Fraudulent Transaction Detection System Using Apriori and Support Vector Machine

Posted on April 9, 2026

Abstract:

The growing use of information technology has made various financial services easily accessible to the users; nonetheless, this has also resulted in an increase in fraudulent transactions. Both the security of online transactions and the user experience can be improved by automatically detecting and identifying fraud. A machine learning (ML) method can be used for identifying fraudulent transactions. From large datasets, ML algorithms can uncover data relationships and implicit hidden patterns. Therefore, there is a possibility for identifying the outlier from each transaction with the use of such a technique, which could aid in identifying fraudulent transactions. Support Vector Machine (SVM) and Apriori algorithm are employed in the suggested study for identifying fraudulent credit card transactions. The suggested system's output is contrasted with that of another ML method currently in use. It has been found that the suggested approach had a greater accuracy rate for detecting fraudulent transactions as well as a lower rate of false positives for fraudulent transactions compared to the Hidden Markov Model-based algorithm.

I. INTRODUCTION

The goal of fraud detection, a data mining (DM) classification problem, is to separate fraudulent transactions from authentic ones. Various banks refuse to allow academics working on the fraud data classification problem to provide or access data because it includes private and sensitive information about individual users. Credit card fraud occurs in the case where a credit card is lost or the case where the credit card's confidential credentials—which are needed to make credit card transactions—are compromised or stolen. Users can now promptly notify banks regarding missing credit cards thanks to enhanced security and banking management; as a result, different credit card fraud cases involving stolen credit card information have been reported. Therefore, banks could be notified of the compromised credit card data right away, and till fraudulent transactions are discovered, it is impossible to tell whether the card is compromised. Because the cardholder won't be aware of the transactions till the credit card bill is generated, this could occasionally result in major problems [1]. The two primary categories related to credit card fraud are determined by whether the fraudulent transaction was conducted offline or online. Online fraudulent transactions are known as "fraud with online transaction processing," or OLTP, since they are typically completed through using internet services [2]. Several terminals, such as retail, financial services, or customer relation management portals, are used in the execution of these OLTP frauds. Also, Point of Service (POS)

transactions are completed online at retail establishments, making them a part of online payment processing methods [3]. Known as the most popular method of payment, payment gateways and POS reduce billing time. Thus, as payment gateways and internet technologies progress, so does the potential of fraud in credit card transactions. The complexity of identifying credit card fraud rises with the number of credit card transactions as well as the number of terminals that accept credit card transactions. Therefore, it's critical to automate the process of detecting fraud by employing computer processing or algorithms that could identify fraudulent transactions. ML and AI algorithms have advanced to the point that numerous medical [4], financial [5], and data analysis applications are using them [6]. ML-based credit card fraud detection is the subject of numerous ongoing studies [7, 8]. Since a fraudulent transaction represents an outlier when compared to all other transactions, it can be identified through the use of ML algorithms for data analysis. The suggested system's goal is to gather and pre-process transaction data from a database that includes both fraudulent and legitimate transactions. Subsequently, an ML model generates a pattern for classifying legitimate and fraudulent transactions. Lastly, transactions are classified in real time with the use of a classification model, and with a high level of accuracy and a low false positive rate, fraudulent transactions are identified. The next sections make up the structure of the suggested study. The use of credit cards and related frauds is discussed in Section I. An overview of relevant literature is provided in Section II. The suggested ML algorithm and system for identifying fraudulent transactions are introduced in Section III. Analysis of the suggested research's findings using different ML algorithms is presented in Section IV. In addition, Section V provides a summary of the suggested study as well as an assessment of its performance.

II. LITERATURE SURVEY

Many studies use data analysis techniques and ML algorithms for identifying fraudulent transactions since they are outliers compared to other transactions. Malini addresses the problem of fraudulent credit card transactions in her study [2017] [9]. In this study, different genetic algorithms, ML algorithms, and fuzzy systems are compared in terms of their ability to identify fraudulent credit card transactions. To enhance credit card fraudulent transaction detection, the outlier detection algorithm and KNN were used on the gathered data, depending on data analysis and research. The study of the results demonstrates the rise in fraudulent transaction detection rate.

