Comparing Machine Learning Models for Credit Card Fraud Detection

Table of contents

Keywords

Machine learning algorithms, credit card, fraud detection, comparison

Introduction

The ratio of credit card fraud has grown in tandem with the rise in digital and online fraud. Many situations have occurred in which an individual gained access to personal and credential information to engage in unethical behavior. Credit card fraud is one of the most prevalent challenges that consumers encounter, in which thieves obtain access to the card and PIN either through stolen information or by hacking for financial activities from individual accounts without authorization (Carcillo *et al.* 2021). It may happen in two ways: physically snatching the card or digitally hacking the data, and it generates major problems for the cardholders. Fraudsters use a variety of methods to gain access to cards, including misplaced cards, skimming cards, stealing information from mail, phishing efforts, and computer hacking.

Credit card fraud has become one of the most significant concerns confronting individuals in this era of digital payment. The world is on its way to becoming a cashless society. Most people currently prefer to pay with a debit or credit card, making the security of their credentials a primary priority. Credit card fraud detection is critical to improve the ecosystem by reducing fraud events. With the passage of time, fraudsters are developing new methods for gaining access to personal information of financial cardholders.

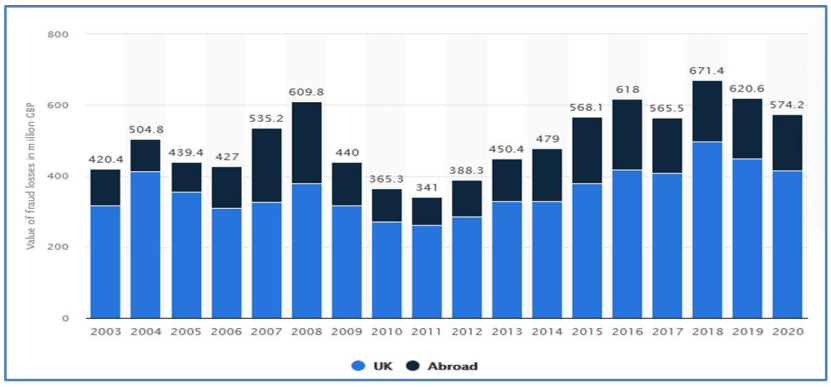
Machine learning may be applied in a range of ways to detect and prevent credit card fraud. The specifics on how to identify fraud using machine learning will be provided in this study. There are several sorts of machine learning algorithms available for credit card fraud detection, just as there are various ways to perpetrate credit card fraud (Ruehle 2020). The many forms of machine learning algorithms will be examined to determine the optimum algorithm for detecting credit card fraud. Once fraud is discovered, it will help to reduce credit card fraud and make cashless transactions more convenient for everyone. The study is interesting because

it outlines specific methods for detecting credit card fraud, which might be useful in the future in a variety of ways. After reading the research, a person will be more aware of the deception

and will be more cautious in the future. Machine learning has the ability to do specified actions using a variety of methods and procedures. The detailed techniques of machine learning will be evaluated and analyzed as part of the fraud detection research.

The challenges that come with using machine learning algorithms are:

- Highly imbalanced datasets: Majority of classifiers are not built to deal with uneven data. When the imbalanced scenario occurs, traditional machine learning algorithms will categorize all occurrences as majority class observations to optimize overall accuracy. As a result, the minority class will have poor accuracy, as seen by a low recall rate.
- Performance measures: Predictive accuracy is a technique of assessing the efficacy of machine learning algorithms; nevertheless, it is not appropriate in all cases, particularly when the dataset is unbalanced or the rates of various mistakes vary dramatically.



Total value of annual debit and credit card fraud in UK

Source: ("Total value of fraud losses on UK-issued credit cards-2020 | Statista", 2022)

The above figure shows the value of fraud losses abroad and in the UK from 2003 – 2020. This study will examine the performance of three machine learning methods, including decision tree, logistic regression, and random forest, to find the best match.

