Online Transaction Fraud Detection using Hidden Markov Model & Behavior Analysis

Niki Patel *patel104@cougars.csusm.edu*
Student / Computer Science and Information System
California State University San Marcos
San Marcos, 92096, USA

Yanyan Li *yali@csusm.edu*
Faculty / Computer Science and Information System
California State University San Marcos
San Marcos, 92096, USA

Ahmad Hadaegh *hadaegh@csusm.edu*
Faculty / Computer Science and Information System
California State University San Marcos
San Marcos, 92096, USA

Abstract

Card payment are mostly preferred by many for transactions instead of cash. Due to its convenience, it is the most accepted payment method for offline as well as online purchases, irrespective of region or country the purchase is made. Currently, cards are used for everyday activities, such as online shopping, bill pays, subscriptions, etc. Consequently, there are more chances of fraudulent transactions. Online transactions are the prime target as it does not require real card, only card details are enough and can be stored digitally. The current system detects the fraud transaction after the transaction is completed. Proposed system in this paper, uses Hidden Markov Model (HMM), which is one of the statistical stochastic models used to model randomly changing systems. Using Hidden Markov Model, a fraud transaction can be detected during the time of transaction itself and after 3 attempts of verification card can blocked at the same time. Behavior Analysis (BA) helps to understand the spending habits of cardholder. Hidden Markov Model helps to acquire high-level fraud analysis with a low false alarm ratio.

Keywords: Fraud Detection, Online Transactions, Hidden Markov Model, Behavior Analysis.

1. INTRODUCTION

Based on the Nielsen study for internet usage,100% of world's population uses internet and has seen a growth of 13,941% from year 2000 to 2020. 95% of global consumers have used online shopping method to make a purchase, according to Nielsen's Connected Commerce report. Globally the eCommerce rate has increased with a booming 600% since 2010 [1].

With the digital economy evolving rapidly, businesses of all sizes need to re-evaluate their position and tools when it comes to fraud management. Transaction fraud obliges genuine threats on e-commerce shopping. Due to increasing popularity of online transaction the types of online transaction frauds related with it are also escalating which disturbs the financial industry. Online payments are a prime target for fraudsters as they do not even need to have the real card, they only need the card details which can be stored digitally. It is also easier to get away with it because it's so much harder for the seller to verify who is really making the purchase.[2]

To overcome these problems numerous fraud detection techniques and algorithms have been proposed, data mining is used by many firms associated with fraud detection. But the data mining alone is not sufficient for detecting the fraud as it depends upon the data set containing history of customer's transaction. The specifics of items bought by any individual transaction are generally not recognized to any Fraud Detection System (FDS) operating at the bank that issues credit cards. BA (Behavior Analysis) is executed for tackling this problem. An FDS operates at a credit card issuing bank. Each inbound transaction is presented to the FDS for authentication. FDS obtains the card details and transaction cost to confirm if the transaction is legitimate or not. The types of commodities bought in that transaction are unknown to the FDS. Bank rejects the transaction if it is found to be a fraud by FDS.

1.1. Problem with Existing System

As reported by Federal Trade Commission (FTC), 1,697,934 fraud reports were made in 2019, out of which 250,678, which is 15% of total frauds, were the reports with Payment Methods [3].

Figure 1 is the representation of frauds reported in various methods of payments with the corresponding total loss. Figure 1 illustrates that Wire Transfer and Credit cards accounts most of the frauds.

