Electronic Credit Card Fraud Detection System by Collaboration of Machine Learning Models

Shiv Shankar Singh

Abstract— In the financial industrial sector the lightning growth and participation of internet-based transactional events give rise to malicious activities like a fraud that result in financial loss. The malicious activities have no continuous pattern their pattern, behavior, working always keep on changing with the increasing growth in technology. Every time a new technology comes in the market the hoaxer study about that technology and implement malicious activity through the learned technology and internet-based activities. The hoaxer analyzes the behavior patterns of consumers to execute the plan of fraud to cause loss to the consumer. So to overcome this problem of fraud, hoax, cheat in the financial sector a fraud identification system is needed to identify the cheating, fraud and alike activities in internet-based money transactions by employing machine learning techniques. This presented paper focuses on fraud activities that cannot be detected manually by carrying out research and examine the results of logistic regression, decision tree and support vector machine. A dataset of electronic payment card is taken from European electronic cardholders, the machine learning techniques are applied on the unstructured and process-free data.

Keywords— Fraud in credit card, data mining, logistic regression, decision tree, SVM.

I. INTRODUCTION

In developing and developed countries the use of electronic smart payment cards has become a general scenario. People use the electronic smart card on a daily basis, to pay shopping bill, school or college fees, etc. However, with a more electronic smart card like credit card customers, the rate of credit card fraud is also increasing. Fraud may be classified as any activity in order to be deceived, without having detailed information of the cardholder and the related issuer bank, to obtain financial gain by any means. Fraud can be done in many ways by electronic smart cards like a credit card by generating false counterfeit, credit card with a changing the magnetic band which is there on the credit card and other electronic payment cards which comprise of cardholder’s information. The current share of theft by demand station, which means that a theft failure proportion of their internet shop was 74% and 49% for their portable apps, according to a 2017 CyberSource report. The lection is to determine anomalies in fraud behaviors, which have changed in comparability with the previous, based on this data. Malicious activities on electronic payment cards can be of following types:

1. Approach frauds: When the scammer gets power over the app scheme with access to delicate customer information such as password and user name and produce a fake account. It is usually done with respect to theft of identification. If the fraudster refers in the presence of the account owner to a loan or fresh loan account. The fraudster robs the records to promote or assist its unauthorized request.
2. Credit card Imprints electronic or manual: If the fraudster skims it off the data on the panel magnetic strip. These are highly private data and the fraudster can use them in the future for personal operations by accessing them.
3. CNP (Card Not Present): When the fraudster knows the termination date and account number of the electronic card, the card can be used without its actual physical ownership.
4. False Card Fraud: The procedure of taking of usually tested. A false plastic swipe board is produced and contains all the initial card information. The false card can be used to undertake operations in the upcoming time and is fully operational.
5. Fraud: When the initial owner of the ticket mistakes the card, he or she may get into fraudsters ' fingers and payment is then made. Lost and Stolen Cards Fraud: It is difficult to do that on the computer, however, because an amount is needed; the fraudster can make internet-based wireless transactions simple enough.

II. RELATED WORK

A paper [1] talks about of the electronic payment like card credit card fraud, which applies data normalization before cluster analysis and which shows that neuronal variables may be reduced by clustering of characteristics, using cluster analysis and artificial neural networks for the identification of fraud. The machine learning program had been trained for successful outcomes through the use of standard information. This study was focused on uncontrolled education. The importance of this article was to discover and improve the precision of outcomes with fresh techniques for detecting fraud.

Jain R [2], this paper talks about an enhanced comparative metric which reflects fairly the profits and casualties caused by the identification of fraud. Price metric has employed to present a price tactful process centered on the Bayes minimum danger, using this technique and other state-of-the-art algorithms, enhancement of up to 23 percent was achieved. The information for that document was built on a big Asian company's real-life transactional information on private information. A pseudo-code was identified and the price estimate was reduced.

Yet another cited paper [3], talks about a distinction

Revised Manuscript Received on 14 September, 2019.

Shiv Shankar Singh, Department of Computer Science and Engineering, Sanskriti University, Uttar Pradesh, India. (E-mail: sanpubip@gmail.com)

which has been created with the overall outline of the fraud detection scheme created, such as the Naive Bavarian Classifier and the Bavarian networks clustering model. A distinction is provided between designs relying on Artificial Intelligence. Findings were drawn regarding the outcomes of the model assessment tests. There were more than or equivalent to 0,65 legal transactions which were determined to be 65% correct using the Bayesian network.

The above-cited papers have some drawbacks based on efficiency, cost and time to deliver output, therefore this research is done to overcome those drawbacks to identify the error in the loan book datasets acquired from a machine learning group by implementing the logistic regression, decision tree, support vector machine, random forest and to assess their exactness, awareness, accuracy, and accuracy with various designs.