Lepovire [2016] used an unsupervised learning algorithm on the collected dataset for creating a fraud detection system [10]. Work packages for this study are created with the use of a classification algorithm, and after that a clustering algorithm is used to combine such packages together. Ultimately, transactions are classified as either legitimate or fraudulent with the use of the K-means clustering algorithm. The study of the results reveals a high rate of precision for the identification of fraudulent transactions.

Sumannet [2013] suggested a new approach for identifying fraudulent transactions involving telecommunication processes and credit cards. To find fraud, a neural network (NN) is applied to the

bank transactions that have been collected. An artificial neural network (ANN) with feed-forward and feed-backward capabilities is made up of interconnected neurons. This aids in the clustering and classification of the collected data. The study of the research's results demonstrates that it can identify fraudulent transactions faster. NNs enhance detection accuracy by supporting non-linear data modeling. In order to determine the data pattern, it can also be utilized to construct models with intricate relations between inputs and outputs [11].

Utilizing the Hidden Markov model for the detection of credit card fraud further improves fraud detection effectiveness. Bhusari (2011) suggested a study that would enhance fraud detection by utilizing a Markov model. This study has a high frequency of fraudulent transaction detection and a low false positive rate. Every state in the hidden Markov model has a probability distribution attached to it, and the model is a finite set. An observation is made, depending on the probability distribution of every finite set, identifying fraudulent transactions. The research findings indicate that the use of the Hidden Markov model improves fraud detection [12].

Bayesian network-based credit card detection improves credit card fraud detection effectiveness as well. Benson suggested a study that would employ a Bayesian network for identifying user behavior [13]. In this study, which makes use of two Bayesian networks, two assumptions are made: one regarding a legitimate user and the other regarding a fraudulent user. Non-fraudulent user data, as well as expert knowledge, are used to form a fraud net and a user net in such instances. The user network is adjusted depending on real-time data. One can specify the measurement probability for both assumptions by introducing evidence and sending it to this network. Determining whether a behavior is legitimate or fraudulent is done with the use of the distribution of probabilities.

III. PROPOSED SYSTEM

The frequent item set mining and matching algorithm are the two key components of the suggested system. The system's goal is to permit legitimate transactions while blocking fraudulent ones with a lower false positive rate. To do this, a legitimate and fraudulent transaction group has been created via frequent item set mining, while each transaction is classified as either fraudulent or legitimate based on a matching algorithm that matches the user's transaction history. The system uses the Apriori algorithm for frequent item set mining and the SVM method for the process of matching.

A priori Algorithm:

The present study uses the Apriori classification algorithm for generating frequent item sets. It is applied to the dataset items to refine them. The Apriori algorithm's goal is to create sets of similar items by grouping them together depending on matching attributes; these sets are referred to as frequent item sets. Also, the Apriori algorithm can be defined as a bottom-up approach association rule mining method that mines the frequent itemsets. The purpose of the Apriori algorithm is to analyze transactional databases. The Apriori algorithm takes a value of threshold ε as input and

creates a frequent itemset, which are subsets of all transactions that contain at least ε. A common subset is expanded through adding one item at a time with the use of the bottom-up technique [14]. We refer to this phase as the “candidate generation step.” The procedure continues iterating till a termination candidate arises, at which point there is no more room for extending the itemset. The Apriori algorithm uses confidence and support values to produce a frequent itemset. Support can be defined as a representation of transactions that include multiple products in a single transaction. When items are similar, transactions are represented by confidence. Items having greater confidence and support values compared to the threshold are found in a frequent item set. For two itemsets Q and P, confidence and support could be computed as follows:

There are two steps in the Apriori algorithm:

Step 1: Find all of the frequent itemsets from the database.

Step 2: Create association rules from the frequent itemsets generated in Step 1.

- Pseudo Code Sets:

F: represents a frequent itemset of size n

C: represents a candidate set of size n

- Join Step:

Ck is produced from the joining of Fk-1 with itself.

- Prune Step:

Itemsets that are not frequent cannot be a subset of the Frequent itemset.

