Journal of Interdisciplinary and Multidisciplinary Research (JIMR)
E-ISSN:1936-6264| Impact Factor: 8.886|
Vol. 18 Issue 03, March- 2023
Available online at: https://www.jimrjournal.com/
(An open access scholarly, peer-reviewed, interdisciplinary, monthly, and fully refereed journal.)

Detection of Credit Card Fraud Using Machine Learning Algorithm Using HYBRID Method

V. Sumalatha[1], Dr. M. Raghavender Sharma[2]

1. V. Sumalatha, Research Scholar, School of Sciences, Career Point University, Kota, Rajasthan
2. Dr. M. Raghavender Sharma, Assistant Professor, Department of Statistics, Osmania University, Hyderabad, Telangana.

Abstract:

Credit card fraud mostly occurs in financial services. Credit card fraud causes a large number of problems every year. The problem is the lack of research on this credit card issue and analysis of real-world credit card fraud. This paper presents the best data mining algorithm called "Machine Learning Algorithm", which is used to identify credit card fraud, so use this algorithm initially and it is one of the standard models. Then, secondly, apply the hybrid methods "AdaBoost and Majority Vote Method". Use the effectiveness of this model, which is evaluated, and then use a publicly available credit card data set. A financial institution involves a real-world data set, so it is being taken and analyzed. Noisy ad data samples are also evaluated in this robustness algorithm. This concept is used in experiments and later results show that hybrid methods, i.e., majority voting, provide good accuracy rates for credit card fraud detection.

Keywords: Machine Learning, Credit Card Fraud, AdaBoost

1. Introduction:

Fraud is a fraudulent or wrongful or criminal activity, its main objective is to make a financial or personal mark. In this proposed system, two mechanisms (i) fraud prevention and (ii) fraud detection are used to prevent losses from fraud, which detect fraud details. First in the fraud prevention system. The most defensive and proactive policy, it prevents misrepresentation from starting. At that point, another fraud detection system guesses the fraudster. This element is required for forged exchanges, but is an approximation of the timing of the exchange made by the fraudster.Credit card fraud refers to the illegal use of credit card data to make purchases that use credit card funds to purchase items. At the time of purchase the user uses the credit card, the fraudster discovers the password or important details the user provides, then it will be applied to our transaction using the easy credit card cash but the person cannot detect that it is a fraud. Credit card transactions completed directly or carefully. A physical exchange-based credit card is used at the exchange, while a physical exchange-based credit card is used only over the phone or the web. Cardholders basically provide important details like card number expiry date and card validation number over phone or web. But in the world of technology credit card is used now so use this credit card every second as credit card transactions increase every day and the rise of e-commerce sector.

The number of credit card businesses increases every year. So the technology is mostly developed and people benefit more, but on the other side it increases the credit card fraud cases. This is the most effective problem in the world. After that, logical and numerical authentication methods are applied in these credit card fraud cases, but this method is mostly undetected, because fraudsters hide their details such as identity and location on the Internet, so this problem has a huge impact on the financial industry. Also, this credit card fraud problem affects both administrator and user side. It affects (a) issuer's charges, (b) charges, (c) administrative charges which are charge losses. Hence traders decide whether to set a higher price on the goods or lower the discount.

The Objective of the study is to reduce the fallout from credit card fraud, eliminating fraud cases. Two machine learning techniques are used (i) artificial networks, (ii) rule-finding techniques, (iii) decision trees, (iv) logistic regression and (v) support vector machines (SVM). This above model is combining several methods i.e. hybrid method. AdaBoost and Ballot Strategies are a big part of casting and identifying credit card fraud.

2. Literature review:

In previous studies, several methods have been applied for fraud detection using supervised, unsupervised algorithms and hybrid. The types and types of fraud are evolving day by day. It is important to have a clear understanding of the technology behind fraud detection. Discuss machine learning models, algorithms and fraud detection models used in previous studies here.

Andreas L. Prodromidis et.al (2000),"Agent-BasedDistributed Learning Applied to Fraud Detection", in their study A risk-based ensemble model is used which can give good results for data with model problems and a Naïve Bayes algorithm is used to remove the fuzzy noise in the transaction.

