ORIGINAL RESEARCH PAPER

Machine Learning

FRAUD DETECTION OF CREDIT CARD USING MACHINE LEARNING

KEY WORDS: Machine Learning, Fraud detection, Confusion matrix, Accuracy, F1 score.

Shubham Damodar Potphode
M.Tech, Department of Computer Science and Engineering, Government College of Engineering, Aurangabad.

Prof. Arjumand Masood Khan
Assistant Professor, Department of Computer Science and Engineering, Government College of Engineering, Aurangabad.

ABSTRACT

In today's world, pressing issue, and it needs to be addressed. In order to conceal the origins of filthy money, a common practice. Detecting credit card fraud on the worldwide market is more difficult due to the constant movement of money. Although (Anti-credit card fraud Suite) has been set up to identify suspicious activity, it only works for transactions performed and not for transactions on other accounts, as we've previously said This is why we've developed a machine learning technique called "Similarity," which searches for characteristics and behaviors that are similar to those of previous banking transactions. We use case-reduction approaches to limit the quantity of data needed to be supplied and then locate pairings of transactions with other bank accounts that have similar traits and behaviors.

INTRODUCTION

Credit card fraud eats up to 5 percent of the global GDP annually (Gross Domestic Product). Using AI to combat credit card fraud is intended to detect suspicious behavior. To combat the majority of businesses that conduct financial transactions must maintain comprehensive records of their clients' accounts and transactions. They are tent to report any suspicious information to the government to inspect. If suspect data is discovered, the transaction records are examined employ AI and Machine Learning Algorithms are used to detect suspicious transactions and resolve them by training on data pertaining to the suspicious activity. We will use both unsupervised and supervised algorithms. Credit card reader has been in existence for a number of years, and with each passing day the model has grown more robust and refined. Machine learning can help to predict fraud in database there are various classification problems that can solve by ML algorithms.

MOTIVATION

Accounts susceptible to credit card fraud transection can be detected and categorized. Using the Haar Cascade Algorithm, the objective is to develop a computer programme that will recognize and segregate tweets based upon text and pictures during catastrophe situations into informative and non-informative categories. System that is user-friendly. Accounts susceptible detected and categorized.

OBJECTIVE

1. Possibility of establishing credit. 2. Earn cash back or miles points as a reward. 3. Security regarding credit card online fraud 4. Access to credit score information with is. There are no such a foreign transaction costs. 5. Enhanced power of purchasing there is no connection to a checking on savings account. 6. Making a reservation for a rental car some time hotel room.

PROBLEM STATEMENT

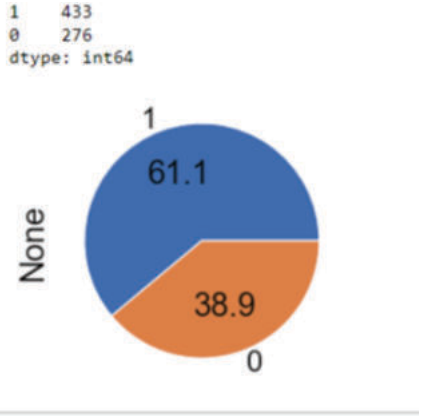
The consumption of the internet, social networking and websites is rising. Social media is source of a wide range of structured and unstructured information. It is difficult for people to obtain efficient, reliable, and timely information. Problems with decision-making. This project compares and predicts based on user feedback in order to make more informed purchasing decisions and save time.

DATASET AND METHODOLOGY

We have taken Capture Image dataset, which used in our machine learning studies, in this section. For this we have collected data and store in database .there are 3 database file which contains card holder details, merchant details and fraud database which consists of 709 transactions and have 9 attributes. The status attribute represent fraud or non-fraud transection.

The value 0 indicates frauds transaction and value 1 non frauds transactions. There are 433 non frauds transections and 276 frauds transactions. We have used some machine learning algorithms to predict fraudulence in dataset. Algorithms like svm, naïve bayse, decision tree. We also comparing accuracy of each algorithm. As we see in Fig 1.the dataset has more non fraud transection compare to fraud once but it will not effect on the result of an algorithm for predicting frauds.

Data prepossessing is an important aspect to manage dataset by doing prepossessing we get quality data for execution. Dataset can have null values we need to deal with it so that we get proper output.



```
1    433
0    276
dtype: int64
```

Figure 1: Graphical View Of Dataset.

The data pre-processing processes that we used before applying machine learning algorithms will be discussed. As shown below in figure 2 features of the dataset need to be processed so it has good understanding of the dataset.