- Pseudo Code:

F_1 represents a frequent itemset

For n=1 and a frequent itemset isn’t null, increment k

C_k represents the Candidate set generated from F_1

For each Transaction t in the database

Increment count of all candidates that belong to C_n+1 and in t

F_n+1 = Candidate with support greater than threshold and belonging to C_n+1

End

Return F_n

SVMs:

SVMs are supervised learning algorithms related to regression and classification. The SVM's hyperplane classifies several classes according to their differences and similarities. A hyperplane used in the suggested study distinguishes between credit card transactions that are fraudulent and those that are legitimate. A hyperplane that correctly classifies two classes with a larger margin is thought to be more accurate. For dividing a given set of data into at least two classes, SVMs have hyperplanes [15]. Credit card transaction data is classified based on the hyperplane's maximal width.

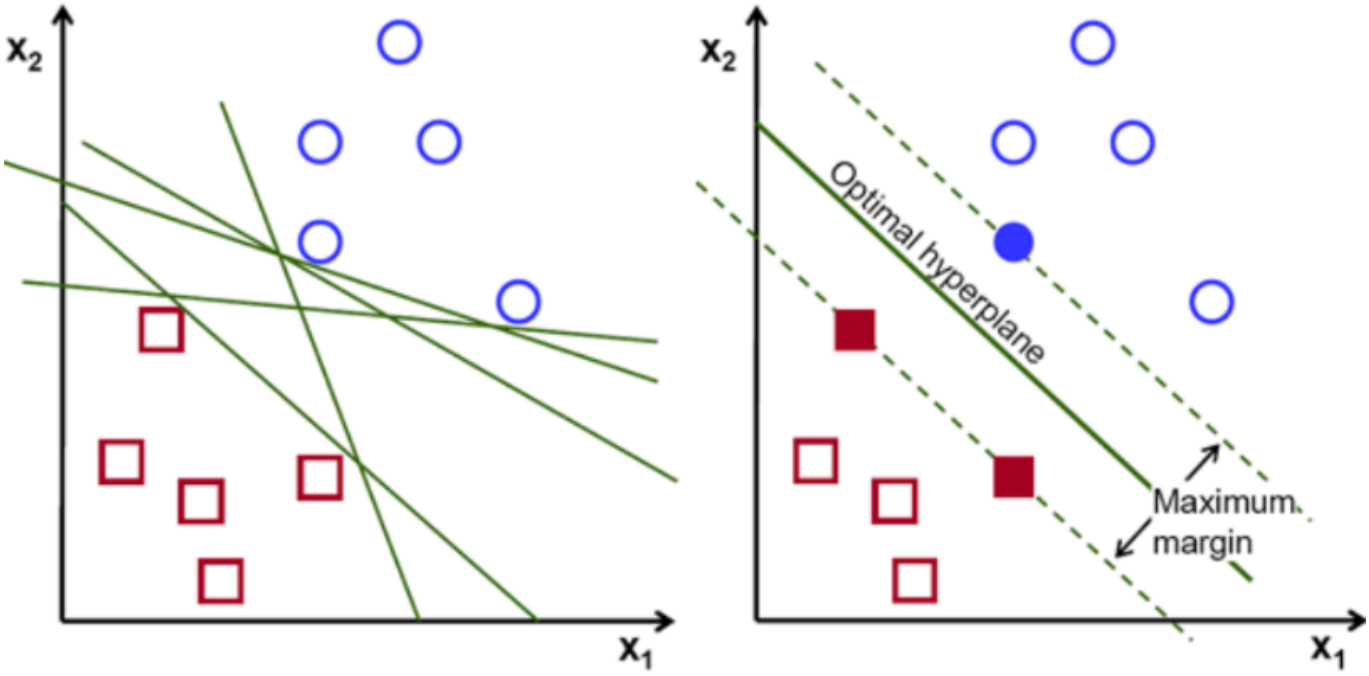

Figure 1: Support Vector Machines

The SVM's working principle is depicted in Figure 1. A line that splits the data is known as the optimal hyperplane. The optimal hyperplane or decision boundary is found using the other two lines shown in the figure. A margin is the separation between two hyperplanes. All data points should be distinct from the boundary when choosing a margin; these points are referred to as support vectors. New credit card transactions are checked for fraud with the use of the trained model.

