Al-Mustansiriyah Journal of Science
ISSN: 1814-635X (print), ISSN:2521-3520 (online) Volume 33, Issue 4, 2022 DOI: http://doi.org/10.23851/mjs.v33i4.1242

Research Article Open Access

Detection and Prevention WEB-Service for Fraudulent E-Transaction using APRIORI and SVM

Shatha Jassim Muhamed

Computer Science Department, College of Science, Mustansiriyah University, Baghdad, IRAQ.

Contact: shonash77@uomustaansiriyah.edu.iq
ORCID: https://orcid.org/0000-0003-4003-5366

Article Info

Received
17/10/2022

Accepted
01/12/2022

Published
30/12/2022

ABSTRACT

With the increased use of information technology, many financial services are available to users at their fingertips. However, this led to many fraud transactions. Automatic fraud identification and detection could improve the user experience and security of online transactions. Using machine learning algorithms, it is possible to detect fraud transactions. Machine learning algorithms have the ability to find the hidden implicit pattern and data relationships from a large dataset. Hence, using this algorithm is possible to detect the outlier from all transactions, which can help in determining the fraud transaction. In this paper, the APRIORI algorithm and Support Vector Machine are used to detect fraud transactions in credit cards via developing a secure web application service enforced the security by standard metrics. We compare the result with the other existing machine learning algorithms. We observed that the accuracy of fraud transaction detection is higher in the proposed algorithm more than 94.56, and the false fraud transaction detection is less than the fraud detection based on the Hidden Markov Model.

KEYWORDS: Fraud Transactions, APRIORI Algorithm, SVM, Confidence, Frequent Item Set, Ecommerce, Secure Web-Service, Hidden Markov Model.

الخلاصـة

مع زيادة استخدام تكنولوجيا المعلومات، تتوفر العديد من الخدمات المالية للمستخدمين في متناول أيديهم، ومع ذلك، فقد أدى ذلك إلى العديد من معاملات الاحتيال. يمكن أن يؤدي التعرف التلقائي على الاحتيال واكتشافه إلى تحسين تجربة المستخدم وأمان المعاملات عبر الإنترنت. باستخدام خوارزميات التعلم الآلي، من الممكن اكتشاف معاملات الاحتيال. تتمتع خوارزميات التعلم الآلي بالقدرة على العثور على النمط الضمني المخفي وعلاقة البيانات من مجموعة بيانات كبيرة. ومن ثم، فإن استخدام هذه الخوارزمية ممكن لاكتشاف ما ينتج من جميع المعاملات، مما يساعد في تحديد معاملة الاحتيال. في هذا البحث، يتم استخدام خوارزمية APRIORI و Support Vector Machine للكشف عن معاملات الاحتيال في بطاقات الائتمان على سبيل المثال من خلال بناء خدمة تطبيق ويب امنة توفر الحماية وفقا لمقاييس في هذا الصدد. تم مقارنة النتائج مع خوارزميات التعلم الآلي الأخرى الموجودة. لاحظنا أن دقة الكشف عن معاملات الاحتيال أعلى في الخوارزمية المقترحة بما يزيد على 94.56، واثبت البحث أن اكتشاف معاملات الاحتيال الخاطئ أقل من اكتشاف عمليات الاحتيال بناءً على نموذج ماركوف المخفي.

INTRODUCTION

Fraud detection is a data mining classification problem in which the aim is to classify the fake transactions from the legitimate transactions. As the data contains user sensitive and private information, many banks do not allow to fetch or provide data to researchers who is working on the fraud data classification problem. Credit card fraud transaction is either be compromised because of lost credit card or because of hacked or stolen credit card confidential credentials which is required while making credit card transactions.

With improved banking management and security users are able to immediately inform banks about credit card lost hence many credit card frauds are observed with stolen credit card confidential data. Hence this comprised credit card data can be immediately informed to bank and hence there is no way to determine whether the card is comprised until fraud transaction are detected. This sometimes can cause a serious issue as the card holder will not come to know the transactions until the credit card bill gets generated [1]. Credit card fraud is generally cauterized into two different types which are based on either the fraud transaction made is online transaction or offline transaction. Fraud transaction made online are mostly carried out using internet services are they are termed as Fraud with Online transaction processing or OLTP [2]. This OLTP frauds are carried out using different terminals like financial services, retail or customer relationship management portals.