III. PROPOSED SYSTEM

Logistic regression is a controlled technique for classifying binary count on a variable that estimates the probability of results with zero or one attributes, yes or no and false or true, based on the independent variable of the dataset, which is logistic regression.

Regression of logistics is alike to linear regression, as the direct row is acquired in the linear regression, logistic regression indicates a curve. The forecast is counted on the use of one or more predictors or autonomous matrix, logical classification generates logistic equations that trace the numbers between null and 1.

Support vector machine is a common regression, classified machine learning algorithm. It is a controlled teaching machine that analyzes the relevant data utilized to classify and regress. The support vector machine design consists of two phases, first to train and get a template, and then to forecast the information of the test data collection with this template. The support vector machine is a common regression, classified machine learning algorithm. It is a controlled teaching machine that analyzes the information used to classify and regress. The support vector machine design consists of two phases, first to train and get a template, and then to forecast the information of the test data collection with this template. In support vector algorithms, a plot is generated because each relevant dataset is used to indicate the significance of each character in an n-dimensional room where n is the valuation of each characteristic. Then the ranking is carried out by finding the hyperplane that very well distinguishes both categories.

A decision tree is an approach using a tree data structure such as a chart or matrix of choices and its feasible results in order to forecast the ultimate choice. It is a pseudo code to approach evaluated objectives. These kinds of algorithms are very popular for interactive learning and have been used effectively for various assignments overseas.

Random Forest is a classifying and regressive algorithm. In short, it's a decision-tab classification set. Spontaneous forests have benefited over the tree, as they actually correct only the practice of overfitting. A small subset of the training set is sampled completely randomly so that each tree is trained, then every node divides on a new feature that is chosen from a completely random subset of the entire feature set.

IV. COMPARATIVE ANALYSIS

Real favorable, real reverse, false positives and false negatives produced by a scheme or procedure in an attempt to combine and evaluate the efficiency and efficiency of distinct processes with the purpose of comparing multiple methods. True Positive, is the amount in operations illicit but ineffective under the scheme. True Negative, is the amount of lawful and lawful operations. False Positive, is the number of operations lawful, but incorrectly considered to be false. False Negation, is, therefore, the number of transactions that have been very completely fraudulent but have been misunderstood as perfectly legitimate system financial transactions.

The main differences in the present cheating tracking designs and methods are:

I. Comprehensive credit card information is unavailable because they are a private estate, and either banks or consumers can not communicate their information in an inappropriate and educated manner.

II. A strong software is not available which can continuously execute throughout all settings and exceed any application.

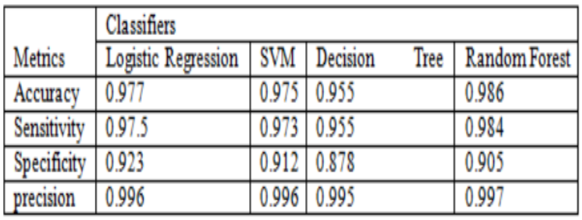
III. The exact nature of the scheme cannot be defined but there may be a production of strong comparison effect between various methods in an organized and effective measurement of configuration.

IV. A system can not efficiently adjust to evolving circumstances, new methods of fraud and real adjustments to a consumer ' s purchasing practices.

V. RESULT

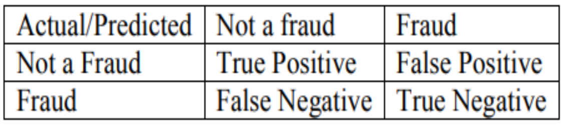
From the studies, it has come to the knowledge that the logistic model is 97.7 percent accurate, while the SVM is 97.5 percent accurate as well as the decision tree is 95.5 percent accurate, however, the Random forest with highest outcomes have achieved. 98.6 percent precision. Thus, the findings show that Random Forest demonstrates the most accurate and accurate issue of money laundering identification by ULB computer education of 98,6 percent.

