An Efficient Techniques for Fraudulent detection in Credit Card Dataset: A Comprehensive study

Akanksha Bansal and Hitendra Garg
Department of Computer Engineering & Applications,
GLA University, Mathura (UP) India
{akanksha.bansal.mtcs19@gla.ac.in, hitendra.garg@gmail.com}

Abstract. Now a day, credit card transaction is one the famous mode for financial transaction. Increasing trends of financial transactions through credit cards also invite fraud activities that involve the loss of billions of dollars globally. It is also been observed that fraudulent transactions have increased by 35% from 2018. A huge amount of transaction data is available to analyze the fraud detection activities that require analysis of behavior/abnormalities in the transaction dataset to detect and ignore the undesirable action of the suspected person. The proposed paper lists a compressive summary of various techniques for the classification of fraud transactions from the various datasets to alert the user for such transactions.

1. Introduction:

Credit card fraud is relatively common among other fraudulent events and can happen in our daily lives. There are various ways credit card fraud can happen. Credit cards can be lost or stolen and then used by a thief [1]. Another way, credit card fraud can occur is that your identity might have been exposed to malicious persons who then use your credentials to access your account. Scammers can even use telephone phishing for credit card fraud [2]. As there are many ways credit card fraud can happen, many credit card holders are exposed to the risk of this type of fraud and have a proper way to prevent them from such happening. Many credit card companies have employed various measures to prevent and detect these types of fraudulent activities, using various Machine Learning (ML) and anomaly-detection technologies [3].

Globally, credit card frauds are the foremost common problem for every individual and financial organization that results from the loss of billions of money. It has been reported in various financial notes that online fraud is 24.26 billion dollars in 2018 that is gradual increase of approx 11%-18% ever year and still climbing [5] The various mode of fraud transaction exist either by stolen the physical card and their credentials or just generating a clone of the card / mobile phone or your device [6]. The non-secure payment portals...
where the customer usually performs online payment also save their payment credentials are one of the main causes of this fraud transactions. The worst part of this is it still no protective technological measures to prevent such transactions. The only way to prevent these to aware people of the phishing activities available over mails, websites, social media sites, and even on direct phone calls [7]. Parallely, these financial instructions are also analyzing the credit card transaction data to find the unusual of an anomaly in a transaction. Therefore, efficient, effective, and prominent technology may be needed to analyze and prevent such transactions. The proposed manuscript lists various Machine Learning (ML) and anomaly-detection technologies reported. The Machine Learning (ML) based techniques are used to classify apprehensive/inapprehensive transactions [8]. The credit card fraud activities can be categorized as [9]: Internal credit card fraud and External credit card fraud.

Internal credit card fraud occurs due to the permission of the cardholder and bank by using a false identity to execute fraud transaction whereas external fraud transaction occurs when card details and other personal credential are stolen by suspicious means [10]. Besides this categorization, there are other means of fraud transactions like the traditional method where application/account details stolen internet frauds where site cloning, card cloning, suspicious merchant sites are shown in Figure 1.

![Types of Credit Card Frauds](image)

**Figure-1:** Types of credit card fraud

### 1.1. Types of credit card fraud:

There are various types of credit card fraud which are as follows:

- **1.1.1. Application Fraud:** When the fraudster is well known and he access the control of the application of authorized user using sensitive user details like email address, user name, password etc. to access the account.

- **1.1.2. Card Not Present (CNP).** This situation arises when the fraudster knew the expiry date and account number of the credit card. So, the card can be used without its physical appearance of the account holder [12].

- **1.1.3. Card id Theft:** This theft is just similar to application fraud. In this the fraudster knew the details of the original card and opens the new account using its previous details [13].

- **1.1.4. Lost/Stolen Card:** This case happens when an account holder misplaces the credit card. It can
1.1.5. **Skimming:** Skimming or credit card forgery both are types of credit card theft where the fraudsters used a small device called a card’s magnetic tape. This device captures all the details of the cardholder and stores them in it.

1.1.6. **Electronic Card:** This electronic card skims the information on the magnetic stripe of the card. It stores the information, and the fraudster used this information whenever it is used in the future [14].

1.1.7. **Phishing:** Phishing attack is the meaning of false merchant sites. It is a type of victim that happened in social media where the customer’s data, login credentials, and credit card numbers are to be taken.

1.1.8. **Account Takeover:** This is the most common fraud. In this, the fraudster may access the account details of the cardholder as well as all the relevant documents related to it [15].

The transactional data of credit cards are primarily described by an uncommon phenomenon. The fraudster implements new ways of copying the profile of the legitimate user. So, there is very difficult to classify both of the transaction from the identified pool of credit card transactions [16].

