IRAQI Academic Scientific Journals

العـــراقيـة المجـلات الأكـاديمـيـة العـلمـيـة

Kut University Journal

ISSN (E): 2616 - 7808 II ISSN (P): 2414 - 7419

www.kutcollegejournal.alkutcollege.edu.iq

k.u.c.j.sci@alkutcollege.edu.iq

Vol. 10 , Issue 1 , June 2025

Credit Card Fraud Detection Using Data Mining Methods

Lec. Zahraa Raji Mohi [1]

Abstract

Focusing on Data-mining role in fraud detection for discovering the threats on credit card transactions by the increase of financial deals and huge information used by it, while the threats change continuously are evaluated which determined by a group of odd and anomalous behaviors, an example of detecting fraud in financial credit card transactions which represent one of the basic things in our life and significant in business and financial deals that is an interested subject for study by its important in life. An important requirement of fraud detection is finding system able to detect a various types of credit card attacks and an effective procedure for detect it, depends on various credit cards type and different fraud in financial deals used in trade, banks and industry, there is many measures can reduce fraud through using data mining techniques, we take two of them and comparing a results to for best reduce and detect fraud trying in financial domain for credit card which are Adaboost classifier and Hidden markov model.

Using these techniques objective is to minimize frauding on credit card. By finding fraudulent clients and merging classification methods focusing on two different algorithms, discovering the degree of fraudulent activities in the financial domain the the used techniques are Adaboost and Hidden Markov model.

Keywords: Fraud Detection, Credit Card Fraud, Data Mining

Affiliation of Author

[1] Department of Invitation Rhetoric and Thought, Alimam aladham University college, Iraq, Baghdad, 10001

[1] zahraarajizs@gmail.com

[1] **Corresponding Author**

Paper Info.
Published: Jun. 2025

كشف احتيال بطاقات الائتمان باستخدام طرق تعدين البيانات

زهراء راجي محي [1]

المستخلص

نقل الاموال بواسطة بطاقات الائتمان يتعرض الى عمليات احتيال نظرا للستعمال الهائل لها فان التركيز على دور تعدين البيانات في كشف الاحتيال بينما تكون هذه العمليات متغيرة ومستمرة من قبل مجاميع شاذة وضارة التصرف، وكمثال لكشف الاحتيال المالي الحاصل على نقل الاموال ببطاقات الائتمان وقد اصبحت من اساسيات حياتنا ولما لها من اهمية في الاعمال والتعاملات المالية وهي مادة مهمة للدراسة لاهميتها في الحياة.

من المتطلبات المهمه لكشف الاحتيال هي ايجاد نظام قادر عدة انواع من الهجوم على بطاقات الائتمان واجراء فعال لكشفه، اعتمادا على انواع بطاقات الائتمان المنوعة وباختلاف انواع الاحتيال في التعاملات المالية المستعملة بالتجارة والمصارف والصناعة، هناك عدة مقاييس بامكانها تقليل الاحتيال باستخدام تقنيات تعدين البيانات سنأخذ اثنان منها ونقارن النتائج الافضل التي تقلل وتكشف الاحتيال في المجال المالي لبطاقات الائتمان والتقنيات هي Adaboostالمصنف و نموذجHidden markov.

الكلمات المفتاحية: تعدين البيانات، كشف الاحتيال، احتيال بطاقات الائتمان

انتساب الباحث

[1] قسم الدعوة والخطابة والفكر، كلية الامام الأعظم الجامعة، العراق، بغداد، 10001

[1] zahraarajizs@gmail.com

[1] **المؤلف المراسل**

معلومات البحث
تأريخ النشر : حزيران 2025

1. Introduction

Fraud detection includes determining a set of malicious actions that compromise the integrity, confidentiality, and availability of information resources. Data mining is one of the important ways to Detect fraud because of its efficiency, which is a well-defined procedure that takes data as input and returns output in the form of patterns. In other words, its role is to resolve a huge quantity of data and to extract useful information that is easy to deal with. A model of data used can predict future behavior through data classification. In fraud detection, data mining techniques can be used to understand data by analyzing it with an appropriate model to detect fraud activities[1].