Transaction made at point of services (POS) are also carried out using the online transactions at shops, and hence involved in the online transaction processing payment modes [3]. As POS and payment gateway minimize the billing time it is accepted as the mostly used payment mode. Therefore, the risk of fraud in credit card transaction is increasing with advancement in the internet technology and payment gateways. With increase in credit card transactions and terminals where credit card transactions are allowed, increases the difficulty to identify the credit card fraud. Hence it is important to automate the fraud detection using algorithm or computer processing which can detect the fraud transactions.

With enhancement in the artificial intelligence and machine learning algorithms, its application is being used in many financial [4], medical [5] and data analysis purpose [6]. There are many ongoing researches on credit card fraud detection using machine learning [7, 8]. As the fraud transaction is an outlier from all the transaction it is possible to detect the fraud transaction using data analysis technique of machine learning algorithm.

The objective of the proposed system is to collects and pre-process the transaction from database which contains all legitimate and fraud transactions. After that using machine learning model creates pattern to classify the fraud and legitimate transactions. Finally using the classification model, classify the transactions in real time and detect the fraud transaction with high accuracy and less false positive rate.

The proposed research paper is organized in the following sections. Section I introduces the credit card usage and frauds associated with its usage. Section II briefs about related work and literature survey. Section III introduces the proposed system and machine learning algorithm to detect the fraud transaction. Section IV contains analysis of result of proposed research with various machine learning algorithm. Section V, conclusion explains conclusion the proposed research work and its performance evaluation.

RELATED WORKS

As the fraud transaction is an outlier from all the transaction it is many researches detect the fraud transaction using data analysis technique of machine learning algorithm. Malini (2017) [9] describes the issue of credit card fraudulent transactions. In this research a comparison is carried out using various machine learning algorithm, genetic algorithms and fuzzy system with respect to detection of fraud transaction in credit card. Based on the research and data analysis KNN and outlier detection algorithm was applied on the collected data to enhance the credit card fraud transaction detection. Result analysis shows the improvement in the fraud transaction detection rate.

Lepovire (2016) [10] developed a fraud detection system by applying unsupervised learning algorithm on the collected dataset. In this research work packages are generated using classification algorithm and then these packages are grouped together using clustering algorithm. Finally using K-means clustering algorithm transaction are classified into fraud or legitimate category. Result analysis shows high precision rate for detection of fraud transaction.

Sumanet (2013) [11] proposed a new method to determine the fraud transaction in the credit card and telecommunication process. Neural network is applied on the collected bank transaction to detect the fraud. Artificial neural network consisting interconnected neurons which provides feed forward and feed backward functionality. This helps in clustering and classification of collected data. The result analysis shows that this research able to detect the fraud transaction with improved detection time. As neural network supports the non-linear data modelling, it improves the detection accuracy. It also be used to build model with complex relationship between output and inputs to get the pattern of the data.

Credit card fraud detection performance is improved when utilizing the Hidden Markov model. To enhance fraud detection, Bhusari (2011) [12] suggested doing research using a Markov model. This study consistently and with a low probability of false positives finds fraud transactions. All states in a hidden markov model, which is a finite set, are accompanied by a probability distribution. An observation is produced that pinpoints the fraudulent transactions based on the probability distribution of each finite collection. The outcome demonstrates increased fraud detection using the Hidden Markov model.

In addition, performance of the credit card fraud detection is improved when Bayesian networks are used for credit card detection. Benson (2011) [13] presented a study that would identify user habits using a Bayesian network. In this study, which employs two Bayesian networks, it is assumed that one user is a fraud and the other is a genuine user in two different scenarios. The fraud net and user net in this study are built utilizing expert knowledge and legitimate user data, respectively. User net is modified in accordance with real-time data. It is possible to determine the probability of the measurement for the two assumptions by introducing evidence and sending it to the network. The determination of whether an action is fraudulent or legitimate is made using the probability distribution.