System Architecture

In system there are mainly four stages firstly we need to upload the dataset. After that we will be doing proses on the dataset live cleaning after that we will apply ML algorithms to classify the fraud transections. Then we will store that on to the system.

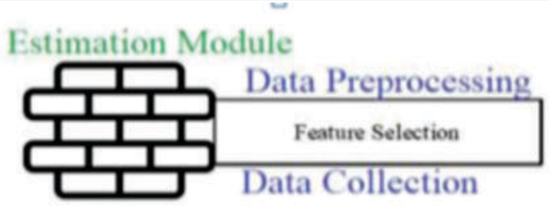


Figure 2: Data Pre-processing Processes

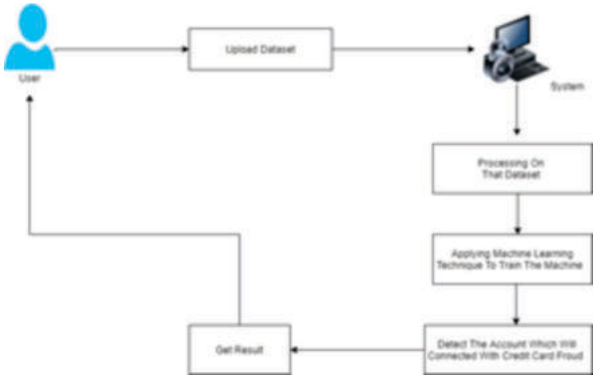

User
Upload Dataset
System
Processing On That Dataset
Applying Machine Learning Technique To Train The Machine
Detect The Account Which Will Connected With Credit Card Fraud
Get Result

Figure 3: System Architecture

The Haar cascade method is being used in this project. The system can communicate via the Internet using the HTTP protocol, while intranet communication will be done using the TCP/IP protocol suite also the fraud will be detected using SVM algorithm. The dataset must be large enough so that the algorithm can be implemented optimistically.

LITERATURE SURVEY

Acording to Tanouz, D ML algorithms are used for detecting various problems of fraud. Various algorithms are used to detect fraud. To find which the best algorithms for classification problem is that can be measure with various aspects. On the basis of accuracy, recall, precision, confusion matrix effectiveness of the model can determine [1].

From a research conducted by Dongxu Huang, Dejun Mu, Libin Yang, Xiaoyan Cai,the finding ewre that, In recently, financial fraud activities, i.e credit card fraud, have gradually expanded. These actions result in the loss of private and/or business property. Worse, they endanger national security by directing fraud earnings to terrorism. Therefore, detecting and tracing financial fraud accurately is both necessary and urgent. Nevertheless, detecting financial fraud is challenging due to the complex trade networks and transactions involved. Credit card fraud, for instance, is defined as the use of commerce to move money/goods with the intention of obscuring the true source of funds [2].

As per Reza Soltani, Uyen Trang Nguyen, Yang Yang, Mohammad Faghani, Alaa Yagoub, Aijun There are a variety of ways to commit credit card theft. Casinos and real estate might be used to conceal the source of the funds, as could the inflation of legal tender. Once someone has obtained it is typically used to rack up additional fraudulent charges. Placing unlawful funds into the financial system occurs in a variety of ways known as placement. In order to conceal the source of one's funds, the practise known as "layering" entails conducting numerous, intricate financial transactions. Finally, the funds will be withdrawn from the designated bank account. Credit card fraud detection software is designed to be fooled via multi-layered deception [3].

Fahimeh Ghobadi has said that Credit card fraud is on the rise because of how quickly e-commerce and electronic payment systems have grown. The goal of this article is to come up with a credit card fraud detection (CCFD) model based on Artificial Neural Networks (ANN) and the Meta Cost approach. This will help reduce loss risk and risk to reputation. The ANN approach has been used to find and stop different kinds of data, it is hard to find fraudulent transactions (Fraud and Non-Fraud cases). The Meta Cost approach data that is not balanced. The Artificial Immune System-based approach is more expensive and doesn't work as well as this model (AIS). The data for this study come from real transaction data that a large Brazilian credit card company gave to the researchers [4].

Tanmay Kumar Behera stated that Because of the rapid advancement of e-commerce and online banking, the usage of credit cards has expanded significantly, resulting in a considerable number of fraud instances. The first phase performs initial user authentication and card credentials verification. If the check clearance is successful cleared, the transaction proceeds to the following phase, where a fuzzy means clustering method is used to determine the nominal usage pattern of given credit card users based on their previous behaviour. When a transaction is suspected of being fraudulent, a neural network-based learning method is used to assess if it was a fraudulent activity or an occasional deviation by a righteous user. Intense work along stochastic models demonstrates that combining the clustering technique with learning aids in effectively detecting fraudulent activity while decreasing the occurrence of false alarms [5].

