Credit Card Fraud Detection using Machine Learning

MS Joitson[1]
[1]Student, Dept. of Computer Science & Engineering,
Mangalam College of Engineering, India,

Preejamol Prasad[2]
[2] Student, Dept. of Computer Science & Engineering,
Mangalam College of Engineering, India,

Rimil Joseph[3]
[3]Student, Dept. of Computer Science &Engineering,
Mangalam College of Engineering, India,

Jayakrishnan B[4]
[4] Assistant Professor,Dept. of Computer Science & Engineering,
Mangalam College of Engineering, India,

***Abstract*—The usage of credit cards for online purchases has grown significantly as a result of the E-Commerce sector's explosive growth, and as a result, there has been an increase in fraud associated to it.In recent years, it has become exceedingly challenging for banks to identify credit card system fraud. For transactions to be free of credit card fraud, machine learning is essential. Banks utilise a variety of machine learning approaches to forecast these transactions. Past data has been gathered, and new elements have been added to increase the predictive power. The choice of variables, sampling strategy for the data set, and detection methods all have a significant impact on the effectiveness of fraud detection in credit card transactions. The study examines the effectiveness of decision trees, random forests, and logistic regression for detecting credit card fraud. 2,84,808 credit card transactions from a European bank are included in the dataset of credit card transactions that was obtained from Kaggle. It views fraudulent transactions as belonging to the "positive class" whereas legitimate ones are in the "negative class." Based on sensitivity, specificity, accuracy, and error rate, the performance of the approaches is assessed for a variety of factors.**

Keywords: Fraud detection, Credit card, Logistic regression, Decision tree, Random forest.

1. INTRODUCTION

To provide e-commerce, information, and communication services to their clients more effectively and more conveniently, businesses and organisations are shifting some or all of their operations online. Worldwide, there is a significant issue with payment card fraud. Annual losses from fraud are enormous for businesses and institutions, and criminals are always looking for new ways to engage in unethical behaviour. We'll strive to spot any fraudulent activity here. When a credit card is cloned, stolen, lost, and found by fraudsters, it is typically utilised up to its available credit limit. Therefore, a solution that reduces the total available limit on cards susceptible to fraud is more important than the quantity of correctly classified transactionsIt uses a genetic algorithm to minimise false alarms by optimising a collection of interval-valued parameters.Generally speaking, fraud is unapproved conduct that occurs in electronic payment systems; this activity ought to be prohibited by law and is treated as unlawful. In many diverse fields, including financial systems, telecommunications, and public and private services, fraud can occur. It is focused on identifying fraudulent credit card transactions and is concerned with financial scams. The classification challenge of fraud detection has been addressed by numerous data mining algorithms using some statistical techniques. more common among decision trees. Fraud detection has typically been a part of data mining and e-commerce. It optimises a group of interval-valued parameters using a genetic algorithm to reduce false alarms.In general, fraud is unapproved behaviour that takes place in electronic payment systems; this behaviour should be illegal and outlawed by law. Fraud can happen in many different industries, such as financial systems, telecommunications, and public and private services. It is concerned with financial fraud and focuses on recognising fraudulent credit card transactions. Numerous data mining algorithms that employ some statistical techniques have been developed to address the classification challenge of fraud detection. more prevalent in decision trees. Data mining and e-commerce have traditionally included fraud detection.

2. LITERATURE REVIEW

Fraud is defined as an illegal or criminal deception meant to produce a monetary or personal profit. It is an intentional action that violates a rule, legislation, or policy with the intention of obtaining unrecognised pecuniary benefit. There is a wealth of publicly accessible material on the topic of anomaly or fraud detection in this field that has already been published. Data mining applications, automated fraud detection, and adversarial detection are among the strategies used in this field, according to a thorough review undertaken by Clifton Phua and his coworkers. Suman, a research scholar with GJUS&T at Hisar HCE, discussed methods for detecting credit card fraud, such as supervised and unsupervised learning, in a different study. Even though some of these techniques and algorithms achieved unexpected success, they were unable to offer a reliable, long-lasting answer to fraud detection. Wen-Fang YU and Na Wang presented a similar area of research in which they employed distance sum algorithms, outlier mining, outlier detection mining, and outlier detection mining to precisely forecast fraudulent transactions in an experiment simulating credit card transaction data from a particular commercial bank. Data mining's field of outlier mining is primarily utilised in the financial and internet sectors. It focuses on identifying detached objects from the main system, or transactions that aren't real. The distance between the observed value of an

attribute and its predetermined value was calculated using customer behaviour attributes and their respective values. Unusual methods like hybrid data mining/complex network classification algorithms, which are based on network reconstruction algorithm and allow creating representations of the deviation of one instance from a reference group, have typically proven effective on medium-sized online transactions. These methods are able to detect illegal instances in a real card transaction data set. There have also been initiatives to advance from an entirely new perspective. In the event of a fraudulent transaction, efforts have been made to improve the alert-feedback interaction. A feedback would be issued to the authorised system to deny the current transaction in the event of a fraudulent transaction. One method that provided new insight into this area dealt with fraud in a different way: Artificial Genetic Algorithm. It worked well at identifying fraudulent transactions and reducing the incidence of false alarms. Even so, there was a categorization issue with fluctuating misclassification costs.