MATERIALS AND METHODS

The suggested system includes two critical phases: frequent item set mining and a matching algorithm. The system's goal is to block fraudulent transactions with a low false positive rate while allowing legal transactions. For this purpose, a legitimate and fraud transaction group is produced using frequent item set mining, and the transaction is defined as a legit or a fraud transaction using a matching algorithm. The Apriori method is employed in the system for frequent item set mining, while the support vector machine approach is used for matching.

APRIORI Algorithm

In this research for frequent item set generation, Apriori classification algorithm is used. It is used for refining the dataset items. The aim of Apriori algorithm is to group the item based on the matching features to generate set of items which are similar, this sets are called as frequent item set. Apriori algorithm is an association rule mining algorithm which uses bottom-up approach to mine the frequent item set.

Apriori algorithm is designed to analyze the database which contains transactions. In this algorithm a threshold value ε is provided and then the frequent item set are generated which are subsets of at least ε transactions in from whole transactions. Using the bottom-up approach frequent subset is extended by adding at a time one item [14]. This step is called as candidate generation step. The process keeps iterating until the termination candidate occur which is no possibility to extend the item set further.

In Apriori algorithm, frequent item set is generated using support and confidence value. Support represents the transactions with items are together in a single transaction. Confidence represents transactions where the items are similar. A frequent item set contains the items with higher support and confidence value than threshold value. Support and Confidence for two item set P and Q can be calculated as:

$$Support(P) = \frac{TotaltransactionwithPitems}{TotalTransaction} \quad (1)$$

$$Confidence(P \rightarrow Q) = \frac{Support(PUQ)}{Support(P)} \quad (2)$$

Apriori algorithm

Step 1: Finding all frequent item set from database.

Step 2: Create association rule from frequent itemset generated in step 1.

- Pseudo Code Sets:

  C: candidate set of size n, F: Frequent item set of size n

- Join Step:

  Ck is produced by joining Fk-1 with itself.

- Prune Step:

  Itemset which is not frequent cannot be subset Frequent itemset.

- Pseudo Code:

```
F₁ = frequent item set
For n=1 and frequent item set is not null, increment k
        Cₖ = Candidate set generated from F₁
        Foreach Transaction t in database
                Increment the count of all candidate belong to Cₙ+1 and in t
                Fₙ+1 = Candidate with support greater than threshold and belong to Cₙ+1
        End
Return Fₙ
```

Support Vector Machine

Speculate about Support vector machine is a supervised learning algorithm which is associated with regression and classification. Hyperplane of the support vector machine categories different classes based on the similarity and differences. In the proposed research support vector machine hyperplane divides the credit card transaction from legitimate transaction and fraud transaction. Hyperplane classifying two classes with higher margin is considered as with higher accuracy. SVM have hyperplane to categorize given data into two or more classes [15, 16, 17]. Maximal width of hyperplane classifies the credit card transaction data [16, 18].

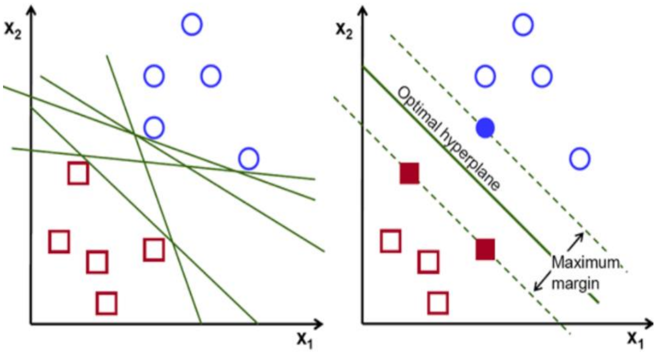

Figure 1: Support Vector Machine

Figure 1 shows the working principle of support vector machine. Optimal hyperplane is a line which divides the data. Other two lines which are present in figure are used to find decision boundary or optimal hyperplane. Margin is a distance between two hyperplanes. Margin is selected in such a way that all data points must be separate from boundary, this point are noted as support vectors. Using the trained model new credit card transaction are validated for fraud.

