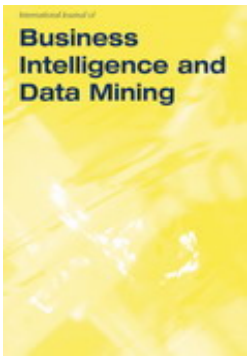


International Journal of Business Intelligence and Data Mining

ISSN online: 1743-8195 - ISSN print: 1743-8187
https://www.inderscience.com/ijbidm

A widespread survey on machine learning techniques and user substantiation methods for credit card fraud detection

T. John Berkmans, S. Karthick

DOI: 10.1504/IJBIDM.2023.10047750

Article History:
Received: 08 September 2021
Accepted: 28 December 2021
Published online: 30 November 2022

A widespread survey on machine learning techniques and user substantiation methods for credit card fraud detection

T. John Berkmans*

Department of Computer Science Engineering,
SRM Institute of Science and Technology (SRMIST),
Kattankulathur, Chennai, India
Email: jt3007@srmist.edu.in
*Corresponding author

S. Karthick

Department of Computational Intelligence,
SRM Institute of Science and Technology (SRMIST),
Kattankulathur, Chennai, India
Email: karthiks2@srmist.edu.in

Abstract: In this modern scientific digital world, credit card usage has enormously increased everyday. Simultaneously a huge amount of credit card misuse also has been expressively popular. It prompts monetary misfortunes for both charge cardholders and monetary associations. To keep away from that, monetary associations created and conveyed Visa extortion discovery techniques. In the upcoming years, everybody will utilise the greatest exchange through online mode just to save their time. So we partition this review into two primary parts. In the first part, we centre around old-style AI models, and in this part we focus on what the client knows (knowledge-based strategy). For the second part, we focus more on the turn of events procedure of client verification, and their conduct biometrics to distinguish an individual remarkable conduct while utilising their electronic gadgets. An outline of the current methodology in this writing review means to grow a more precise, dependable, versatile, super-fast, effective, and modest model of charge card extortion identification.

Keywords: credit card transaction; machine learning; bio-metrics; XGBoost; SVM; random forest.

Reference to this paper should be made as follows: Berkmans, T.J. and Karthick, S. (2023) 'A widespread survey on machine learning techniques and user substantiation methods for credit card fraud detection', *Int. J. Business Intelligence and Data Mining*, Vol. 22, Nos. 1/2, pp.223–247.

Biographical notes: T. John Berkmans has completed his MTech and PhD and he is a full time research scholar doing research in the field of machine learning domain in the School of Computing from SRM Institute of Science and Technology (SRMIST), Kattankulathur, Chennai, India.

S. Karthick holds an ME and PhD and is working as an Associate Professor. He has expertise in data mining and machine learning.

1 Introduction

These days as we likely know the whole gang has, all things considered, discard online portion procedures; consequently the risk of customer accreditations is in harm's way. This is a regular kind of deception these days. Conspicuous evidence of distortion is an essential discussion standing up to gigantic financial associations (Tounekti et al., 2019) and banking territory which has caused given the increase in control card portions. There exist seven unmistakable kinds of MasterCard blackmail:

1 internet assistance

2 credit card

3 healthcare

4 TV set and microelectronic broadcasting

5 overseas money bids and phony stunts

6 computer hardware and programming

7 investment-related.

Thusly, we have endeavoured to handle this key issue for the end customers. Electronic business locales and banking zones need to challenge certifiable safety problems as an effect of the inadequacy of virtual extortion, hamstring, spyware, and stealing of secretive information. The safety of a virtual store is not the single basis for the proprietors to track business training efficiently; anyway, it is additionally critical for clients to remain calm. To give an ensured shopping stage, e-stores use different sorts of approval systems, and biometrics is seen as the best of all. Not in the least like customary check measures, does not biometrics anticipate that clients should set a mysterious key, keep on recalling it and hide it from others.

Figure 1 Types of frauds (see online version for colours)

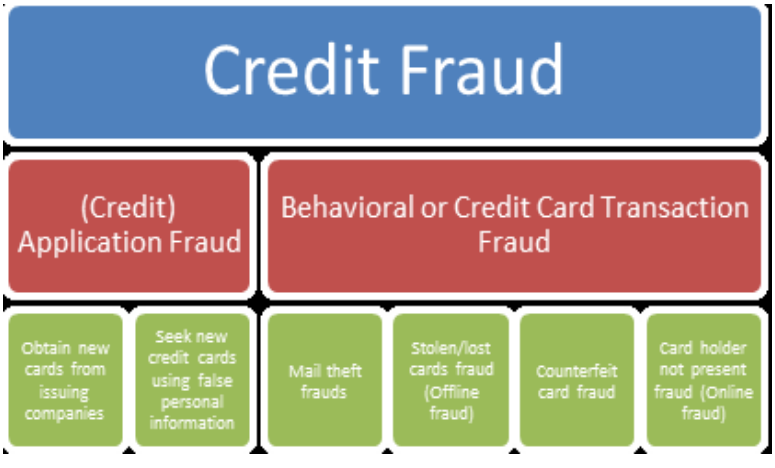


Figure 2 Financial fraud estimation graph (see online version for colours)

1.1 Initiations

1.1.1 Machine learning

ML is the study of causing PCs to learn and act like people by nourishing facts without being programmed. Machine learning makes predictions and decisions based on past historical data.

Machine learning-based scam recognition:

- discovering scam spontaneously
- instantaneous gushing
- fewer intervals wanted for confirmation approaches
- classifying unseen connections in statistics.

1.1.2 Super-fast

Machine learning is suited for scam finding. When it comes to fake choices, we need the results as super-fast. Numerous teams of specialists run thousands of queries and connect to discover the finest result. It is all done in the present scenario and takes a fraction of seconds. As well as at present, it is evaluating distinct consumer activities. It is continuously studying usual consumer action, so once commercials are irregular this can spontaneously chunk or else banner an expense on behalf of expert analysis.

1.1.3 Ascendable

Every internet business needs to raise its agreement estimations. Yet, with ML it is very inverse. ML frameworks increase with prevalent data file since provides the framework extra samples of upright and awful example. Honest and misleading clients. From this

perfect anticipate extortion in impending exchanges quickly as conceivable by activities of cardholders spending design.

1.1.4 Effective and cheap

It runs enough instalments each second. The expense is additionally reasonable is only the expense of the worker running. Humans can cause mistakes however machines anticipate information examination in a division of time.

1.1.5 Extra perfect

It gives more accuracy from predictive data scrutiny. These prototypes are astute to gain from examples of common exercises. They are exceptionally quick if any abnormality occurs during exchange it will anticipate the fake one and not consider an exchange.

Officially, there are various sorts of learning systems like managed, semi-regulated, unaided, supported, transduction. The two most broadly received AI strategies are managed to realise, which prepares the calculation on predefined named datasets, and unaided realising, which gives the calculation unlabelled preparing information to permit learning the examples and relationships in the info information. In the accompanying, our methodology these two learning systems in more detail.

Progressed charge card extortion recognisable proof techniques are parted into:

- Unaided. For example, PCA, LOF, one-class SVM, and isolation forest.
- Directed. Like decision trees (for example ExtremeGradientBoost and LightGBM), random forest, and K-nearest neighbour (KNN).

1.2 Supervised learning

1.2.1 Supervised

Regulated learning incorporates ML calculations that learn under the event of a boss. Supervised ML methods can predict with the help of a labelled dataset. Among regulated ML extortion ID techniques, we characterise choice trees, arbitrary backwoods, KNN, and innocent Bayes.

The properly administered knowledge measure contains variable factors, the first variable is (x), and the second variable is (y). We apply an estimation to contemplate the preparation volume of contribution to the profit. In straightforward arithmetic, the yield (y) is a reliant variable of info (x) as shown by:

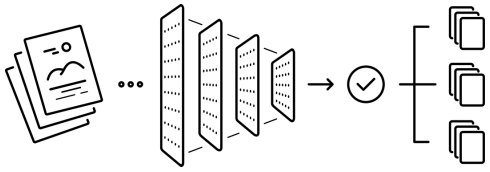
$$y = f(x)$$

Here, our ultimate objective is to attempt to surmise the planning capacity (f), with the goal that we can anticipate the yield factors (Y) when we have new information (X).

