as well as its accuracy.

K. Karthikeyanet al. [4] detects credit card abuse utilizing machine learning algorithms. Firstly regular versions are included. Then hybrid techniques are implemented that utilize AdaBoost as well as methods of bulk vote. To test the feasibility of the model, a sample of credit card data is used and is publicly accessible. Then a Financial institution's real-world payment card data collection is assessed. Additionally, noise is applied to the data samples to help test algorithm robustness. Favorably, the investigational findings show that the plurality voting system reaches strong accuracy levels in the identification of credit card fraudulent events.

Diwakar Tripathi et al. [6] proposed usage of the local outlier element (LOF) for credit card fraud identification. We used the buying number as fraud verification. The planned program is introduced in MATLAB code, and device output is measured in terms of the system's factual unhelpful, fake optimistic rate as well as consistency over the various closest neighbours. Accurateness of the suggested solution varies among 60-69 per cent for dataset 1 and 96 per cent for dataset 2 with community variance.

Kuldeep Randhawa et al. [7] Suggested a strategy for the analysis of card fraudulent utilizing machine learning. Normal versions were first used following alternative versions were imagined using AdaBoost as well as plurality voting systems. The publicly accessible data collection was used for assessing the performance of the model. The experiments were conducted on the basis of the theoretical results which show that the majority of voting methods achieve good accuracy rates so that notice the fraudulent in the credit cards. The experimental data was applied to the noise of between 10 per cent to 30 per cent for further assessment of the hybrid versions. A strong result of 0.942 for 30 per cent applied noise was obtained through multiple voting processes. And it was assumed that in the midst of noise, the voting system displayed more reliable efficiency.

Abhimanyu Roy et al. [8] Deep learning topologies suggested for the analysis of online transaction abuse. This method is adapted through hierarchical artificial network of built-in time as well as remembrance elements for instance short long-term memory and many other parameters. I used centralized cloud storage system for strong efficiency. The researchers' proposed study offers an appropriate guide to the sensitivity analysis of the proposal criteria, based on the fraudulent detecting results. The researchers have suggested a method for the parameter tuning of topologies for detecting falsification in deep learning. This helps the financial company to popular the damages by eliminating illegal practices.

Shiyang Xuan et al. [9] used 2 forms of random woods, that educate regular and irregular transaction behavioral distinctiveness. The investigator contrasts these 2 random woods, that are distinguished by success in identify card fraudulent based on their classificators. The data used is from

a China e-commerce business that is used to evaluate the output of the random forest model of these two forms. In this article the researcher utilized the B2C dataset to define as well as diagnose credit card fraud. Thus, the researcher concludes through the outcome that perhaps the proposal random forests have fine consequences on a limited dataset however are still troubles for instance unfair data that render it less successful than other dataset.

Johannes Jurgovskyet al.[12] phrased the issue of fraud identification as a sequence recognition function, utilizing Large Short-Term Memory (LSTM) networks to integrate sequences of transactions. The article also incorporates hi-tech aggregation techniques for apps as well as utilizes conventional retrieval measures to monitor our performance.

Dastgir Pojee et al. [13] Proposed a new method to avoid paying the invoice or the bill. This technique is referred to as the mobile "NoCash" programme, which is predominantly used by traders to facilitate payment of customers. This method requires no NFC-Enabled Sales Point (PoS), and only mobile telephones. The only purpose for which the system is built is to reduce customers' burden of carrying cards outdoors by providing simple transfer mechanisms. When the NoCash framework with several features is introduced based on the rise in multiple NFC cell phones, client shopping experience is enhanced. Fraud behaviours are reduced by this proposed application in order to provide merchants with benefits. The customers of the application can be connected to the cost history and reduce unnecessary costs using this proposed process.

Zahra Kazemi et al. [15] planned deep auto encoder that is utilize to retrieve the good credit card transaction detail. It would also incorporate softmax functionality to address issues with the class names. An above-complete automatic encoder is being used to map the data into a high-dimensional domain, as well as a sparse model could be used in a concise fashion that offers advantages in classifying a form of deception. This is among the majority driven or efficient methods used to identify credit card fraudulent. Such kinds of networks have a dynamic data delivery that is really hard to understand. In certain points, Deep autoencoder was used to select the best data functionality as well as for identification needs. Better precision as well as low variance within such networks are also obtained.

