ISSN: 2454-9940

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

E-Mail :
editor.ijasem@gmail.com
editor@ijasem.org

www.ijasem.org

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

ISSN2454-9940

www.ijasem.org

Vol 18, Issue 2, 2024

A Novel Approach for Credit Card Fraud Detection using Decision Tree and Random Forest Algorithms

[1]MR.A N L KUMAR, [2]AKKARABOTHU BHARGAV SAI KRISHNA

[1](Head of the Department), MCA, Swarnandhra College

[2]MCA, Scholar, Swarnandhra College

ABSTRACT

New methods of doing business emerged in the financial sector with the expansion of technology. One of them is the credit card system. However, several issues have arisen in this system as a result of credit card frauds due to numerous vulnerabilities. Credit card companies and their clients are taking a major financial hit as a result of this. When it comes to privacy concerns, there is a lack of investigation lessons on looking at real credit card numbers. An effort to detect credit card fraud using algorithms that use machine learning approaches is presented in the article. Regarding this matter, two algorithms are employed: one for credit card fraud detection using Decision Tree and the other for random forest fraud detection. By sampling from a set of publicly available data, we may determine the model's efficiency. The next step is to look at a real-life credit card data set that a bank has. In addition, the data samples are enhanced with additional noise to further verify the systems' robustness. The first technique builds a tree against the user's behaviors and uses this tree to detect frauds; this is the most important method employed in the study. In the second approach, a user activity based forest is built and then used to try to identify the culprit. Results from the study show that the popular alternative method may detect credit card fraud with a reasonable degree of accuracy.

1.INTRODUCTION

If a trader takes precautions to prevent the theft of money, goods, or services obtained via an unauthorized credit card operation, they have committed a credit card scam. The consumer or a third party could be involved in a credit card fraud. A great deal of effort has gone into developing methods to foil such scams. In the event that such scams occur, methods for tracing the fraudulent transactions are also devised on the fly. Digital data transfers may be protected against unwanted access with the use of a variety of new and innovative algorithms. Nonetheless, there are downsides in some form or another. The main focus of this study is on the approaches used to identify credit card fraud.

A Machine Learning-Based Decision Tree-Based System for Detecting Credit Card Fraud [A].

When a company has to aggregate the unusual occurrences from an approved client, they apply the algorithm. In statistics, it is one method for predictive modeling. An important benefit of this algorithm is that it generates a thorough study of the results, tracks each possible route to a conclusion, and ensures that all possible choice outcomes are considered. Practical Example: Consideration is given to any scenario in which a client transacts. In order to foretell the likelihood of a business-related fraud, the decision tree is built.

B. Random Forest based Credit Card Fraud Detection Algorithm utilizing Machine Learning.

This method employs a mixture of decision trees to provide superior results, making it an upgraded variant of the decision tree algorithm. All decision tree draughts for the divers condition are applicable to decision trees and any data collection. There is a potential scam and a non-scam company around every corner. Random forests and random decision forests are two types of group learning methods used for classification, prediction, and other tasks. They work by training a large number of decision trees at runtime and then producing a class that represents the mean prediction (regression) or mode of the modules (classification). Over fitting is a problem with decision trees, but random decision forests avoid it. Example of Use: Picture a

situation when a purchase is made. Here we can see an example of how fraud detection algorithms employ the random forest in Machine Learning.

2.LITERATURE SURVEY

Uncovering Credit Card Fraud employing Hidden Markov Models (HMM):

Authors: A. Kundu, S. A. Srivastava Sural and A. Majumdar

Technology: Hidden Markov Model

Objective: Despite several advancements in detection methods, credit card theft continues to rise. Since fraudsters are always coming up with new methods to perform fraudulent transactions, there is a continuing need for innovative detection strategies. A multitude of methods have developed for identifying different types of credit card fraud, drawing from many fields such as AI, data mining, ML, SG, decision trees, neural networks, logistic regression, naïve Bayesian, Bayesian networks, metalearning, Genetic Programming, etc. An effective system for detecting credit card fraud may be achieved with a consistent use of all three tactics. Credit card fraud detection methods and the Hidden Markov Model (HMM) are thoroughly covered in this paper's examination of numerous methodologies. Based on the amount spent, HMM classifies card holders as either low, medium, or high spenders. Every cardholder is being given a set of odds based on the value of the transaction. After that, we compare the amount of each incoming transaction to the category of the card owner. If it meets a certain threshold, we consider the transaction legal; otherwise, we label it as fraudulent.