Proposed System:

The matching algorithm and frequent item set mining are the two key components of the suggested system. To do this, a legitimate and fraudulent transaction group is created via frequent itemset mining, while the transaction is classified as either legitimate or fraudulent based on the matching algorithm that matches the user's transaction history. The system uses the Apriori algorithm for frequent itemset mining and the SVM method for matching. The suggested system's process is

depicted in Figure 2.

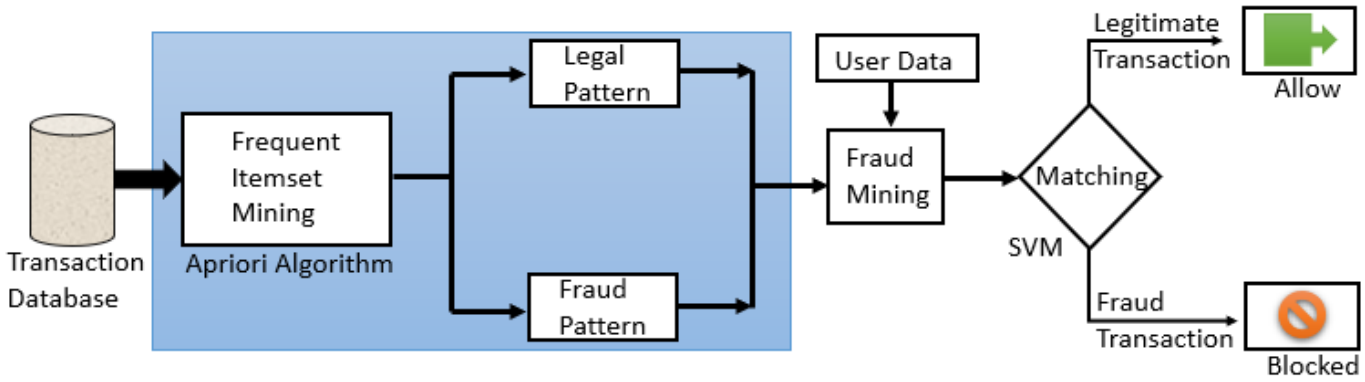

Figure 2: Workflow of the suggested system

A set of transactions has been gathered from the dataset, as seen in Figure 2. Transactions are represented by a row in this example, and attributes are represented by columns. After that, a frequent itemset mining procedure is carried out with the use of the Apriori algorithm for identifying frequent itemsets in credit card transactions. Items are classified into legitimate and fraudulent patterns with the use of frequent item set mining, and a transaction count is produced for each one of the patterns. Total user-specific transactions could be analyzed using count. Two groups—legitimate transaction pattern and fraudulent transaction pattern—are produced depending on the user's prior transactions. After that, it groups users' transactions by employing bank account numbers to analyze each user's transactions. The previously established legitimate and fraudulent group is used to validate new transaction data whenever the user attempts to complete a new transaction. A prediction is made during matching using the new transaction as well as the ML model. This prediction is used to classify the transaction as fraudulent or legitimate. In the event that a transaction is flagged as fraudulent, the system blocks it, notifies the user and administrator, and allows the transaction to proceed in the case of a legitimate transaction.

IV. RESULT ANALYSIS

The suggested work is subjected to an experimental investigation with the use of the UCI ML Repository. Transaction data related to credit cards is taken from this repository. Transaction data that has been verified is contained in the UCI ML Repository. There are 10,000 transactions with 23 attributes in the data. The transaction includes information about the total credit amount, the account holder's gender, age, marital status, and education. It also includes six attributes that include the account holder's credit history for the previous six months, other repayment information, bill statements, and default history for that particular month. This data is utilized for testing and training the suggested model in order to assess the system's performance. The accuracy, F-score, and precision of several algorithms are compared in order to analyze the outcome of the suggested method against an existing ML algorithm. The comparison between the suggested algorithm and the

current ML algorithm is displayed in Table 1.