3. PROBLEM DEFINITION

Clarifying the issues with the current system is the goal of the early stages of the research. According to the preliminary analysis of the current system, it is discovered to have various problems, such as an outdated database for storing information that is inflexible and inefficient. Updating and maintaining is difficult. It takes some time to look for a certain entry or record. The database is accessible to a large number of individuals. As a result, it offers less security for the database's data.Thus, the proposed system is keeping an eye on all of these concerns to improve the efficiency and security of information recording within the organisation.

4. PROPOSED SYSTEM

Simple system is offered in the system. which is able to identify frauds by taking into account a cardholder's spending patterns even though it does not require fraud signatures. Any Fraud Detection System (FDS) operating at the bank that gives credit cards to the cardholders often does not have access to the specifics of things purchased in Individual Transactions. Therefore, we believe that this is the best option for solving this issue. Another significant benefit is a sharp decline in the amount of transactions labelled as malicious by an FDS despite being legitimate. A bank that issues credit cards operates an FDS. The FDS receives every incoming transaction for verification. To determine if the transaction is legitimate or not, FDS receives the card information and the purchase amount. The FDS is unaware of the products that were purchased in that transaction. Based on the erroneous pin entry exceeding three, it looks for any anomalies in the transaction and attempts to block the card if there are more than three. If a card is blocked three times, it is considered to be a fraudulent card, and the block of a fraudulent card cannot be removed by the administrator. Every card that has been registered will send mail.

5. IMPLEMENTATION

The system must be put into use and the relevant documentation must be given to users and operation staff so they can use and maintain the new system. The process of converting from the old system involves all of the activities that are involved in implementation. A reliable system must be implemented properly in order to satisfy organisational needs. Improvement in the organisation using the new system may not be guaranteed by successful implementation, and poor installation will hinder this.There are 4 methods

- Parallel operation: The new system is used in conjunction with the old system.
- Direct cut over: The old system is switched out for the new one.
- Pilot approach: Based on the input, a working version of the system is implemented in one area of the company. Changes are then made, and the system is then installed using one of the other ways in the remaining areas of the company.
- Phase-in method: Rolls out the system gradually to all users.

6. SYSTEM SPECIFICATION

The software requirements specification (SRS) is a tool for formalising the concepts that clients have in their heads. The development and software validation are formed by this document. The fact that there are three parties interested—the clients, the end users, and the software developer—is the primary cause of the difficulty in defining software requirements. The requirements document needs to be clear enough for the client, the user, and the developers to utilise as a starting point for software development. There is a communication gap in software requirement formulation because of the variety of parties involved. This communication gap occurs when either the client or the software development processor does not grasp the client's problem. SRS's application area fills this communication gap.
An in-depth grasp of the clients' and users' needs as well as the specific requirements for the programme is obtained through problem analysis. The actual specification is the result of analysis. Analysts, who carry out the study, are also in charge of outlining the criteria. The needs of the client are what drive the software project. These needs are initially on the minds of several individuals within the customer organisation. By speaking with these individuals and comprehending their needs, the requirement analyst must determine their requirements. The analyst's primary sources of knowledge are these persons and the records already in existence that describe the current style of functioning.

7. METHODOLGY & ALGORITHMS

7.1 METHODOLGY

Design methodology describes the process of creating a system or approach for a certain circumstance.The importance of brainstorming in design approach is emphasised, along with the use of cooperative thinking to sort through each concept put out and find the optimal answer. The most important consideration is meeting the demands and requirements of the end user. In order to implement design approach and satisfy the desired user needs, numerous studies and tests have been conducted. In this software, each input made by the user is tested. This means that each piece of data is verified for

accuracy, and if necessary, prompting and warning messages are shown. The output forms are also thoroughly examined to determine whether or not the required output is achieved. Additionally, the output forms are improved for user comprehension by being made clearer and more meaningful.

