REDVET - Revista electrónica de Veterinaria - ISSN 1695-7504
Vol 25, No. 1S (2024)
http://www.veterinaria.org
Article Received: Revised: Accepted:

One Step Ahead Using Data Mining Techniques Loan Fraud Investigation/Analytics

Himani, Dr. Lakshmi Shankar Singh[1*]

[1*]***Department of Sir Chhotu Ram Institute of Engineering And Technology, Chaudhary Charan Singh University (Campus), Meerut India. himanichoudharybca623@gmail.Com, Shanker.laxmi@gmail.Com***

Abstract

Background: loan fraud refers to any deceptive or dishonest activity to obtaining or misusing a loan it involves intentionally providing false information or manipulating the loan process for personal gain. Loan fraud is illegal and can have serious consequences, including financial loss, damaged credit, and potential legal action. Lenders employ various measures, such as background checks, verification processes, and fraud detection systems, to mitigate the risks of loan fraud.

Many businesses across various industries have expressed worry about loan fraud; this causes billion- rupees losses every year. Thus, in order to deal with this ongoing and expanding issue, firms use data mining approaches. The purpose of this paper is to analyze studies that have been done in the last ten years to identify financial fraud using data mining techniques and to inform academic researchers and business professionals of the latest developments.
Financial organizations place a high value on loan fraud detection because of the losses that may be prevented by using an effective fraud detection system. This is a topic that is regularly brought up in scientific study.
Objective: The aim of this research is to describe a machine learning method to loan identification of fraud using data mining techniques while comparing a machine learning algorithm using statistical frameworks in fraud detection.
Methods: The study suggests a method for enhancing performance through machine learning and data mining approaches. The writers employed two datasets for the training phase. The first set of data is unprocessed, whereas the second set has been pre-processed using techniques for feature engineering and selection. All classifiers' performance significantly improved after being evaluated on pre-processed data, according to the results.
Result: Finally, they draw the conclusion that, when compared to machine algorithms for learning, using machine learning algorithms straight to raw data produces poor results for statistical models, and that the suggested approach—which makes use of feature design and selection techniques—helps to enhance performance. A variety of financial applications, including credit cards and health insurance, were found to have fraud using data mining techniques.

Key words: Data analysis methods, a scam, loan fraud, loan identification, data mining.

Introduction

When people or organizations use dishonest or fraudulent methods to get a loan, this is known as loan fraud. To obtain a loan, this may entail submitting inaccurate data on loan applications, representing one's assets or financial situation fraudulently, falsifying documents, or undertaking other illegal actions. This is a severe offense that can lead to legal repercussions, criminal prosecution, and harm to one's character and credit.
The process of identifying loan fraud usually combines sophisticated technology tools with manual assessment procedures.

REDVET - Revista electrónica de Veterinaria - ISSN 1695-7504
Vol 25, No. 1S (2024)
http://www.veterinaria.org
Article Received: Revised: Accepted:

Loan fraud detection technologies have been introduced to the scene to deal with this issue and offer dependable company solutions. Loan fraud is typically uncovered by an outlier detection procedure made possible by data mining tools, which also disclose important information by illuminating hidden linkages, trends, and patterns within a sizable database. One of the main tools for identifying various kinds of fraud is data mining, which is described as "a method that gathers and recognizes relevant information from a large database through statistical, mathematical, artificial intelligence, and data mining techniques." financial fraud using a range of techniques, such as naïve Bayes model, decision trees, logistic regression, support vector algorithms (SVM), and artificial neural networks (NN).

While many organizations perceive the identification of financial fraud as a top concern, there is currently a dearth of current, In-depth and comprehensive literature that might help companies select the most effective data analysis method.

A significant and expanding issue is loan fraud. Loan fraud causes billions of rupees ' worth of losses every year. According to Cyber Source's 10th monthly online fraud report, growing online sales have caused the real amount of lost income to increase even if the percentage has remained constant at 1.4% of payments made online over the previous three years.

Some common features used in loan fraud detection?

In loan fraud detection, several common features are used to analyze loan applications and identify potential fraudulent activity. Here are some examples:

1. Credit Score: The applicant's credit score is a crucial factor in determining their creditworthiness and the likelihood of fraudulent behavior.

2. Income: The applicant's income can help determine whether their financial status is in line with the loan request and can also be used to evaluate the applicant's ability to repay the loan.

3. Employment History: Evaluating the applicant's employment history provides insights into their stability and reliability as a borrower.

4. Loan Amount: The requested loan amount is considered to ensure it aligns with the applicant's financial profile and doesn't indicate an excessive or suspicious request.

5. Ratio of Debt to Income: This ratio evaluates the applicant's financial stability and capacity to handle further debt by comparing their overall debt to their income.

6. Loan Purpose: Examining the purpose of the loan can help identify potentially fraudulent activities, such as using the loan for illegal or suspicious purposes.

7. Geographic Location: Analyzing the applicant's location can provide insights into regional fraud patterns or identify high-risk areas.