A classification model can be used for predicting fraudsters, which involves a set of machine learning methods able to extract important types from data and discover unknown information hidden in the given data. This is known as knowledge discovery. The mined information may be very helpful for companies that use data mining, as the consequence can help them make significant decisions to excel in their competition. In the fraud process, data mining enables companies to focus on important information to reveal fraudsters and predict new behaviors to avoid them [2].

Furthermore, the only method to identify credit card fraud is to use data mining techniques to analyze client spending patterns in order to determine which clients are most likely to default on their obligations. Data mining can also be used to uncover fraud in credit card financial transactions. Then, distinguish between questionable transactions and a usual expenditure profile. The machine learning algorithms that use data mining techniques often learn models from the data. It is a procedure that involves examining the facts from several angles and condensing them into insightful knowledge. It may also be used to validate different facts in credit card fraud detection and management [3].

Predicting whether a customer is in default or whether giving a credit card to a specific client would result in a low credit score are other useful insights. In order to identify fraud, the data mining tool will translate all of the data banks and create a number of new rules. Additionally, the branch where these issues arise is indicated. Analyzing data patterns such as client behavior and dependability, client transaction history, client diagrams, and transactions that may result in fraud is also helpful. Every customer records credit card information on a single mainframe system; these files hold all of the fundamental data associated with a credit card. These details are fundamental to the investigation of any unlawful activity related to the credit card procedure. [3].

One of the most important problems in internet financial dealings is that when doing the transaction, the card and the cardholder do not need to be ready. This makes it hard to test whether or not the client who is making a transaction is the true cardholder. That makes it easy and possible for a fraudster to conduct a fraudulent transaction. This paper is arranged as follows: Section 2 describes a credit card definition and different types of fraud in the financial field, Section 3 discusses data mining technology, Section 4 discusses a proposed methodology and the comparison of methods, and finally, the results and conclusion are highlighted in Section 5.

2. Credit card and types of fraud in the financial domain:

A credit card is a basic, appropriate plastic card that authorizes the person listed on it to make purchases or manage charges to their account, for which they may occasionally be charged. It may also contain personal data, such as a signature, image, card number, or magnetic chip data. Robotic teller machines (TMs), retail bar code scanners, banks, and internet web banking frameworks may all access card data. They have a very rare and distinct code number. The credit card number must be protected for it to be secure. Using a credit card as a fictitious source of base in a transaction is known as credit card fraud, which is a broad category of theft [4].

Payment information is transferred via the internet between the customer's PC and the dealer's shop when selling credit cards online. Identity theft and credit card cybersecurity are raised by this. The majority of online stores have security measures in place to stop hackers from accessing customer data, and your browser should indicate this by displaying a secure site icon. The transfer of credit card information and other personal information may be considered theft if there is no indication that the website is secure [3].

Before entering the data, one should exercise caution and thoughtfulness. Three broad categories can be used to further categorize credit card fraud: traditional card fraud (fake, application, stolen, account takeover, and counterfeit), merchant fraud, and Internet fraud (site cloning, credit card generators, and fraudulent merchant sites). Additionally, customers using credit cards can carry interest-free balances for roughly two months because they can do so not just during the credit cycle but also for a "grace period" that lasts for at least twenty days after the credit period ends [4].

Fraudsters typically fall into one of three categories:

- **Pre-planned fraudsters:** those who plan to commit fraud from the outset. These can be longer-term agents, such as those who commit intricate money laundering schemes and bankruptcy fraud, or they can be short-term agents, such as many who use stolen credit cards or incorrect social security numbers.
- **Intermediate fraudsters:** These are people who are initially honest but turn to fraud in hard times or when life events alter their usual course, such as being thrilled to be passed over for a promotion or having to pay for a family member.
- **Slippery-slope fraudsters:** These individuals can be either large-scale businesspeople or regular traders who simply continue trading even when they are objectively unable to pay their debts [5].