7.2 .DECISION TREE

The process of learning decision trees from class-labeled training tuples is known as decision tree induction. In a decision tree, each internal node (non-leaf node) symbolises a test on an attribute, each branch reflects the test's result, and each leaf represents the test itself. A class label is stored in a node (or terminal node). The root node is the node at the top of a tree.Decision tree classifier building is suitable for exploratory knowledge discovery because it doesn't require parameter configuration or domain understanding. Multidimensional data can be handled via decision trees. A. Their depiction of learned knowledge in a tree is not intuitive and is generally simple for humans to understand. Decision tree induction's learning and classification phases are easy and quick. Decision tree classifiers typically have good accuracy. However, the data at hand may have an impact on how well it is used. Numerous application areas, including medicine, manufacturing and production, financial analysis, astronomy, and molecular biology, have used decision tree induction algorithms for classification. The foundation of several commercial rule induction systems are decision trees.

8. MODULES

8.1 GUEST MODULE

- Registration of users

This module assists the visitor in entering their information on the website to register for additional purchases and other services.

8.2 ADMIN MODULE

- Add Category

The administrator of this module may add new product categories. The category might be Computers,Laptops & Gaming, Mobiles & Tablets, etc

- Add Product

The administrator of this module may add new products and their data. Each product may be added in conjunction with the relevant category and subcategory. The product information contains the quantity, unit cost, picture, etc.

- Managing Products

Administrators can change the supply, price, name, and image of a product in this module.The administrator may receive a request from a user for a new credit card.

8.3 USER MODULE

- View All Products

The user can view the product image, model name, price,stock, etc. with the aid of this module. The user adds the necessary goods to the shopping cart.

- Acquire Products

This module assists the user in making an online purchase by allowing them to pay with a shopping card. If the right card number and pin number are input, the purchase of the items in the shopping basket will be completed.

- Obtain a fresh shopping card

In this module, users can request new cards and purchase up to 3 cards by paying from their wallet balance. The administrator grants the request and issues the user a new card can receive mail. That mail will contain the shopping card number, pin code, and card validity.

- Unblock Demand

The incorrect entry of a pin number results in the blocking of a shopping card; if an unauthorised user enters the pin number incorrectly, the warning "the pin is wrong" is displayed the first time. When he enters the wrong pin a second time, it displays "Continues wrong entry of pin number cause Card Block" and the card is blocked. Then the user sends the admin a request to unblock.