Algorithm	Data-set	Precision	Recall	F-measure	Accuracy
Random forest classifier	10,000	79.980%	88.230%	82.100%	84.320%
K-means Clustering	10,000	73.780%	83.450%	77.250%	78.340%
Hidden Markov model	10,000	84.670%	93.110%	87.980%	89.430%
Proposed Algorithm	10,000	89.550%	96.570%	92.330%	94.560%

Table 1: Comparison of the suggested algorithm against the existing ML algorithm

Table 1 presents a comparison of the suggested method with the present algorithm. It can be noticed that the accuracy of the suggested algorithm is higher compared to that of the existing algorithm, with improvement in other metrics such as recall, precision, and F-measure. This investigation analysed a total of 10,000 transactions, finding that 9,456 of them were successfully identified. This demonstrates that the proposed algorithm can identify fraudulent transactions in real time.

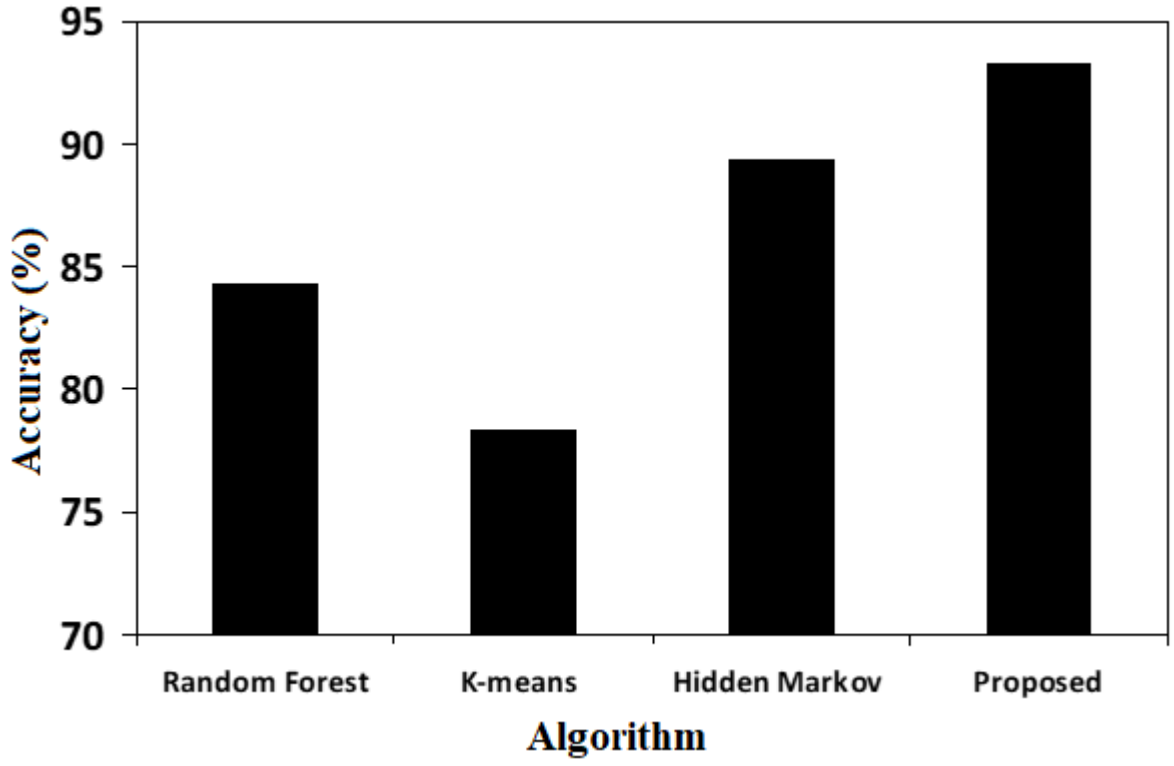

Chart 1: Accuracy Comparison

The suggested approach outperforms K-means, random forest (RF), and the hidden Markov model in terms of accuracy, as demonstrated by Chart 1. When it comes to credit card transactions, accuracy is just as crucial as the algorithm's false positive rate. The credit card users or administrators will become frustrated if there are several false triggers for fraudulent transactions, which could lead to the real fraud going unnoticed. The test results of various ML algorithms are compared with the suggested approach in order to ascertain the ratio of false transactions. Chart 2 compares the suggested algorithm with RF, hidden Markov model, and K-means, taking into account the total number of fraudulent transactions discovered and the number of false fraud detections.