KNN is a grouping calculation that checks similitude's dependent expanse in multi-dimensional space (Hines and Youssef, 2018). It is used for the social affair of MasterCard blackmail recognisable proof by manipulative nearest point. This methodology is not unprotected against uproar and lost file centres which resources making greater data files in the fewer period. Additionally, it is definite and needs less work as a planner to adjust the figurine. The distance in KNN between two data

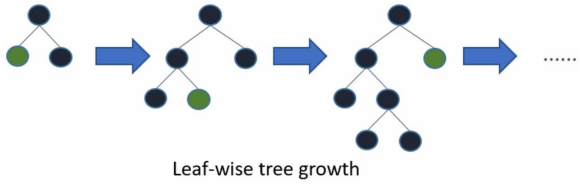
portrayals can be considered by using a grouped technique, anyway generally by using the Euclidean distance. KNN is important. The results show that well-constructed functions and majority class undersampling can successfully detect fraud using ANN, SVM.

Figure 3 Formal supervised learning



XGBoost and *light gradient boosting machine* are a solitary kind of slope that helped decision trees calculation, made for quickness just an augmenting the effectiveness of processing interval and memorial assets (Samy and Parthiban, 2018). PCA for representation and SVM and XGBoost for classification. This mixing calculation strategy of new models is included to repair the mistakes brought about by surviving figure.

Figure 4 Gradient-boosted decision trees algorithm (see online version for colours)



To exchange a fake charge the outcome (possibility) of numerous decision trees is summed up – through each upcoming tree improves their outcomes dependent mistakes prepared by its archetypes. Subjective random forest area algorithm is a request count that is included various decision trees. Every tree has centre points with circumstances, which describe an authority conclusion subject to most vital worth.

The random forest area figuring for distortion distinguishing proof and shirking had both fundamental factors have extraordinary at expecting a thing. Anomaly is the first one, concluding the lines and fragments of information picked discretionarily from the datasets and embedded into various decision trees, number 1 gets underlying 2,000 lines, number 2 gets lines 3,000 to 4,000, and number 3 gets segments 5,000 to 6,000.

The 2nd thing forest area of trees that add to an authority end as opposed to just a single decision tree. The best benefit of assortment reduction in the model overfitting, while the tendency remains as in past.

Particular ML models are used to recognise threats; they have their benefits and disservices. Some models are exceptionally difficult to interpret, explain, and study, but they are incredibly accurate (for instance neural associations, boosting, social events, etc.) others are less troublesome, so it can be successfully unravelled and envisioned as plenty of rules (for instance decision trees).

It is indispensable to setup the coercion revelation model interminably every time different data appears, a different distortion structures/models should learn false data recognised exactly on schedule as could truly be considered typical.

1.3 Unsupervised learning

Unaided ML techniques utilise unlabelled information to discover examples and conditions in the charge card extortion recognition dataset, making it conceivable to accumulate information tests by likenesses without manual marking, anonymised charge card information of one of the main banks in Egypt (Umuhoza et al., 2020).

PCA empowers the performance of an examining information investigation to uncover the internal construction of the information and clarify its varieties. Principle components analysis is perhaps the best well-known procedure on behalf of inconsistency identification (Arun and Lakshmi, 2021) genetic algorithm's oversampling approach improves error prediction performance and reduces the rate of false positives.

PCA looks for connections among highlights – (Randhawa et al., 2018) which on account of Visa exchanges could be time, area, and measure of cash spent – and figures out which mix of qualities adds to the inconstancy in the results. Such joined element esteems permit the formation of a tighter element space named head segments.

Local outlier factor perceives in what manner a particular data test to be a special case (irregularity). It is a greater amount of the most standard peculiarity area procedures. To process the outlier factor, the amount of neighbouring data centres are measured by the model of thickness, modification from their thickness of new information centres. Uncertainty a particular information point has a liberally small thickness that appeared differently concerning its close neighbours it is a special case.

One-class support vector machine is a blueprint calculating assists with perceiving inconsistencies in information. This calculation licenses one to administer imbalanced information-related issues like extortion location.

The thought behind one-class SVM is to prepare just on a strong measure of authentic exchanges and afterward distinguish peculiarities or oddities by contrasting every new information points with them.

Isolation forest is an inconsistency location technique with decision tree families. The fundamental thought of IF, which separates it from extra mainstream exception discovery calculations, is that it unequivocally identifies inconsistencies as opposed to reporting the progressive information focuses. IF is worked of decision trees where the detachment of information focuses happens first on account of haphazardly choosing a divided worth amid base and most extreme estimation of the picked highlight.

In this way, if we have a bunch of authentic exchanges, the isolation forest calculation will characterise deceitful MasterCard exchanges in light of their qualities –which are frequently altogether different from the qualities positive exchanges have (for example they occur further away from the ordinary information focuses in the component space). Necessities for instalment extortion recognition with simulated intelligence-based techniques.

To run a computer-based intellect-driven methodology for Visa extortion investigation, various necessities ought to be encountered. It guarantees that the typical arrives at its finest discovery total (Zheng et al., 2020; Taha and Malebary, 2020).

1.3.1 Quantity of information

Preparing great machine learning models requires critical interior recorded information. That implies if you need more past fake and typical exchanges, it is difficult to execute an ML model. Because the nature of the preparation cycle relies upon the nature of the data

sources. Since it is an infrequent situation the preparation set holds an equivalent measure of information tests in both classes, how many attributes a dataset has decreased, or else information increase methods may be utilised.

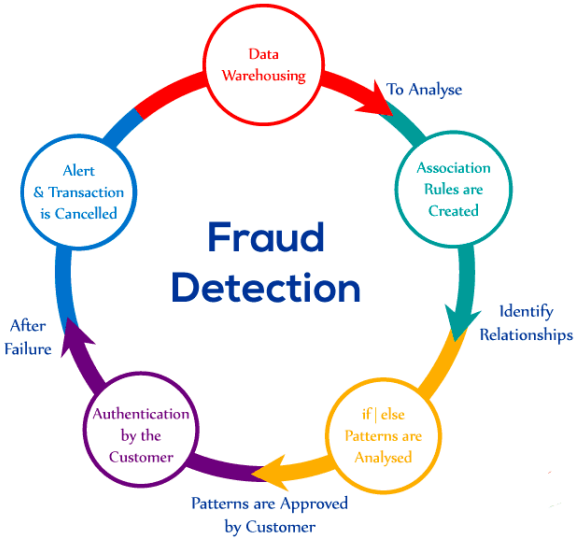
1.3.2 Nature of information

Models might be liable to inclination dependent on the nature of verifiable information. This assertion implies that if the stage maintainers did not gather and sort the information perfectly and appropriately or even blended the data of fake exchanges in with the data of typical ones that are probably going to cause a significant predisposition in the model's outcomes.

1.3.3 The honesty of components

If we have sufficient information that is all around organised and balanced and is combined pleasantly probabilities may countless threats recognition will function well for clients and their business.

Figure 5 Progressive visa extortion ID techniques and their benefits (see online version for colours)



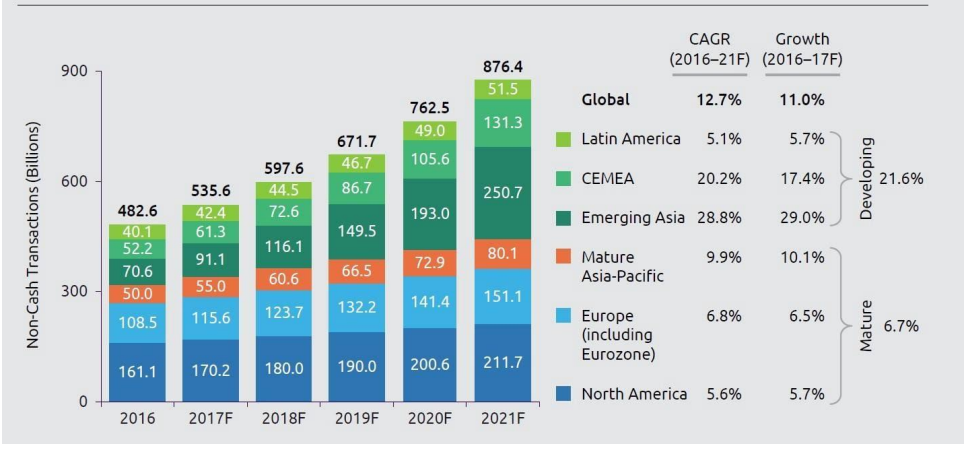
1.4 How does MasterCard extortion occur?