Baoping Cai et al. [16] Provides a bibliographic analysis of the usage of BNs in fault diagnosis across the past years, with a emphasis on engineering. This research also provides general method for the modeling of fault diagnosis for BNs; procedures involve modeling of BN configuration, modeling of BN parameters, BN inference, detection of fault, confirmation and testing. The article presents a set of categorization systems for BNs for error analysis, BNs paired with additional methods as well as the defect diagnosis area for BN. At last, this review examines existing differences & problems, as well as a variety of avenues for potential studies.

N. Balasupramanian et al. [17] Proposed a machine learning methodology as well as Big Data Analytics to spot along with avoid fraudulent electronic purchases. The model requires the vast amount of electronic transaction data to be processed, and is then clean as well as functionality removed as well as the using the primary method of review of components. The decreased functions may be utilized to instruct the model of machine learning, that detects as well as recognizes consumer habits relevant to e-transactions.

III. RESULT ANALYSIS

3.1 Result Analysis

https://www.kaggle.com/mlg-ulb/creditcardfraud collect from this link and This dataset presents transactions that occurred in two days, where we have 492 frauds out of 284,807 transactions. The dataset is highly unbalanced, the positive class (frauds) account for 0.172% of all transactions.

3.2 Result Analysis

This part incudes the details of the experiment on the basis of different classifiers as represented below:

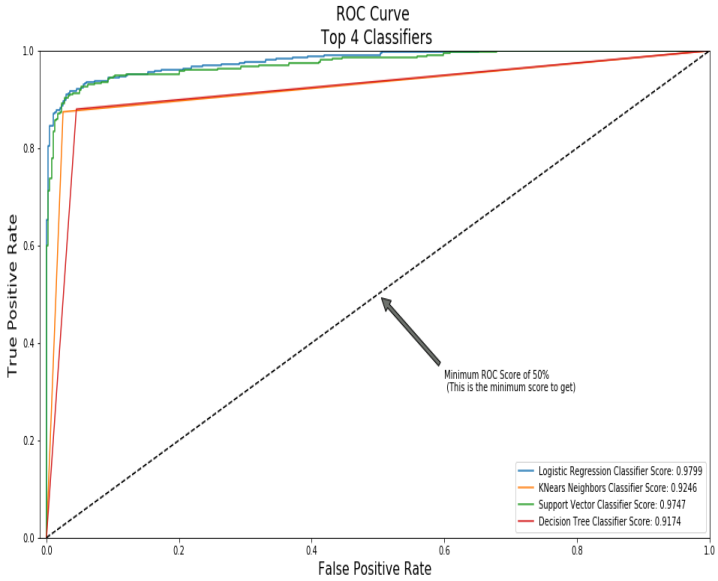

Figure 3.1: ROC Curveof different classifier

In fig 3.1 show the ROC curve comparsion of different existing approaches.in comaprsion use logistic regression,knn,SVM and decision tree andlogistic regression improve 0.97.

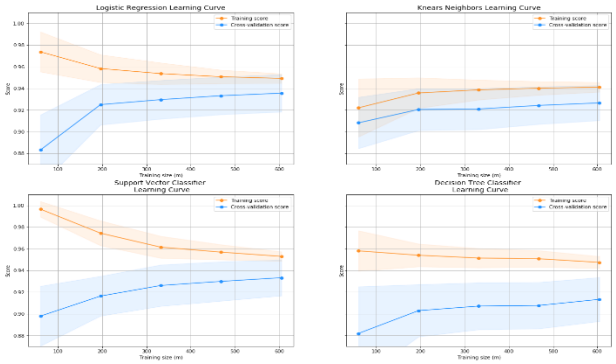

Figure 3.2: Comparsion Curveof different classifier

In fig 3.2 show the accuracy curve comparsion of different existing approaches.in comaprsion use logistic regression,knn,SVM and decision tree and also analysis cross validation results for finding overfitting and under fitiing comparsion.cross validation less score than training score so wecan say all are inherit with overfitting.

3.3 Confusion Matrix

Positive/Negative: Class form (label) ["No", "Yes"] True / False: listed correctly or wrongly in the model.

True Negatives (Top-Left Square): This is the number of classifications of the class No.

The number of wrongly categorized groups "No"(not known as fraud) is False Negative (Top-Right Square).

Bottom-left (false positives): This is the number of erroneously labelled "yes" groups (fraud detected)

True Positives (Bottom-Right Square): that is the correct number of classifications for the "Yes" class.

