Journal of Advanced Multidisciplinary
Research Studies and Development

ISSN : 2583-6404
Sep - Oct 2023

AN ADABOOST AND MASS VOTING METHOD FOR CREDIT CARD FRAUD DETECTION

Manisha Goyal
PhD Research Scholar, Department of Engineering, Thapar Institute of Engineering and Technology, Patiala, India.

Abstract

One significant risk element for the financial services industry is credit card theft. Countless millions of dollars are lost to credit card theft each year. Due to privacy concerns, investigations into the credit card industry data analysis have not been finished. This paper uses artificial machine learning algorithms to detect credit card fraud. Standard models are the first to be used. After that, hybrid strategies are applied using the majority vote and AdaBoost. The model's effectiveness is assessed using a publicly available credit card dataset. Next, real-world credit card data is used to analyse financial institution data. Noise added to the samples in order to assess the approach more thoroughly. The outcomes of the experiment demonstrate that identifying credit card fraud with a high degree of precision is possible using the majority voting technique.

Keywords: Adaboost, credit card, fraud.

► *Corresponding Author: Manisha Goyal*

Introduction

The article comprises a total of 12 number credit card fraud research algorithms. The approaches span from basic neural networks to deeper learning models study. These are assessed using the world's current standards and credit card sets. In the hybrids are also employed the Models of AdaBoost and majority voting. The noise in the actual world data set is included to further check the strength and dependability of models. An evaluate of many machine learning models using actual data for fraud detection is the major contribution of this work. The Whilst earlier studies have employed various approaches for data sets, the data set utilised is generated from real credit card transaction data for three months.

Existing System

Losses due to credit card fraud damage dealers who shoulder all costs, Includes issuer card fees, administrative costs and fees. As traders are forced to bear losses, some goods are forced to receive higher pricing or reduced rebates and income. Therefore the loss has to be reduced, and an effective method of identification of fraud is required if fraud incidences are to be decreased or eliminated. Various credit card screening studies have been conducted. The most frequently utilized methods and learning of machines are artificial neural networks, Inducing rules and strategies, taking decisions, logistical regression and support machines for vector systems. These methods are either used separately or combined to build hybrid models with other techniques.

Proposed System

The article includes a total of 12 credit card fraud research algorithms. The approaches vary from basic neural networks to models of deeper learning. These are assessed by means of existing credit card and benchmarks in the globe. In the hybrids are also employed the AdaBoost and the majority voting models. Noise is added to the actual world data set to further check the strength and dependability of models. An evaluate of many machine learning models with actual data for fraud detection is the major contribution of this research.

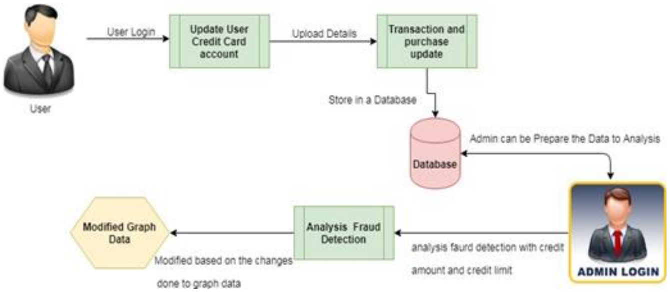
User
User Login
Update User Credit Card account
Upload Details
Transaction and purchase update
Store in a Database
Database
Admin can be Prepare the Data to Analysis
ADMIN LOGIN
analysis faurd detection with credit amount and credit limit
Analysis Fraud Detection
Modified based on the changes done to graph data
Modified Graph Data

Fig 1: Architecture Block Diagram

While previous research employ various techniques for data sets, during three months the data set utilized are based on actual information on a credit card transaction.