Incorrect Visa information is generally due to the cardholder's carelessness with their information or breaking into the security of a website (Kalid et al., 2020). Here are some models:

- a buyer reveals his MasterCard number to new people
- if the card is missing or has been taken by someone else, it will be used

- suggested recipient takes the e-mail and hooligans use it
- sales reps double the amounts of their cardholder cards
- creation of a counterfeit visa.

Figure 6 Number of worldwide non-cash transactions (billions), by regions, 2016–2021F (see online version for colours)



Once your card has been misused, an unauthorised charge will be made; as such, the who discovers it uses it to make a purchase. Criminals can also make their name and use the card or order person some products through a cell phone or computer. There is also the problem of using a counterfeit MasterCard, a counterfeit card with the actual data taken from the cardholders. Particularly insecure as victims have their real cards but do not know someone has a fake card. The fake cards look very authentic and have the logos and attractive coded parts of the original. Fake MasterCard cards are usually destroyed by thieves after a few successful instalments, not long before an injured person understands the problem and reports it.

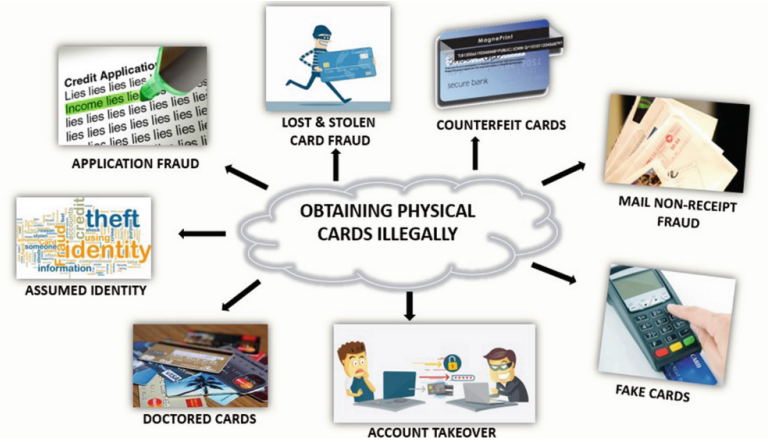
1.5 Problem statement

Charge card cheats are expanding intensely due to misrepresentation monetary misfortune is expanding definitely. Consistently because of misrepresentation billions of sums lost. To dissect the extortion there is an absence of exploration. Many ML calculations are carried out to identify genuine charge card extortion. ANN and crossover calculations are applied (Omar et al., 2021). Recommended the use of three proposed assessments ROA, ROS, and RONS, on both diversion and genuine information.

The main challenges involved in credit card fraud detection are:

Colossal information is prepared each day and the model form should be sufficiently quick to react to the trick on schedule. Imbalanced: information for example the majority of the exchanges (99.8%) are not false which makes it truly hard for distinguishing the deceitful ones.

Figure 7 Types of fraud (see online version for colours)



1.5.1 Data availability as the data is mostly private

Misclassified Information can be another significant issue, as few out of every odd false exchange is gotten and revealed. Versatile strategies utilised against the model by the tricksters.

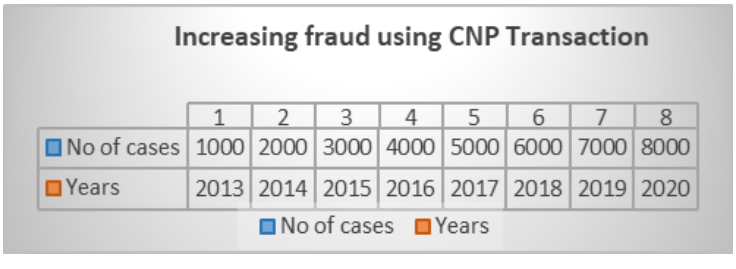
1.5.2 How to tackle these challenges?

The model utilised should be straightforward and quick enough to recognise the oddity and group it as a false exchange as fast as could be expected. Awkwardness can be managed by appropriately utilising some calculations. For securing the protection of the client the dimensionality of the information can be decreased.

A more dependable source should be taken which two fold check the information, in any event for preparing the model.

We can simplify the model and interpretable so when the con artist adjusts to it with simply a few changes we can have another model going to send.

Figure 8 Fraud using card not present (CNP) transaction (see online version for colours)



2 Literature survey

In previous research, we have seen all systems revolve around online fakes since it is to be a broadcaster. In any case, a huge segment gives the outcome subject to an exact dataset, which is the issue depicted by unstable data. As furthermore a couple of works have mishandled the astounding advances in the simulated intelligence methodologies for estimate and cataloguing. From this model what the client knows (knowledge-based system) and what the customer access to (object-based method). We concentrate on the greater improvement technique of customer confirmation, and their lead biometrics to recognise an individual novel direct while using their electronic contraptions. In this paper, we base on that results and endeavour to assess the identical dataset.

Figure 9 How an ML framework functions (see online version for colours)

2.1 Info information

With regards to misrepresentation recognition, data should as much as possible.

For managed ML, the information should be named as great (real clients who have never dedicated misrepresentation) or terrible (clients with a chargeback related to them or have been physically named as fraudsters) (Can et al., 2020) misrepresentation recognition framework could choose to utilise the most suitable identification model and lift execution. Just as producing numerous discovery models.

2.2 Concentrate highlights

We bunch highlights into five fundamental classifications, every one which has hundreds or thousands of single highlights: character No of digits in the client's e-mail address, age record, gadgets client was seen on, misrepresentation client's IP address pace. In their 1st week of month more no of orders request, items exchanges, normal request esteem, hazardous container substance. Instalment strategies Misrepresentation pace of giving bank, the likeness between client names and charging name, cards from various nations.

Areas transportation address (Randhawa et al., 2018) coordinates with the charging address, dispatching nation matches nation of client's IP address, extortion rate at client's area. Organisation number of messages, telephone numbers, or instalment strategies shared inside an organisation, age of the client's organisation.

2.3 Train calculation

A count is a lot of rules to be followed when dealing with multipart issues, like a mathematical condition or even a method. The estimation uses customer data depicted by our features to sort out some way to make conjectures, e.g., distortion/not coercion.

Table 1 Restricting the covering degree to improve (Omar et al., 2021) class-imbalanced learning under pitiful component choice

