INTERNATIONAL JOURNAL OF TECHO-ENGINEERING

ISSN : 0975-4520

A SURVEY ON MACHINE LEARNING ALGORITHMS IN CREDIT CARD FRAUD DETECTION

[1]Channapragada Chinmayi Sree Chitra, Associate Software Engineer, Optum Global Solutions, Hyderabad, ccsreechitra@gmail.com

[2]C. Rama Seshagiri Rao, Professor, CSE Department, Vignana Bharathi Engineering, College, Hyderabad, crsgrao@gmail.com

[3]Dr.Yeligeti Raju, Assoc.Professor in IT, Vignana Bharathi Institute of Technology, raju.yeligeti@gmail.com

Abstract: Credit card plays a fundamental rule in today's marketplace. Various machine learning methods have been introduced to face credit card fraud increases. However, all of these methods have the same goal of avoiding credit card fraud; everyone has its disadvantages, advantages, and features. The main aim is to defend credit card transactions; so people can use e-banking securely and efficiently. Financial fraud is growing significantly with the advancement of new technology and the global superhighways of communication, following in the loss of billions of dollars worldwide every year. In recent years, various deep learning techniques implemented to detect fraud in credit card transactions, which are Logistic Regression (LR), Naïve Bayesian(NB), Support Vector Machine (SVM), Neural Network (NN), Artificial Immune System(AIS), K Nearest Neighbour(KNN), Decision Tree (DT), Fuzzy logic based System, Genetic Algorithm, etc. This paper provides a brief survey of various techniques used in credit card fraud detection and estimates every methodology based on particular design criteria.

Keywords: Credit card fraud detection, fraud transaction, deep learning techniques.

1. Introduction

Credit card fraud can be defined as the unlawful use of any system or criminal activity by using the physical card or card information without the cardholder's knowledge. A credit card is a small plastic card issued to the user as a payment system. With the rapid growth in credit card transactions, fraudulent activities are also increasing. The credit card can be physical or virtual [1]. On a physical card, the cardholder physically presents his card to the merchant to make the payment. To make fraudulent transactions in this type of purchase, the attacker must steal the credit card. In the second type of investment, only certain important

ISSN : 0975-4520

information about the menu, such as card number, expiration date, security code, etc., is required to make the payment. These purchases are usually made online or over the phone. The fraudster simply needs to know the card details to commit fraud on these investments. Most of the time, the original cardholder doesn't know that someone else has seen or stole their card information. In real life, fraudulent transactions are familiar with real trades and simple pattern matching.

Often the methods are not sufficient to accurately detect such scams. External disclosure is a commonly used data extraction technique for fraud detection [2]. Outliers are data points that do not align with the data set call or deviate from other observations that they raise suspicion that a different mechanism generated them. External detection can be achieved by technologies such as neural networks, SOM, HMM, etc.

Because of the financial fraud involving credit card transactions, both merchants and buyers suffer financial losses [3]. It is a critical topic. To solve this problem, banks and card manufacturers pay a high cost. Online shopping and payment services make electronic payment convenient, smooth, convenient, comfortable, and user-friendly. Still, we cannot ignore the economic losses that also increase with e-commerce—calling criminals for a new type of fraud.

There are two ways to conduct credit card transactions: physically and physically, i.e., CNP (card not present). Physically, the card is required to make a pass. While you are on the virtual card, there are some details to pass a card like CVV number, cardholder name, password, security question, etc., for online banking [4].

Both fraud prevention and detection are one way to deal with fraud. In preventing fraud, the main objective is to prevent fraudulent activity; it detects the transaction and blocks authorized transactions. In fraud detection, the goal is to distinguish a fraudulent transaction from legitimate transactions. Through historical data, the user's pattern and behaviour are used to verify and verify whether a transaction is fraudulent or not. Sometimes the system fails to prevent fraudulent activities, and at that moment, fraud detection takes the lead.