Satvik Vats, et.al (2013), "A Tool forEffective Detection of Fraud in Credit Card System""Improving a credit card fraud detection system using geneticalgorithm", in their study Fraudulent transactions are very less compared to normal transactions. A normal transaction appears legitimate when it looks like a fraudulent or fraudulent transaction. Also discuss the difficulties in dealing with categorical data. Many machine learning algorithms will not support categorical data. Discuss search costs and optimization as a challenge. Prevention costs and the cost of fraudulent behavior are taken into account.

S. Xuan, G. Liu, et.al (2018),"Random Forest for credit card fraud detection", "Credit Card FraudDetection & Prevention of Fraud Using Genetic Algorithm",in their study data mining techniques are discussed and these methods take time to handle huge data. Overlapping is another problem when generating credit card transaction data. Unbalanced data distribution is overcome by using sampling methods.

Monika andAmarpreetKaur."Fraud Prediction forcredit card using (2018)classification method", in their study observed that Entertainment results showed a true

Journal of Interdisciplinary and Multidisciplinary Research (JIMR)
E-ISSN:1936-6264| Impact Factor: 8.886|
Vol. 18 Issue 03, March- 2023
Available online at: https://www.jimrjournal.com/
(An open access scholarly, peer-reviewed, interdisciplinary, monthly, and fully refereed journal.)

positive rate of 98%. A modified Fisher discriminant function was used for the detection of charge card extortion. The change gradually weakened traditional capabilities in significant instances. A weighted normal was used to calculate the variances, allowing learning of beneficial exchanges. The results of altered capacity confirm that they can provide more benefits.

Three techniques for identifying misrepresentations are shown. Initially, a grouping model is used to arrange legitimate and fraudulent exchanges using information parameters. Also, a Gaussian mixture model of past behavior and current behavior can be determined to identify any anomalies in the past. Finally, Bayesian systems are used to illustrate specific client insights and measurements of various misrepresentation scenarios.

3. Material and Methods:

Hybrid clones are combinations of clones from different animals. A hybrid replication consisting of Multilayer Perceptron (MLP) neural system, SVM, LOR and Harmony Search (HS) headway was used for corporate duty avoidance. HS was useful in finding the best parameters for the course of the models. MLP with HS streamlining obtained the most remarkable accuracy rate of 90.07% using information from nutrition and physical sector in Iran. A semi-classified clustering framework with outlier detection capability was used to distinguish misrepresentations in lottery and internet recreations.

The framework aggregated online calculations with factual data from information sources to differentiate between different extortion types. Preparatory information index was packed into the basic memory of current period information trials, gradually incorporated into data blocks. The framework achieved an extreme location rate of 98% with a false alarm rate of 0.1%. Twelve machine learning algorithms are used in combination to detect credit card fraud. Computation runs from qualitative neural systems to deep learning models. Also, AdaBoost and large-scale casting ballot strategies are combined to build crossbreed models. The main commitment of this paper is the assessment of classification of AI models with real charge card information index for extortion location.

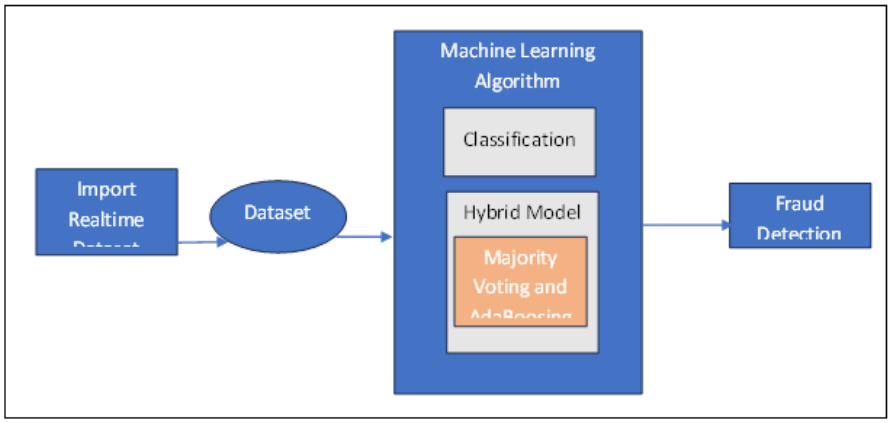

Figure 3.1: System Architecture Diagram

Journal of Interdisciplinary and Multidisciplinary Research (JIMR)
E-ISSN:1936-6264| Impact Factor: 8.886|
Vol. 18 Issue 03, March- 2023
Available online at: https://www.jimrjournal.com/
(An open access scholarly, peer-reviewed, interdisciplinary, monthly, and fully refereed journal.)