Who are the people who steal credit cards?

I. **I. Buyers of credit card information: These** are the unscrupulous individuals who purchase credit card information on websites through illegal means, often lacking any computer expertise (such as in computer programming or networking). They purchase this credit card information with the goal of using it to electronically pay for goods and services that they find online.

II. **Black hat hackers:** those who compromise computer security for their own gain or with malevolent intent. They use two procedures referred to as the "pre-hacking stage" to select their targets: targeting, research, information gathering, and carrying out the attack.

III. **Black hat hackers** are those who compromise computer security for their own gain or with malevolent intent. They use two procedures referred to as the "pre-hacking stage" to select their targets: targeting, research, information gathering, and carrying out the attack. These hackers are very proficient in computer programming and networking, and they have the ability to disrupt a computer network. The goal of their hacking or intrusion is to take personal data, including bank account and credit card details, among other things.

IV. **Those who physically steal credit cards** and write information on them are known as physical credit card thieves. They physically take these plastic credit cards (possibly by pick-pocketing in a busy area) and write down the credit card details with the goal of using these details to make an electronic payment for a good or service, and services on the internet [6].

3. Data mining

In order to find legitimate and hidden relationships among large data sets, data mining employs data analysis. This includes machine learning techniques like AdaBoost and Hidden Markov Model, which are the ones used in this paper, as well as statistical models and algorithms. Data mining is the process of taking given data sets and using them for administration, analysis, and prediction. Since credit card fraud (CCF) is a common task when following standard procedures, the credit card fraud detection model has recently gained importance in both the business and academic communities. These models, which are primarily artificial intelligence or statistics-driven, have the potential benefit of not imposing arbitrary assumptions on the input variables [2].

Information on fraudulent activities is a primary objective and a wise tactic for banks and industries both have enormous databases. Sometimes, however, accessing databases is challenging. Data stores, where data has been temporarily stored, can yield valuable business information. The process of identifying fraudulent transactions and dividing databases into two classes—legitimate and fraudulent—is known as credit card fraud detection [5].

One of the four detection techniques used to find malware nowadays is data mining. Data mining techniques are used by developers of security applications to enhance malware detection speed and quality and to detect more zero-day attacks. Three methods exist for identifying malware[7]:

a) **Anomaly detection:** The process of finding unusual occurrences or observations that cause suspicion because they deviate noticeably from the bulk of the data is known as anomaly detection. Finding deviations from typical usage patterns entails modeling a system's or network's normal behavior. Additionally, anomaly-based detection can identify previously unidentified attacks and be used to define signatures for misuse detectors. Anomaly detection's primary flaw is that it reports any deviation from the norm as an anomaly, even when it's a valid behavior. This leads to a high number of false positives.

b) **Misuse detection:** using samples of their signatures, it only recognizes known attacks. It describes the process of identifying attacks by searching for particular patterns, like byte network traffic, or malware's known malicious instruction sequences. Although the false positive rate is lower with this method, zero-day attacks cannot be detected.

c) **A hybrid approach**, which reduces the amount of false positives while increasing the number of detected intrusions by combining anomaly and misuse detection techniques. Instead of creating any models, it uses data from both safe and dangerous programs to build a classifier, which is a collection of guidelines or a detection model produced by the data mining algorithm. Next, the misuse detection system looks for malware signatures in the code, and the anomaly detection system looks for departures from the typical profile.

The most widely used and well-liked data mining tools are listed below:

- **Rapid Miner:** This widely used software provides advanced analytics, is ready-made, is open source, and requires no coding knowledge. Constructed in Java, it encompasses diverse data mining features like preprocessing, visualization, and predictive analysis. It can be effortlessly integrated with R-tool and WEKA to generate models directly from scripts written in the former two.
- **WEKA:** This free customization tool has a Java foundation. Clustering, association, regression, classification, and predictive analysis and modeling techniques are all included.
- **R-Programming Tool:** Scripts can be written by data miners using this C and FORTRAN-written tool, just like they would on a programming language or platform. As a result, statistical and analytical software for data mining is made with it. It allows for both linear and graphical analysis.