Author	*Title*	*Source*	*Findings*
Omar et al.	Minimising the overlapping degree to improve class-imbalanced learning under sparse feature selection: application to fraud detection.	IEEE Access (2021) February 1, Vol. 9, pp.28101–28110.	Recommended the use of three proposed assessments ROA, ROS, and RONS, on both diversion and genuine information. Those calculations hope to diminish the cover degree and circle the figuring out some approach to isolated datasets to upgrade the social event of class-imbalanced information. The outcomes show that RONS includes an accomplishment in unmistakable the basic highlights in model choice, while both ROA and ROS get a huge transcendent to the degree F-measure and Gmean. In future work, they will add more imbalanced datasets and arrange re-enacted knowledge models in their assessments to support their strategy.
Arun and Lakshmi	Genetic algorithm-based oversampling approach to prune the class imbalance issue in software defect prediction.	Soft Computing (2021) August, Vol. 29, pp.1–7.	The proposed algorithm produces a minority class of synthetic samples based on a distribution measure, ensuring that the samples in the class are diverse and efficient. The proposed oversampling algorithm was compared to a non-sampling approach using 20 error prediction datasets from SMOTE, BSMOTE, ADASYN, random oversampling, MAHAKIL, and promise repositories and five prediction models. The results show that the genetic algorithm's oversampling approach improves error prediction performance and reduces the rate of false positives.
Zheng et al.	Improved transfer AdaBoost and its application to transaction fraud detection	IEEE Transactions on Computational Social Systems (2020) August 27, Vol. 7, No. 5, pp.1304–1316.	Proposed appointment of Visa trade data can change with the changes in the trade practices of customers, yet the movements are moderate usually. These movements are yet huge for perceiving trade distortion since they achieve a supposed thought skim issue.
Alam et al.	An investigation of credit card default prediction in the imbalanced datasets	IEEE Access (2020) October, Vol. 8, pp.201173–201198.	Proposed the forecast exactness pace of the GBDT (inclination lift choice tree) model is higher than the customary AI-based models. The GBDT technique gave the best precision of 88.7% while using the K-implies destroyed resampling strategy on Taiwan customers' credit datasets.
Umuhoza et al.	Using unsupervised machine learning techniques for behavioural-based credit card users segmentation in Africa	SAIEE Africa Research Journal (2020), September, Vol. 111, No. 3, pp.95–101.	Characterise and portray the means that can be taken to fabricate a conduct-based division model that separates African credit cardholders dependent on their buying information. They show the proposed approach at work utilising anonymised charge card information of one of the main banks in Egypt, the Business Global Bank of Egypt.
Can et al.	'Closer look into the characteristics of fraudulent card transactions'	IEEE Access (2020) September, Vol. 8, pp.166095–166109.	Proposed irregular woods and multi-facet perceptron classifiers showed a critical improvement over the expert model for the 0 to 10 and 1,000 to 100,000 sum range in Turkish Lira. Consequently, a shrewd misrepresentation recognition framework could choose to utilise the most suitable identification model and lift execution. Just as producing numerous discovery models dependent on the exchange attributes contrarily influenced the model's presentation.
Feng et al.	'Using cost-sensitive learning and feature selection algorithms'	IEEE Access (2020) April, Vol. 8, pp.69979–69996.	Proposed an expense touchy component choice general vector machine (CFGVM) calculation dependent on GVM and BALO calculations to handle the imbalanced grouping issue, conveying distinctive expense loads to various classes of tests. In our strategy, the BALO calculation decides the expense loads and concentrates more critical highlights to improve the order execution.

Table 1 Restricting the covering degree to improve (Omar et al., 2021) class-imbalanced learning under pitiful component choice (continued)

Author	*Title*	*Source*	*Findings*
Li et al.	'Deep representation learning with full centre loss for credit card fraud detection'	IEEE Transactions on Computational Social Systems (2020) April, Vol. 7, No. 2, pp.569–579, date of publication: 17 February 2020.	Proposed better detachability and separation of highlights so it can improve the exhibition of their extortion identification model and keep its soundness. They proposed another sort of misfortune work, full focus misfortune (FCL), which thinks about the two distances and points among highlights and, hence, can exhaustively regulate the profound portrayal of learning. Later on, plan to consider the idea of float issue from the part of misfortune work.
Taha and Malebary	'An intelligent approach to credit card fraud detection using an optimised light gradient boosting machine'	IEEE Access (2020) February, Vol. 8, pp.25579–25587	Proposed a shrewd methodology for distinguishing misrepresentation in Visa exchanges utilising an improved light slope boosting machine (OLightGBM). Later on, they intend to consider the idea of float issue from the part of misfortune work.
Arun and Lakshmi	'Class imbalance in software fault prediction data set' in artificial intelligence and evolutionary computations in engineering systems'	Springer (2020) pp.745–757	This paper attempt to study different techniques proposed to solve the class imbalance problem. Under-sampling, oversampling, and synthetic sampling techniques such as SMOTE, borderline-SMOTE, SMOTE Boost, cluster-based oversampling, ADASYN, and ensemble techniques such as bagging, boosting, and GASEN.
Kalid et al.	'A multiple classifiers system for anomaly detection in credit card data with unbalanced and overlapped classes'	IEEE Access (2020) February Vol. 8, pp.28210–28221.	Utilised a numerous classifiers framework (MCS) on these two informational indexes: 1 visa cheats (CCF) 2 charge card default instalments (CCDP). The MCS utilises a successive choice mix system to deliver exact peculiarity identification. Presently, specialists had endeavoured profound learning calculations, for example, long transient memory (LSTM) and profound conviction organisations for identifying oddities in charge card exchanges.
Shaikh and Nazir	'A novel dynamic approach to identifying suspicious customers in money transactions'	International Journal of Business Intelligence and Data Mining (2020) Vol. 17, No. 2, pp.143–158.	The paper introduced an original methodology that utilises the client's dynamic practices (i.e., profile highlights, client exchange history) to adequately recognise unusual and dubious exchanges. The proposed approach has been tried by executing a trial with reasonable information and approved the outcome with affirmed dubious clients. The proposed model has accomplished in general exactness 92.11% which is viewed as much better as contrasted and the measurable models that utilisation the pre-characterised set of rules.
Zhang et al.	'Fraud detection method for low-frequency transaction'	IEEE Access (2020) Vol. 8, pp.25210–25220, date of publication: 31 January 2020	Proposed a multi-conduct identification model dependent on new exchange conduct is proposed. As indicated by the consequence of every conduct, the naive Bayes model is utilised to ascertain the likelihood that the current exchange has a place with misrepresentation, lastly decide if the current exchange is extortion. Later on, we will focus closely on the issue of online model refreshing and making the strategy more powerful.

Table 1 Restricting the covering degree to improve (Omar et al., 2021) class-imbalanced learning under pitiful component choice (continued)

Author	*Title*	*Source*	*Findings*
Samy et al.	'Intelligent web-history based on a hybrid clustering algorithm for future-internet systems'	Artificial Intelligence and Evolutionary Computations in Engineering Systems (2020) pp.571–581, Springer, Singapore.	Suggests a viable recommended structure for logic and web usage mining. A fundamental advance in technology is to separate functionality from web files and generate compelling ideas. Until then, collect the logic of your site and use the ideas and basic terms that are removed from the report and used for analysis.
Tounekti et al.	'Users supporting multiple (M)EPS in online purchases'	IEEE Access (2019) Vol. 8, pp.735–766, date of publication: 23 December 2019	Proposed factual investigation (Chi-square test), customers can pay utilising mobile electronic payment system (MEPS) during their online exchange; have a favoured instalment framework dependent on its security, expenses, convenience, and usability just as on their number one internet browser for these exchanges. In the future clients who support numerous instalment strategies – improvements that would encourage trades among shippers and clients.
Dawood et al.	'Improve profiling bank customer's behaviour using machine learning'	IEEE Access (2019) August, Vol. 7, pp.109320–109327.	Proposed better k-mean, fuzzy c-means, and neural network. The used data file is noticeable and creating a new name as an independent neural organisation arrangement is the principal part of this investigation, which assists with decreasing the grouping execution time and getting the best precision results. In future work, they attempt to improve the viability and execution of their proposed approach by applying some profound learning calculations.
Makki et al.	'An experimental study with imbalanced classification approaches for credit card fraud detection'	IEEE Access (2019) July, Vol. 7, pp.93010–93022.	Paper proposed moved toward SVM and ANN, considering that different model' presumptions and conditions are difficult to meet. Their exploratory investigation uncovered that the methodologies typically used to take care of lop-sidedness issues may have undesirable outcomes when the unevenness is outrageous, for example, creating countless bogus positives. Consequently, they intend to build up a versatile climate utilising large information innovation.
Zhang et al.	'GMM-based under sampling and its application for credit card fraud detection'	Proc. Int. Joint Conf. Neural Netw. (IJCNN), July 2019, pp.1–8.	This white paper focuses on sampling unbalanced data. Based on the idea of subsampling, Gauss mixed subsampling (GMUS) has been proposed to make the data distribution uniform. Unlike mainstream random under sampling, GMUS uses a Gaussian mixed model to fit most of the training set and select samples by combining the log-likelihood values of a small number of samples. Experiments with 16 public records and real credit card transaction records have shown that GMUS works well in both AUC and F1 values and can improve the detection rate of minority samples.
Wu and Liu	'A hybrid model on learning cross features for transaction fraud detection'	Proc. ICDM (2019), pp.88–102	The proposed concept has been extended to include an online learning model. Using the internet guide, credit card fraud. The proposed system is detection and before preventing fraudulent transactions and activities reduce the unification of waste in the economy and industry.
Kim et al.	'Champion-challenger analysis for credit card fraud detection: Hybrid ensemble and deep learning'	Expert Syst. Appl., (2019) August 15, Vol. 128, pp.214–224.	They conducted a comparative study of deep learning and hybrid ensembles. Introduced the champion challenger framework to compare models. We took advantage of various realistic metrics, taking into account real constraints. The model was compared in two phases: offline testing and post-release testing. The winning model is now available on our partner's system.