2. Literature Survey

Many techniques exist based on analytical and computational. Those techniques were implemented to detect credit card fraud.

ISSN : 0975-4520

Fatima Zohra et al.[2020] Due to the increasing digitization of banking offerings and mobile banking software's dominance, the rate of credit card bills is growing every 12 months, among the billions of transactions recognized as fraudulent. Data mining algorithms have played a fundamental role in detecting fraudulent transactions and fighting attacks by fraudsters who work around traditional fraud prevention systems. This article attempts to detect fraudulent transactions using the artificial neurotransmitter classifiers, Multilayer Perceptron (MLP) and Extreme Learning Machine (ELM), performed on a credit card fraud dataset. These classifiers' overall performance is evaluated based on accuracy, computation, accuracy, and class time. The results showed that the MLP and ELM classifiers' accuracy reached 97.84% and 95.46%. Otherwise, ELM can be very fast to predict new fraudulent transactions.

Petr Mrozek et al. [2020] the rapid rise in e-commerce and online shopping has resulted in an unprecedented surge in the amount of money credit card fraudsters lose annually. Researchers are leveraging the utility of various devices to learn strategies for effectively detecting and stopping fraudulent credit card transactions from dealing with credit card fraud. A typical general problem surrounding the analysis of credit card transactions is the somewhat unbalanced data sets' nature, often associated with binary issues. This article aims to study, examine, and implement a branch of excellent algorithm learning tools, including Logistic Regression, Random Forest, K-Nearest Neighbours, and Stochastic Gradient Descent. With the incentive to empirically evaluate their competencies in manipulating unbalanced data sets and, at the same time, detecting fraudulent credit card transactions. A publicly available dataset of 284,807 European cardholder transactions is analysed and trained using well-thought-out device mastery techniques to identify fraudulent transactions. This article also evaluates the incorporation of brilliant re-modelling techniques, specifically the methods of random sub-sampling and synthetic majority over-sampling (SMOTE) techniques within the algorithms mentioned above, to examine their efficiency in dealing with unbalanced data sets. . The proposed reconfiguration methods significantly increased the detection capacity, and the most successful way of combining random forest and random resampling made the consideration score of 100% in contrast to the 77% retrieval score of the model without the resampling technique.

Pawan Kumar et al. [2019] Due to the rapid growth of the network, purchasing products online is an essential part of everyone's lifestyle, and most of the time, MasterCard is

ISSN : 0975-4520

contracted to pay for goods online. It is sincere thanks for the research, and humans get their required product for a visual display unit or realistic cell phone. As for buying online, the use of MasterCard will go up. However, there may be some loopholes in the internet search engine leading to internet or credit card fraud. Therefore, fraud detection structures have become vital for all MasterCard delivery banks to reduce their losses. The main fraud detection strategies commonly used are the neural network (NN), bases induction techniques, fuzz systems, call trees, support vector machines (SVM), logistic regression, local outlier factor (LOF), closest K neighbors, and algorithms genetic. These strategies are often used alone or in conjunction with co-abuse or meta-study strategies for the classifiers' work. This document provides an examination of the many methods applied in fraud detection at MasterCard and evaluates each criterion associated with technical support.

Deepti Dighe et al. [2018] online transactions in everyday life have been increasing since the end of the decade due to advancements in technology and community communication. Due to the efficiency, simplicity, and ease of using a network transaction device, new customers are continually becoming members of the prevailing community that benefits from such a device. Credit card fraud due to the misuse of the system is defined as the theft or misuse of credit card records used for personal gain without permission from the cardholder. It is essential to test consumer usage patterns in past transactions to hit such scams. By comparing the usage pattern with the current transaction, we will classify it as either a fraud or a valid transaction. In this article, the technologies used are KNN, Naïve Bayes, Logistic Regression, Chebyshev Functional Link Artificial Neural Network (CFLANN), Multilayer Perceptron, and Decision Trees that are evaluated based on their scores assessed in terms of several accuracy measures.