  And nonlinear modeling, classification, clustering, and time-based data analysis.
- **Orange and NTLK:** based on Python. Python is widely used because of its robust features and ease of use. Orange is an open-source Python tool with a visual programming interface that includes helpful machine learning, text analysis, and data analytics features. NTLK, which is also written in Python, is a potent language processing data mining tool that includes machine learning and data mining and can be readily expanded to meet specific requirements.
- **Knime:** A powerful tool with a graphical user interface (GUI) that displays the network of data nodes. Knime is primarily used for preprocessing, data extraction, transformation, and loading. It is well-liked by financial data analysts and features modular data pipelining, data mining concepts, and machine learning. Generously for creating reports on business intelligence [8].

4. Proposed Methodology(*Detection process):*

Data mining fraud detection consists of two steps (Extracting features, Classifying/clustering). The proposed method is seen as part of the payment gateway that would test whether or not a transaction is fraudulent. It should function on the bank data and the payment data. Figure 1 below shows the proposed fraud identification scheme. The functions of the fraud identification module are as follows:

- ✓ The payment entrance has to provide a data set of credit card details, like the card number and expiration date. Additional details like the delivery date and time, sales number, mailing address, etc.
- ✓ The Fraud Detection module needs to receive the necessary information from the payment entrance, such as transaction history.
- ✓ Using the most efficient classification algorithm, it trains itself using effective data mining techniques.
- ✓ Depending on the information provided, the payment entrance would be notified of the outcome of a transaction, whether it was fraudulent or not.
- ✓ The two suggested algorithms are used to evaluate the credit card information, and the output provides the judge and other relevant details to the payment entrance administrator as a final report. Both suggested methods are used to evaluate credit card information, the

information transaction from entered data to the output decision shown in Figure (1).

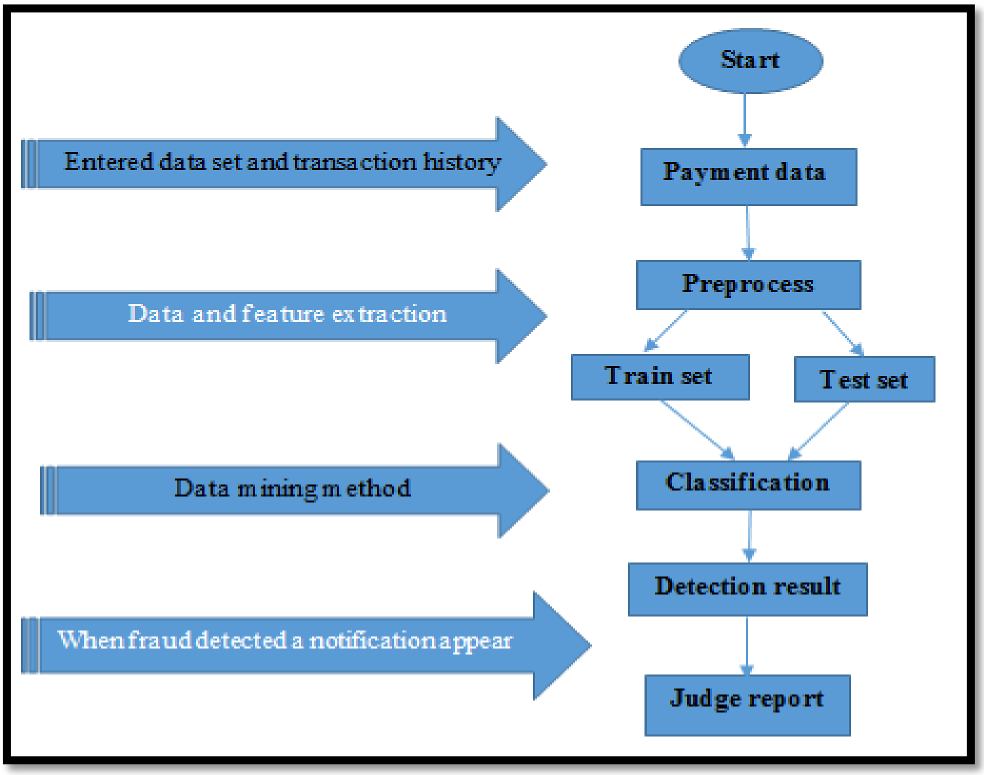

Start
Entered data set and transaction history
Payment data
Data and feature extraction
Preprocess
Train set
Test set
Data mining method
Classification
Detection result
When fraud detected a notification appear
Judge report

