International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 08 Issue: 04 | Apr 2021 www.irjet.net p-ISSN: 2395-0072

Credit Card Fraud Detection using Unsupervised Machine Learning

V Uma Mahesh[1], Sunil Kumar Reddy[2], Leelakar Reddy[3], Sai Krishna[4], Jangam Dilleshwar[5], Savleen Kaur*[6]

[6]Assistant Professor, School of Computer Science and Engineering, Lovely Professional University, Phagwara, Punjab, India

[1-5]Research Scholars, School of Computer Science and Engineering, Lovely Professional University, Phagwara, Punjab, India

---

Abstract- The advancement of correspondence innovations and online business has made the credit card the most widely recognized installment procedure for online and ordinary buys. Protection is supposed to avoid fraud transactions in this framework. Credit card data purchase fraud transactions are growing annually. The challenges should be solved by unsupervised machine learning. In this aspect, scientists are still working to identify and avoid such fraud using new techniques. However, such methods are often required to detect such fraud correctly and effectively [1]. Our aim here is to identify fraudulent transactions while reducing incorrect classifications of fraud. Credit card Fraud Identification is a typical standard sample of variety. However, specific approaches are still required to detect fraud correctly and efficiently. The proposed process exceeds the approach to Auto Encoder (AE), Local Outlier Factor (LOF), Isolation Forest (IF), Neural Network (NN), and K-mean clustering.

Keywords- *Credit card fraud, Unsupervised Learning, Auto-Encoder, Anomaly Detection*

I. INTRODUCTION

Fraud is unauthorized and undesired using an account by someone other than the account holder in credit card transactions. The owner of the card, the card agent issued, and may not be notified even the card guarantee before the record is used for shopping. As online shopping and payment of accounts have come true, a physical card for creating purchases is no longer required. Required measures of prevention can be taken to stop this abuse[2]. To reduce it and protect against similar occurrences, the conduct of these fraudulent practices can be studied in the future. Many challenges are complicated when it comes to successful fraud management. This includes vast and data volume, the flexibility of the applications, improvements in business practices and operations, and the continuing introduction of new evasion methods to prevent current techniques of detection[3]. It is difficult to identify false financial statements using standard auditing techniques because of insufficient knowledge of financial statement features, a lack of expertise, and rapidly shifting fraudster tactics. This is a significant problem and requires public attention, like machine learning, to automate this issue's solution. A safety alert against such fraud should therefore always be provided.

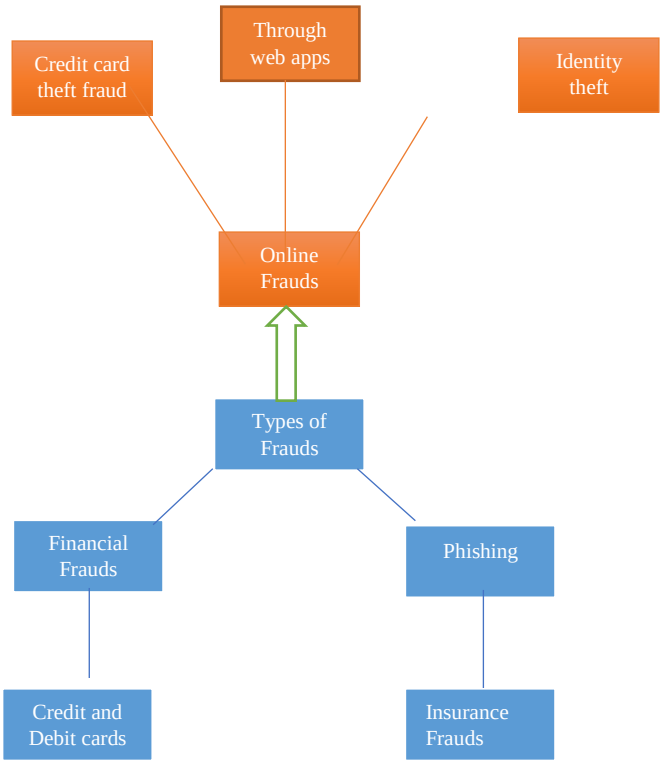


Figure 1: Types of Frauds

Figure 1 shows the types of frauds. Frauds may be classified as financial fraud, communication fraud, and market-induced fraud in three distinct ways. Financial fraud is the product of credit card fraud. These frauds should be stopped and identified in due course. Various researchers perform several studies in this direction to establish reliable and productive techniques. Hackers and intruders seek new forms of violating on safety. So, a security warning against such fraud should always be issued[4]. Machine learning has proposed many algorithms. A structure is proposed for investigating credit card theft.