Data mining for Credit Card Fraud:

Authors: Zhang, R et al

Technology: Data mining

Objective: This research looked at how well logistic regression, random forests, and support vector machines—two sophisticated data mining techniques—detected credit card fraud. Our assessment was based on a real-world dataset of credit card transactions that occurred between January 2006 and January 2007. Two methods that have become more popular recently because to their improved performance in many

applications are random forests and support vector machines (SVM). Credit card fraud has persisted around their use. The prevalence and severity of credit card theft are major concerns. Although predictive models are already seeing widespread application in the credit card fraud detection industry, there has been surprisingly little study on data mining techniques specifically for this purpose, which may be attributable to a dearth of relevant data. In an effort to better identify (and subsequently prevent and prosecute) credit card fraud, this article assesses three sophisticated data mining techniques: logistic regression, support vector machines, and random forests. The research relies on actual credit card transaction data from a global business.

Using Logistic Regression to Detect Credit Card Fraud:

Author: Hosmer, Lemeshow, and Sturdivant

Technology: Logistic Regression

Objective: Hosmer, Lemeshow, and Sturdivant are respected authors known for their expertise in logistic regression modeling. David W. Hosmer, Jr., a prominent statistician, has contributed significantly to the application of logistic regression in epidemiology and health sciences. Stanley Lemeshow, a renowned biostatistician, is recognized for his work in logistic regression modeling, particularly in clinical and epidemiological research. Rodney X. Sturdivant, a statistician, has collaborated with Hosmer and Lemeshow on the influential textbook "Applied Logistic Regression," which is a standard reference in the field. This book is praised for its clear explanations, practical examples, and guidance on model interpretation and validation, scholars and practitioners across fields will find it an invaluable resource. Their collective work has had a profound impact on the field of logistic regression and its applications, shaping the way researchers approach and utilize this statistical method in real-world scenarios.

Credit Card Fraud Detection using SVM:

Author: Nello Cristianini and John Shawe-Taylor

Technology: Support Vector Machine

Objective: Nello Cristianini and John Shawe-Taylor are notable figures in the field of machine learning and support vector machines (SVMs). Their collaborative work has significantly influenced the theory and application of SVMs, particularly in classification tasks. Cristianini is known for his contributions to kernel methods and their application in various domains, including bioinformatics and text analysis.

Shawe-Taylor's research spans machine learning theory and algorithms, with a focus on SVMs and their use in pattern recognition and data mining. Together, Cristianini and Shawe-Taylor have authored the influential book " An Overview of Kernel-Based Learning Techniques, Including Support Vector Machines," which provides a comprehensive overview of SVMs and their theoretical foundations. The book is widely regarded as a seminal work in the field, offering insights into the principles of SVMs and their practical implementation. Their collaboration has helped bridge the gap between theoretical research and real-world applications, making SVMs more accessible to researchers and practitioners in machine learning and related fields.

3.EXISTING SYSTEM

An in-depth investigation on fraud detection by natural observation of customer-side events has been conducted by A. A. Akinyelu and O. Adewumi [1]. An extensive investigation of the use of hidden Markov models for the purpose of detecting credit card fraud was conducted by A. Kundu, S. A. Srivastava Sural, and A. Majumdar [2]. Since the algorithms covered in the article are involved with machine learning approaches, the work of Singh, P. K. Saraswat et al. [3] on swarm intelligence directed to machine learning might be considered. Jung, J. J. et al. [4] has addressed the field's problems by developing ways for gathering social media data and describing it in terms of big data models. Apache Spark, which employs fuzzy-based clustering logic for massive data analysis, has been the subject of extensive research by Bharill N et al. [5]. A thorough nominal cost model for detecting scams in the business area was presented by Y. Sahin et al. [6]. The Nilson Report [7] provides a comprehensive analysis of the many ways in which credit card fraud and scams may happen, as well as ways to spot

them and the damage they do to businesses. A story was told by J. T. Quah et al. [8] on how they worked on an automated approach to identify corporate fraud. By keeping track of transactions and constructing a model using data mining methods, S. Jha et al. [9] created a system that helps identify commercial scams and frauds. For the purpose of detecting fraud in communication networks, S. Panigrahi et al. [10] used Bayesian inferencing and Dempster-Shafer theory. As previously shown by T. Fawcett et al. [11], Adaptive fraud detection, Data Mining, and Knowledge Discovery may be advantageous. Distributed intrusion detections using data fusion have been the subject of an in-depth investigation by Y. Wang et al. [12]. Using GA Feature Selection on Naive Bayes Random Forest and SVM for Credit Card Fraud Detection was proven by Yakub K et al. [13]. Online fraud detection using sequential behavioral data processing with deep learning and the markov transition field was extensively discussed by Zhang, R et al. [14]. One method for building various data sets is Attributed Sequence Embedding, which was shown by Zhong fang Zhuang et al. [15].