Proposed System Workflow

The proposed system contains two important phase which are frequent item set mining and matching algorithm. For this purpose, using frequent item set mining legitimate and fraud transaction group is prepared and using matching algorithm user specific transaction history is matched and the transaction is defined as a legit or a fraud transaction. For frequent item set mining, Apriori algorithm is used in the system while for matching, support vector machine algorithm is used. Figure 2 shows the workflow of the proposed system.

As shown in Figure 2, set of transaction is collected from the dataset. In this row represents as transactions and columns are represent as attributes. After this using Apriori algorithm a frequent item set mining is performed to find out the frequent item set in the credit card transactions. Using the frequent item set mining, items are classified into legal pattern and fraud pattern and transaction count for each pattern is generated. Using the count, total user specific transaction can be analyzed. Based on the user’s previous transaction, two group are generated as fraud transaction pattern and legal transaction pattern.

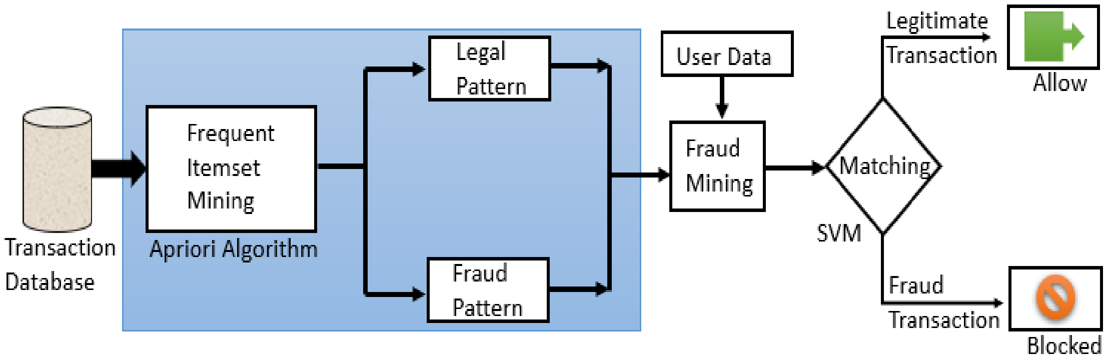

Figure 2: Workflow of the proposed system

It then groups user's transaction by analyzing all user's transaction using bank account number. When the user tries to perform new transaction, new transaction details are validated using the previously created fraud and legal group. In the matching process, a prediction based on the new transaction and machine learning model is made. Using this prediction, the transaction is either marked as a fraudulent or legal transaction. If the transaction is marked as a fraudulent, then it is blocked by the web-service and administrator and user is notified with transaction details, and in case of legitimate transaction, it is system allow to transact.

RESULTS AND DISCUSSIONS

Using UCI Machine Learning Repository an experimental analysis of proposed research is performed. From this repository, credit card transaction data is used. UCI Machine Learning Repository, contains an authenticated transaction data. In this data 10k transactions are present with 23 attributes. This transaction contains data like total credit amount of transaction, age and gender of the account holder, his education, marital status, with 6 attribute contains credit history of past 6 month, 6 attributes contain bill statements, other the repayment, defaulter history of that specific month. For performance evaluation of the proposed system, this data is used for training and testing of the proposed model.

For analysing the result of proposed algorithm with existing machine learning algorithm, F-score, precision and accuracy of different algorithm is compared. Table 1 shows the comparison of proposed algorithm with existing machine learning algorithm.

Table 1: Comparison of proposed algorithm with existing machine learning algorithms

Algo.	Dataset	Precision	Recall	F-measure	Accuracy
Random forest	10000	79.98%	88.23%	82.10%	84.32%
K-means Clustering	10000	73.78%	83.45%	77.25%	78.34%
Hidden Markov	10000	84.67%	93.11%	87.98%	89.43%
Proposed	10000	89.55%	96.57%	92.33%	94.56%

Table 1 shows the comparison of proposed algorithm, from this table, it is observed that the accuracy of the proposed algorithm is greater than the existing algorithm, also other measures like precision, recall and F-measure of the proposed algorithm shows enhancement. Total 10k transaction dataset was tested in this analysis, and it observed that the 9456 transactions were correctly classified. This shows that the proposed algorithm is capable to detect the fraud transactions in the real time.