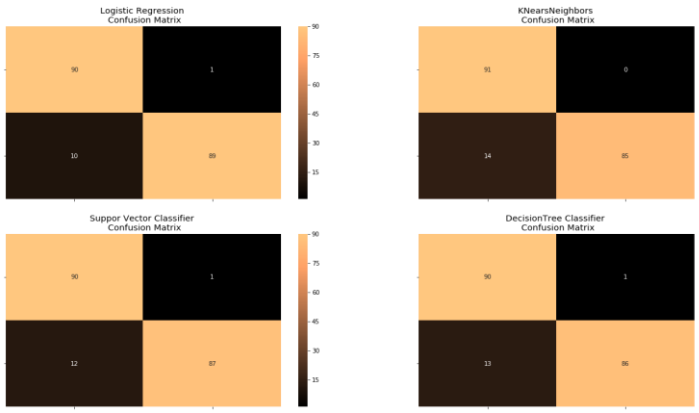

Figure 3.3: Confuion matrix of different classifier

Here is again, how the confusion matrix works:

Upper Left Square: the sum of our model of no fraud transactions correctly identified.

High Right Square: the number of transactions falsely reported as fraud, but no fraud is the actual mark.

Lower left Square: the number of transactions wrongly identified as not fraudulent, but fraud is the real mark.

Lower Right Square: the number of fraud transactions correctly categorized as our guide.

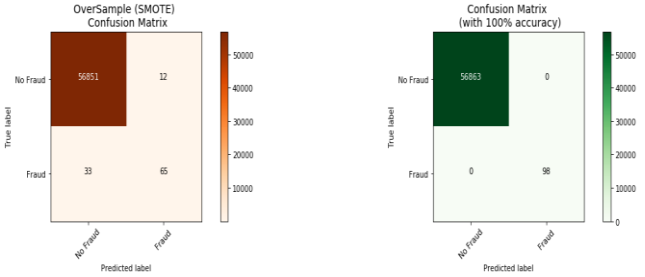

Figure 3.4: Confuion matrix of Proposed of different classifier

Approaches	Accuracy	Precision	Recall
SVM	93.23	94.23	96.23
KNN	94.12	93.45	95.34
Logistic regression	93	92.13	96.23
Decision Tree	92.34	90.23	91.34
SMOTE-CNN	97.12	96.23	98.12

Table 3.1 Comparsion of different parameters of proposed and existing approaches

In fig 5.5 and table 5.a show the accuracy different parameters anlysis of existiong and proposed approch and improve the all parameters significant improve all parameters.

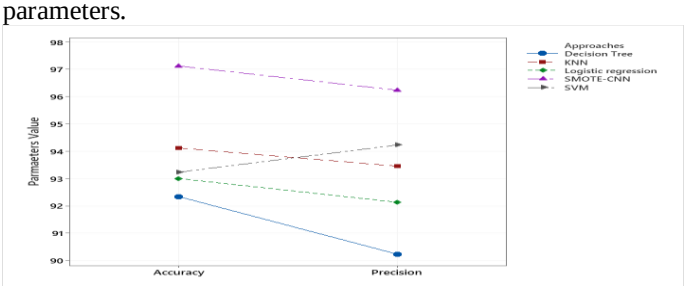

Figure 3.5: Comparsion of accuracy and precision

IV CONCLUSION

The credit card fraud detection methods have gained the popularity in the past decade with the evolution of the statistical models. These models are used to automate the process of pattern recognition, which takes comparatively lesser time and can handle many transactions per day. The statistical model based upon the binary classification such as support vector machine (SVM), stochastic gradient descent (SGD), smote-CNN, etc. to improve the accuracy of the credit card fraudulent pattern detection. The malicious patterns are also known as outlier or anomaly, which must be detected correctly in order to minimize the bank losses caused by fraudulent transactions. The performance of the proposed model would be evaluated using the precision, recall, F1-measure and accuracy-based parameters.

V REFERENCES

[1] Darwish, S. M. (2020). An intelligent credit card fraud detection approach based on semantic fusion of two classifiers. *Soft Computing*, *24*(2), 1243-1253.

[2] Taha, A. A., & Malebary, S. J. (2020). An Intelligent Approach to Credit Card Fraud Detection Using an Optimized Light Gradient Boosting Machine. *IEEE Access*, *8*, 25579-25587.