Disadvantages:

1. The system doesn't have technique to analyze large number of datasets.
2. There is no technique Random decision forests and Random forests which are the group learning techniques for categorization, prediction and additional jobs that function by building a gigantic volume of decision trees at exercise time and outputting the class.

3.1 PROPOSED SYSTEM

The proposed system defines the procedure used to hostage the credit card scam. The numerous competent approaches like arrangement orientation, device learning, neural networks, artificial intelligence, fuzzy logic are employed to detect and encounter scams in credit card businesses. Credit card fraud has become progressively widespread in modern years. In Current day, the fraud is one of the key causes of excessive business losses, not only for merchants, distinct clients are also affected. So there are some methods to detect such kind of frauds. Initially, clustering model was adopted to categorize the authorized and deceitful operation by means of data clusterization of

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

ISSN2454-9940

www.ijasem.org

Vol 18, Issue 2, 2024

areas of factor value. Furthermore, Gaussian mixture model is used to model the possibility thickness of credit card operator's past performance such that the chance of present actions can be intended to perceive any irregularities from the historical behavior. Finally, Bayesian networks are used to define the measurements of a specific user and the pointers of different scam circumstances.

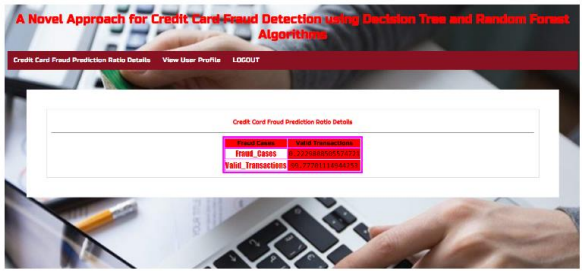
Advantages:

1. The proposed system offered several innovative approaches that have vastly increased the efficiency of cyber threat identification.
2. The system is more effective due to presence of Random Forest based Credit Card Fraud Detection Algorithm using Machine Learning.

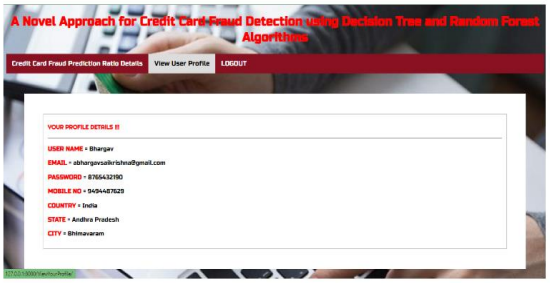
4.OUTPUT SCREENS

REMOTE USER:

Credit Card Fraud Prediction Ratio Details



View User Profile



Service Provider:

Browse and Train & Test Data Set

Find Credit Card Ratio Details

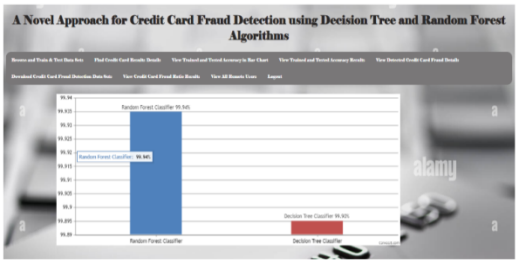
INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