M. Seera et al. [2017] Transcranial Doppler (TCD) is a reliable method with the advantage of being non-invasive to diagnose cerebral vascular disease is through measurements of blood flow rate in cerebral arterial segments. In this study, a repetitive neural network (RNN) was used to classify the TCD signals captured from the mind. A total of 35 real and anonymous counts of affected people were compiled, and a series of trials were being run to diagnose stenosis. The capabilities extracted from TCD signals are used for classification using some forms of RNN with repetitive reactions. In addition to the person's RNN scores, a version of the RNN cluster is formed where most people vote using a casting approach to combine RNN predictions of a character into one combined prediction. The effects, consisting of quotes of accuracy, sensitivity, specificity, and place under the receiver operating characteristic curve,

ISSN : 0975-4520

were compared with those of the Random Forest Ensemble. The results positively indicate the RNN group's utility as a robust approach to detecting and classifying changes in blood flow velocity due to brain problems.

Bahnsen et al. [2016] every year billions of Euros are lost globally because of credit score card fraud. Thus, forcing economic institutions to enhance their fraud detection structures continuously. In current years, numerous research has proposed using machine mastering and statistics mining techniques to address this problem. However, the maximum study used some misclassification degree to evaluate the one of a kind answer and no longer consider the actual economic charges related to the fraud detection technique. Moreover, when constructing a credit score card fraud detection version, its miles essential to extract the transactional facts' useful features. This is normally done via aggregating the transactions to observe the clients' spending behavioural patterns. In this paper, we expand the transaction aggregation approach and recommend creating a brand new set of functions based on analysing the periodic conduct of the time of a transaction using the von Mises distribution. Then, using an actual credit card fraud dataset furnished by using a sizeable European card processing employer, we examine contemporary credit card fraud detection models and compare how the exclusive sets of capabilities affect the consequences. Together with the proposed periodic functions into the methods, the results show a median boom in savings of 13%.

Mahmoudi et al. [2015] Parallel to the boom in credit card transactions, the economic losses resulting from fraud expanded. Therefore, credit card fraud detection has doubled in popularity for both academics and banks. Several supervised mastery techniques have been incorporated into the credit card fraud literature, and some have very complex algorithms. Compared to the complex algorithms that can overwhelm the data set on which they are generated, you will have less complex algorithms that can also show better overall performance on various data sets. Although the linear discriminant characteristics are less complicated classifiers and may reflect larger issues such as credit card fraud detection, they have not received much attention yet. This test investigates linear discrimination, called Fisher's Discriminant Function for the first time in credit card fraud detection hassles.

On the other hand, in this field and some different areas, the price of false negatives can be much higher than the cost of false positives, and it varies for each transaction. Therefore, it is essential to increase writing styles biased towards the most important examples. A modified

ISSN : 0975-4520

Fischer discriminant function is proposed that makes the classical characteristic more sensitive to critical times to address this issue in this study. This way, the profits that can be made from a legitimate/fraudulent sorter are maximized. The experimental results confirm that a modified Fisher characteristic can generate higher gains.

Halvaiee et al. [2014] the volume of online transactions these days is evolving into an enormous variety. A large component of these transactions includes credit card transactions. On the other hand, the growth of online fraud is stellar, usually the result of everyone's easy access to parts technology. Studies have been conducted on several trends and strategies to prevent and detect credit card fraud. An artificial immune system is one of them. However, agencies want accuracy alongside the tempo within their fraud detection systems, which are not always fully received. This document addresses credit card fraud detection using Artificial Immune Systems (AIS) and introduces a new model known as the AIS Based Fraud Detection Model (AFDM). They were useful and enhanced the Immune System Stimulation Rules (AIRS) suite to detect fraud. Increased accuracy by up to 25%, decreased value by up to 85%, and reduced device response time by up to 40% compared to the lesser rule set.