Figure (1): proposed method system

Data mining techniques aid in the analysis of these types of transactions and patterns that point to fraud. The banking industry works harder to detect fraud. Determining which transactions are not ones the user would be performing is crucial to the detection of fraud. There are two categories for fraud detection: supervised and unsupervised. Supervised methods utilise a database of known fraudulent/legitimate cases, from which a model has been built. When a new case has a different score after past behaviour analysis, the model generates a suspicion score. Rule-based techniques known as supervised learning algorithms generate classifiers by applying the following types of rules:

***If* {certain conditions}, *Then* {a consequent}**

It is possible to employ combinations of some or all of these algorithms as meta-learning algorithms, which enhance prediction in fraud detection. Uneven class sizes and varied costs for various forms of misclassification are important factors to take into account when developing a supervised tool for fraud detection. However, the costs of looking into observations and the benefits of spotting fraud must also be taken into account [9].

When there are no previous sets of authentic and fraudulent observations available, unsupervised methods are employed. Often, a combination of outlier detection and profiling techniques is used in certain situations. A model exists that uses a baseline distribution to represent normal behavior, and it looks for comments that deviate most from the norm. There would be a specific form to the distribution of the first significant digits of numbers selected from a wide range of arbitrary distributions. As of right now, this law is considered to be nothing more than a mathematical curiosity with no apparent practical use.

Legitimate user accounts may alter their behavior over a longer period of time, and it is important to prevent false alarms. This law can be used to detect fraud in accounting data [1,9].

4.1 Hidden Markov Model (HMM)

It is an embedded, hybrid saved procedure model. The Markov model is not as complex as this generalized process. This is a measured risk and a fake transaction if the learner with a high potential probability rejects the hidden Markov model bank transaction. The model divides transactions into three categories: high, average, and low. The most basic models that can be applied to sequential data modeling. In Markov models, the state is directly visible to the observer; however, in hidden Markov models, the output is visible because it depends on the state and is not directly visible.

The HMM consists of a limited number of states, each of which has a probability distribution attached to it. A collection of probabilities known as transition probabilities controls changes in state. The associated probability distribution can be used to generate an outcome or observation in a given state. States are therefore "hidden" from the outside since it is only the result that is visible, not the state itself, to an outside observer. HMM implementation involves grouping training sets according to the cardholder's spending profile. The kinds of things that are bought serve as the model's states. The probability distribution controls the change from one state to another. A minimum of ten prior transactions is needed, and the basis for the upcoming transaction is chosen as fraud or genuine [10].

Let X_n and Y_n be discrete-time stochastic processes, and $n \geq 1$. The pair (X_n, Y_n) is a hidden Markov model if

- X_n is a Markov process whose behavior is not directly observable ("hidden");
- $P(Y_n \epsilon A | X_1 = x_1, \ldots.., X_n = x_n) = P(Y_n \epsilon AX_n = x_n)$, for every $n \geq 1, x_1, \ldots, x_{n,}$ (1)

And every set, let X_t and Y_t be continuous-time stochastic processes. The pair (X_t, Y_t) is a hidden Markov model if:

- (X_t is a Markov process whose behavior is not directly observable ("hidden");
- $P(Y_{t0} \epsilon A \mid \{ X_t \epsilon B_t \} t \leq t0) = P(Y_n \epsilon A \mid X_{t0} \epsilon B_{t0})$ for every t0 every set A and every family of sets $\{B_t\}$ $t \leq t0$ [11]. (2)

The experiments showed that HMM could detect anomaly data quickly and at a lower mismatch rate. However, HMM training needs multiple passes through the training data, which takes a great deal of time. HMM training also requires extensive memory to store transition probabilities during training, especially for long sequences.