ISSN2454-9940

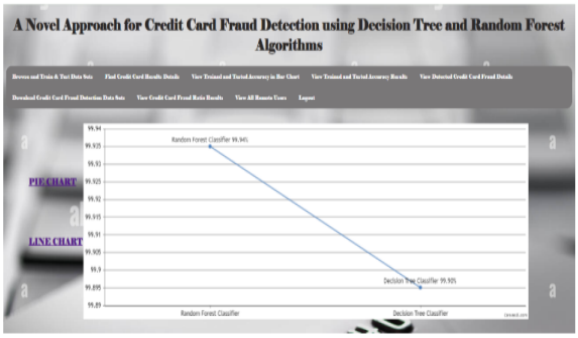
www.ijasem.org

Vol 18, Issue 2, 2024

Watch the Bar Chart Showing the Results of Trained and Tested Accuracy

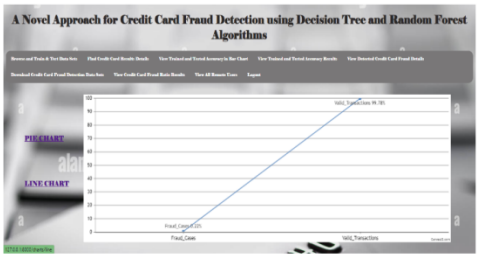


Examine the Results of Trained and Tested Accuracy



Check the Details of Indicated Credit Card Fraud

View Credit Card Fraud Ratio Results



Check Out Every Remote User

5.CONCLUSION

An examination of credit card fraud using AI equations has been included into this broadsheet. Experimental assessment has made use of similar template mockups employing DL, NB, and SVM. Using the lion's share of the vote mixing strategies and the single (normal) model-half breed model, we evaluated publicly accessible Master card knowledge indexes. Since it discusses the real and fake optimistic and terrible outcomes anticipated, the MCC metric has been set up as an exhibition ration. Algorithms can forecast

credit card theft to a certain extent, and the likelihood of An examination of credit card fraud using AI equations has been included into this broadsheet. Experimental assessment has made use of similar template mockups employing DL, NB, and SVM. Using the lion's share of the vote mixing strategies and the single (normal) model-half breed model, we evaluated publicly accessible Master card knowledge indexes. Since it discusses the real and fake optimistic and terrible outcomes anticipated, the MCC metric has been set up as an exhibition ration. In addition to these metric assessment techniques for assessing the suggested algorithms' effectiveness, the algorithms may anticipate credit card fraud to a certain extent, even if there are several intermediary routes via which credit card fraud might occur. It is challenging to construct classified data amongst misleading data and to discover relationships among all of them. The suggested method has limitations, as stated in the conclusion.

6.REFERENCES

[1] O. Adewumi and A. A. Akinyelu, "A survey of machine learning and nature-inspired based credit card fraud detect ion techniques," International Journal of System Assurance Engineering and Management, vol. 8, pp. 937–953, 2017J. Clerk Maxwell, A T reatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68-73.

[2] A. Srivastava, A. Kundu, S. Sural, A. Majumdar, "Credit card fraud detect ion using hidden Markov model," IEEE Transact ions on Dependable and Secure Comput ing, vol. 5, no. 1, pp. 37–48, 2008K. Elissa, "Tit le of paper if known," unpublished.

[3] Bansal, J. C., Singh, P. K., Saraswat, M., Verma, A., Jadon, S. S., and Abraham, A. (2011). Inert ia weight strategies in part icle swarm optimization. In Nature and Biologically Inspired Computing (NaBIC), (Salamanca, Spain, October 19 - 21, 2011).IEEENaBIC'11,633--640.

[4] Bello-Orgaz, G., Jung, J. J., & Camacho, D. (2016). Social big data: Recent achievements and new challenges. Information Fusion. 28 (Mar. 2016), 45--59

[5] Bharill, N., T iwari, A., and Malviya, A. (2016). Fuzzy Based Clustering Algorithms to Handle Big Data with Implementation on

Apache Spark. In Proceedings of the IEEE 2nd Internat ional Conference on Big Data Comput ing Service and Applications, (Oxford, UK, March 29-April 01, 2016). IEEE BigDataService '16, 95--104.

[6] Y. Sahin, S. Bulkan, and E. Duman, "A cost -sensit ive decision tree approach for fraud detection," Expert Systems with Applications, vol. 40, no. 15, pp. 5916–5923, 2013.

[7]TheNilsonReport(October2016)[Online]. Available:https://www.nilsonreport .com/upload/content_promo/The_Nilson _Report_10-17-2016.pdf

[8] J. T . Quah, and M. Sriganesh, "Real-t ime credit card fraud detection using computational intelligence," Expert Systems with Applications, vol. 35, no. 4, pp. 1721–1732, 2008.

[9] S. Bhattacharyya, S. Jha, K. Tharakunnel, and J. C., "Data mining for credit card fraud: A comparative study," Decision Support Systems, vol. 50, no. 3, pp. 602–613, 2011.

[10] S. Panigrahi, A. Kundu, S. Sural, and A. K Majumbar, "Use of Dempster-Shafer theory and Bayesian inferencing for fraud detection in communication networks", Lecture Notes in Computer Science, Spring Berlin/ Heidelberg, Vol. 4586, , 2007, p.446-460