Table 1 Restricting the covering degree to improve (Omar et al., 2021) class-imbalanced learning under pitiful component choice (continued)

Author	*Title*	*Source*	*Findings*
Randhawa et al.	'Credit card fraud detection using AdaBoost and majority voting'	IEEE Access (2018), Vol. 6, pp.14277–14284.	AI calculations are utilised to identify Visa misrepresentation. Standard models are first utilised. At that point, crossbreed strategies that use AdaBoost and larger part casting ballot techniques are applied. This shows that the dominant part of casting a ballot strategy offers hearty execution within the sight of clamour. This thus will help recognise and forestall false exchanges before they occur, which will diminish the number of misfortunes brought consistently in the monetary area.
Samy and Parthiban	'A novel feature extraction approach using principal component analysis and quantum behaved particle swarm optimisation-support vector networks for enhancing face recognition'	Journal of Computational and Theoretical Nanoscience (2018) Vol. 15, Nos. 9–10, pp.3012–3016.	The proposed calculation turned out to be a good confirmation, with fewer highlights selected. Based on our understanding, all relevant concepts are explained in this paper. Use QPSO and PCA for representation and SVM and XGBoost for classification.
Dal Pozzolo et al.	'Credit card fraud detection: a realistic modelling and a novel learning strategy'	IEEE Transactions on Neural Networks and Learning Systems (2018) August, Vol. 29, No. 8.	Exhibit the effect of class unbalance and idea float in a true information stream containing more than 75 million exchanges, approved throughout a period window of three years.
Yee et al.	'Credit card fraud detection using machine learning as data mining technique'.	Journal of Telecommunication, Electronic and Computer Engineering (JTEC) (2018), Vol. 10, Nos. 1–4, pp.23–27.	This paper describes the Bayes network classifier, that is, the K2, tree augmented naive Bayes (TAN) and naive Bayes, Logistics, and the J48 – monitor-based classification using the classifier. To do. Pre-processing 444 normalisation and principal component records all classifiers achieved pre-processing results for datasets compared to with an accuracy greater than 95.0.
Wu and Wang	'Customer segmentation of credit card default by self-organising map'	American Journal of Computational Mathematics (2018) Vol. 8, No. 3, p.197.	In this paper, they applied a self-organising map (SOM) technique to segment individuals based on credit information. It reduces data complexity and dimensions. Especially when there are unclear non-linear relationships between data features increase. This method provides clearer and more intuitive segmentation that cannot be achieved with other traditional methods.
Cody et al.	'A utilitarian approach to adversarial learning in credit card fraud detection'	Proc. Syst. Inf. Eng. Design Symp. (SIEDS) (2018) April, pp.237–242	In the context of credit card fraud, the framework is derived from decision theory. In addition, They showed how to use the utility function to identify the best economically fraudulent strategy.
Hines and Youssef	'Machine learning applied to rotating check fraud detection'	1st Int. Conf. Data Intell. Secur. (ICDIS), (2018) April, pp.32–35.	The results show that well-constructed functions and majority class undersampling can successfully detect fraud using ANN, SVM, or RF by 90% or more. However, given the long RF duration, restaurant owners will find it more convenient to use SVMs or KNNs.
Roy et al.	'Deep learning detecting fraud in credit card transactions'	Proc.Syst. Inf. Eng. Design Symp. (SIEDS), (2018) April, pp.129–134	In their analysis, they kept the number of neurons in each hidden layer constant. By varying the number of neurons in each layer, and can gain additional insight into the effect of network size on the performance of their model using ANN.

Table 1 Restricting the covering degree to improve (Omar et al., 2021) class-imbalanced learning under pitiful component choice (continued)

Author	*Title*	*Source*	*Findings*
Zheng et al.	'Transaction fraud detection based on total order relation and behaviour diversity'	IEEE Trans. Comput. Social Syst., (2018), September, Vol. 5, No. 3, pp.796–806.	In this paper, proposed a logical graph of BP (LGBP), which is a complete order-based model, and show the logical relationship of transaction record attributes. You can calculate the path-based transition probability from one attribute to another based on LGBP and the user's transaction record.
Jiang et al.	'Credit card fraud detection: a novel approach using aggregation strategy and feedback mechanism'	IEEE Internet Things J., (2018) October, Vol. 5, No. 5, pp.3637–3647.	It uses a classifier set to detect rogues online, and if a new transaction is rogue, a feedback mechanism is included in the detection process to resolve the misunderstanding problem. The results of our experiments show that our approach is superior to other approaches.
Carneiro et al.	'A data mining-based system for credit-card fraud detection in e-tail'	Decis. Support Syst., (2017) March, Vol. 95, pp.91–101.	This white paper explores a combination of manual and automatic classification, provides insight into the complete development process, and compares different machine learning methods. Therefore, helps researchers and practitioners design and implement data mining-based systems for fraud detection or similar issues.
Adewumi and Akinyelu	'A survey of machine-learning and nature-inspired based credit card fraud detection techniques'	Int. J. Syst. Assurance Eng. Manage., (2017) November, Vol. 8, No. S2, pp.937–953.	This white paper focuses on recent machine learning and the naturally inspired credit card fraud detection technology proposed in the literature. This review outlines some of the limitations and contributions of existing methods for detecting credit card fraud and provides the necessary background information for researchers in this area.
Awoyemi et al.	'Credit card fraud detection using machine learning techniques: a comparative analysis'	ICCNI, Lagos, Nigeria, October (2017), pp.1–9.	In this article, we'll look at the performance of naive Bayes, nearest neighbours, and logistic regression of highly biased credit card fraud data. The credit card transaction dataset is from a European cardholder and contains 284,807 transactions. A hybrid method of undersampling and oversampling is performed on distorted data.
Correa Bahnsen et al.	'Feature engineering strategies for credit card fraud detection'	Expert Syst. Appl. (2016) June, Vol. 51, pp.134–142.	This paper proposes to extend the transaction aggregation strategy to create a new set of functions based on the analysis of periodic behaviour during transactions using the von Mises distribution.
Carminati et al.	'BankSealer: a decision support system for online banking fraud analysis and investigation'	Comput. Secur., (2015) September, Vol. 53, No. 1, pp.175–186.	Develop a semi-supervised outlier detection framework based on the online banking fraud detection model. Provides scores with clear statistical significance and can handle concept drift, rare and non-stationary data. Develop a detailed analysis of real online banking datasets and focus on key challenges. Run tests on real data and launch a realistic series of attacks. The system evaluates fraud up to a detection rate of 98%.
Selvakumarasamy and Dekson	'Architecture of adaptive e-learning ecosystem'	Journal of Emerging Trends and Technology in Computer Science (IJETTCS) (2013).	The purpose of this white paper is to introduce you to the cutting edge of adaptive e-learning systems by explaining its dimensions, design, architecture, and theoretical approach as an alternative to traditional learning.

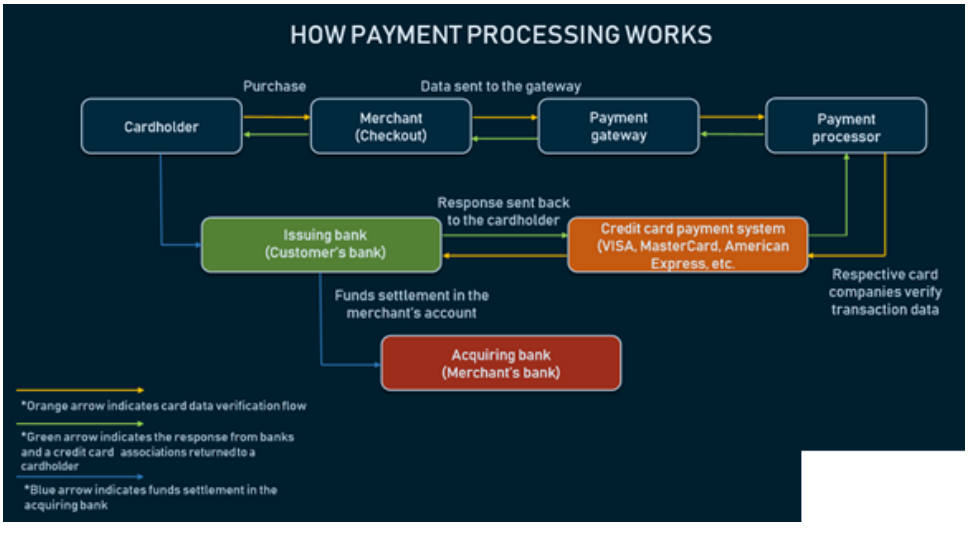
At first, we'll train the computation on an online seller's own recorded data; we call this a planning set. The more coercion in this planning set the better, so the machine has lots of advisers for acquire from.