Figure 3 shows the accuracy of the proposed system with other systems.

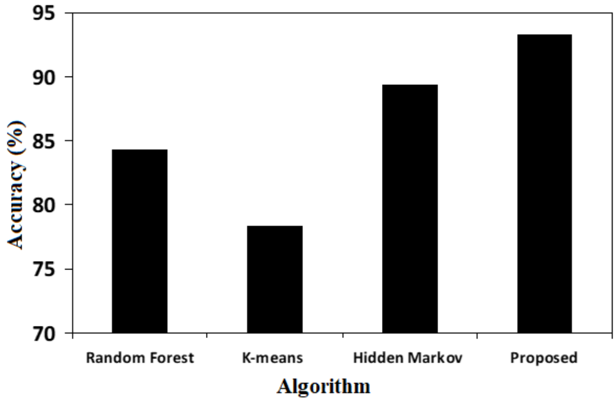

Figure 3: Accuracy Comparison

Figure 3 shows that the accuracy of the proposed algorithm is greater than random forest, K-means and hidden Markov model. However, in case of credit card truncation, accuracy is equally important as false positive rate of the algorithm. As if there are multiple false triggers for fake transaction, they administrator or credit card users will get annoyed and this may cause an ignorance towards the fraud alters.

For determining the ratio of false transaction, the proposed algorithm is test result is compared with different machine learning algorithm. Figure 4 shows the comparison of the proposed algorithm with random forest, K-means and hidden Markov model by considering the total detected fraud transactions and false fraud transactions.

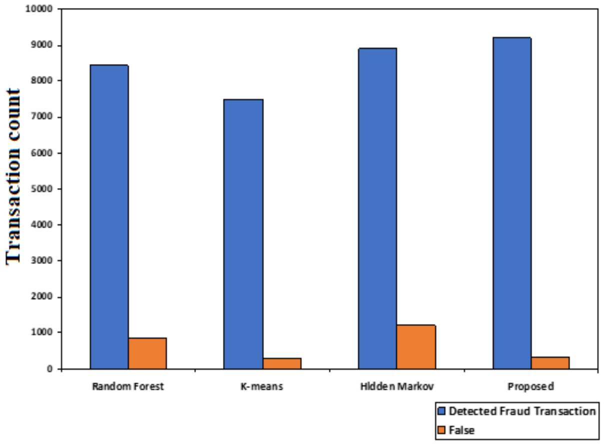

Figure 4: False Fraud Detection Comparison

Figure 4 shows that the detection of fraud in Random Forest and hidden Markov model is higher but it also triggers false fraud transaction with high frequency, which can reduce the overall performance and user experience of user, in case of K-means algorithm it is observed that the false fraud detection is less however it also misses the true fraud transaction, which will reduce the accuracy of the system. On the other hand, it is observed that the fraud detection accuracy of the proposed algorithm is high as well as the proposed algorithm produces less false trigger for fraud transactions.

Comparisons of Time Complexity

Along with accuracy and false trigger rate, time complexity is one of the important aspects when it comes to credit card fraud detection. As increased or more time consumption cloud provides a horrible user experience to user and they might prefer other option for payment.