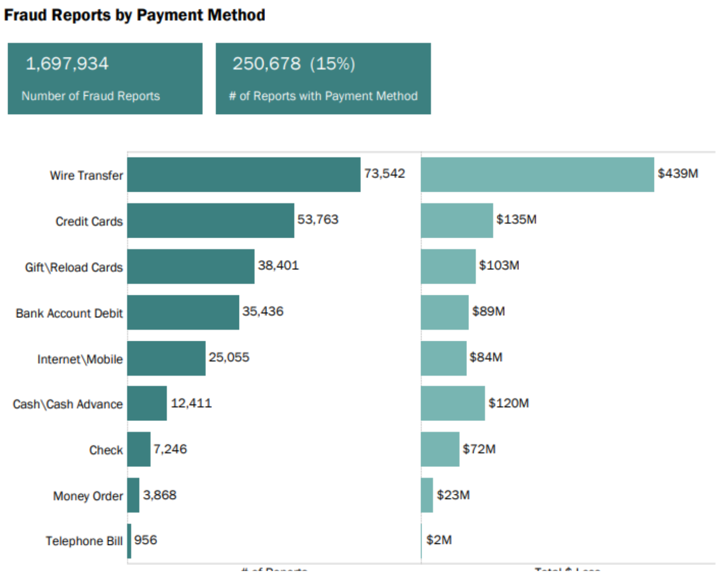
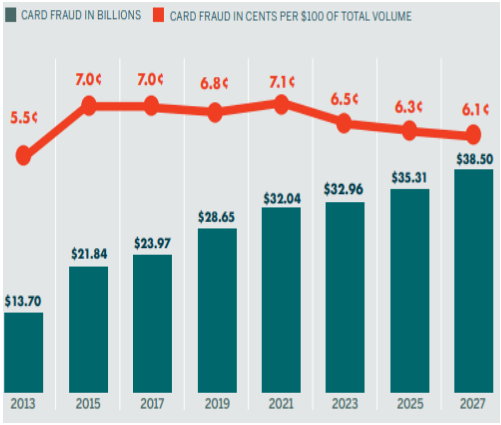
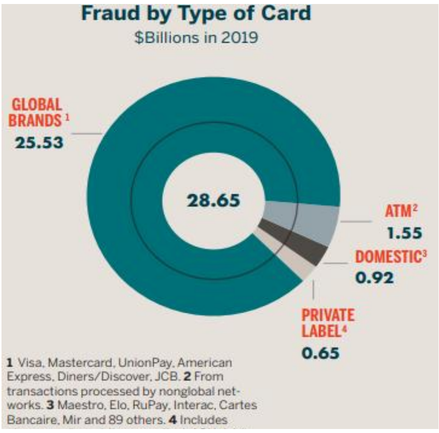


FIGURE 1: Wire Transfer and Credit cards accounts.





1 Visa, Mastercard, UnionPay, American Express, Diners/Discover, JCB. 2 From transactions processed by nonglobal networks. 3 Maestro, Elo, RuPay, Interac, Cartes Bancaire, Mir and 89 others. 4 Includes

FIGURE 2 **FIGURE 3**

According to the Nilson Report [4], a total of $28.65 billion losses was generated worldwide in 2019 through card-based payment systems. Based on the previous years they have also predicted the below graph till 2027. (Figure 2).

Global Brand cards (American, Diners club/Discover, JCB, Mastercard, Visa and UnionPay) together lost $25.53 billion in 2019 due to frauds [4]. This is 2.7% more than in 2018. This totals to 89.11% gross fraud losses worldwide in 2019 as shown in Figure 3.

Most payment related frauds take place in eCommerce. Identity theft is the most typical types of frauds in eCommerce.[5]

Identity Theft is when an attacker or an imposter offense a user account, makes alteration to the personal data, and then tries to spend the money in buying products using the fake information. This type of frauds can be detected using Behavior Analysis (BA). BA can analyze any suspicious activity and can find any inconsistencies in the past personal data.

To give the customer the best experience in online payments and an increase in the mobile payments, banks have reduced the amount of verification procedures. Current Conventional (Rule-based) fraud Detection methods are failing to differentiate among fault or an odd transaction from real fraud. This method does not process the real time data, which in turn takes a huge amount of time to detect any fraud. The performance also degrades when system has to process a huge amount of data. Conventional (Rule-based) method may be challenging to cope with as the algorithms and rules are written manually and to configure them in a proper manner involves accurate, arduous, and labor-intensive programming for every thinkable fraud prospect. Multiple fraud methods are invented and to cope with those new methods, these rules have to be constantly adapted to the existing, evolving, and future fraud methods.