3. Various machine learning methods for credit card fraud detection system

3.1 Artificial Neural Networks (ANNs)

An Artificial Neural Network simulates the way the human brain works. It consists of a series of nodes, called neurons, and the edges that connect neurons. Neurons are computational units that process some input data and provide some output. The output of one neuron is delivered as input to any other cell. A neuron in a human brain is activated if the acquired signal is merely strong enough. Likewise, artificial neurons not only receive a few input stimuli, but they also receive a weight that determines whether the input signals are strong enough or not. If the signs are vital enough, the playback feature will start to supply the output.

3.2 Artificial immune system (AIS)

The Artificial immune system is a highly complex system consisting of a complex network of specialized tissues, organs, cells, and chemical molecules. These elements are interconnected and work in a coordinated and specific way when they recognize and remember and eliminate the foreign cells that cause disease. Anything that the immune system can recognize is called

ISSN : 0975-4520

an antigen. Immune system detection devices are the antibodies that can recognize and destroy harmful and dangerous antigens.

The immune system consists of two primary immune and defense responses: the innate immune response and the acquired immune response. The body's first defense response consists of the healthy outer skin and the "mucous membranes" that line internal ducts such as the respiratory and digestive systems. If harmful cells penetrate the innate immune defence, the acquired immunity will defend itself.

3.3 Support Vector Machines (SVMs)

Support Vector Machines is a binary classification methodology. This is how an input sample can be classified into possible exercises. It is suitable for credit card fraud detection because only two classes are required, in particular, "legitimate" and "fraudulent." SVM tries to calculate the most desirable superscript level if you want to separate samples from the two lessons. Several hyper planes can perform this process, but a more satisfying super-level can increase the two classes' samples' margins. The blue and black bullets corresponded to samples from two different training sessions. Support vectors draw each elegance's boundaries by trading on the model closest to the hyper level. The dielectric superscript is located within the main auxiliary vectors, which increases the margin between them. A new pattern is classified with the help of super-plane distance measurement.

3.4 Bayesian Belief Networks (BBNs)

For the purpose of fraud detection, two Bayesian networks to describes the configured consumer behaviour. First, the Bayesian community of behaviour modelling is built on the belief that the consumer is a fraudulent (F) and any other version under the notion that the person is legitimate (NF). The 'fraud net 'was created with the help of experts. The 'user net' is created through non-fraudulent customer records. During operation, the consumer grid is specially designed for a specific person based on the increased records. By introducing the test into these networks and spreading it through the community, the probability that the measurement x is smaller than the two hypotheses mentioned above is obtained.

3.5 Decision Trees (DTs)

The decision tree is a machine learning method that can solve classification and regression problems. Each internal node represents a feature check in the decision tree, each section

ISSN : 0975-4520

indicates the test result, and the leaf nodes indicate instructions or effects. Decision trees are based on input records because due to their computational complexity and sequential nature, any small surrogate can affect the tree structure.

3.6 Hidden Markov Model (HMM)

A Hidden Markov Model is a double built-in random procedure generates more complex random strategies than the traditional Markov model. If the incoming credit card transaction is not typical through the learned Hidden Markov Model with a sufficiently high probability, it is considered fraudulent transactions. HMM [AKS08] Baum Welch algorithm released for teaching purposes and K-path algorithm for clustering. HMM, stores stats in groups based on three tariff scores: low, medium, and high. It determined the potential of the initial batch of transactions, and the FDS examined whether the transaction was genuine or fraudulent. Because HMM keeps a transaction log, it reduces the staff's hard work but produces a high false alarm and an excess of pleasant error. The initial selection of parameters affects the performance of this set of rules, and as a result, it should be chosen with caution. It's like taking into account the unique case of a completely connected HMM where it can access every country in the model in one step from each different kingdom.

3.7 Naïve Bayes algorithm

To solve the classification problems. The method works well with a small set of training data to estimate the required parameters for purpose of classification. It applies Bayes theorem to perform classification by calculating the probability of true class.