4.2 The AdaBoost classifier

This classifier takes internal metrics of misclassification value when calculating theories to detect fraud. Figure (2) shows the algorithm. Let S = ((x_1,c_1,y_1), …, (x_m, c_m,y_m)) be a sequence of training examples where each *instance xi* belongs to a *domain* X, each *cost factor* c_i belongs to the nonnegative real domain R^+, and each *label* y_i belongs to a finite *label space* Y. We only focus on binary classification issues in which Y = {–1, +1}. *H is a little hypothesis, it has the form h: X – R.* The sign of $h(x)$ is translated as the predicted label, and the magnitude $|h(x)|$ is the “confidence” in this prediction. Let t be an index to show the round of boosting, and $D_t(i)$ be the weight given to (x_i, c_i,y_i) at the tth round. $0 \leq D_t(i) \leq 1$, and $\sum Dt(i) = 1$ is the chosen parameter as a weight for weak hypothesis *ht* at the *t*th round.

We assume α_i *t > 0.* β(sign(y_i h_t (x_i)), c_i) is a cost-adjustment function with two arguments: sign(y_i h_t(x_i)) to show if h_t(x_i) is correct, and the cost factor c_i. Where it is obvious in context, we use either β(*i*) or β(c_i) as a shorthand for β(sign($y_i h_t$(x_i)), c_i). Also, we use β_+ when sign (y_i h_t(x_i)) = +1 and β_- when sign(y_i h_t(x_i)) = –1. For an instance with a higher cost factor, β(*i*) increases its weights "more" if the instance is misclassified, but decreases its weight "less" otherwise. further, we require β–(c_i) to be non-decreasing with respect to c_i,β_+ (c_i) to be non-increasing, and both are nonnegative. We proved that Adaboost minimizes cost on the training data. 1. Logically, we can assign a cost factor *c* of transmit overhead to frauds and a factor *c* of overhead to non-frauds. This shows how the prediction errors will add to the total cost of a hypothesis [12].

```
Given :( x1,c1,y1), ....,($x_m$,$c_m$ ,$y_m$):$x_i$ ϵ R+$y_i$ϵ {-1,+1}
Define D1(i) (such as $D_1$(i) = (such as $D_1$(i) = c| $\sum_j^m c_j$) for t=1, ...,T:
    1. Train weak learner using distribution $D_t$.
    2. Compute weak hypothesis $h_i$: x→R.
    3. Choose $\alpha_t$ϵ R and β(i) ϵ $R^+$
    4. Update
       $D_{t+1}$(i)= $\frac{\text{Dt+1(i)exp}(-\propto tytht(x_i)\beta(sign(yiht(xi)),ci))}{Z_t}$
       where β(sign(yiht(xi)),ci) is a cost-modification function. $Z_t$ is a
       normalization factor chosen so that $D_{t+1}$ will be a distribution .
Output the final hypothesis:
H(x) =sign(f(x) where f(x)= ($\sum_{t=1}^{T} \propto tHt(x)$)
```

Figure (2): Adaboost algorithm

5. Experimental results:

There are three features to calculate classification accuracy: precision, recall, and F1-measure.

For class ci, precision can be obtained as shown in equation (3) [13]:

Pi =T Pi/ (T Pi+ FPi) (3)

Which is specificity, is the normal patterns that were correctly detected as normal, *precision* It is the ration of actions correctly classified as *attack.*

Recall is calculated as shown in equation (4):

Ri = T Pi/ (T Pi+ FNi) (4)

This is known as the false alarm rate. It is the normal patterns that were falsely classified as an attack, where:

TPi refers to true positive.

FNi refers to false negative.

FPi refers to false positive.