Machine Learning Algorithms

Machine learning is the science of algorithms that learn from past cases and apply them. It uses complicated algorithms, which iterate across massive data sets and evaluate data patterns. The method allows the machines to respond to many scenarios in which they were not expressly programmed. This is used for spam detection, image identification, product recommendation, predictive analysis and other purposes. The fundamental objective of data scientists in the implementation of ML is a significant decrease of human labour. It takes many times for humans to read, collect, classify and evaluate data even with contemporary analysis technologies. ML teaches robots to determine and measure the significance of human patterns. In circumstances of use in particular where data must be examined, supported by machines enables people to be more effective and to act on confidentially within a short period of time.

Module Implementation

Fraud Detection

Decision Tree (DT)

For user understanding it is beneficial to display facts in the form of a tree structure. The Decision Tree (DT) is a series of nodes that determine specific class properties. Each node is a rule that separates a feature. Until the condition halt is met, new nodes are generated. The name of the class of a certain leaf is based on most samples. Just like a DT operator, a Random Tree (RT) function is provided for all splits but just the Random Feature Sample. Samples of nominal and numerical data are used to learn this. The subset is determined by a sub-set ratio parameter.

Naïve Bayes (NB)

Naïve Bayes (NB) applies the theorem for classification of Bayes with strong or ingenuous assumptions of independence. Certain class characteristics are not believed to correspond with others. It just requires a minimal piece of training information to estimate methods, and classification variances.

The Random Forest (RF)

The Random Forest (RF) consists of a random arboreal ensemble. The number of trees will be selected by the user. In the resulting model, the final classification result is based on all created trees. The Boosted Gradient Tree (GBT) is a classification or regression model collection. It employs ensemble models for advanced learning that provide forecast results by incrementally improving estimations. Enhanced precision of the tree helps. The Stump Decision (DS) only produces a split decision tree. It can be used for classification of unequal data sets.

AdaBoost and Majority Voting

To increase their performance together with various sorts of algorithms, adaptive boosts or adaboost are used. The results are mixed with a high volume representing the combined output of the enhanced classifier. AdaBoost adjusts weak students to misclassified data. However, it is susceptible to sound and bands. AdaBoost can improve the results of diverse algorithms as long as classification performances are not random. AdaBoost helps to raise the rate at which fraud is detected and there's a remarkable difference in the accuracy rate for NB, DT and RT.

Objectives

1. Input Design is the process through which the input is user- oriented and transformed into a computer system. This is vital to ensure the avoidance of data input errors and to provide the correct management direction to receive correct information from the computerized system.
2. It is done by building user-friendly data entry windows for handling massive data volumes. The objective of input design is to facilitate data entering and to avoid any errors. The data entering page has been built to perform all manipulations of the data. It also offers recording watching equipment.
3. The validity of the data is checked when it is entered. With the help of displays, data may be entered. Suitable messages are sent whenever necessary in order to avoid instantaneous maize. The goal of the input design is to easily follow the input scheme.

Results

Input design: The input design is the link between the user and the information system. It involves the preparation of specifications and methods for read data from a written or printed document to translate data into usable form or by people directly incorporated into the system by means of a computer examination.The input design is designed to control the amount of input necessary, to control errors, to eliminate delays, to avoid additional stages and to keep it simple. The input is such designed to ensure the retention of privacy by providing security and convenience of usage. Input Design took into consideration the following:

- ➢ What data should be given as input?
- ➢ How the data should be arranged or coded?
- ➢ The dialogue to guide the operating staff to provide information.
- ➢ Methods for preparing validations and follow-up steps in the event of a mistake.

Output design:

A quality results fulfil the requirements of the end user and display data clearly.Any processing results are transmitted by means of outputs to users and other systems. In the output design the information must be moved to urgent need and the hard-copy output is also determined. It is the most crucial information of the user's direct source. Efficient and intelligent output development enhances the interaction between the system and helps users make decisions.

- ➢ Computer output should be built in an orderly, thorough way; The correct result should be determined while ensuring that each output item is intended to quickly and efficiently discover the system. If analysing the computer output, the exact result needed to satisfy the requirements should be identified.