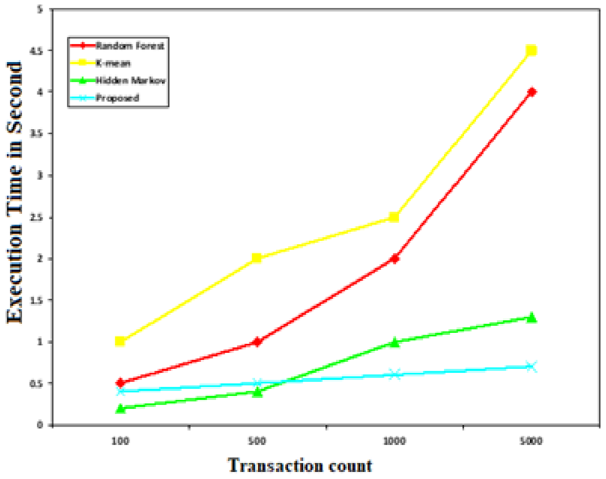

Figure 5: Comparison of Time complexity

Figure 5 illustrates the comparison of the proposed algorithm with random forest, K-means and hidden Markov model with variable transaction count. It is observed that the time complexity of the proposed algorithm is very much similar with compared algorithm for transaction count less than 100, but as the transaction count increases the time complexity of other algorithm reduces sharply, however in case of proposed algorithm the complexity is maintained and it increases in very tiny amount which shows that the time complexity of proposed algorithm is better for small or large transaction counts as compared to random forest, K-means and hidden Markov model. The result analysis from Figures 3, 4 and 5, it is observed that the proposed algorithm produces fraud transaction result with high accuracy, less false positive rate and enhanced time complexity as compared to existing systems.

CONCLUSIONS

The proposed system's objective is to gather transactions from a database that comprises both valid and fraudulent activities. A machine learning algorithm builds patterns to differentiate between fraudulent and legitimate transactions. Finally, using the classification model, categorize the transactions in real time and detect fraud transactions with high accuracy and a low false positive rate. The suggested system includes two critical phases: frequent item set mining and matching algorithm. Using frequent item set mining, a valid and fraudulent transaction group is created, and the transaction is defined as real or fraudulent using a matching algorithm. The Apriori method is employed in the system for frequent item set mining, while the support vector machine approach is used for matching.

For analyzing the result of proposed algorithm with existing machine learning algorithm, F-score, precision and accuracy of different algorithm is compared. In result analysis it is observed that the accuracy of the proposed algorithm is greater than the existing algorithm, also other measures like precision, recall and F-measure of the proposed algorithm shows enhancement. For determining the ratio of false transaction, the proposed algorithm is test result is compared with different machine learning algorithm. It is observed that the fraud detection accuracy of the proposed algorithm is high as well as the proposed algorithm produces less false trigger for fraud transactions. Along with accuracy and false trigger rate, time complexity is one of the important aspects when it comes to credit card fraud detection. It is observed that the time complexity of the proposed algorithm is very much similar with compared algorithm for transaction count less than 100, but as the transaction count increases the time complexity of other algorithm reduces sharply, however in case of proposed algorithm the complexity is maintained. This shows the enhancement in accuracy, false positive rate and time complexity with proposed algorithm. This provides features like dependency on the system for fraud transaction detection, less chances of getting false alert trigger and less fraud transaction detection time.

ACKNOWLEDGMENT

The author wishes to thank Mustansiriyah University (www.uomustansiriyah.edu.iq) and the Computer Science department for their support.

Disclosure and conflict of interest: The author declares no conflicts of interest.

REFERENCES

[1] K. Modi and R. Dayma, "Review on fraud detection methods in credit card transactions," 2017 International Conference on Intelligent Computing and Control (I2C2), 2017, pp. 1-5. https://doi.org/10.1109/I2C2.2017.8321781

[2] I. M. Mary, M. Priyadharsini, K. K and M. S. F, "Online Transaction Fraud Detection System," 2021 International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE),2021, pp. 14-16 https://doi.org/10.1109/ICACITE51222.2021.9404750

[3] B. K. Jha, G. G. Sivasankari and K. R. Venugopal, "Fraud Detection and Prevention by using Big Data Analytics," 2020 Fourth International Conference on Computing Methodologies and Communication (ICCMC), 2020, pp. 267-274, https://doi.org/10.1109/ICCMC48092.2020.ICCMC-00050

[4] R. A. Kamble, "Short and long term stock trend prediction using decision tree," 2017 International Conference on Intelligent Computing and Control Systems (ICICCS), 2017, pp. 1371-1375, https://doi.org/10.1109/ICCONS.2017.8250694