The last calculated the F-measure: the harmonic means of recall and precision, also known as f-value or f-score, is obtained as the following function (5) and (6) :

$$F1= (2 *Recall *Precision=(Recall+ Precision)) \qquad (5)$$

$$= (2TP / (2TP+ Fp+ FN)) \qquad (6)$$

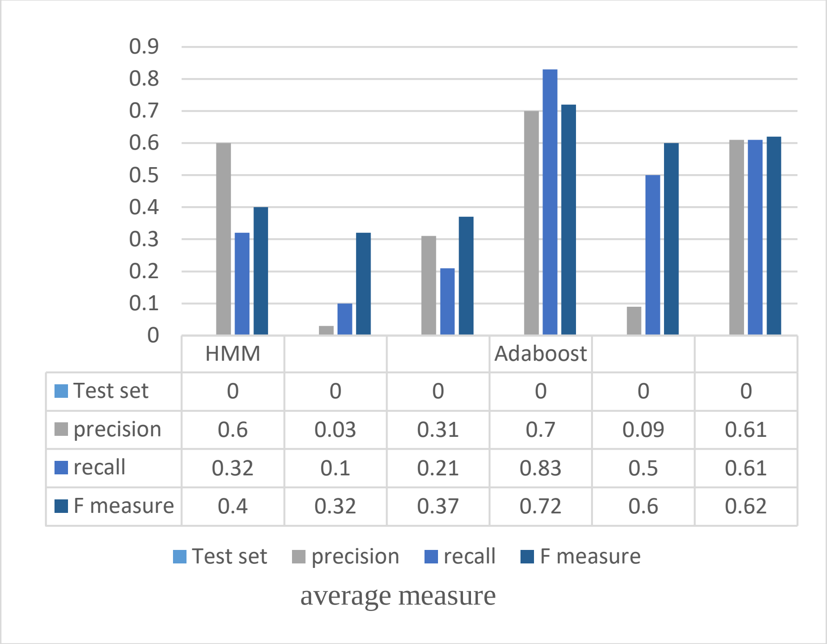
The experimentation dataset was obtained from "kaggle.com," which provides a vast dataset for researchers studying a variety of subjects. A credit card transaction with 3075 rows and 11 important features makes up the selected dataset. The performance is assessed using three metrics: accuracy, recall, and precision, which are computed as previously indicated. The results of the experiment demonstrate that the HMM classifier's performance is more likely to accurately suggest the level of fraudulent activity. The evaluation results of identifying fraudulent transactions have improved when employing HMM instead of an Adaboost classifier, as Tables (1) and (2) demonstrate. The suggested system uses a fraud identifier transaction algorithm that is more accurate in order to function better. As illustrated in Figure (3), we observed in this figure the performance comparison measure between HMM and Adaboost classifiers algorithims show that HMM performance score accurate and more significant, shows the precision, recall and accuracy for classifiers parameters that give the best classification performance of the algorithm, our proposed classification algorithm is applied on chosen data set, experimental results means that the significant increase in prediction performance may be achieved not only by the itemsets with high precision but also the itemsets with high recall and accuracy.

Table (1): Comparison of accuracy in HMM and Ada cost models

Model	Test set	accuracy	FB rate	FN rate
HMM	Normal	94.19	5.8	0.0
	Attack1	66.60	0.0	33.24
	Attack2	65.76	0.0	0.08
	Real	99.70	0.09	
Ada boost	Normal	94.44	5.23	20.80
	Attack1	80.10		11.12
	Attack2	87.80		
	Real	96.15	6.38	

Table (2): Average of HMM and Adaboost

model	Test set	precision	recall	F measure
HMM	Normal	0.6	0.32	0.4
	Attack	0.03	0.10	0.32
	Real	0.31	0.21	0.37
Ada boost	Normal	0.7	0.83	0.72
	Attack	0.09	0.50	0.60
	rael	0.61	0.61	0.62



	HMM			Adaboost		
Test set	0	0	0	0	0	0
precision	0.6	0.03	0.31	0.7	0.09	0.61
recall	0.32	0.1	0.21	0.83	0.5	0.61
F measure	0.4	0.32	0.37	0.72	0.6	0.62

Figure (3): Comparative Analysis for both classifiers

Conclusion

The likelihood of using data mining as a tool to detect fraud is high. It enables you to mine vast data sets for new knowledge and analyze them. The research community is very interested in credit card fraud detection when it comes to effective methods, and several approaches have been put forth to measure credit card fraud. Credit card fraud has become more significant as credit card usage spreads throughout all facets of daily life. As a result, it is necessary to automatically and effectively improve the security of financial transaction systems by developing an effective credit card fraud detection system.