### 1.2 Transaction Flow:

![Transaction Flow Diagram](image)

**Figure-2:** Transaction Flow

In most of the online shopping transactions, there are no requirements of the physical card, CVV number, and expiry date to perform the transaction. This information is usually taken by the fraudsters [17] [18]. The frauds are stolen from this valuable information of the customer. They placed the order and
used the customer account for the online fraudulent transaction. This transaction workflow station represents in Figure 2.

2. Dataset:
The credit card transaction datasets play an important role in identifying suspected transactions [19]. Such various datasets are available on Kaggle, UCI machine learning repository, etc. In this proposed paper, the dataset available on Kaggle is considered for analyzing the various parameters available as shown in Table-1. The dataset is too large and highly unbalanced.

Table 1: Various Parameters of the Datasets

| S.No | Dataset Attributes Name          | Explanation                                                  |
|------|----------------------------------|--------------------------------------------------------------|
| 1.   | Posting Date                     | Whenever our account is posted of the further date.          |
| 2.   | Merchant Category Code           | A code given for the respective goods.                       |
| 3.   | Transaction Date & Time          | The actual timing, day and date of the transaction.          |
| 4.   | Transaction Status               | The current status after transaction of the money.           |
| 5.   | Transaction Place                | From which place we debit/credit the money.                  |
| 6.   | Money Amount                     | The money amount which is to be transact via bank.           |
| 7.   | Transaction Type                 | What mode of payment for the transaction?                    |
| 8.   | Customer Identification          | Identification code to confirm the details                   |
| 9.   | Scheme                           | The type of credit card (VISA/MASTER)                        |

3. Machine Learning (ML) and Anomaly – Detection Technologies:

3.1. Data Analysis:
There are different dataset available for experimental purpose as discusses in section-2 having various transactions having some of them are fraudulent transaction and rest are normal transactions. The supervised and unsupervised techniques are used to classify the fraudulent transaction in order to determine the risk in credit card transaction [20-21].

In the proposed summary of credit card fraud detection, data visualization techniques like histograms and correlations matrix are better techniques for data analysis. Credit card fraud detection is an important application of outlier detection implemented in batch mode but these techniques are not so effective as for huge amount of dataset [22]. Therefore, PCA is used that requires low computation and low memory requirement. In the analysis, using Principal Component Analysis (PCA) and using Anomaly Detection Algorithm (ADA) are used. Principal Component Analysis (PCA) is the common method in detecting the frauds in various financial organizations, insurance companies and government agencies [23-25]. The occurrence of such transactions are comparatively less in the dataset of credit card transactions therefore, its detection is little bit difficult and tricky [46]. PCA is used to shrink the dimensionality of the dataset maintaining the correlation and variation present in the dataset. PCA transform the original dataset with new reduced dataset. Anomaly Detection Algorithm are used to identify the datasets with unusual patterns in the datasets. These anomalous have prospective of transforming into some problems like errors, structural defects and frauds. The Machine Learning (ML) techniques enhance the speed of classification of suspected transactions [26-28].
3.2. Machine Learning based Techniques:

Manual Inspection of dataset is difficult due to its volume, variety and velocity that mean data increase in volume rapidly. Therefore, machine learning based techniques are much efficient and efficient in such type cases [45]. Various classification techniques are in practices depending upon the nature of dataset [29].

3.1.1 Artificial Neural Network (ANN):

An artificial neural network is the combination of supervised, unsupervised as well as hybrid learning methods. ANN is a set of interconnected nodes which gives the functionality of the neural brain. Basically, the nodes having its corresponding weights, and its adjacent point of the neural network layer [29]. This model tells “an input having an activation function with its weights which finally gives the output”. In supervised technique, it must have both fraudulent and non-fraudulent records’ while; in unsupervised technique, there is no previous knowledge of the fraudulent and non-fraudulent records’ instead, the hybrid technique, including both supervised and unsupervised learning methods [30].

3.1.2 Genetic Algorithm (GA)-

Genetic algorithms are the methods of searching which is based on some rules, principles of natural selection and genetics. Firstly, encode the decision and then the finite length strings of alphabet is taken for searching the variables based on the decisions. Here, the strings called as “chromosomes”, the alphabets called as “genes” and the values of genes are called as “alleles” [31] [32].

3.1.3 Random Forest (RF)-

Random forest is a supervised learning algorithm. It produced a predictive model that solves both classification and regression problems. This algorithm works very fast and it handles the unbalanced database efficiently. We call this as a forest because this forest taking more number of decision trees. This random forest combines the output of the multiple decision trees to generate the best output. This model gives more accurate results [33] [34].