[5] F. Lv, "Data Preprocessing and Apriori Algorithm Improvement in Medical Data Mining," 2021 6th International Conference on Communication and Electronics Systems (ICCES), 2021, pp. 1205-1208, https://doi.org/10.1109/ICCES51350.2021.9489242

[6] H. Wu, Q. Liu and Z. Zhang, "Analysis of University Students Employment Recommendation System Based on Apriori Algorithm," 2020 Asia-Pacific Conference on Image Processing, Electronics and Computers (IPEC), 2020, pp. 262-265, https://doi.org/10.1109/IPEC49694.2020.9115188

[7] S. Khatri, A. Arora and A. P. Agrawal, "Supervised Machine Learning Algorithms for Credit Card Fraud Detection: A Comparison," 2020 10th International Conference on Cloud Computing, Data Science & Engineering (Confluence), 2020, pp. 680-683,

https://doi.org/10.1109/Confluence47617.2020.9057851

[8] M. R. Dileep, A. V. Navaneeth and M. Abhishek, "A Novel Approach for Credit Card Fraud Detection using Decision Tree and Random Forest Algorithms," 2021 Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV), 2021, pp. 1025-1028, https://doi.org/10.1109/ICICV50876.2021.9388431

[9] N Malini and M Pushpa, "Investigation of Credit Card Fraud Recognition Techniques based on KNN and HMM", IJCA Proceedings on International Conference on Communication, Computing and Information Technology ICCCMIT 2017(1):9-13, June 2018. https://doi.org/10.1109/AEEICB.2017.7972424

[10] Maria R. Lepoivre, Chloé O. Avanzini, Guillaume Bignon, Loïc Legendre, and Aristide K. Piwele, "Credit Card Fraud Detection with Unsupervised Algorithms", Vol. 7, No. 1, pp. 34-38, February, 2016. https://doi.org/10.12720/jait.7.1.34-38

[11] Suman, Nutan, "Review Paper on Credit Card Fraud Detection", International Journal of Computer Trends and Technology (IJCTT) - volume 4 Issue 7- July 2013.

[12] V. Bhusari and S. Patil, "Study of Hidden Markov Model in credit card fraudulent detection," 2016 World Conference on Futuristic Trends in Research and Innovation for Social Welfare (Startup Conclave), 2016, pp. 1-4, https://doi.org/10.1109/STARTUP.2016.7583942

[13] S. Benson Edwin Raj and A. Annie Portia, "Analysis on credit card fraud detection methods," 2011 International Conference on Computer, Communication and Electrical Technology (ICCCET), 2011, pp. 152-156, https://doi.org/10.1109/ICCCET.2011.5762457

[14] J. Du, X. Zhang, H. Zhang and L. Chen, "Research and improvement of Apriori algorithm," 2016 Sixth International Conference on Information Science and Technology (ICIST), 2016, pp. 117-121, https://doi.org/10.1109/ICIST.2016.7483396

[15] P. Naveen and B. Diwan, "Relative Analysis of ML Algorithm QDA, LR and SVM for Credit Card Fraud Detection Dataset," 2020 Fourth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), 2020, pp. 976-981, https://doi.org/10.1109/I-SMAC49090.2020.9243602

[16] Marwa M. Saadoon, Bashar M. Nema, "Security and Efficiency of Information Retrieval System: Survey Study", Gongcheng Kexue Yu Jishu/Advanced Engineering Science Journal, Volume 54, Issue 6, 2022.

[17] Bashar. M. Nema, " Automatic passkey generator using speech biometric features", AIP Conference Proceedings, Vol. 2290, Issue 1, Dec. 2020. https://doi.org/10.1063/5.0027417

[18] Rand Mohammed Rafee, Bashar M Nema. (2022) Secure E-Learning System Based on ZNP and AES. Al-Mustansiriyah Journal of Science 33:1, 39-44. Online publication date: 10-Mar-2022. https://doi.org/10.23851/mjs.v33i1.1016

How to Cite

S. J. Muhamed, “Detection and Prevention WEB-Service for Fraudulent E-Transaction using APRIORI and SVM”, *Al-Mustansiriyah Journal of Science*, vol. 33, no. 4, pp. 72–79, Dec. 2022.