II. LITERATURE REVIEW

L. Bhavya, V. Sasidhar Reddy, U. Anjali Mohan, and S. Karishma A [26]: Today, vast amounts of internet purchases have increased. There is an enormous share of

online credit card purchases. Credit card fraud detection technologies in banking and financial enterprises are also more needed. Credit card theft may be used to obtain products from an account without payment or obtain illegal money. Fraud occurrences were popular with the demand for the money credit card. The financial loss of the cardholder is also enormous.

Renjith, Shini. B [27]: The e-commerce share of worldwide retail expenditure has been steady over the years, suggesting that customer focus has shifted from bricks and mortar to retail clicks. In recent times, online markets have played a key role. Details and several methods to monitor and avoid fraudulent e-Commerce customers and their purchases are under discussion. Another area of consumer manipulation, called commercial fraud, is on the seller's side. Goods/services provided and advertised at low prices but never delivered are a clear example of such fraud.

Rai, Arun Kumar, and Rajendra Kumar Dwivedi. C [19]: Development of communication technologies and e-commerce has made the credit card the most common technique of payment for both online and regular purchases. So, security in this system is highly expected to prevent fraudulent transactions. Fraud transactions in credit card data transactions are increasing each year. In this direction, researchers are also trying novel techniques to detect and prevent such frauds. However, there is always a need for some techniques that should precisely and efficiently detect these frauds. This paper proposes a scheme for detecting frauds in credit card data that uses a Neural Network (NN) based unsupervised learning technique. The proposed method outperforms the existing approaches of Auto Encoder (AE), Local Outlier Factor (LOF), Isolation Forest (IF), and K-Means clustering. THE proposed NN-based fraud detection method performs with 99.87% accuracy, whereas existing methods AE, IF, LOF, and K Means give 97%, 98%, 98%, and 99.75% accuracy, respectively.

III. PROPOSED WORK

Here, the unsupervised learning approaches are used for fraud detections. The study proposed is described in various ways of flow Charts. The flow diagrams indicate the steps required to identify fraud. It shows that the data is cleaned and features are extracted before processing. We then train the algorithm and test it with different unsupervised learning models. And output on other metrics is evaluated. We used five algorithms, NN, AE, IF, LOF, K-means, one by one for the identification of outliers and tested the achievement. We have seen NN and the other schemes are outperforming.