3.1.4 Support Vector Machine (SVM)-

Support Vector Machine is a supervised machine learning algorithm that can be used for both classification and regression problems [35]. This model having one distribution of line that called as hyperplane. This hyperplane divides the separation areas in N-dimensional space of the given data points. For training, the machine model we fed the supervised data to check the closest points of the known results [36]. So, it learns the behavior of the transaction either genuine or fraud transactions and then it classifies on which class it belongs. It classifies the pattern into fraudulent and non-fraudulent [37].

3.1.5 Bayesian Network-

It is a graphical model that represented the way taking the random variables so, that we have the right decisions which is based on some conditional values [38]. It is an acyclic graph having the known probabilities in the presence of either uncertainty or certainty cases regarding their problems. At last, we measure the maximum and minimum value of the transaction whether it is legal or fraud [39].

3.1.6 Decision Tree (DT)-

Decision tree is the most powerful and popular tool for classification. This is the best partition algorithm because it does not overlap the values and never disjoint to each other [40]. This algorithm is shown by the representation of the flowchart structure where the branch shown by the output of the given test, its internal nodes shown by the attributes, leaf shown by the labeled class. So, the representation of decision tree must be shown by the diagrammatic form and reveals the output of it [41].
3.1.7: Logistic Regression (LR)-

The logistic regression used for classification problems of predicting the multi-varied value outcomes. It is used for clustering process and it examines the attributed value and check our transaction is further proceeding or not. It is used for binary classification which based on statistically model [42].

3.1.8 K-Nearest Neighbor (KNN)-

K-Nearest Neighbor is a supervised machine learning technique. This algorithm is used for both classification process and regression model to predict the desired problems. Basically, it is based on three factors that is, distance metrics, distance rules and the value of K [43]. The value of K should be used for all predictions to give the optimal value with respect to its corresponding task then we calculate its dominant class for any new transactions [44].

Table 2: Analysis of Credit Card Fraud Detection Techniques.

| S.No | Reference | Methodology & Tools | Dataset | Merits | Demerits | Accuracy |
|------|-----------|---------------------|---------|--------|----------|----------|
| 1    | [29][30] | Artificial neural network (ANN) | UCI website repository. | Better result is achieved. Neural networks are sensitive to noise. | Take large amount of time to train the model. Neural network is very challenging task. | It gives a good accuracy. 92% correct fraud cases. |
| 2    | [31][32] | Genetic Algorithm | - | Better performance. Efficient and secure | Fail to identify the self-transaction. | 75% accuracy. |
| 3    | [33][34] | Random Forest (RF) | Chinese e-commerce company website. | It gives the best results to detect the credit card fraud transaction. | The dataset is fully imbalanced. | 96.77% accuracy. |
| 4    | [35-37] | Support Vector Machine (SVM) | Financial institution dataset. | It detects the real time credit card fraud. | This algorithm takes lot of time and a detailed study. | Best results. 91% accuracy |
| 5    | [38][39] | Bayesian | European company | Good performance. | Requires balancing data. | Good accuracy. 75% accuracy (approx.) |
| 6    | [40][41] | Decision Tree (DT) | European cardholder | This algorithm shows the lower mean square value. | DT is inaccurate. | 75% accuracy. |
| 7    | [42] | Logistic Regression (LR) | European cardholder | This approach gives better | It requires more time. | Less accurate. 50% accuracy. |
4. Evaluation:

There are numbers of measures for evaluating different assessments by means of False Positive (FP), False Negative (FN), True Positive (TP), True Negative (TN). \( FP \) indicates the non-fraudulent transactions falsely being classified as actual fraudulent transactions. \( FN \) indicates the fraudulent transactions falsely being classified as normal transactions, \( TP \) indicates the fraudulent transactions rightly being classified as actual fraudulent transactions and \( TN \) indicates the normal transactions rightly being classified as the normal transactions.

Based on this accuracy, precision, sensitivity, specificity, False Positive Rate, ROC, Cost and F1 measure is calculated. The brief summary of these parameters is shown in Table-3.