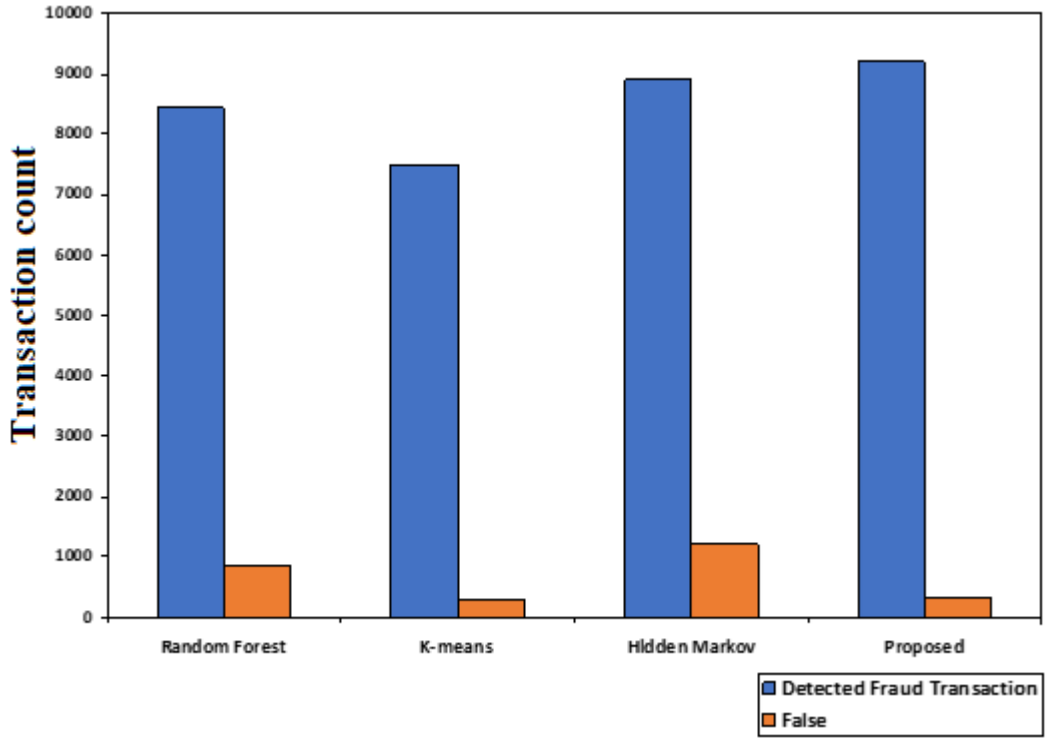

Chart 2: Comparison of False Fraud Detection

Chart 2 demonstrates that while fraud detection in the Hidden Markov model and RF is higher, it frequently results in false fraudulent transaction flags, which could negatively impact user experience and system performance. In contrast, the K-means algorithm is observed to have a lower false fraud detection rate, yet it fails to identify true fraudulent transactions, which decreases system accuracy. However, it is noted that the suggested method has high accuracy in detecting fraudulent transactions and generates fewer false triggers.

Comparisons of Time Complexity:

Regarding credit card fraud detection, time complexity is just as crucial as false trigger rate and accuracy. Because longer processing times can frustrate users, they may find alternative payment methods, resulting in a worse user experience. The suggested technique is compared with RF, hidden Markov model, and K-means with variable transaction counts in Chart 3.