2.4 Create a model

When preparing is finished we have a model explicit to their business, which can recognise extortion in milliseconds.

We continually watch out for the model to ensure it is carrying on as it ought to, and we're continually searching for approaches to improve it. We consistently improve, refresh and transfer another model for each customer so the framework will consistently recognise the most recent extortion strategies.

Figure 10 Flow of payment process (see online version for colours)

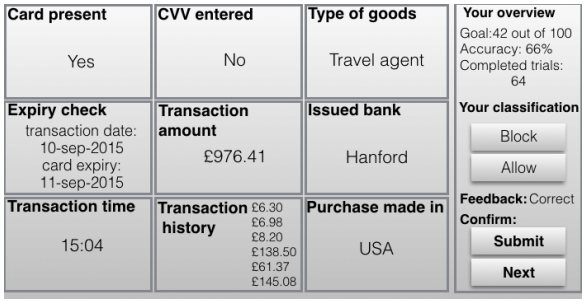


Utilising ML strategies to forestall misrepresentation and stop honest individuals losing cash. There are sure fraudy circumstances which the model ought to consistently get on – a few models are:

- high speed of new instalment strategies, e.g., a client adds new ten instalment cards in 60 minutes
- suspicious e-mail address, e.g., a confound between the record name or name on the card, or discourteous/mischievous words in the e-mail
- a client submitting bunches of requests of high worth great, e.g., extravagance things
- orders from an especially fraudy area, transportation to a known misrepresentation area of interest or a PO Box as opposed to a private location.

These models ought to be hailed as fraudy – so what happens when the machine makes an expectation?

Figure 11 Block diagram for credit card transaction



The ML score is basically a number between 0–100. This score depicts our gauge of how deceitful a client is. To score a client, we remove many signs about a specific client: the quantity of cards they have added as of late, the distance between the letters on the console in their e-mail, regardless of whether they have reordered a card, and so forth.

On the off chance that the model identifies the extortion accurately, we can convey it to be utilised against the online deals'. We additionally do some programmed presence of mind investigation on ongoing information for which we do not have misrepresentation names to guarantee the model will carry on effectively when it is sent.

1 customer places order

2 ML generates features

3 model predicts risk score – block-review-allow.

2.5 Greater security

Internet for the client's new online exercises like instalment conduct, web-based media, and government managed retirement, IP area, gadget movement and charging address. Authentication:

1 billing verification

2 individuality verification.

Transaction monitoring

1 biometrics

2 fraud

3 database.

Anti-money laundering (AML) screening

1 signals.

Architecture flow

1 while registering fingerprint scanning.

2 joint authentication.

Mix of at least two above confirmation plans can be used to validate clients and upgrade the precision of verification framework. A portion of the blend techniques incorporate iris and contact validation; example and contact confirmation; PIN and contact verification; and open example and development-based touch verification.

Table 2 Comparison table

Models	*Advantages*	*Disadvantages*
One class support vector machine	Small changes in the data do not affect the stability of the model.	Long training time, a lot of memory is required.
Logistic regression	Provides information on the size of the coefficient with direction (positive/negative).	If (number of observations <=““b=”” style=”box-sizing: inherit;”> may be an overfitting.
Gaussian naïve Bayes	The estimation and preparing process is quick and gives the best outcomes with little informational indexes.	All attributes of the training data are assumed to be independent of other characteristics.
K-neighbour classification	Since no training period is required, new data can easily be added.	The estimation and preparing process is quick and gives the best outcomes with little informational indexes.
Random forest	Bagging and set learning, which reduces overfitting and improves the accuracy of a model.	It is complex in nature.

3 Comparison of different machine learning algorithm

Here five algorithms are used in this study, namely (svm), lgr, GNB, NB classifier K, and random forest. After experimenting with this algorithm, we applied the XGBoost and Isolation forest to find the best algorithm parameters that would give our model a good accuracy rating.

- Support vector machine (Makki et al., 2019): we use the SVM algorithm because it works well for problems based on nonlinear classification, it also works with uneven data structure, and because it reduces the risk of overfitting in a model to a very low level.
- Logistic regression (Dal Pozzolo et al., 2018): works best when applied to data that contains mapped attributes. The need for computing resources is very low. Since it is easy to implement, we can mark it as our benchmark and then work on other algorithms. It usually has the best informational output for the classification method.
- GNB (naïve Bayes): it is a conditional probability algorithm, so it is good to work with datasets in real time. It can lead to a good recommendation system design. It can be applied to large datasets. Use a formula to find the conditional probability. This formula is:

$$P(A \mid B) = P(B \mid A) * P(A) \mid P(B)$$

where $P(A \mid\mid B)$ = posterior probability, $P(B \mid A)$ = prior probability, $P(A)$ = probability, $P(A)$ = evidence. The result is a probabilistic prediction with less training dataset. It can also process continuous and discrete data.

- K-neighbour classifier: it is useful for taking care of loud information. It is a memory-based methodology where we can utilise the two sorts of arrangement (twofold class and various class) without extra exertion. The assessment and getting ready cycle is fast and gives the best results with minimal enlightening lists.
- The random forest: the information does not need to be rescaled or changed into this. It very well may be applied to arrangement and relapse issues. The calculation isolates the information dependent on its trademark and each tree has a high difference and a low inclination, which prompts a decent outcome. It prepares the model at rapid, is likewise simple to carry out and can deal with a considerable lot of element misfortune and dataset blunders (Li et al., 2020). They proposed another sort of misfortune work, full focus misfortune (FCL), which thinks about the two distances and points among highlights.

3.1 Methodology

A systematic literature review is the most commonly used method for the systematic literature review used to compose this systematic literature review. In accordance with this methodology, we investigated and discussed the relevant articles in this literature review.

3.2 Inclusion and exclusion

We follow a criterion for the inclusion of articles in this literature review. This criterion ensures that the added works are complete and written in English.

3.3 Quality inspection

Before articles are added to this literature review, their quality is assessed. The quality assessment process is carried out based on the researcher's work presented in this article.

In our study, we will discuss different types of machine learning algorithms used by different researchers for the purpose of CCF detection. Our main objective is to provide a complete literature review on this problem. This literature search also identifies various attributes of the clients that can be used to train the algorithm for this purpose (Wu and Liu, 2019; Carneiro et al., 2017; Adewumi and Akinyelu, 2017; Carminati et al., 2015). The proposed system is detection and before preventing fraudulent transactions and activities.

4 Method used

4.1 Dataset

In this assessment the charge card deception revelation dataset was used, which can be downloaded from Kaggle.

This dataset contains trades, occurred in two days, made in September 2013 by European cardholders. The test size of informational index which is in csv design contains 144 mb (Awoyemi et al., 2017).

The dataset contains 31 mathematical highlights. Since a portion of the information factors contains monetary data, the PCA change of these data factors were acted to keep these data obscure. Three of the given features were not changed. Feature 'time' shows the time between first trade and every single trade in the dataset. Feature 'sum' is the proportion of the trades made with Visa. Feature 'class' addresses the imprint, and takes only two characteristics: regard 1 if there ought to be an event of blackmail trade and 0 regardless.

Dataset contains 284,807 trades where 492 trades were cheats and the rest were affirmed (Awoyemi et al., 2017). Contemplating the numbers, we can see that this dataset is uncommonly imbalanced, where only 0.173% of trades are named as cheats. Since the circulation proportion of classes assumes a significant part in model exactness and accuracy, pre-preparing of the information is essential.