Table 2: Performance matrices



Metrics	Classifiers			
	Logistic Regression	SVM	Decision Tree	Random Forest
Accuracy	0.977	0.975	0.955	0.986
Sensitivity	0.97.5	0.973	0.955	0.984
Specificity	0.923	0.912	0.878	0.905
precision	0.996	0.996	0.995	0.997

Table 3: Confusion matrix format



Actual/Predicted	Not a fraud	Fraud
Not a Fraud	True Positive	False Positive
Fraud	False Negative	True Negative

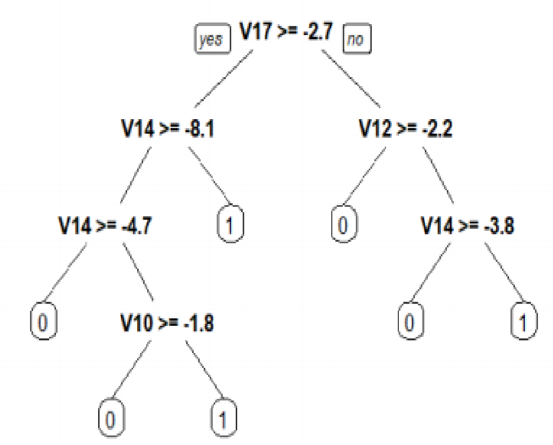


Figure 1 Decision Tree References

VI. CONCLUSION

Though there are many identity verification methods available today none is able to identify all frauds entirely while they are actually occurring, they generally detect it until the fraud has been perpetrated. This happens because a very minuscule number of transactions from the total transactions are actually fraudulent in nature.With more learning information, the Random Forest Algorithm will do faster, but velocity will be impaired in experimentation and implementation. It would also assist to implement more pre-processing methods. The support vector machine software already comes from unbalanced data sets issue and needs a higher preliminary processing rate to achieve superior outcomes at the outcomes as seen by Support vector machine. The requisite to develop a successful hybrid system is to combine costly training techniques with incredibly precise and exact outcomes with an enhancement method to reduce system costs and rapidly train the machine. The selection of hybrid methods depends on how the fraud sensing device works and the workplace.

REFERENCES

1. Raj S.B.E., Portia A.A., Analysis on credit card fraud detection methods, Computer, Communication and Electrical Technology International Conference on (ICCCET) (2011), 152-156.
2. Jain R., Gour B., Dubey S., A hybrid approach for credit card fraud detection using rough set and decision tree technique, International Journal of Computer Applications 139(10) (2016).
3. Dermala N., Agrawal A.N., Credit card fraud detection using SVM and Reduction of false alarms, International Journal of Innovations in Engineering and Technology (IJIET) 7(2) (2016).
4. Phua C., Lee V., Smith, Gayler K.R., A comprehensive survey of data mining-based fraud detection research. arXiv preprint arXiv:1009.6119 (2010).
5. Bahnsen A.C., Stojanovic A., Aouada D., Ottersten B., Cost sensitive credit card fraud detection using Bayes minimum risk. 12th International Conference on Machine Learning and
6. Applications (ICMLA) (2013), 333-338.
7. Carneiro E.M., Dias L.A.V., Da Cunha A.M., Mialaret L.F.S., Cluster analysis and artificial neural networks: A case study in credit card fraud detection, 12th International Conference on
8. Information Technology-New Generations (2015), 122-126.
9. Hafiz K.T., Aghili S., Zavarsky P., The use of predictive analytics technology to detect credit card fraud in Canada, 11th Iberian Conference on Information Systems and Technologies (CISTI) (2016), 1-6.
10. Sonepat H.C.E., Bansal M., Survey Paper on Credit Card Fraud Detection, International Journal of Advanced Research in Computer Engineering & Technology 3(3) (2014).
11. Varre Perantalu K., Bhargav Kiran, Credit card Fraud Detection using Predictive Modeling (2014).
12. Wang, Deshen, Bintong Chen, and Jing Chen. "Credit card fraud detection strategies with consumer incentives." Omega(2018).
13. Westerlund, Fredrik. "Credit Card Fraud Detection (Machine learning algorithms)." (2017).
14. Bhattacharyya, Siddhartha, et al. "Data mining for credit card fraud: A comparative study." Decision Support Systems 50.3 (2011): 602-613.
15. Readshaw, Neil Ian. "Method and System for Identification By A Cardholder of Credit Card Fraud." U.S. Patent Application 12/496,239, filed January 6, 2011.
16. Saia, Roberto. "A discrete wavelet transform approach to fraud detection." International Conference on Network and System Security. Springer, Cham, 2017.
17. Dal Pozzolo, Andrea, et al. "Credit card fraud detection: a realistic modeling and a novel learning strategy." IEEE transactions on neural networks and learning systems 29.8 (2017): 3784-3797.
18. Demla, Nancy, and A. Aggarwal. "Credit Card Fraud Detection using SVM and Reduction of False Alarms." International Journal of Innovations in Engineering and Technology (IJIET)7.2 (2016): 176-182.
19. Alowais, Mohammed Ibrahim, and Lay-Ki Soon. "Credit card fraud detection: Personalized or aggregated model." 2012 Third FTRA International Conference on Mobile, Ubiquitous, and Intelligent Computing. IEEE, 2012.