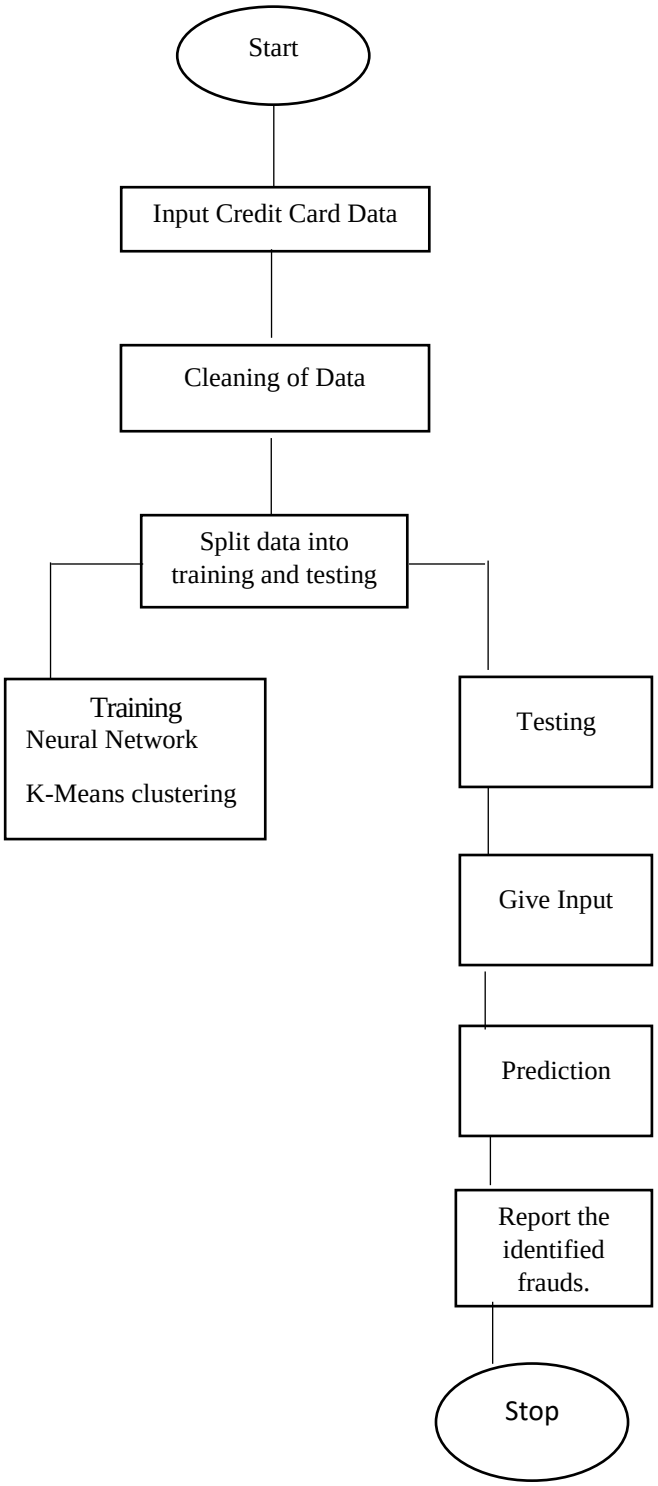


Figure 3: Flow Chart of Fraud Detection

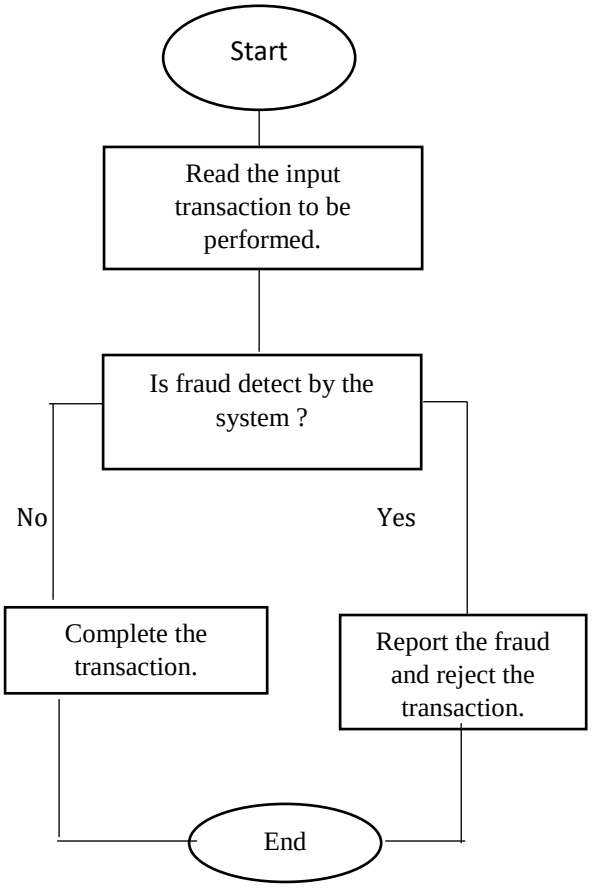


Figure 4: Flow chart for performing transaction.

Here Figure 3 illustrates the fraud detection model's flow chart. Model. Figure 4 shows the flow chart of performing the transaction by our system.

IV. IMPLEMENTATION

The proposed analysis on a credit card scheme dataset is carried out in Python, and performance tests are carried out on some measures explained in this section.

A. Unsupervised learning approaches to detect frauds.

We have used unsupervised learning methods to identify fraud in card details.