3.1. Machine Learning Algorithm:

A total of twelve calculations are used in this test. They are used in conjunction with AdaBoost. The dominant part of the ballot is a lot of time spent in the information group, which includes a coupled model of something like two counts. Each calculation makes its own prediction for each test. The last income that receives the most ballots is as follows. Check the classes (or points) selected by K, Ci, with K .K speaks to the target class expected by the classifier. Given a particular information x, each classifier presents a prediction corresponding to an objective class, submitting a sum of K expectations, i.e., PK. When casting a ballot, a large part is expected to distribute the cumulative expectation from all K estimates for information

x, P (x) j, jK, i. e. , pk (x) jk.

Double capacity can be used to speak with opinions.

$$\text{if} p_k(x) = i, j \in K, V_k(x \in C_i) = 0$$

At that point, the total votes from all K classifiers for each Ci, and the name with the most votes is the last (joined) expected category.

3.2. AdaBoost:

Versatile boosting or AdaBoost is used to upgrade the implementation of various types of computations. The yield is added using triviality as a whole, which speaks to the combined yield of the supported taxonomy, e.g.

$$FT(x) = ft(x), t = 1$$

where each foot is a classifier (weak learner) that gains the expected class corresponding to the input x. Each weak student gives an income estimate, h(xi), for each preparation test. In each cycle t, a weak learner is selected, and a coefficient, αt, is distributed with the aim that the preparation error total, Et, of the subsequent t-systematic aid classification is bounded,

$$Et = F[Ft - 1(xi) + \alpha th(xi)]$$

where $Ft - 1(x)$ is the supported classifier employed in the previous step,

$E(F)$ is the error potential, and ft (x) $\alpha th(x)$ is the unpowered classifier considered for the last classifier.

AdaBoost replaces low power for misclassified information trials. Nevertheless, it is sensitive to confusion and outliers. For as long as the classifier's execution is not arbitrary, AdaBoost can improve the results of individuals.

4. Experimental Result:

4.1. Study System:

In the Credit Card Information Index, the amount of fraudulent transaction activity is generally very low and the overall number of exchanges. With skewed information collection, ex post accuracy does not present an accurate depiction of the implementation of the framework. Disrupting the actual exchange leads to poor customer administration and neglecting to identify

extortion cases causes misfortune to money related foundations and customers. This information inconsistency causes implementation problems in AI computation. A lion's share of class tests affects results. Under-inspection has been used to deal with misinformation issues. In that capacity, this paper uses under-checking to deal with distorted information indices. Although there is no single most perfect strategy to delineate valid and false practical and obstructive using a single marker, the largest wide compute Matthews Correlation Coefficient (MCC). MCC computes the notion of a bi-class problem considering the error producer and constraint. This is a reasonable calculation even when classes are of different sizes.Matthews Correlation Coefficient (MCC) can be deployed as follows:

$$\mathrm{MCC} = \frac{\mathrm{TPXTN} - \mathrm{FPXFN}}{(\mathrm{TP}+\mathrm{FP})(\mathrm{TP}+\mathrm{FN})(\mathrm{TN}+\mathrm{FP})(\mathrm{TN}+\mathrm{FN})}$$

4.2. Bench Mark:

A freely accessible information archive has been established since. It has a total of 284,915 exchanges made by Indian cardholders in April 2021. There are 503 trades in the instruction file, which is increasing. Due to characteristic problems, a total of 37 key parts depending on the change are given.

4.3. True Data:

A genuine master card submitted by a budgetary institution in India is pre-owned in the examination. It depends on the customer in India locally. A total of 286,348 trades were registered out of which 107 were cases of forgery. Figures include specific trades. To comply with the client's security needs, no data is used near home.