Andrea Dal Pozzolo, Giacomo Boracchi Credit card theft is a great approach to determine if artificial intelligence is up to the task. Concept drift (consumer behaviour changes over time, and fraudsters' schemes improve) and class imbalance (legal transactions are significantly more numerous than fraudulent transactions) are only a few of the major issues with this issue (only a small set of transactions are timely checked by investigators). As compared to these algorithms, most of the proposed methods for learning to detect fraud are predicated on assumptions that are extremely unlikely to be true in the real world (FDS) [6].

As per Chunzhi WangYichao. The research shows a technology for detecting fraud that is based on an optimised whale algorithm BP neural network. The goal is to fix the BP neural network's problems, such as its slow convergence speed, its tendency to fall into local optimum, network faults, and poor system stability. To use the whale swarm optimization technique to find the best weight for a BP network, we first use the WOA algorithm to find the best initial value. Then, we use the BP network method to fix the wrong initial value and find the best weight [7].

Gokula Krishnan., Dhinesh Raj tell us about Technological advancements have altered our way of life. Credit cards have been introduced by banks. Credit card use has expanded as electronic commerce technology has advanced, and common means of payment for both online and offline purchases. Despite their immense appeal, the cards are not without risk. However, the great majority of learning algorithms proposed assumptions that are unlikely to remain true in a real-world scenario of fraud-detection system. Our project was primarily concerned with detecting credit card theft in the real world. Initially, we will collect credit card datasets for training. Then, for the testing data set, we will offer the user with credit card queries. Following final optimization, the findings show that the optimal accuracy for the Random Forest Algorithm is 98.6[8].

As per Anusorn Charleonnan Currently, corporate systems are focusing on credit card expenditure services in generally because it is an efficient method of paying for goods and services. Thus, the goal of this work is to employ RUS and MRN machine learning to identify credit card payment fraud. The proposed approach employs three basic classifiers: the NB, MLP and Nave Bayes algorithms. It can also judge if dealing with imbalanced datasets is correct. The information was then used to create a prediction of whether the payment risks were correct. The findings show that the suggested approach has the most accurate and sensitive classification performance [9].

According Liyun He-Guelton a lot of people use machine learning and data mining to find credit card fraud. But buying habits and ways to scam people will change over a period of time. It is called "dataset shift" or "concept drift" within the industry of fraud detection. From this study, we give to figure out how much our database of face-to-face credit card transactions changes every day (Owner of card located in the shop). Practically, we compare the days with each other and judge how well they were put into groups. The more varying the buying habits are between two days, the more accurate the classification, and vice versa [10].

As Anuruddha Thennakoon tell frequent credit card fraud results in enormous financial losses. As the online transactions has grown up by leaps and bounds, so has the number of online credit card transactions. Because of this, banks and other financial institutions put a lot of stock in detection programmes and put a lot of pressure on them to work well. Fraudulent transactions can look different and fit into a number of different groups. Each fraud is stopped by a set of machine learning models, and the best solution is found by evaluating the models. This evaluation shows a good performance metric and gives detailed instructions on how to choose the best algorithm based on the Another important part of our project is that it will help find in real time[11].

Confusion Matrix:

A confusion matrix is used to measure performance of various techniques for machine learning algorithms or classifiers. It is a type of table that allows you to determine how well a classification model performs on a set of test data so that the real values may be determined.

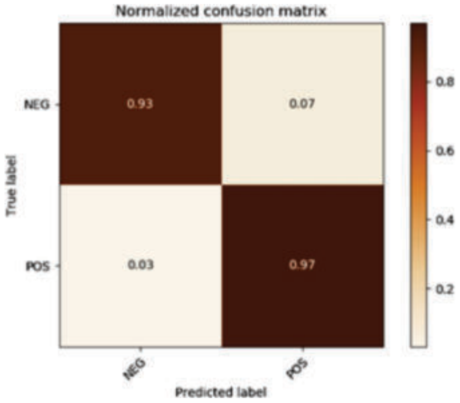


Figure 4: Normalized confusion matrix

F1 Score

A substitute for Accuracy, F-score is a machine learning model performance statistic that equally weights Precision and Recall when assessing how accurate the model is. The model score as a function of recall and accuracy is represented by the model F1 score

Formula for calculating F1 score is

F1 Score = 2 * Precision * Recall Score/ (Precision + Recall Score/)

F1 scores directly implies how good the accuracy of the model. It's like a performance matrix to calculate the performance of the classifier.