Table 3: Definitions of various evaluation metrics

| Metrics                          | Notation                  | Explanation                                                                 |
|---------------------------------|---------------------------|----------------------------------------------------------------------------|
| Accuracy/Decision Rate          | \( \frac{TN+TP}{TP+FP+FN+TN} \) | Accuracy is the truly or correctly classified metric which measures the classification performance. |
| Precision/Hit Rate              | \( \frac{TP}{TP+FP} \)    | Precision gives the number of positive rate metrics to classify as the actual fraudulent transactions. |
| True Positive Rate/Sensitivity   | \( \frac{FP}{TP+FN} \)    | TPR gives the number of positive rate metrics to classify as abnormal instances. It is also called as sensitivity measure. |
| True Negative Rate/ Specificity | \( \frac{TN}{TN+FP} \)    | TNR gives the number of negative rate metrics to classify as actual normal transactions. It is also called as specificity measures. |
| False Positive Rate (FPR)       | \( \frac{FP}{FP+TN} \)    | Ratio of detection the fraudulent incorrectly. |
| ROC                             | Curve between TP&FP       | ROC stands Relative Operating Characteristic Curve, comparison of TPR & FPR. |
| Cost                            | \( 100*FN+10*(FP+TP) \)   | Cost is the function measuring the cost of all |
The evaluation is also done by the confusion matrix. The table 4 shows the 2*2 confusion matrix. This matrix is based on either fraudulent label or genuine label.

|                  | Fraudulent Label | Genuine Label |
|------------------|------------------|---------------|
| Fraudulent Label | True Positive    | True Negative |
| Genuine Label    | False Positive   | False Negative |

5. Conclusion

In the last decades, online transactions are growing rapidly and the most common tool for financial transactions. The increasing growth of online transactions also increases threats. Therefore, in keeping in mind the security issue, nature, an anomaly in the credit card transaction, the proposed work represents the summary of various strategies applied to identify the abnormal transaction in the dataset of credit card transaction datasets. This dataset contains a mix of normal and fraud transactions; this proposed work classifies and summarizes the various classification methods to classify the transactions using various Machine Learning-based classifiers. The efficiency of the method depends on the dataset and classifier used. The proposed summary will be beneficial to the banker, credit card user, and researcher to analyze to prevent credit card frauds. The future scope of this credit card fraud detection is to explore the things in each and every associations and banks to live safe and happily life. The data must be balanced in each place and we are getting the best results.

References

[1] Zheng, Lutao, et al. "Transaction fraud detection based on total order relation and behavior diversity." IEEE Transactions on Computational Social Systems 5.3 (2018): 796-806.
[2] Zojaji, Zahra, Reza Ebrahimi Atani, and Amir Hassan Monadjemi. "A survey of credit card fraud detection techniques: data and technique oriented perspective." arXiv preprint arXiv:1611.06439 (2016).
[3] Bhattacharyya, Siddhartha, et al. "Data mining for credit card fraud: A comparative study." Decision Support Systems 50.3 (2011): 602-613.
[4] Delamaire, Linda, Hussein Abdou, and John Pointon. "Credit card fraud and detection techniques: a review." Banks and Bank systems 4.2 (2009): 57-68.
[5] Gupta, Gaurav. "Analysis of Various Credit Card Fraud Detection Techniques." (2019).
[6] Dutta, Sharmistha, Ankit Kumar Gupta, and Neetu Narayan. "Identity Crime Detection Using Data Mining." 2017 3rd International Conference on Computational Intelligence and Networks (CINE). IEEE, 2017.
[7] Jain, Rajni, Bhupesh Gour, and Surendra Dubey. "A hybrid approach for credit card fraud detection using rough set and decision tree technique." International Journal of Computer Applications 139.10 (2016): 1-6.

[8] Suman, Mitali Bansal, and M. Bansal. "Survey Paper on Credit Card Fraud Detection." International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) 3.3 (2014): 827-832.

[9] Mantri, Arpita, Chelsi Sen, and Sunil Kumar. "An Overview of Credit Card Fraud Detection Using Data Mining Techniques."

[10] Taha, Altyeb Altaher, and Sharaf Jameel Malebary. "An Intelligent Approach to Credit Card Fraud Detection Using an Optimized Light Gradient Boosting Machine." IEEE Access 8 (2020): 25579-25587.

[11] Patel, Yogesh, et al. "Remote banking fraud detection framework using sequence learners." Journal of Internet Banking and Commerce 24.1 (2019): 1-31.

[12] Kumari, Priyanka, and Smita Prava Mishra. "Analysis of credit card fraud detection using fusion classifiers." Computational Intelligence in Data Mining. Springer, Singapore, 2019. 111-122.

[13] Niu, Xuetong, Li Wang, and Xulei Yang. "A comparison study of credit card fraud detection: Supervised versus unsupervised." arXiv preprint arXiv:1904.10604 (2019).

[14] Jain, Yashvi, Shripriya Dubey Namrata Tiwari, and Sarika Jain. "A Comparative Analysis of Various Credit Card Fraud Detection Techniques." International Journal of Recent Technology and Engineering (2277-3878) 7.5S2 (2019): 402-407.