In current system a Fraud Detection is carried out after the fraud is already performed. Existing system observes a series of transactions made and then tries to classify them as a legitimate or fraudulent transaction.

2. RERLATED WORKS

Increasing frauds in eCommerce have led many researchers to perform a research in this area. Many different techniques and methods are invented and talked about which uses Machine Learning, Neural networks, and data mining.

Andrea Pozzolo and Olivier Caelen [6] in their recent study have mentioned about two methods of detecting fraud. One in Static method, which trains the detection model during every fixed interval of time (once in a month or year). The other method is Online method, which updates the model as soon as a new data of transaction enters. As the conduct of fraud changes every now and then, they have indicated that the Online method is way better than the Static method. Andrea Pozzolo and Olivier Caelen [7] in their work suggested that Random Forest is by far the best measure for detecting frauds.

Kuldeep Randhawa, et al. [8] in their comparative study of machine learning for fraud detection have used standard models in their initial phase and later on made use of hybrid models which created with the help of AdaBoost and voting methods. Based on the experiments they concluded that the voting method had most stable performance even with the presence of noise in them.

John Awoyemi et al. [9] operated an analysis using K-nearest neighbor, Naïve Bayes and logistic regression. Their work was done in Python. These were applied to transaction data and then was additionally resampled with the help of Synthetic Minority Over-Sampling Technique. Their results showed that K-nearest performed the best among other techniques which were calculated based on sensitivity, precision, accuracy, correlation, and specificity.

You Dao et al. [10] in their work “Online Credit Card Fraud Detection: A hybrid Framework with Big Data Technologies” designed a fraud detection framework with the help of big data. They were able to accomplish three most important objectives: capacity to merge several detection models for an enhanced accuracy, capability to process huge amount of data, and detecting fraud in real time.

Abhimanyu Roy et al. [11] made use of deep learning for the fraud detection in online transactions. They used artificial neural network approach where they were able to pre-label almost 80 million transactions as fraudulent and legitimate. High performance cloud computing was also used by them for this approach.

Shiyang Xuan et al. [12] took help of random forest for detecting frauds. For their experiments they used two types of random forests namely Random-tree-based random forest and CART-based, to train the normal and fraud behavior features. Data of an e-commerce company from China was taken into consideration for efficacy of two methods. As a result, random forest was able to get good results for small dataset, it faced problems with the imbalanced data.

W.N. Robinson and A.Aria [13] in their store-centric approach detected sequential anomalies with the help of hidden markov model. They introduced store-model divergence method (SMDM) that was able to work great with various parameters. SMDM pertains sequence analysis to specified records of sequential typed transactions which will then uncover anomalies as probable fraud. SMDM proved to be beneficial where all the transactions are profoundly aggregated and had several, typed and continuous transactions.

2.1. Difference between ML Fraud Detection and Conventional (Rule-Based) Fraud Detection

Recently Machine Learning (ML) Fraud Detection has come into limelight and due to its more accurate results, the fraud detection industry is moving from Conventional (Rule-based) Fraud Detection to ML Fraud detection.

Here are some differences between ML Fraud Detection and Conventional Fraud Detection [14]:
<u>Conventional (Rule-based) Fraud Detection:</u>

- Conventional (Rule-based) Fraud Detection_systems comprise of algorithms that are written and set manually by analysts that detect fraud. Due to these manual algorithms, Conventional method is straightforward. This also requires manually adjusting/adding scenarios which can scarcely identify implied correlations.
- Conventional Fraud Detection make use of legacy software which can barely handle real-time data. Even though this scheme is still in use [15], it is an old method, technology, computer system, or application program, “of, relating to, or being a previous or outdated computer system”.
- Further, since they are not capable of managing real-time data, they take immense amount of time to detect any fraud. This can lead to delays in detecting frauds.
- Manual algorithms are set in Conventional method which makes them to detect only the most obvious fraud pursuits.
- Conventional Fraud detection uses several verification procedures, and this may be troublesome for the user.