Literature Review:

Machine learning techniques are important in a range of effective data processing applications, including card fraud detection. Prior studies have proposed supervised approaches, unsupervised methods, and a hybrid strategy for identifying fraud; this needs knowledge of some of the technology used in credit card fraud detection, as well as a better awareness of the many forms of card fraud .Several options were proposed and investigated, three of which will be discussed below.

Credit card frauds have become all too common in the digital world. Humans who stay at home like to have complete control over their surroundings. It encourages the use of e-commerce, which has always given attackers and fraudsters greater opportunities to conduct fraud. The scammers often employ a variety of methods in their deception. Recognizing the strategy is necessary in order to prevent further fraud. A number of other earlier research have also been conducted on a range of methodologies to uncover answers related to card fraud identification. Just a few examples of such strategies include neural network models (NN), Naïve bayes, intelligent decision engines (IDE), optimization algorithms, meta-learning agents, artificial intelligence, image processing, Constitution-based systems, logistic regression (LR), support vector machine (SVM), decision tree, k-nearest neighbour (kNN), meta-learning strategy, adaptive learning, and so on. Both the neural network and its topology are largely employed in real-time payment processing applications employing an unsupervised method. A self-organizing graph of both neural networks, using optical classification, may be able to answer this question for each related community. Ensemble cast learning (also known as meta-

classifier) enhances statistical findings by mixing many learning techniques and optimization algorithms. Ensemble cast learning (also known as meta-classifier) improves results by integrating multiple learning algorithms optimization algorithms to improve statistical results with over 95% of the overall detection system of ROC demand curve fraud without really creating any other false alarm (Bahnsen et al. 2014).

Because there are more number of genuine transactions when compared to fraud ones, a credit card data collection looks to be substantially biased. This means that without recognizing the fraudulent transaction, prediction might get a high precision grade. Class allocation, or sampling minority classes, appears to be one effective strategy to address such a problem. Even with oversampling large minorities, class learning instances might be increased in an acceptable proportion to the substantial majority class, increasing the likelihood of such right prediction using the new technique. This article also contains a full explanation of such supervised and otherwise unsupervised tools and procedures. Using supervised optimization approaches, it will always be impossible to detect incidents of fraud. The restricted Boltzmann machine (RBM) is a design in deep autoencoders that can create conventional transfers to identify anomalies from regular trends. It's not just about establishing the hybrid approach with so many different AdaBoost and possibly Majority Voting techniques (Fiore et al. 2019).

In a data analysis paper (Adewumi and Akinyelu 2017), specialists primarily assessed a hybrid data model in which functionality selection and heuristic classification were accomplished on three levels. During the first stage, the usual preprocessing was completed. In the second and third phases, four functionality selection techniques were applied, including a genetic algorithm, a data gain ratio, and an assessment of recovery characteristics. Variables with functionality selection strategies were selected based on both the precision of distinct classification machine learning and then the feature selection technique, which is the most appropriate for a certain classification. The results of such a hybrid model were quite precise.

Aims and S.M.A.R.T objectives

The primary aim of the study is to identify classifiers with higher accuracy predictions for credit card fraud detection. The research objectives may be described as follows using the S.M.A.R.T principles:

S- To have a better understanding of the various machine learning algorithm models used to prevent credit card fraud.

M- To assess the efficiency of machine learning in detecting credit card fraud.

A- To improve the efficiency of cashless banking by reducing fraud.

R- To apply machine learning methods to address existing credit card fraud problems.

T- Within five years, improve credit card security and decrease the fraud rate.

Research Questions

A complete literature review must address one or more research questions. These are the research questions that will be addressed in the conclusion and discussion section.

1. What are the many types of credit card frauds and how can machine learning algorithms be trained to detect and prevent them?
2. What are the various machine learning techniques used to identify credit card fraud?
3. What are the difficulties of employing a machine learning algorithm for credit card fraud detection?

Research Methodology

Steps in Machine learning algorithm:

Machine learning algorithms, according to Vabalas et al. (2019), have become the ultimate key for tackling many forms of contemporary challenges in this generation.

Figure: Steps in Machine learning algorithm

Source: Author

1. The initial stage in machine learning is data acquisition. In this stage, all necessary information is gathered. The modelling process necessitates the acquisition of constructive data. In this phase, data is automated, enhanced, and normalized.