[15] Kumari, Priyanka, and Smita Prava Mishra. "Analysis of credit card fraud detection using fusion classifiers." Computational Intelligence in Data Mining. Springer, Singapore, 2019. 111-122.

[16] Niu, Xuetong, Li Wang, and Xulei Yang. "A comparison study of credit card fraud detection: Supervised versus unsupervised." arXiv preprint arXiv:1904.10604 (2019).

[17] Vimala, S., and K. C. Sharmili. "Survey Paper for Credit Card Fraud Detection Using Data Mining Techniques."

[18] Patel, Yogesh, et al. "Remote banking fraud detection framework using sequence learners." Journal of Internet Banking and Commerce 24.1 (2019): 1-31.

[19] Shirodkar, Nikita, et al. "Credit Card Fraud Detection Techniques–A Survey." 2020 International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE). IEEE, 2020.

[20] Mantri, Arpita, Chelsi Sen, and Sunil Kumar. "An Overview of Credit Card Fraud Detection Using Data Mining Techniques."

[21] Iranmehr, Arya, Hamed Masnadi-Shirazi, and Nuno Vasconcelos. "Cost-sensitive support vector machines." Neurocomputing 343 (2019): 50-64.

[22] Sasank, JVV Sriram, et al. "Credit Card Fraud Detection Using Various Classification and Sampling Techniques: A Comparative Study." 2019 International Conference on Communication and Electronics Systems (ICCES). IEEE, 2019.

[23] Zhou, Hua, et al. "Credit Card Fraud Identification Based on Principal Component Analysis and Improved Adaboost Algorithm." 2019 International Conference on Intelligent Computing, Automation and Systems (ICICAS). IEEE, 2019.

[24] Das, Tamojit. "Credit Card Fraud Detection." Methodology 9.11 (2019): 1-4.

[25] John, Hyder, and Sameena Naaz. "Credit card fraud detection using local outlier factor and isolation forest." Int. J. Comput. Sci. Eng. 7 (2019): 1060-1064.

[26] Shukur, Hamzah Ali, and Sefer Kurnaz. "Credit Card Fraud Detection Using Machine Learning Methodology." Int. J. Comput. Sci. Mob. Comput. 8 (2019): 257-260.

[27] Sailusha, Ruttala, et al. "Credit Card Fraud Detection Using Machine Learning." 2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS). IEEE, 2020.
[28] Shirgave, Suresh K., et al. "A Review On Credit Card Fraud Detection Using Machine Learning."

[29] Niu, Xuetong, Li Wang, and Xulei Yang. "A comparison study of credit card fraud detection: Supervised versus unsupervised." arXiv preprint arXiv:1904.10604 (2019).

[30] Saraswathi, E., et al. "Credit Card Fraud Prediction And Detection using Artificial Neural Network And Self-Organizing Maps." 2019 3rd International Conference on Computing Methodologies and Communication (ICCMC). IEEE, 2019.

[31] Alotaibi, Asmaa, et al. "Identifying Credit Card Fraud Using Genetic and Clonal Selection Algorithms."

[32] Olabode, Anthony Onaolapo. "PERFORMANCE EVALUATION OF GENETIC ALGORITHM AND COUNTER PROPAGATION NEURAL NETWORK FOR ANOMALY DETECTION IN ONLINE TRANSACTION." Annals. Computer Science Series 17.2 (2019).

[33] Kumar, M. Suresh, et al. "Credit card fraud detection using random forest algorithm." 2019 3rd International Conference on Computing and Communications Technologies (ICCCT). IEEE, 2019.

[34] Xuan, Shiyang, et al. "Random forest for credit card fraud detection." 2018 IEEE 15th International Conference on Networking, Sensing and Control (ICNSC). IEEE, 2018.

[35] Maes, Sam, et al. "Credit card fraud detection using Bayesian and neural networks." Proceedings of the 1st international naiso congress on neuro fuzzy technologies. Springer, Cham, 2019.

[36] Dighe, Deepti, Sneha Patil, and Shrikant Kokate. "Detection of credit card fraud transactions using machine learning algorithms and neural networks: A comparative study." 2018 Fourth International Conference on Computing Communication Control and Automation (ICCCUBE4). IEEE, 2018.

[37] Eweoya, I. O., et al. "Fraud prediction in bank loan administration using decision tree." Journal of Physics: Conference Series. Vol. 1299. No. 1. IOP Publishing, 2019.

[38] Gowthami, K., et al. "CREDIT CARD FRAUD DETECTION USING LOGISTIC REGRESSION."