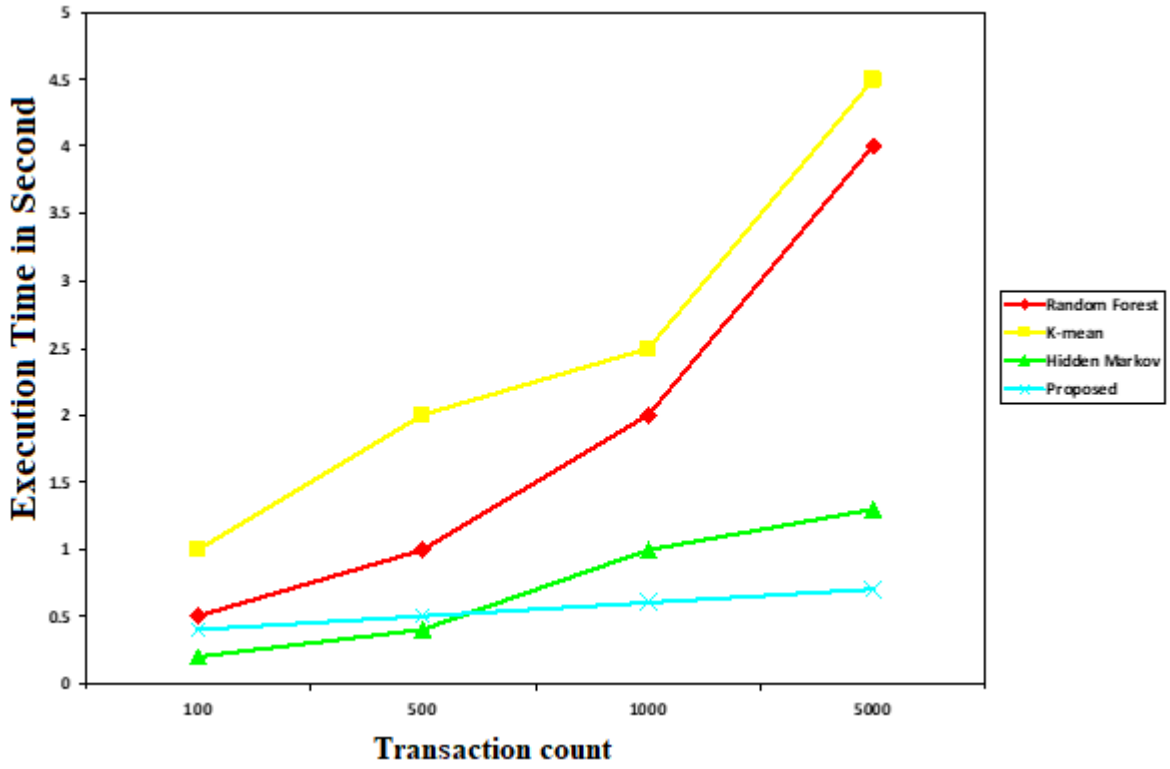

Chart 3: Comparison of Time Complexity

The suggested technique is compared with K-means, RF, and the hidden Markov model with variable transaction counts in Chart 3. When compared with RF, hidden Markov, and K-means models, the time complexity of the suggested algorithm is extremely comparable for transaction counts under 100. Yet, when transaction counts rise, the time complexity of other algorithms decreases sharply, while in the suggested algorithm, complexity is maintained and increases only slightly. This suggests that the suggested algorithm's time complexity is superior for both large and small transaction counts. According to the results of Charts 1, 2, and 3, the suggested algorithm generates fraudulent transaction results with high accuracy, manageable time complexity, and a lower false positive rate compared to current systems.

V. CONCLUSION

The proposed system aims at gathering transactions from a database that includes both fraudulent and legitimate transactions. Subsequently, an ML model generates a pattern for classifying legitimate and fraudulent transactions. Lastly, transactions are classified in real time with the use of a classification model, and with a high degree of accuracy and a low rate of false positives, fraudulent transactions are identified. The frequent item set mining and matching algorithm are the two key components of the suggested system. A fraudulent and legitimate transaction group is created by frequent itemset mining. A matching algorithm is then used to match user-specific transaction histories, classifying each transaction as either fraudulent or legitimate. The system uses the Apriori algorithm for frequent item set mining and SVM for matching. The accuracy, F-score, and precision of