Due to a lack of readily available data sets, two of the chosen models are challenging to implement. Additionally, in order for data mining to produce a distinct and efficient model for every specific use case, which is intricate, resource-intensive, and costly, we process massive datasets more quickly. It can be difficult to build a suitable classifier because it must be updated frequently to incorporate samples of fresh fraudulent data. It is necessary to increase the accuracy of fraud detection and guarantee the dependability of the agents by testing the system on an actual bank server to assess acceptability and performance.

Acknowledgement

First and forever to the god, my family, and for every friend who supports and encourages me.

References

[1] K. Sai Manoj, P. S. Aithal , “Data Mining and Machine Learning Techniques for Cyber Security Intrusion Detection ,” International Journal of Engineering and Advanced Technology, Vol.9, 2020.

[2] J. Waruwu1, W. Hadinata, “Data Mining Techniques in Detecting and Predicting Cyber Crimes In Marketplace Sector ,” JURNAL INFORMATIKA , Vol. 1 No. 1, 2022 .

[3] Niranjan A, Nitish A, P Deepa Shenoy and Venugopal K R, “Security in Data Mining- A Comprehensive Survey By Exploration of Data mining techniques in Fraud Detection:

Credit Card," Global Journals Inc., Vol. 16, 2016.

[4] Philip K. Chan, Wei Fan, Andreas L. Prodromidis, and Salvatore J. Stolfo, "Distributed Data Mining in Credit Card Fraud Detection,", IEEE, pp. 67-74, 1999.

[5] R. Patidar, L. Sharma, "Credit Card Fraud Detection Using Neural Network", International Journal of Soft Computing and Engineering (IJSCE), Vol.1, 2011.

[6] Amanze B.C., and Onukwugha C.G., " Data Mining Application in Credit Card Fraud Detection System," International Journal of Trend in Research and Development, Vol. 5,2018.

[7] VISHALAKSHI, N. S. and Deepika, N, "SURVEY ON DATA MINING METHODOLOGIES FOR CYBER CREDIT CARD AND CREDIT card FRAUD DETECTION SYSTEM" , *International Journal of Current Research, Vol. 8, pp.27479-27484, 2016.*

[8] Lawrence Borah1, Saleena B2, Prakash B3, "Credit Card Fraud Detection Using Data Mining Techniques", Journal of Seybold Report, pp. 2431-2436, 2020.

[9] P. Santhosh Raj1, G. Silambarasan, M.Phil Scholar, " Role of Data Mining in Cyber Security,", International Journal of Engineering Science and Computing, vol. 7, pp. 13931-13935, 2017.

[10] Emmanuel Prestat, Maude M. David and Jenni Hultman, "FOAM (Functional Ontology Assignments For Metagenomes): A Hidden Markov Model (HMM) Database With Environmental Focus", Nucleic Acids Research, vol.42, Issue 19, pp. e145-153,2014.

[11] Walter zucchini, lain l.macdonald and ronald rangrock, "Hidden Markov models for time series", CRC press a champion and hall books,2017.

[12] Yuanshen Zhao, Liang Gong and Bin Zhou,"Detecting tomatoes in greenhouse scenes by combining AdaBoost classifier and colour analysis" , Biosystems Engineering Elsevier,Vol.148, pp. 127-137,2016.

[13] A. Ahmim, M. Derdour, and M. A. Ferrag, "An intrusion detection system based on combining probability predictions of a tree of classifiers," ,Int. J. Commun. Syst, vol. 31, no. 9, pp. 3547–3547, 2018.