9. EXPERIMENT AND RESULTS

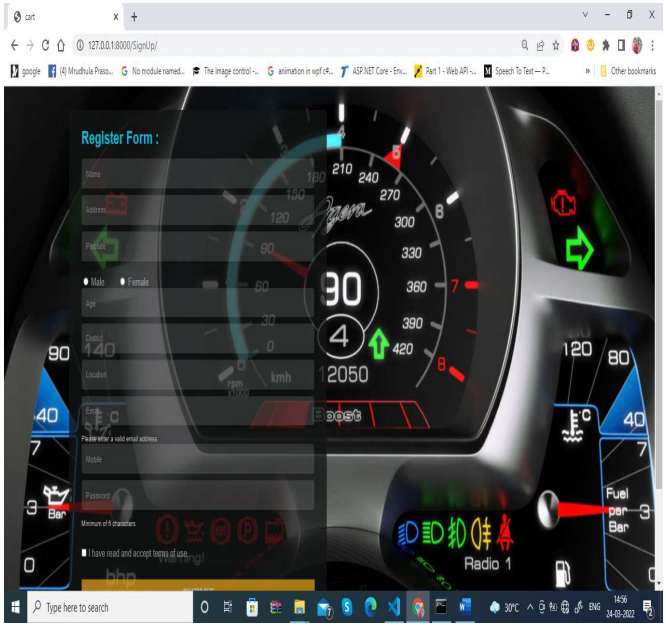

Fig 1. Registration Form

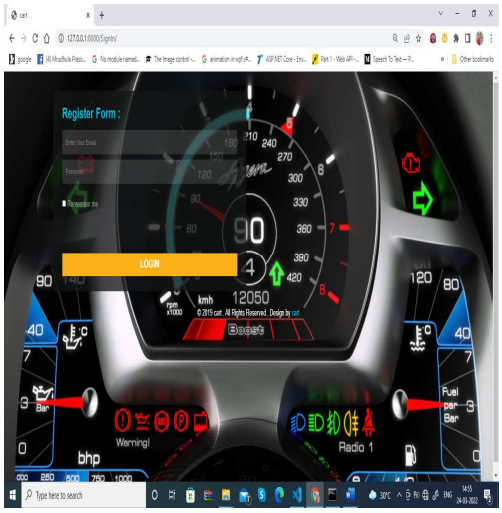

Fig 2.Login page

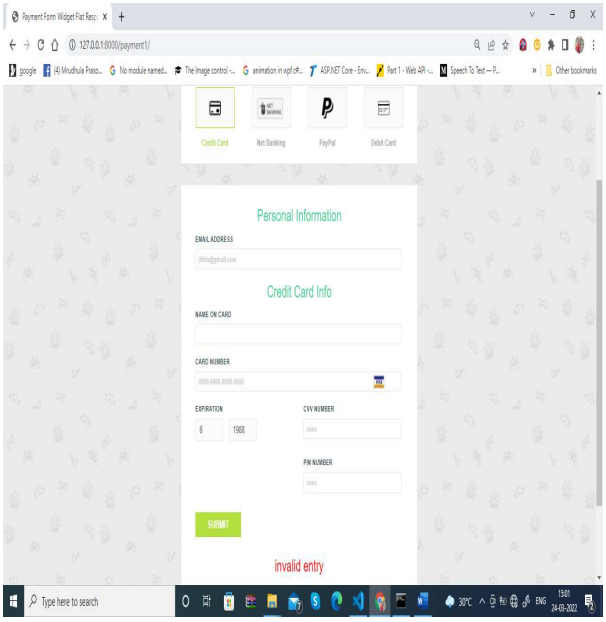

Fig 3.Creditcard information

10 . CONCLUSION

Unquestionably, using a credit card fraudulently is a criminal act of dishonesty. The most popular fraud schemes, as well as how to spot them, are listed in this article, which also reviews recent research in the area. Additionally, a thorough explanation of how machine learning can be applied to improve fraud detection outcomes, along with the algorithm, pseudocode, justification for its use, and results of experimentation. The algorithm does achieve over 99.6% accuracy, however when only a tenth of the data set is considered, its precision is still only 28%. The precision increases to 33% when the system is fed the whole dataset, though. This high accuracy rate is expected given the vast disparity between the number of transactions that are valid and those that are genuine. If this initiative were to be used on a commercial scale, a small portion of the data may be made public.Based on machine learning techniques, the program's effectiveness will only grow as more data is sent into it.

REFERENCES

[1] Andrew. Y. Ng, Michael. I. Jordan, "On discriminative vs. generative classifiers: A comparison of logistic regression and naive bayes", Advances in neural information processing systems, vol. 2, pp. 841-848, 2002.

[2] A. Shen, R. Tong, Y. Deng, "Application of classification models on credit card fraud detection", Service Systems and Service Management 2007 International Conference, pp. 1-4, 2007.

[3] A. C. Bahnsen, A. Stojanovic, D. Aouada, B. Ottersten, "Cost sensitive credit card fraud detection using Bayes minimum risk", Machine Learning and Applications (ICMLA). 2013 12th International Conference, vol. 1, pp. 333-338, 2013.

[4] B.Meena, I.S.L.Sarwani, S.V.S.S.Lakshmi," Web Service mining and its techniques in Web Mining" IJAEGT,Volume 2,Issue 1 , Page No.385-389.

[5] F. N. Ogwueleka, "Data Mining Application in Credit Card Fraud Detection System", Journal of Engineering Science and Technology, vol. 6, no. 3, pp. 311-322, 2011.

[6] G. Singh, R. Gupta, A. Rastogi, M. D. S. Chandel, A. Riyaz, "A Machine Learning Approach for Detection of Fraud based on SVM", International Journal of Scientific Engineering and Technology, vol. 1, no. 3, pp. 194-198, 2012, ISSN ISSN: 2277-1581.

[7] K. Chaudhary, B. Mallick, "Credit Card Fraud: The study of its impact and detection techniques", International Journal of Computer Science and Network (IJCSN), vol. 1, no. 4, pp. 31-35, 2012, ISSN ISSN: 2277-5420.

[8] M. J. Islam, Q. M. J. Wu, M. Ahmadi, M. A. Sid Ahmed, "Investigating the Performance of Naive-Bayes Classifiers and KNearestNeighbor Classifiers", IEEE International Conference on Convergence Information Technology, pp. 1541-1546, 2007.

[9] R. Wheeler, S. Aitken, "Multiple algorithms for fraud detection" in Knowledge-Based Systems, Elsevier, vol. 13, no. 2, pp. 93-99, 2000.

[10] S. Patil, H. Somavanshi, J. Gaikwad, A. Deshmane, R. Badgujar, "Credit Card Fraud Detection Using Decision Tree Induction Algorithm", International Journal of Computer Science and Mobile Computing (IJCSMC), vol. 4, no. 4, pp. 92-95, 2015, ISSN ISSN: 2320-088X.

[11] S. Maes, K. Tuyls, B. Vanschoenwinkel, B. Manderick, "Credit card fraud detection using Bayesian and neural networks", Proceedings of the 1st international naiso congress on neuro fuzzy technologies, pp. 261-270, 2002.

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

[13] Y. Sahin, E. Duman, "Detecting credit card fraud by ANN and logistic regression", Innovations in Intelligent Systems and Applications (INISTA) 2011 International Symposium, pp. 315-319, 2011.