   The author has claimed that the second phase is data preparation, in which the acquired data is reviewed and compared to other sources, and then the right data is selected for modelling. The flawed data is corrected and supplied for machine learning algorithms to use. Algorithms can be better applied with a well-presented data model.

2. Another important stage in a machine learning algorithm is choosing the appropriate model for the task. It is critical to use a proper model for data modelling while applying a machine learning method. Machine learning methods such as image recognition, speech recognition, decision trees, and logistic regression are utilized depending on the needs.
3. Another stage that must be completed to carry out the procedure properly is training the model. The models are reused for the data in this stage to create patterns and predictions. The models can help to enhance prediction and productivity in the future.
4. However, to do the analysis, not only training but also evaluation is required. Individuals must examine whether the model is working or not after modelling. Based on historical and current data, the model's performance should be evaluated. It will be easier for individual to determine the model performance if testing is done on the same level of data.
5. The variables that examine and decide input value are referred to as parameters. Hyper-parameter tuning stage of machine learning method is performed after the models have been evaluated and the accuracy of the models has been determined. Tuning the model's parameters will help to increase the algorithm's accuracy.
6. The final stage is to make a prediction, which is done when all the preceding processes have been completed. After analyzing the past stages, the algorithm predicts possible future output.

Machine learning algorithms:

Several types of machine learning algorithms are used for varied reasons, according to (Ray 2019). The following are three of the Machine Learning algorithms that have been investigated for their performance in this study.

a) Decision Trees:

Decision Trees are a type of machine learning algorithm that may be used to address classification and regression issues. Each internal node in a decision tree is an attribute test, each branch is 8a test's result, and the leaf nodes show classes or outcomes. Because of the computational complexity and sequential nature of decision trees, every minor change in the input data might have an influence on the tree structure (Rajora et al. 2018).

b) Logistic Regression:

Logistic regression is a classification system that divides data into discrete outcomes. When the dependent variable is categorical, logistic regression is used to create a regression model. There are three types of logistic regression: binary, which is used when the response variable is binary, multinomial, which is used when the dependent variable has more than two non-ordered categories, and ordinal, which is used when the categories are ordered (Rajora et al. 2018).

c) Random Forest:

Random Forest (RF) is an ensemble learning classifier that consists of numerous decision trees and outputs a class that is obtained from aggregating the outputs of each decision tree (Rajora et al. 2018).

After all of the processes have been completed and examined, the final algorithm that will be most effective in determining credit card fraud will be approved and further reviewed.

Feasibility, Significance and Scope of Innovation

According to the findings, credit card fraud is a big concern in the financial industry that is becoming more prevalent over time. More and more businesses are migrating to an online method that allows clients to conduct transactions online. Criminals can use this chance to steal other people's information or credit cards in order to conduct online transactions. To detect such acts, a fraud detection system is required (Paruchuri 2017).

The performance comparison of the machine learning algorithms provided in this work gives a quantitative insight of their applicability for credit card fraud detection applications.

For most individuals, the increasing use of technology has generated a necessity for cashless transactions, which has expanded the usage of credit and debit cards. Different algorithm types will be examined in this study to detect credit card fraud and prevent it from occurring. Predictive analytics will be performed by the algorithm. It will detect fraud tendencies in order to predict potential future frauds and will notify the majority of people about the fraud instances. In developing markets, outlier models are necessary to anticipate data findings that do not yet exist. If the data indicate anything unusual, the outlier model will establish a card flag and detect fraud before it occurs.

It is critical for any credit card detection solution to be able to accommodate cardholders without causing them any inconvenience. To avoid fraud and offer custom rule administration, it will be measured that vital transactions are not banned, even if they come from outside sources.

The technology has the potential to evolve into mobile card controls in the future, as giving cardholders authority over the device allows financial individuals to share fraud liability.

Credit card fraud detection using a machine learning algorithm guarantees the effectiveness of fraud detection, allowing consumers to manage their cash flow, minimize utility payments, and receive greater cash card rewards. It would enable a risk-free transaction for customers, minimizing economic friction, and fraudsters will not be able to conduct many fraudulent purchases with stolen identities since cardholders will always be aware of the activities made with the card.