Therefore, a credit card is the most common way to get into a line of credit. Usually, it is provided by a bank or financial institution. Credit card fraud is on the rise these days with easy access to a person's financial account details. So further the user can complain to the bank to block the card or account. And the methods used for fraud detection are: Decision Tree, Naïve Bayes and Random Forest. The X-axis represents the different methods and the Y-axis represents the year. Dark blue represents maximum theft occurrence and red represents average theft. while green and blue represent minimum theft. Finally, the purple represents the thefts that occurred throughout the year.

5. Conclusion:

This proposed system proposes the best concept of data mining, machine learning algorithm for credit card fraud. Then, quality replication characters such as NB, SVM and DL are used for evaluation terms. Credit card data is publicly available, used for evaluation, i.e. use standard model and hybrid model. Hybrid replication, such as AdaBoost and majority voting, are also model mixing techniques. MCC metrics only calculate performance measures and take into account and predict true or false results of credit card transactions. The best MCC score majority voting is used for majority voting. A financial institution sets up a credit card figure for evaluation. But the perfect MCC score is only to be obtained using the combination of AdaBoost

and majority voting, as it represents the combination method and gives strong and robust performance. In this the proposed concept is extended to online learning models. Use Internet Alerts to quickly raise awareness of credit card fraud. The proposed system helps to detect and prevent fraudulent transactions and activities earlier, thereby reducing the decline in the financial industry.

6. References:

1. Abhinav Srivastava, Amlan Kundu, Shamikr Sural andA.K. Majumadar, "Credit Card Fraud Detection UsingHidden Markov Model", IEEE Transactions onDependable and Secure computing, Vol. 5, No. 1, pp. 37-48, 2008.
2. Andreas L. Prodromidis and Salvatore J. Stolfo; "Agent-BasedDistributed Learning Applied to Fraud Detection"; Department ofComputer Science- Columbia University; 2000.
3. Chuang-Cheng Chiu and Chich-Yuan Tsai, "A WebServices-Based Collaborative Scheme for Credit CardFraud Detection", Proceedings of the IEEE InternationalConference on e-Technology, e-Commerce and e-Service,pp. 177-181, 2004.
4. Cristianini N and Shawe-Taylor J, "An Introduction toSupport Vector Machines and other Kernel-based LearningMethods", Cambridge University Press, Cambridge, UK,2000.
5. Deepa V. and Dhanpal R. "Behavior Based Credit Card Fraud Detection using Support Vector Machines", ICTACT Journal on Soft Computing, (2012), Vol-2, Issue-4, pp. 391-397.
6. Monika andAmarpreet Kaur."Fraud Prediction forcredit card using classification method". InternationalJournal of Engineering and Technology, (2018); 7(3)1083-1086.
7. Satvik Vats, Surya Kant Dubey, Naveen Kumar Pandey, "A Tool forEffective Detection of Fraud in Credit Card System", published inInternational Journal of Communication Network Security ISSN:2231 – 1882, Volume-2, Issue-1, 2013.
8. SuvasiniPanigrahi, Amlan Kundu, Shamikr Sural and A.K.Majumadar, "Credit Card Fraud Detection: A FusionApproach Using Dempster Shafer Theory And BayesianLearning", Information Fusion, Vol. 10, No. 4, pp. 354-363, 2009.
9. Salvatore J Stoflo, David W Fan, Wenke Lee and AndreasL Prodromidis and Philip K Chan, "Cost –Based Modelingfor Fraud and Intrusion Detection: Results from the JAMProject", Proceedings of the DARPA InformationSurvivability Conference and Exposition, Vol. 2, pp. 130-144, 2000.
10. Xuan S., Liu G., Li Z., Zheng L., Wang S., and SurnameG. N., "Random Forest for credit card fraud detection", IEEE 15th International Conference on Networking, Sensing and Control (ICNSC),2018.
11. Zhang yongbin, You Fucheng and Liu Huaqum,"Behavior–Based Credit Card Fraud Detection Model",Fifth International Joint Conference on INC, IMS and IDC,pp. 855-858, 2009.