3.8 Logistic Regression

Logistic regression is an algorithm for classification to classify the data as different results. Logistic regression is used to develop a regression model when the dependent variable is categorical. It was created in 1958 by David Cox. There are three types of logistic regression: (1) binary, for a binary response variable, (2) polynomial - where the dependent variable contains more than two unordered classes, and (3) ordinal - when the studies are ordered.

3.9 K-Nearest Neighbour (K-NN)

The concept of K-nearest neighbour analysis has been used in several defect detection techniques. One of the best classifier algorithms used in credit card fraud detection is the k-nearest neighbour algorithm. In this moderated learning algorithm, the new instance query

ISSN : 0975-4520

results are ranked based on the majority of the K-Nearest Neighbour. It was first introduced by Aha, Kibler, and Alber.

The performance of KNN algorithm is influenced by three main factors:

• The distance metric used to locate the nearest neighbours.

• The distance rule used to derive a classification from k-nearest neighbour.

• The number of neighbours used to classify the new sample.

Among the various credit card fraud detection methods to identify supervised statistical patterns, the KNN achieves a consistent high performance, with no preconceived assumptions about the distributions from which training examples are drawn. K- Nearest neighbour credit card fraud detection techniques require a specific distance or similar procedure between two data instances.

3.10 Gradient Boosting Machines

The algorithm of Gradient Boosting Machines refers to optimize a cost function through function space by iteratively choosing a function which selects the negative gradient direction.

Techniques	Advantages	Disadvantages
Artificial neural network (ANN)	• Ability to learn from the past/lack of need to be reprogrammed • High accuracy • high speed in detection	• Difficulty to confirm the structure • High processing time for large neural networks
Artificial Immune System (AIS)	• High capability in pattern recognition • Self-organization • Multi layered/ has diversity	• Need high training time in NSA • Poor in handle missing data in ClonalG and NSA
Genetic Algorithm	• Works well with noisy data • Easy in build and operate • Fast in detection	• Requires extensive tool knowledge to set up • Difficult to understand

ISSN : 0975-4520

Hidden Markov Model	• Fast in detection	• Highly expensive • Low accuracy • Not scalable to large size data sets
SupportVector Machine (SVM)	• SVMs deliver a unique solution, since the optimality problem is convex • SVMs can be robust, even when the training sample has some bias	• Poor in process large dataset • Medium accuracy • lack of transparency of results
Naïve Bayes (NB)	• High processing and detection speed • High accuracy	• Excessive training need • Expensive
Decision Tree (DT)	• High flexibility • Good haleness • Easy to implement	• Requirements to check each condition one by one • In fraud detection condition is Transaction.
K-nearest neighbour	• Speed of detection is good	• Accuracy is medium • Expensive

Limitations observed in literature survey

- Imbalanced data
- Different misclassification importance
- Overlapping data
- Lack of adaptability
- Fraud detection cost
- Lack of standard metrics

4. CONCLUSION

Credit card fraud detection is an interesting domain. We analysed machine learning as best compared to prediction, clustering, and outlier discovery from the survey. This paper provided a brief study of credit card fraud detection and various machine learning algorithms to detect credit card fraud. The literature study summarized that the large volumes of data in the Credit Card Fraud system are growing. In turn, increases the need for advanced and efficient techniques for analysis. Also, defined limitations in current methodologies to detect fraud in credit card. Finally, discussed the advantages and disadvantages of various methods. In the future, we need to consult a better methodology to detect credit card fraud.

5. REFERENCES

1. Anshul Singh, Devesh Narayan "A Survey on Hidden Markov Model for Credit Card Fraud Detection". International Journal of Engineering and Advanced Technology (IJEAT), (2012). Volume-1, Issue-3; (49-52).

2. E.W.T. Ngai, Yong Hu, Y.H. Wong, Yijun Chen, Xin Sun "The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature". Elsevier-Decision Support Systems (2011). 50; (559–569).