several algorithms are compared in order to analyze the outcome of the suggested method against an existing ML algorithm. According to the results of the investigation, the suggested method outperforms the current approach in terms of accuracy, recall, precision, and F-measure. It has been noted that the suggested algorithm has a high accuracy rate for detecting fraud and generates fewer false positives for fraudulent transactions. With regard to credit card fraud detection, time complexity is just as crucial as false trigger rate and accuracy. When compared with other algorithms, the time complexity of the suggested algorithm is found to be extremely comparable for transaction counts under 100. Yet, when transaction counts rise, the time complexity of other algorithms decreases noticeably, while the complexity of the suggested algorithm remains constant. This demonstrates how the suggested approach improves the false positive rate, accuracy, and time complexity, providing faster and more reliable fraudulent transaction detection.

REFERENCES

[1] Review on fraud detection methods in credit card transactions, Krishna Modi; Reshma Dayma, DOI: 10.1109/I2C2.2017.8321781, International Conference on Intelligent Computing and Control (I2C2), Coimbatore, India, 2017

[2] Online Transaction Fraud Detection System, I. Mettildha Mary; M. Priyadharsini; Karuppasamy. K; Margret Sharmila. F, DOI: 10.1109/ICACITE51222.2021.9404750, International Conference on Advanced Computing and Innovative Technologies in Engineering (ICACITE), Greater Noida, India, 2021

[3] Fraud Detection and Prevention by using Big Data Analytics, Bineet Kumar Jha; G G Sivasankari; K R Venugopal, Fourth International Conference on Computing Methodologies and Communication (ICCMC), DOI: 10.1109/ICCMC48092.2020.ICCMC-00050, Erode, India, 2020

[4] Short and long term stock trend prediction using decision tree, Rupesh A. Kamble, International Conference on Intelligent Computing and Control Systems (ICICCS), Madurai, India, DOI: 10.1109/ICCONS.2017.8250694, 2017

[5] Data Pre-processing and Apriori Algorithm Improvement in Medical Data Mining, Feng Lv, 6th International Conference on Communication and Electronics Systems (ICCES), India, DOI: 10.1109/ICCES51350.2021.9489242, 2021

[6] Analysis of University Students Employment Recommendation System Based on Apriori Algorithm, Hao Wu; Qian Liu; Zhifang Zhang, Asia-Pacific Conference on Image Processing, Electronics and Computers (IPEC), China, DOI: 10.1109/IPEC49694.2020.9115188, 2020

[7] Supervised Machine Learning Algorithms for Credit Card Fraud Detection: A Comparison, Samidha Khatri; Aishwarya Arora; Arun Prakash Agrawal, 10th International Conference on Cloud Computing, Data Science & Engineering (Confluence), DOI: 10.1109/Confluence47617.2020.9057851, India, 2020

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

[9] Investigation of Credit Card Fraud Recognition Techniques based on KNN and HMM, N. Malini; M. Pushpa, IJCA Proceedings on International Conference on Communication, Computing and Information Technology, 2017

[10] Credit Card Fraud Detection with Unsupervised Algorithms, Maria R. Lepoivre; Chloé O. Avanzini; Guillaume Bignon; Loïc Legendre; and Aristide K. Piwele, Journal of Advances in Information Technology Vol. 7, No. 1, February 2016.

[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; S. Patil, "Study of Hidden Markov Model in Credit Card Fraudulent Detection", International Journal of Computer Applications (0975 – 8887) Volume 20– No.5, April 2011

[13] Analysis on credit card fraud detection methods, Benson Edwin Raj; A. Annie Portia, DOI:10.1109/ICCCET.2011.5762457, Computer, Communication and Electrical Technology (ICCCET), 2011 International Conference on, 2011

[14] Research and improvement of Apriori algorithm, Jiaoling Du; Xiangli Zhang; Hongmei Zhang; Lei Chen, Sixth International Conference on Information Science and Technology (ICIST), Dalian, China, DOI: 10.1109/ICIST.2016.7483396, 2016

[15] Relative Analysis of ML Algorithm QDA, LR and SVM for Credit Card Fraud Detection Dataset, P Naveen; B Diwan, Fourth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), 2020