[3] Saraswathi, E., Kulkarni, P., Khalil, M. N., & Nigam, S. C. (2019, March). Credit Card Fraud Prediction and Detection using Artificial Neural Network and Self-Organizing Maps. In *2019 3rd International Conference on Computing Methodologies and Communication (ICCMC)* (pp. 1124-1128). IEEE.

[4] Karthikeyan, K., Raj, K. S., Ramaganesh, S., Parthasarathi, P., & Suguna, N. (2019). Credit Card Fraud Detection Using Machine Learning, 6(2). DOI: 10.32628/IJSRST196271

[5] Tinubu, C. O., Aborisade, D. O., Sodiya, A. S., Onashoga, S. A., & Ganiyu, M. A. (2019). Towards detecting credit card frauds using Hidden Markov Model. *Journal of Computer Science and Its Application*, *26*(2), 54-63.

[6] Tripathi, D., Sharma, Y., Lone, T., & Dwivedi, S. (2018). Credit card fraud detection using local outlier factor. *International Journal of Pure and Applied Mathematics*, *118*(7), 229-234.

[7] Randhawa, K., Loo, C. K., Seera, M., Lim, C. P., & Nandi, A. K. (2018). Credit card fraud detection using AdaBoost and majority voting. *IEEE access*, *6*, 14277-14284.

[8] Roy, A., Sun, J., Mahoney, R., Alonzi, L., Adams, S., & Beling, P. (2018, April). Deep learning detecting fraud in credit card transactions. In *2018 Systems and Information Engineering Design Symposium (SIEDS)* (pp. 129-134). IEEE.

[9] Xuan, S., Liu, G., Li, Z., Zheng, L., Wang, S., & Jiang, C. (2018, March). Random forest for credit card fraud detection. In *2018 IEEE 15th International Conference on Networking, Sensing and Control (ICNSC)* (pp. 1-6). IEEE.

[10] Campus, K. (2018). Credit Card Fraud Detection Using Machine Learning Models and Collating Machine Learning Models. *International Journal of Pure and Applied Mathematics*, *118*(20), 825-838.

[11] Gyamfi, N. K., & Abdulai, J. D. (2018, November). Bank Fraud Detection Using Support Vector Machine. In *2018 IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)* (pp. 37-41). IEEE.

[12] Jurgovsky, J., Granitzer, M., Ziegler, K., Calabretto, S., Portier, P. E., He-Guelton, L., & Caelen, O. (2018). Sequence classification for credit-card fraud detection. *Expert Systems with Applications*, *100*, 234-245.

[13] Pojee, D., Zulphekari, S., Rarh, F., & Shah, V. (2017, October). Secure and quick NFC payment with data mining and intelligent fraud detection. In *2017 2nd International Conference on Communication and Electronics Systems (ICCES)* (pp. 148-152). IEEE.

[14] Oberoi, R. (2017). Credit Card Fraud Detection System: Using Genetic Algorithm. *International Journal of Computer & Mathematical Sciences*, *6*(6).

[15] Kazemi, Z., & Zarrabi, H. (2017, December). Using deep networks for fraud detection in the credit card transactions. In *2017 IEEE 4th International Conference on Knowledge-Based Engineering and Innovation (KBEI)* (pp. 0630-0633). IEEE.

[16] Cai, B., Huang, L., & Xie, M. (2017). Bayesian networks in fault diagnosis. *IEEE Transactions on Industrial Informatics*, *13*(5), 2227-2240.

[17] Balasupramanian, N., Ephrem, B. G., & Al-Barwani, I. S. (2017, July). User pattern based online fraud detection and prevention using big data analytics and self-organizing maps. In *2017 International Conference on Intelligent Computing, Instrumentation and Control Technologies (ICICICT)* (pp. 691-694). IEEE.

[18] Arora, S., & Kumar, D. (2017, May). Selection of optimal credit card fraud detection models using a coefficient sum approach. In *2017 International Conference on Computing, Communication and Automation (ICCCA)* (pp. 482-487). IEEE.

[19] 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)* (pp. 1-9). IEEE.

[20] Adewumi, A. O., & Akinyelu, A. A. (2017). A survey of machine-learning and nature-inspired based credit card fraud detection techniques. *International Journal of System Assurance Engineering and Management*, *8*(2), 937-953.

[21] Modi, K., & Dayma, R. (2017, June). Review on fraud detection methods in credit card transactions. In *2017 International Conference on Intelligent Computing and Control (I2C2)* (pp. 1-5). IEEE.