Features in credit card data

1 credit card 16-digit number

2 transaction amount

3 cardholder billing address

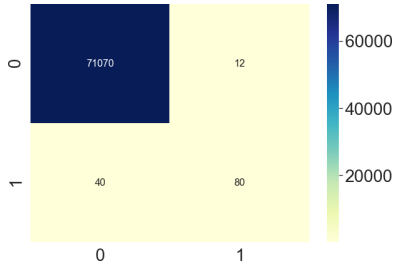
4 date local transaction

5 time local transaction

6 system trace audit number

7 authorisation identification response.

Figure 12 Confusion matrix of logistic regression and naïve Bayes (see online version for colours)



4.2 Algorithm implementation

At the point (support vector algorithm) to the informational collection, we partition it into preparing information and test information by applying a boundary to our preparation test division strategy. As a boundary, we name the size of the test, which is 0.25 for the informational collection. This implies that the all-out information is partitioned into two sections where it is 0.75 for preparing purposes and the rest for testing purposes 0.25 as referenced. We additionally pass every one of the characteristics and the objective class

in two unique factors. Then, we fabricate a model of our SVC technique wherein the preparation information is fit to prepare the model. When the model is prepared, it predicts the qualities from our test information. Therefore, we get the arrangement report and the confusion matrix.

Figure 13 In this 0 represents legitimate and 1 represents fraud transactions

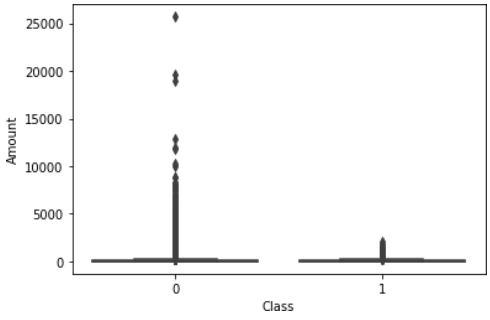
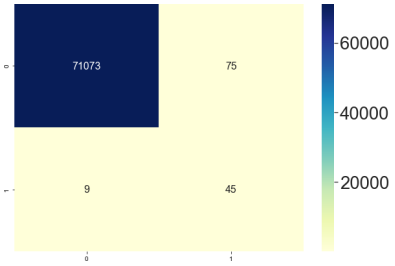


Figure 14 Confusion matrix of support vector machine (see online version for colours)



The above network is the eventual outcome of the SVM calculation with certifiable positive is 71,073, fake negative is 75, fake positive is 9, authentic negative is 45. In this model, according to our chaos lattice, the right expected characteristics are 71,118 and the Incorrect expected characteristics are 84. If we use the standard show measure Matthews Association Coefficient (MCC) for twofold course of action, the best motivator for SVC is 0.558. The collaboration for any leftover estimation is something almost identical, yet the procedures have been changed. Additionally, for the calculated relapse.

The above network is a consequence of strategic relapse. In which genuine positive is 71,070, bogus positive is 40, bogus negative is 12, genuine negative is 80. In this model, the right prescient qualities are 71,150 and wrong prescient qualities are 52. Furthermore, the MCC score for the strategic relapse model is 0.761. Since the best score of MCC is 1 when contrasted with MCC's most noteworthy score it is a further developed calculation for Visa misrepresentation location.

Guileless Bayes, K-closest neighbour and irregular backwoods utilised strategies are calculated relapse, Gaussian NB, K-neighbours classifier, and arbitrary woods classifier individually. Accordingly, we have produced the disarray lattice for every one of the calculations and on the best resultant calculation.

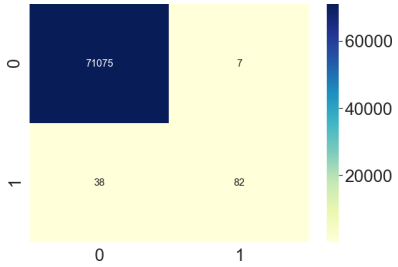
The lattice shaped by the Zhang et al. (2020) and Samy et al. (2020) naive Bayes calculation compares to the strategic relapse for this informational index. This implies

that the credulous Bayes calculation MCC score for the recorded informational collection is 0.761. The calculations can deliver various outcomes for an alternate Mastercard misrepresentation recognition informational index, which can change in size dependent on the f1 score and recuperation esteems in the two calculations.

Table 3 For comparing the models MCC values

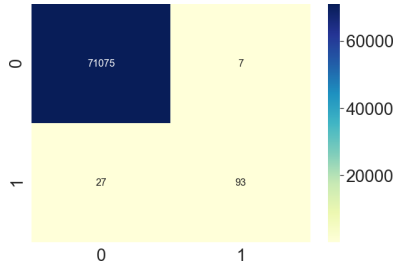
Models	*Recall value*	*F1-score*	*MCC value*
SVM	0.82	1.00	0.568
Naïve Bayesian	0.81	0.98	0.771
Logistic regression	0.93	1.00	0.771
Knn	0.74	1.00	0.783
Isolation forest	0.79	1.00	0.858

Figure 15 Confusion matrix for K nearest neighbour (see online version for colours)



The matrix displayed above is the after-effect of the KNN calculation. Where genuine positive qualities are 71,075, bogus positive qualities are 38, bogus negative worth is 7, genuine negative. The qualities are 82. For this model, the right anticipated qualities are 71.157 and the inaccurate anticipated qualities are 45. Furthermore, the MCC score for this KNN model is 0.793.

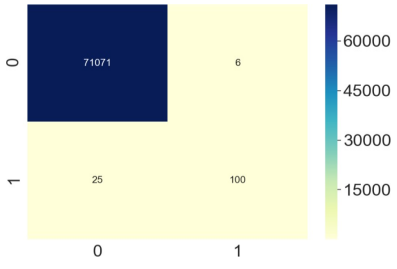
Figure 16 Confusion matrix for isolation forest (see online version for colours)



The matrix displayed above is the consequence of the arbitrary backwoods calculation. The boundaries utilised are n_estimator's 190, the rule utilised is 'entropy', max_depth is 10. For the genuine positive qualities 71,075, the bogus positive qualities 27, the bogus negative qualities 7 and the bogus positive qualities are 38. The right forecasts made by this calculation are 71,168 and the bogus expectations are 34. At long last, the irregular woods MCC score with the above boundaries is 0.848.

Presently on the off chance that we check out the correlation table, the best score for MCC is 0.848 given by segregation woodland with the arbitrary boundaries. Then, at that point, we pick the separation backwoods calculation and apply the XGBoost strategy to track down the best boundaries and again with the new boundaries we produce a model and think about the outcomes.

Figure 17 Isolation forest with XGBoost algorithm with confusion matrix (see online version for colours)



The matrix created above is the consequence of the arbitrary woodland with the network search boundaries and the boundaries are n_estimators are 500, the most extreme attributes utilised are auto, max_depth is 10, the measures is entropy. Because of the disarray lattice, you have 71071 genuine positive qualities, 6 bogus negative qualities, 25 bogus positive qualities, and 100 genuine negative qualities. That is, the right conjecture esteems are 71.171 and the erroneous estimate esteems are 31. Furthermore, the MCC score of the new coming about calculation is 0.89.

Table 4 After applying XGBoost generated parameters

Models	*Recall value*	*F1-score*	*MCC value*
XGBoost with random parameters	0.89	1.00	0.848
XGBoost with Grid Search parameters	0.90	1.00	0.89

In the table, we can see that the best worth diverged from the best MCC regard is 1. The accompanying worth is made by the IF calculation with new boundaries created by the framework search calculation which is 0.89. From here we can see since we are further developing results.

5 Conclusions

Is representation recognition utilising MasterCard is an intense issue in monetary administrations. The misfortune because of Visa misrepresentation is expanding with the increment in online business. This investigation manages methods that assist with discovering the MasterCard extortion. There are various technique used in machine learning to find out the online credit card fraud and to generate a fraud risk score Using XGBoost, Isolation forest, KNN, Logistic regression, SVM, and to solve classification problem.

References

Adewumi, O. and Akinyelu, A.A. (2017) 'A survey of machine-learning and nature-inspired based credit card fraud detection techniques', *Int. J. Syst. Assurance Eng. Manage.*, November, Vol. 8, No. S2, pp.937–953.

Alam, T.M. et al. (2020) 'An investigation of credit card default prediction in the imbalanced datasets', *IEEE Access*, October, Vol. 8, pp.201173–201198, DOI: 10.1109 /ACCESS.2020.3033784.