There are several data mining algorithms commonly used for loan fraud detection

1. Decision Trees: These algorithms create a tree-like model to classify loan applications as either fraudulent or legitimate based on different features and criteria.

2. Random Forest: In order to discover variations in loan fraud and produce forecasts that are more accurate, this algorithm integrates numerous decision trees.

3. Logistic Regression: This algorithm analyzes the relationship between various factors in loan applications to estimate the probability of fraud occurrence.

4. Neural Networks: These algorithms simulate the human brain's neural connections to identify complex patterns and detect loan fraud based on historical data.

5. Support Vector Machines (SVM): SVM algorithms separate loan applications into different categories, helping to distinguish between fraudulent and non-fraudulent cases.

REDVET - Revista electrónica de Veterinaria - ISSN 1695-7504
Vol 25, No. 1S (2024)
http://www.veterinaria.org
Article Received: Revised: Accepted:

Some other method used in loan fraud detection?

In addition to decision trees, there are several other methods used in loan fraud detection.

1. The Logistic Regression Model: A mathematical modeling method called logistic regression estimates the likelihood of an uncertain result, such as whether or not an application for funding is fraudulent.. It analyzes various features and calculates the likelihood of fraud based on their coefficients.

2. The Neural Networks: Another type of artificial intelligence model that draws inspiration from the human brain is the neural network. They are made up of layers upon layers of computer-generated neurons that are connected and process information. Using previous loan data, neural networks may be taught to identify trends and spot possible fraud.

3. Unorganized Forests: To create predictions, random forests mix several decision trees. A final prediction is arrived at by combining the output of all the decision trees, each of which is trained on a distinct subset of the data. The accuracy and capacity to handle big datasets of random forests are well established.

4. Supporting Vector Computers (SVM): SVM is a method of supervised learning that finds the best hyper plane to divide data into distinct classes. SVM can be used in loan detection for fraud to categorize loan applications as authentic or fraudulent according to a number of criteria.

Specific techniques for spotting loan fraud. Every approach has advantages and disadvantages, and the best approach will depend on the particular needs and attributes of the data set. Financial organizations can improve their capacity to detect fraud and make better loan decisions by integrating several strategies.

Proposed Methodology

It is well known that a financial gateway is a safe and secure platform that accepts transactions and sends data to verify the transaction's validity and permit the transfer. Still, in the case that an exchange is generated, it is critical to comprehend how the program operates in order to identify fraud. It is believed that the proposed program would be a part of the payment system that would confirm the authenticity of transactions.

It is essential to use data mining tools to detect loan fraud for a number of reasons:

The following factors make data mining tools for loan identification of fraud so important:

1. Identifying Patterns: Financial organizations can analyze enormous quantities of data and find underlying trends or patterns that can point to fraudulent activity by using data mining methods. By examining historical loan data, patterns of fraudulent activities can be identified, enabling the development of effective fraud detection models.

2. Improving Accuracy: Neural networks, logistic regression, and decision trees are examples of data analysis approaches, can analyze multiple variables simultaneously to make accurate predictions. More accurately than classic rule-based systems, these techniques consider a variety of parameters, including application data, credit history, and income, to determine the possibility of fraud.

3. Real-time Detection: Financial organizations can identify fraudulent activity in real-time by utilizing data mining tools. By continuously monitoring loan applications and comparing them to known patterns of fraud, suspicious activities can be flagged promptly, preventing potential losses.

4. Adaptability: Fraudsters are constantly evolving their techniques, making it necessary for fraud detection systems to adapt and stay ahead. Data mining techniques can be updated and refined to incorporate new patterns or emerging fraud trends, ensuring

that financial institutions remain proactive in detecting and preventing fraud.

5. Cost Reduction: Financial institutions can reduce the requirement for a manual evaluation of each loan application by automating the fraud detection process through the use of data mining tools. This automation saves time and resources while improving the efficiency and accuracy of fraud detection efforts.

Data mining techniques play a vital role in loan fraud detection by leveraging historical data, identifying patterns, improving accuracy, enabling real-time detection, adapting to evolving fraud techniques, and reducing costs for financial institutions.

Loan fraud detection requires the application of data mining tools for a number of reasons

Using data mining techniques for loan fraud detection is crucial for several reasons:

1. Identifying Patterns: Financial organizations can analyze vast amounts of data and find underlying developments or patterns that can point to fraudulent activity by using data mining techniques. By examining historical loan data, patterns of fraudulent activities can be identified, enabling the development of effective fraud detection models.

2. Improving Accuracy: Data mining methods include logistic regression, decision trees, and neural networks, can analyze multiple variables simultaneously to make accurate predictions. More accurately than classic rule-based systems, these techniques consider a variety of parameters, including application data, credit history, and income, to determine the possibility of fraud.

3. Real-time Detection: Financial organizations are able to identify fraudulent acts as they happen by using data mining tools in real-time. By continuously monitoring loan applications and comparing them to known patterns of fraud, suspicious activities can be flagged promptly, preventing potential losses.