Neural Network (NN): The idea is to use neural network approaches to prepare the model with a vast number of renowned samples and implement a query.
Auto Encoder (AE): It is an unsupervised, highly efficient artificial network learning scheme. An autoencoder is to learn Presentation on behalf of a set of data Reduction of dimensionality. It trains the network by disregarding Chaos, Noise.

Isolation Forest (IF): A random selection of a function isolates the findings. It randomly chooses the values.

Local Outlier Factor (LOF): LOF is an unsupervised learning method model that calculates the local compactness variation of the data point with neighboring circumstances.

K-Means Clustering: This is an algorithm of unsupervised machine learning used for unlabelled data. It is a clustering-based learning approach.

B. Confusion Matrix

The output of a classifier is measured in detail by a confusion matrix. A confusion matrix provides a table of the various consequences of a classification problem forecast and results and helps to see the results. Table 1 presents the classification matrix for NN, AE, IF, LOF, and K- Means algorithms. We can see here that NN dominates other algorithms.

Approaches used	True Negative	False Positive	False Negative	True Positive
NN	56851	13	18	80
AE	55491	1373	20	78
IF	55789	1076	48	50
LOF	55091	773	52	46
K-Means	55734	130	13	85

Table 1: Confusion Matrix for various unsupervised learning techniques

C. Performance

We can perform various calculations using this matrix. The equations are as follows.

Precision is the ratio of the number of true positives to the sum of a genuinely positive and false positive. It can be said to be the measure of the quality of the positive feedback data. The equation for precision can be seen in Equation 1.

$$\text{Precision} = TP/TP+FP \quad (1)$$

The sensitivity, which is also called recall, is the measure of accurate optimistic predictions to the sum of a genuinely positive and false negative. The recall evaluates the completeness of the program, examining how many true positives were detected as positive. As seen in equation 2.

$$\text{Sensitivity (Recall)} = TP/TP+FN \quad (2)$$

Specificity is the ratio of true negative to the sum of true negative and false positive, as seen in equation 3.

Specificity = TN/TN+FP (3)

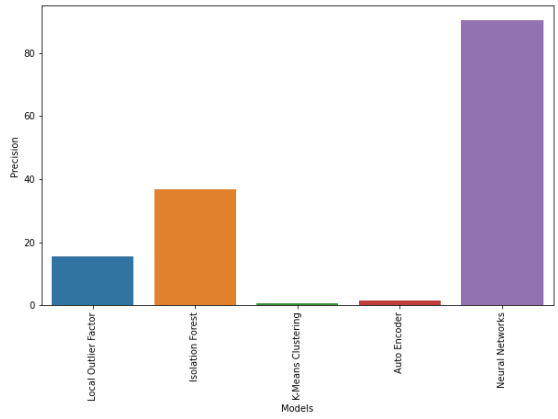
Accuracy is the ratio of the sum of true positive and true negative to the sum of all the predicted samples, as seen in equation 4.

Accuracy = TP+TN/TP+TN+FP+FN (4)

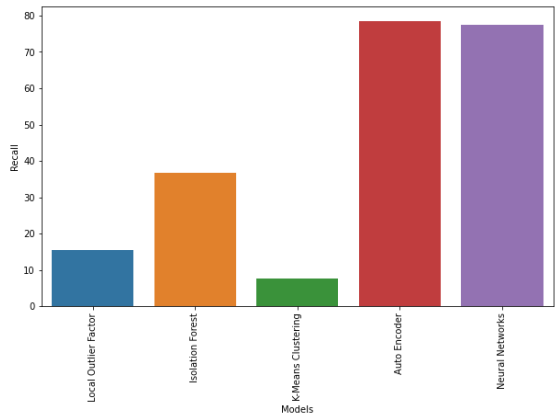
D. Results

Metrics	NN	AE	IF	LOF	K-MEANS
Precision	0.86	0.053	0.51	0.056	0.39
Recall	0.81	0.79	0.044	0.46	0.86
Specificity	0.99	0.97	0.83	0.98	0.090
Accuracy	0.99	0.97	0.98	0.98	0.99