Recall

Recall is the ability of the machine learning model to predict positive in actual positive value. Formula for calculating recall is

Recall=TP/ (TP+FP)

Recall is depends on true positive and false positive predictions then that decide performance of ML model.

From given table we sees that svm model is more accurate and has more recall time which make most suitable model to apply on the dataset and get work done.

We compare some ML models base on their recall score, F1 score and accuracy as shown in table 1.

Table – 1 Statics Of Models

	SVM	**Decision Tree**	**Naive Bayes**
Recall Score	96%	90.66%	92.34%
F1 Score	88.5%	80%	89%

Accuracy:

The ratio of true positives and true negatives to all positive and negative observations is referred to as the model accuracy, which is a performance statistic for machine learning classification models. In other words, accuracy indicates the proportion of times our machine learning model will predict a result accurately out of all the predictions it has made.

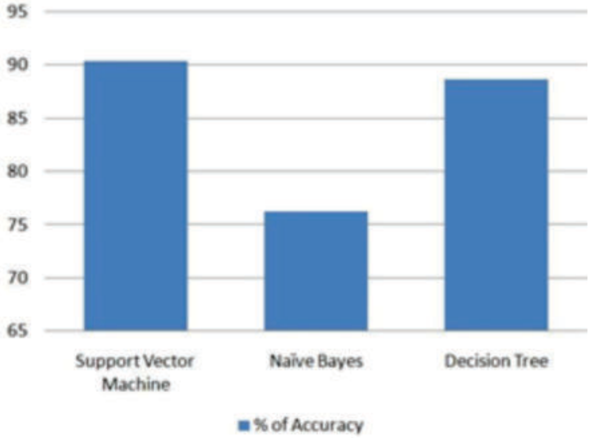


Figure 5: Accuracy of models with Graphical represent

CONCLUSIONS

Last but not least, it is said that credit cards are inherently safe. Credit cards have worked well in the media business. In the long run, they will replace everything we carry in our pocket now, including plastic money. Credit cards could be part of the answer to a security problem in today's world. To predict fraudulence transection we need to consider more machine learning algorithm and al system.

REFERENCES:

[1] Tanouz, D., et al. "Credit card fraud detection using machine learning." *2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS)*. IEEE, 2021..

[2] Dhanapal, R., and P. Gayathiri. "Credit card fraud detection using decision tree for tracing email and IP." *International Journal of Computer Science Issues (IJCSI)* 9.5 (2012): 406.

[3] Patidar, Raghavendra, and Lokesh Sharma. "Credit card fraud detection using neural network." *International Journal of soft computing and Engineering (IJSCE)* 1.32-38 (2011).

[4] Srivastava, Abhinav, et al. "Credit card fraud detection using hidden Markov model." *IEEE Transactions on dependable and secure computing* 5.1 (2008): 37-48.

[5] Talekar, Dinesh L., and K. P. Adhiya. "Credit card fraud detection using hmm and image click point authentication." *International Journal of Advanced Studies in Computers, Science and Engineering* 4.3 (2015): 1.

[6] Bhusari, V., and S. Patil. "Study of hidden markov model in credit card fraudulent detection." *International Journal of Computer Applications* 20.5 (2011): 33-36.

[7] Patel, Rinky D., and Dheeraj Kumar Singh. "Credit card fraud detection & prevention of fraud using genetic algorithm." *International Journal of Soft Computing and Engineering* 2.6 (2013): 292-294.

[8] RamaKalyani, K., and D. UmaDevi. "Fraud detection of credit card payment

system by genetic algorithm." *International Journal of Scientific & Engineering Research* 3.7 (2012): 1-6.

[9] Charleonnan, Anusorn. "Credit card fraud detection using RUS and MRN algorithms." *2016 Management and Innovation Technology International Conference (MITicon)*. IEEE, 2016.

[10] Lucas, Yvan, et al. "Dataset shift quantification for credit card fraud detection." *2019 IEEE second international conference on artificial intelligence and knowledge engineering (AIKE)*. IEEE, 2019.

[11] Thennakoon, Anuruddha, et al. "Real-time credit card fraud detection using machine learning." *2019 9th International Conference on Cloud Computing, Data Science & Engineering (Confluence)*. IEEE, 2019.