4. Adaptability: Fraudsters are constantly evolving their techniques, making it necessary for fraud detection systems to adapt and stay ahead. Data mining techniques can be updated and refined to incorporate new patterns or emerging fraud trends, ensuring that financial institutions remain proactive in detecting and preventing fraud.

5. Cost Reduction: Financial institutions can reduce the requirement for a manual evaluation of each loan application by automating the fraud detection process through the use of data mining tools. This automation saves time and resources while improving the efficiency and accuracy of fraud detection efforts.

Related work

Contrast to the three other recent research articles, which are often application-or technique-oriented, in this paper, fraud detection is examined using a feasible performance-driven, data-oriented methodology. Also, compared to other previous survey articles, this one covers a wider range of pertinent fraud kinds, methods, and tactics and provides a comprehensive definition of the underlying technical issues. The several hybrid techniques and internal fraud are presented below, for instance. A few critiques of the fraud detection field as it is today are also provided, and potential future contributions from adjacent fields to data mining-based fraud detection are emphasized.

Conclusion & Future work

This survey has looked at most published fraud detection research. It defines terms related to adversaries, fraud types and subtypes, technical aspects of data, performance measurements, and strategies. This study highlights the limitations of current fraud detection techniques and protocols and then shows how this field could benefit from additional related subjects, In

particular, real surveillance systems and text mining from law enforcement, semi-supervised and game-theoretic methods from the intrusion and spam detection fields and unsupervised approaches from counterterrorism work can all contribute to future research on fraud detection. Future research will focus on detecting fraud in credit applications.

Reference

1. V. G. T. Costa, A. C. P. L. Carvalho, S. Barbon, "Strict Very Fast Decision Tree: a Memory conservative algorithm for data stream mining", May 2018.
2. A. Aye Khine, H. Wint Khine, "A Survey of Decision Tree for Data Streams to Detect Credit Card Fraud", PROMAC-2019.
3. F. A. Sunny, M. I. Khan, M. S. Satu and M. Z. Abedin, "Investigating external audit records to detect fraudulent firms employing various machine learning methods", Proceedings of the Seventh International Conference on Mathematics and Computing, pp. 511-523, 2022.
4. R. Van Belle, B. Baesens and J. De Weerdt, "Catchm: A novel network- based credit card fraud detection method using node representation learning", Decision Support Systems, pp. 113866, 2022.
5. Brabazon, J. Cahill, P. Keenan and D. Walsh, "Identifying online credit card fraud using artificial immune systems", IEEE Congress on Evolutionary Computation., pp. 1-7, 2010.
6. vAbbott, D., Matkovsky, P. & Elder, J. (1998). An Evaluation of High-End Data Mining Tools for Fraud Detection. Proc. of IEEE SMC98.
7. Aleskerov, E., Freisleben, B. & Rao, B. (1997). CARDWATCH: A Neural Network-Based Database Mining System for Credit Card Fraud Detection. Proc. of the IEEE/IAFE on Computational Intelligence for Financial Engineering, 220-226.
8. Goldenberg, A., Shmueli, G., Caruana, R. & Fienberg, S. (2002).Early Statistical Detection of Anthrax Outbreaks by Tracking Overthe-Counter Medication Sales. Proc. of the National Academy of Sciences, 5237-5249
9. Jensen, D., Rattigan, M. & Blau, H. (2003). Information Awareness: A Prospective Technical Assessment. Proc. Of SIGKDD03, 378- 387.
10. Shah. M. (2019). A Case Study on Increasing of Banking Frauds in India. Parichay: Maharaja Surajmal Institute Journal of Applied Research, 2(1), 20-23.
11. Dogra. A. Banking Frauds in India: Case Studies of Nirav Modi and Vijay Mallya Case. International Journal of Creative Research Thoughts, 6(1), 855-860.
12. R. Goyal, A. Kumar, "Review on Credit Card Fraud Detection using Data Mining Classification Techniques & Machine Learning Algorithms", International Journal of Research and Analytical Reviews, Vol. 7, No. 1, 2020, pp. 972-975.
13. G. Marinakos, S.Daskalaki, "Imbalanced customer classification for bank direct marketing", Journal of Marketing Analytics, Vol. 5, No. 1, 2017, pp. 14-30.

A. Abdallah, M. A. Maarof, A. Zainal, "Fraud detection system: A survey", Journal of Network and Computer Applications, Vol. 68, 2016, pp. 90-113.

14. K. Janaki, V. Harshitha, S. Keerthana, Y. Harshitha, "A Hybrid Method for Credit Card Fraud Detection Using Machine Learning Algorithm", International Journal of Recent Technology and Engineering, Vol. 7, No. 6S4, 2019, pp. 235-239.
15. M. A. Sheikh, A. K. Goel, T. Kumar, "An approach for prediction of loan approval using machine learning algorithm", Proceedings of the IEEE International Conference on Electronics and Sustainable Communication Systems, 2-4 July 2020, pp. 490-494.