- “Select methods for presenting information.
- Create documents, reports or other formats which contain information created by the system.

The result form of an information system should be one or more of the following objectives.

- Provide information on historical events, present conditions or future projections.
- Signal important events, opportunities, problems, or warnings.
- Trigger an action.
- Confirm an action

Conclusion

This article includes a study on the detection of fraud by credit card using machine learning techniques. In the empirical examination, a variety of standard models, such as NB, SVM and DL, were utilized. A credit card data set that is available in public has been used to evaluate AdaBoost and the majority voting combination technique using individual (standard) templates and Hybrid models. The MCC was considered as a performance metric, taking into consideration actual and erroneous positive and negative outcomes. 0.823 is the best MCC for the system. An actual data of the credit card has also been used for assessment by a financial institution. The same hybrid and individual models were used. In order to get a perfect MCC score of 1, the Adaboost algorithms and the majority vote were used. A noise of 10 to 30% in data samples was introduced to adequately assess hybrid models. The MCCscore of 0.942 was voted upon by the majority for 30 percent extra noise in the data set.This indicates that the majority voting approach provides strong noise performance

References

1.Y. Sahin, S. Bulkan, and E. Duman, “A cost-sensitive decision tree approach for fraud detection,” Expert Systems with Applications, vol. 40, no. 15, pp. 5916–5923, 2013.

2. A. O. Adewumi and A. A. Akinyelu, “A survey of machinelearning and nature-inspired based credit card fraud detection techniques,” International Journal of System Assurance Engineering and Management, vol. 8, pp. 937–953, 2017.

3. A. Srivastava, A. Kundu, S. Sural, A. Majumdar, “Credit card fraud detection using hidden Markov model,” IEEE Transactions on Dependable and Secure Computing, vol. 5, no. 1, pp. 37–48, 2008.

4. The Nilson Report (October 2016) [Online]. Available: https://www.nilsonreport.com/upload/content_promo/The_Nil son _Report_10-17-2016.pdf

5. J. T. Quah, and M. Sriganesh, “Real-time credit card fraud detection using computational intelligence,” Expert Systems with Applications, vol. 35, no. 4, pp. 1721–1732, 2008.

6. 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.

7. N. S. Halvaiee and M. K. Akbari, “A novel model for credit card fraud detection using Artificial Immune Systems,” Applied Soft Computing, vol. 24, pp. 40–49, 2014

8. S. Panigrahi, A. Kundu, S. Sural, and A. K. Majumdar, “Credit card fraud detection: A fusion approach using Dempster–Shafer theory and Bayesian learning,” Information Fusion, vol. 10, no. 4, pp. 354–363, 2009.

9. N. Mahmoudi and E. Duman, “Detecting credit card fraud by modified Fisher discriminant analysis,” Expert Systems with Applications, vol. 42, no. 5, pp. 2510–2516, 2015.

10. D. Sánchez, M. A. Vila, L. Cerda, and J. M. Serrano, “Association rules applied to credit card fraud detection,” Expert Systems with Applications, vol. 36, no. 2, pp. 3630– 3640, 2009

11. E. Duman and M. H. Ozcelik, “Detecting credit card fraud by genetic algorithm and scatter search,” Expert Systems with Applications, vol. 38, no. 10, pp. 13057–13063, 2011
12. P. Ravisankar, V. Ravi, G. R. Rao, and I. Bose, “Detection of financial statement fraud and feature selection using data mining techniques,” Decision Support Systems, vol. 50, no. 2, pp. 491–500, 2011
13. E. Kirkos, C. Spathis, and Y. Manolopoulos, “Data mining techniques for the detection of fraudulent financial statements, ”Expert Systems with Applications, vol. 32, no. 4, pp. 995–1003,2007.
14. F. H. Glancy and S. B. Yadav, “A computational model for financial reporting fraud detection,” Decision Support Systems,vol. 50, no. 3, pp. 595–601, 2011