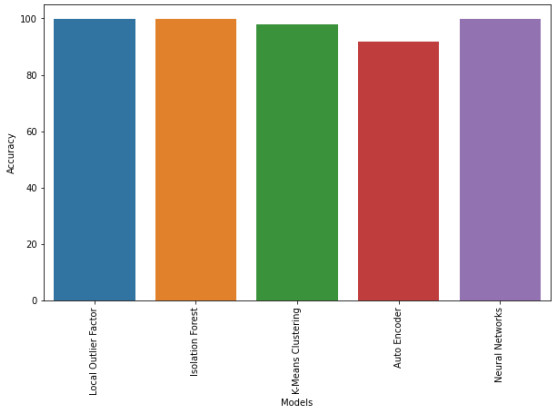
Table 2: Performance comparison



Comparison of Precision



Comparison of Recall



Comparison of Accuracy

Figure 5: Performance measurements of NN, AE, IF, LOF, and K-Means on different met rice

Fraud Detection algorithms are implemented in Python. The performance is compared in Table 2. Various performance analysis results of NN, AE, IF, LOF, K-Means clustering algorithms on different metrics.

V. CONCLUSION

We introduced a neural network-based fraud detection system for credit card data that uses unsupervised learning to detect fraud. The performance of the proposed work with the current systems is comparable. Neural network-based solution is different than the conventional methods. It can be observed. Experimental findings reveal that the suggested solution based on a neural network offers 99.98% accuracy while Auto Encoder, Isolation Forest, Local Outlier Factor, and K Accuracies are clustered to 97%, 98%, respectively, 98%, and 99.97%.

REFERENCES

[1]. Sánchez, D., M.A. Vila, L. Cerda, and J.M. Serrano. 2009. "Association Rules Applied To Credit Card Fraud Detection." Expert Systems With Applications 36 (2): 3630-3640. doi:10.1016/j.eswa.2008.02.001.

[2]. Dada, Emmanuel Gbenga, Joseph Stephen Basse, Haruna Chiroma, Shafii Muhammad Abdulhamid, Adebayo Olusola Adewunmi, and Opeyemi Emmanuel Ojibwa. "Machine Learning for Email Spam Filtering: Review, Approaches, and Open Research Problems." Heliyon 5, no. 6 (2019) doi:10.1016/j.heliyon.2019.e01802.

[3]. Zhang, Liangwei, Jing Lin, and Ramin Karim. 2016. "Sliding Window-Based Fault Detection From High-Dimensional Data Streams". IEEE Transactions On Systems, Man, And Cybernetics: Systems, 1-15. doi:10.1109/tsmc.2016.2585566.

[4]. Pumsirirat, Apapan, and Liu Yan. 2018. "Credit Card Fraud Detection Using Deep Learning Based On Auto-Encoder And Restricted Boltzmann Machine". International Journal Of Advanced Computer Science And Applications 9 (1). doi:10.14569/ijacsa.2018.090103.

[5]. "Kanna Velusamy | Alagappa University, Karaikudi, Tamil Nadu, INDIA - Academia.Edu". 2021. Alagappauniversity.Academia.Edu. https://alagappauniversity.academia.edu/KannaVelusamy.

[6]. "Credit Card Fraud Prediction Using Machine Learning".2021. Medium.https://pub.towardsai.net/credit-card-fraud-prediction-using-machinelearning-f47e52a0dbc2?gi=4f3899ec852a.

[7]. "UnsupervisedLearning".2021. En.M.Wikipedia.Org.https://en.m.wikipedia.org/wiki/Unsupervised_learning.

[8]. "Best 120 Data Science Interview Questions Pdf In 2020.". 2021. E-Learning Portal For IT Courses.https://svrtechnologies.com/120-data-science-interview-questions-pdf/.

[9]. Cai, Zi, and Jinguo Liu. 2018. "Approximating Quantum Many-Body Wave Functions Using Artificial Neural Networks". Physical Review B 97 (3). doi:10.1103/physrevb.97.035116.

[10]. Radovanovic, Milos, Alexandros Nanopoulos, and Mirjana Ivanovic. 2015. "Reverse Nearest Neighbors In Unsupervised Distance-Based Outlier Detection". IEEE Transactions On Knowledge And Data Engineering 27 (5): 1369-1382. doi:10.1109/tkde.2014.2365790.

[11]. Blakely, Tony, and Clare Salmond. 2002. "Probabilistic Record Linkage And A Method To Calculate The Positive Predictive Value". International Journal Of Epidemiology 31 (6): 1246-1252. doi:10.1093/ije/31.6.1246.