3. A. Vellidoa, P.J.G. Lisboaa, J. Vaughan "Neural networks in business: a survey of applications". Elsevier, Expert Systems with Applications, (1999). 17; (51–70).

4. M. Zareapoor, K. Seeja, M. Alam, "Analysis of credit cardfraud detection techniques: based on certain design criteria", International Journal Compututer Application, 2012, pp. 35–42.

5. Fatima Zohra El hlouli, Jamal Riffi, Mohamed Adnane Mahraz, Ali El Yahyaouy, Hamid Tairi, "Credit Card Fraud Detection Based on Multilayer Perceptron and Extreme Learning Machine Architectures", Intelligent Systems and Computer Vision (ISCV) 2020 International Conference on, pp. 1-5, 2020.

6. Petr Mrozek, John Panneerselvam, Ovidiu Bagdasar, "Efficient Resampling for Fraud Detection During Anonymised Credit Card Transactions with Unbalanced Datasets", Utility and Cloud Computing (UCC) 2020 IEEE/ACM 13th International Conference on, pp. 426-433, 2020.

7. Pawan Kumar, Fahad Iqbal, 2019, "Credit Card Fraud Identification Using Machine Learning Approaches", pp.1-4.

ISSN : 0975-4520

8. Deepti Dighe, Sneha Patil, Shrikant Kokate, 2018, “Detection of Credit Card Fraud Transactions Using Machine Learning Algorithms and Neural Networks: A Comparative Study”, pp.1-6.

9. M. Seera, C. P. Lim, K. S. Tan, and W. S. Liew, ``Classification of transcranial Doppler signals using individual and ensemble recurrent neural networks," Neurocomputing, vol. 249, pp. 337- 344, Aug. 2017.

10. A.Bahnsen, D. Aouada, A. Stojanovic, B. Ottersten, “Feature Engineering Strategies for Credit Card Fraud Detection”, ELSEVIER Expert System with Applications, 2016, pp. 134-142.

11. N. Mahmoudi and E. Duman, ``Detecting credit card fraud by modified fisher discriminant analysis," Expert Syst. Appl., vol. 42, no. 5, pp. 2510-2516, 2015.

12. N. S. Halvaiee and M. K. Akbari, ``A novel model for credit card fraud detection using artificial immune systems," Appl. Soft Comput., vol. 24, pp. 40-49, Nov. 2014.

13. Y. Sahin, S. Bulkan, and E. Duman, ``A costsensitive decision tree approach for fraud detection," Expert Syst. Appl.

14. M. Zareapoor, K. Seeja, M. Alam, “Analysis of credit card fraud detection techniques: based on certain design criteria”, International Journal Computer Application, 2012, pp. 35–42.

15. Y. Sachin, E. Duman, “Detecting Credit Card Fraud by Decision Tree and Support Vector Machine”, In Proceedings of the international multi Conference of Engineers and Computer Scientists, Hong Kong, 2011, pp. 1-6.

16. A.Brabazon, J. Cahill, P. Keenan, D. Walsh, Identifying online credit card fraud using artificial immune systems, in: 2010 IEEE Congress on Evolutionary Computation (CEC) [Proceedings], IEEE Press, 2010.

17. S. Panigrahi, A. Kundu, S. Sural, A. Majumdar, Credit card fraud detection:a fusion approach using Dempster–Shafer theory and Bayesian learning, Inf.Fusion 10 (2009) 354–363.

18. M. Gadi, X. Wang, A. do Lago, Credit card fraud detection with artificial immune system, Artif. Immune Syst. (2008) 119–131.

ISSN : 0975-4520

19. E. Turban, J.E. Aronson, T.P. Liang, R. Sharda, “Decision Support and Business Intelligence Systems”, Eighth ed, Pearson Education, 2007.

20. S. Kotsiantis, E. Koumanakos, D. Tzelepis, and V. Tampakas, “Forecasting fraudulent financial statements using data mining, International Journal of Computational Intelligence, vol. 3, no. 2, pp. 104-110, 2006