Project Timetable

		Task Name	Duration	Start_date	End_date	Cost	Predecessors	Resource Names								
		Credit card fraud detection with machine learning algorithms	**98 days**	**18/01/2022**	**26/04/2022**											
		Activities	**Duration**	**Week 1**	**Week 2**	**Weekn 3**	**Week 4**	**Week 5**	**Week 6**	**Week 7**	**Week 8**	**Week 9**	**Week 10**	**Week 11**	**Welek 12**	**Week 13**	**Week 14**
	1	**Project Initiation**	**6 days**														
	1.1	Setting the proejct	2days														
	1.2	Creating the project charter	1days														
	1.3	Identifying the stakeholders	1days														
	1.4	Assembling the teams	2 days														
	2	**Project planning**	**15 days**														
	2.1	Determining the requirements	4 days														
	2.2	Determining the sources	4 days														
	2.3	Identifying the key success factors	5 days														
	2.4	Creating the budget	2 days														
	3	**Execution**	**30 days**														
	3.1	Selecting resource for algorithms	5 days														
	3.2	Negotiation for resources	5 days														
	3.3	Obtaining the team based on availability	7 days														
	3.4	Assigning the roles to team members	4 days														
	3.5	Using each algorithms	6 days														
	3.6	Identifying results of each algorithm	2 days														
	3.7	Training the individuals about new algorithm	1days														
	4	**Monitoring**	**15 days**														
	4.1	Checking result quality of each algorithm	2 days														
	4.2	Assuring the efficiency of the plan	5 days														
	4.3	Controlling the cost	4 days														
	4.4	Implementing the change requests as per the requirement	4 days														
	5	**Closure**	**6 days**														
	5.1	Transferring all the deliverables	2 day														
	5.2	Ensuring the deliverables	1day														
	5.3	Confirming the project completion	2 day														
	5.4	Reviewing all the contracts and documentations	1day														

Milestones 1: Assembling teams, **Milestone 2:** Budget creation, **Milestone 3:** Traning the staff about new algorithm, **Milestone 4:** Implementing the change

Figure: Project Time plan

Source: Author

References

- Carcillo, F., Le Borgne, Y.A., Caelen, O., Kessaci, Y., Oblé, F. and Bontempi, G., 2021. Combining unsupervised and supervised learning in credit card fraud detection. *Information sciences*, *557*, pp.317-331.
- Ruehle, F., 2020. Data science applications to string theory. *Physics Reports*, *839*, pp.1-117.
- Bahnsen, A.C. et al., 2014. Improving credit card fraud detection with calibrated probabilities. *In Proceedings of the 2014 SLAM international conference on data mining*, Society for Industrial and Applied Mathematics, (pp. 667-685).
- Fiore,U., et al., 2019. Using generative adversarial networks for improving classification effectiveness in credit card fraud detection. Information Sciences,479, pp.448-455.
- Adewumi,A.O. and Akinyelu,A.A.,2017.A survey of machine-learning and nature-inspired based credit card fraud detection techniques. International Journal of System Assurance Engineering and Management,8(2), pp.937-953.
- Vabalas, A., Gowen, E., Poliakoff, E. and Casson, A.J., 2019. Machine learning algorithm validation with a limited sample size. *PloS one*, *14*(11), p.e0224365.
- Ray, S., 2019, February. A quick review of machine learning algorithms. In *2019 International conference on machine learning, big data, cloud and parallel computing (COMITCon)* (pp. 35-39). IEEE.
- Rajora, S.et al., 2018. A Comparative Study of Machine Learning Techniques for Credit Card Fraud Detection Based on Time Variance. *IEEE Symposium Series on Computational Intelligence (SSCI)*, 2018, pp. 1958-1963.
- Paruchuri, H. (2017) "Credit Card Fraud Detection using Machine Learning: A Systematic Literature Review", *ABC Journal of Advanced Research*, 6(2), pp. 113-120.