[12]. "Beyond Accuracy: Precision And Recall". 2021. Medium.https://towardsdatascience.com/beyond-accuracy-precision-and-recall-3da06bea9f6c?gi=abdf756a67c3.

[13]. Zamini, Mohamad, and Seyed Mohammad Hossein Hasheminejad. 2019. "A Comprehensive Survey Of Anomaly Detection In Banking, Wireless Sensor Networks, Social Networks, And Healthcare". Intelligent Decision Technologies 13 (2): 229-270. doi:10.3233/idt-170155.

[14]. Gu, Zhiwei, Shah Nazir, Cheng Hong, and Sulaiman Khan. 2020. "Convolution Neural Network-Based Higher Accurate Intrusion Identification System For The Network Security And Communication". Security And Communication Networks 2020: 1-10. doi:10.1155/2020/8830903.

[15]. West, Jarrod, and Maumita Bhattacharya. 2016. "Intelligent Financial Fraud Detection: A Comprehensive Review". Computers & Security 57: 47-66. doi:10.1016/j.cose.2015.09.005.

[16]. know, 11. 2021. "Evaluation Metrics Machine Learning". Analytics Vidhya. https://www.analyticsvidhya.com/blog/2019/08/11-important-model-evaluation-error-metrics/.

[17]. "Accuracy, Precision, And Recall In Deep Learning | Paperspace Blog". 2021. Paperspace Blog. https://blog.paperspace.com/deep-learning-metrics-precision-recall-accuracy/.

[18]. "Can Someone Help Me To Calculate Accuracy, Sensitivity,... Of A 6*6 Confusion Matrix?". 2021. Researchgate. https://www.researchgate.net/post/Can_someone_help_me_to_calculate_accuracy_sensitivity_of_a_66_confusion_matrix.

[19]. Rai, Arun Kumar, and Rajendra Kumar Dwivedi. "Fraud Detection in Credit Card Data Using Unsupervised Machine Learning Based Scheme." 2020 International Conference on Electronics and Sustainable Communication Systems(ICESC),2020.doi:10.1109/icesc48915.2020.9155615.

[20]. "Precision And Recall".2021. En.Wikipedia.Org.https://en.wikipedia.org/wiki/Precision_and_recall.

[21]. Zhang, Y. and Hamori, S., 2020. Forecasting Crude Oil Market Crashes Using Machine Learning Technologies. Energies, 13(10), p.2440.

[22]. Medium. 2021. The Case of Precision v. Recall. [online] Available at: <https://towardsdatascience.com/the-case-of-precision-v-recall-1d02fe0ac40f> [Accessed 19 March 2021].

[23]. Minallah, Nasru, et al. “On the Performance of Fusion Based Planet-Scope and Sentinel-2 Data for Crop Classification Using Inception Inspired Deep Convolutional Neural Network.” PLoS One, vol. 15, no. 9, Public Library of Science, Sept. 2020, p. e0239746.

[24]. Carcillo, Fabrizio, Yann-Aël Le Borgne, Olivier Caelen, Yacine Kessaci, Frédéric Oblé, and Gianluca Bontempi. 2021. "Combining Unsupervised And Supervised Learning In Credit Card Fraud Detection". Information Sciences 557: 317-331. doi:10.1016/j.ins.2019.05.042.

[25]. Sánchez, D., M.A. Vila, L. Cerda, and J.M. Serrano. 2009. "Association Rules Applied To Credit Card Fraud Detection". Expert Systems With Applications 36 (2): 3630-3640. doi:10.1016/j.eswa.2008.02.001.

[26]. L. Bhavya, V. Sasidhar Reddy, U. Anjali Mohan, and S. Karishma. 2020. "Credit Card Fraud Detection Using Classification, Unsupervised, Neural Networks Models". International Journal

Of Engineering Research And V9 (04). doi:10.17577/ijertv9is040749.

[27]. Renjith, Shini. 2018. "Detection Of Fraudulent Sellers In Online Marketplaces Using Support Vector Machine Approach". International Journal of Engineering Trends And Technology 57 (1): 48-53. doi:10.14445/22315381/ijett-v57p210.