Arun, C. and Lakshmi, C. (2020) 'Class imbalance in software fault prediction data set' in *Artificial Intelligence and Evolutionary Computations in Engineering Systems*, pp.745–757, Springer, Singapore.

Arun, C. and Lakshmi, C. (2021) 'Genetic algorithm-based oversampling approach to prune the class imbalance issue in software defect prediction', *Soft Comput*., https://doi.org /10.1007/s00500-021-06112-6.

Awoyemi, J.O., Adetunmbi, A.O. and Oluwadare, S.A. (2017) 'Credit card fraud detection using machine learning techniques: a comparative analysis', in *Proc. ICCNI*, Lagos, Nigeria, October, pp.1–9.

Can, B. et al. (2020) 'Closer Look into the characteristics of fraudulent card transactions', *IEEE Access*, September, Vol. 8, pp.166095–166109, DOI: 10.1109/ACCESS.2020.3022315.

Carminati, M., Caron, R., Maggi, F., Epifani, I. and Zanero, S. (2015) 'BankSealer: a decision support system for online banking fraud analysis and investigation', *Comput. Secur.*, September, Vol. 53, No. 1, pp.175–186.

Carneiro, N., Figueira, G. and Costa, M. (2017) 'A data mining-based system for credit-card fraud detection in e-tail', *Decis. Support Syst.*, March, Vol. 95, pp.91–101, https://doi.org/10.1016/j.dss.2017.01.002.

Cody, T., Adams, S. and Beling, P. A. (2018) 'A utilitarian approach to adversarial learning in credit card fraud detection', in *Proc. Syst. Inf. Eng. Design Symp. (SIEDS)*, April, pp.237–242.

Correa Bahnsen, A., Aouada, D., Stojanovic, A. and Ottersten, B. (2016) 'Feature engineering strategies for credit card fraud detection', *Expert Syst. Appl.*, June, Vol. 51, pp.134–142, DOI:10.1016/j.eswa.2015.12.030.

Dal Pozzolo, A. et al. (2018) 'Credit card fraud detection: a realistic modeling and a novel learning strategy', *IEEE Transactions on Neural Networks and Learning Systems*, August, Vol. 29, No. 8, pp.3784–3797.

Dawood, E.A.E. et al. (2019) 'Improve profiling bank customer's behavior using machine learning', *IEEE Access*, August, Vol. 7, pp.109320–109327, DOI: 10.1109/ACCESS .2019.2934644.

Feng, F. et al. (2020) 'Using cost-sensitive learning and feature selection algorithms', *IEEE Access*, April, Vol. 8, pp.69979–69996, DOI: 10.1109/ACCESS.2020.2987364.

Hines, C. and Youssef, A. (2018) 'Machine learning applied to rotating check fraud detection', in *Proc. 1st Int. Conf. Data Intell. Secur. (ICDIS)*, April, pp.32–35.

Jiang, C., Song, J., Liu, G., Zheng, L. and Luan, W. (2018) 'Credit card fraud detection: a novel approach using aggregation strategy and feedback mechanism', *IEEE Internet Things J.*, October, Vol. 5, No. 5, pp.3637–3647.

Kalid, S.N. et al. (2020) 'A multiple classifiers system for anomaly detection in credit card data with unbalanced and overlapped classes', *IEEE Access*, February, Vol. 8, pp.28210–28221, DOI: 10.1109/ACCESS.2020.2972009.

Kim, E., Lee, J., Shin, H., Yang, H., Cho, S., Nam, S-K., Song, Y., Yoon, J-A. and Kim, J-I. (2019) 'Champion-challenger analysis for credit card fraud detection: hybrid ensemble and deep learning', *Expert Syst.Appl.*, August, Vol. 128, pp.214–224, https://doi.org/10.1016 /j.eswa.2019.03.042.

Li, Z., Liu, G. and Jiang, C. (2020) 'Deep representation learning with full center loss for credit card fraud detection', *IEEE Transactions on Computational Social Systems*, 17 February, April, Vol. 7, No. 2, pp.569–579.

Makki, S. et al. (2019) 'An experimental study with imbalanced classification approaches for credit card fraud detection', *IEEE Access*, July, Vol. 7, pp.93010–93022, DOI: 10.1109/ACCESS.2019.2927266.

Omar, B., Rustam, F., Mehmood, A. and Choi, G.S. (2021) 'Minimizing the overlapping degree to improve class-imbalanced learning under sparse feature selection: application to fraud detection', *IEEE Access*, February, Vol. 1, No. 9, pp.28101–28110.

Randhawa, K. et al. (2018) 'Credit card fraud detection using AdaBoost and majority voting', *IEEE Access*, Vol. 6, pp.14277–14284, DOI: 10.1109/ACCESS.2018.2806420.

Roy, A., Sun, J., Mahoney, R., Alonzi, L., Adams, S. and Beling, P. (2018) 'Deep learning detecting fraud in credit card transactions', in *Proc. Syst. Inf. Eng. Design Symp. (SIEDS)*, April, pp.129–134.

Samy, S.S. and Parthiban, L. (2018) 'A novel feature extraction approach using principal component analysis and quantum behaved particle swarm optimization-support vector networks for enhancing face recognition', *Journal of Computational and Theoretical Nanoscience*, Vol. 15, Nos. 9–10, pp.3012–3016.

Samy, S.S., Sivakumar, V., Sood, T. and Negi, Y.S. (2020) 'Intelligent web-history based on a hybrid clustering algorithm for future-internet systems', in *Artificial Intelligence and Evolutionary Computations in Engineering Systems*, pp.571–581, Springer, Singapore.

Selvakumarasamy, S. and Dekson, D. (2013) 'Architecture of adaptive e-learning ecosystem', *International Journal of Emerging Trends & Technology in Computer Science (IJETTCS)*, ISBN No: 978-93-80609-14-0, ISSN 2278-6856.

Shaikh, A.K. and Nazir, A. (2020) 'A novel dynamic approach to identifying suspicious customers in money transactions', *International Journal of Business Intelligence and Data Mining*, Vol. 17, No. 2, pp.143–158.

Taha, A.A. and Malebary, S.J. (2020) 'An intelligent approach to credit card fraud detection using an optimized light gradient boosting machine', *IEEE Access*, February, Vol. 8, pp.25579–25587, DOI: 10.1109/ACCESS.2020.2971354.

Tounekti, O. et al. (2019) 'Users supporting multiple (M)EPS in online purchases', *IEEE Access*, 23 December, Vol. 8, pp.735–766, DOI:10.1109/ACCESS.2019.2961785.

Umuhoza, E. et al. (2020) 'Using unsupervised machine learning techniques for behavioral-based credit card users segmentation in Africa', *SAIEE Africa Research Journal*, September, Vol. 111, No. 3, pp.95–101.

Wu, H. and Liu, G. (2019) 'A hybrid model on learning cross features for transaction fraud detection', in *Proc. ICDM*, pp.88–102.

Wu, H. and Wang, C-C. (2018) 'Customer segmentation of credit card default by self-organizing map', *American Journal of Computational Mathematics*, Vol. 8, No. 3, p.197.

Yee, O.S., Sagadevan, S. and Malim, N.H.A.H. (2018) 'Credit card fraud detection using machine learning as data mining technique', *Journal of Telecommunication, Electronic and Computer Engineering (JTEC)*, Vol. 10, Nos. 1–4, pp.23–27.

Zhang, F., Liu, G., Li, Z., Yan, C. and Jiang, C. (2019) 'GMM-based undersampling and its application for credit card fraud detection', in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, July, pp.1–8.

Zhang, Z. et al. (2020) 'Fraud detection method for low-frequency transaction', *IEEE Access*, 31 January, Vol. 8, pp.25210–25220, DOI: 10.1109/ACCESS.2020.2970614.

Zheng, L., Liu, G., Yan, C. and Jiang, C. (2018) 'Transaction fraud detection based on total order relation and behavior diversity', *IEEE Trans. Comput. Social Syst.*, September, Vol. 5, No. 3, pp.796–806.

Zheng, L., Liu, G., Yan, C., Jiang, C., Zhou, M. and Li, M. (2020) 'Improved TrAdaBoost and its application to transaction fraud detection', *IEEE Transactions on Computational Social Systems*, 27 August, Vol. 7, No. 5, pp.1304–1316.