<u>Machine Learning (ML) based Fraud Detection:</u>

- Machine Learning Fraud Detection can create algorithms which are able to process a huge number of datasets.
- This method does not require manual work for detecting possible frauds, and so they work automatically and detects potential fraud situations. Since they are automatic and not manual, they can find hidden and implied correlations.
- ML based Fraud Detection are efficient for processing real-time data contrasting Conventional based method.

- Machine Learning algorithms are smart enough to get along with the Behavior Analysis (BA) , which in turn aids in cutting the all over number of verification procedures.
- The time required for verification procedures is also less compared to the Conventional Fraud Detection.

2.2. Proposed System

2.2.1. Intro to HMM

The Hidden Markov Model is based on enhancing the Markov Chain. Markov Chain is basically a model which depicts about the probabilities of sequences of *states*, random variables. Each state takes up some values from some sets. Sets can comprise of tags, words, or symbols such as weather. In a Markov chain, it makes a powerful hypothesis that to forecast the future, all that is required is current state. This means the state that is before the current state has no influence on the future state, and the only state that matters is the current state. Figure 4 explains the Markov chain to predict tomorrow's weather. For that we can check on today's weather, but we are not able to check yesterday's weather [16].

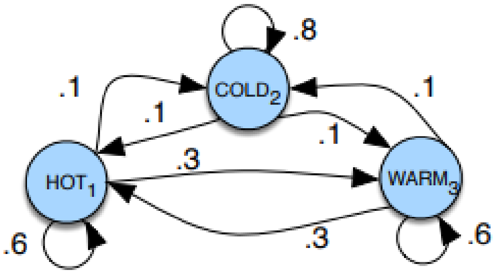


FIGURE 4

In Figure 4, the states are represented as nodes, and transactions with corresponding probabilities as edges. The represented values are probabilities whose sum must be equal to 1. As shown in figure 4, if we start from COLD (2) state, there is a probability of 0.8 that it will be COLD again tomorrow and 0.1 probability to become either HOT or WARM. Same thing can be set for the HOT and WARM states.

Formulation of Markov assumption for a sequence of state variables $q_1, q_2, \ldots\ldots q_i$, where the current state matters but not the past is as follows:
P $(q_i = a|q_1 \ldots.. q_{i-1}) = P(q_i = a|q_{i-1})$

A Markov chain can formally consist of following components.

- S = $s_1, s_2, \ldots\ldots s_N,$ are a set of N states
- A= $a_{11}, a_{12}, \ldots\ldots a_{n1} \ldots\ldots a_{nm}$, are Transaction probability matrix where each a_{ij} states the probability of transacting from state i to j.
- π = $\pi_1, \pi_2, \ldots\ldots, \pi_N$, are initial probability distribution over states.

When we need to calculate probability for a sequence that is observable events, Markov chain is useful. But in some cases, for which we want to calculate probabilities, they are hidden and are not observable. For example, we are not able to observe a transaction taking place. Therefore, for those events, *Hidden Markov Model (HMM)* is used, which allows us to find probabilities for both observed and hidden events.

A Hidden Markov Model can formally consist of following components [17]:

- S = $s_1, s_2, \ldots\ldots s_N$ refer to a set of N states
- A = $a_{11}, a_{12}, \ldots\ldots a_{n1} \ldots\ldots a_{nm},$ are Transaction probability matrix A, where each a_{ij} states the probability of transacting from state i to j.

- O = O_1, O_2, ……O_N, are sequence of T observations where V = V_1, V_2,, ……,V_N are individual symbols
- B = b$_i$(o_t), are sequence of observations likelihoods, each one having the probability of observation generated from a state *i.*
- π = π_1, π_2, ……, π_n , are initial probability distribution over states.

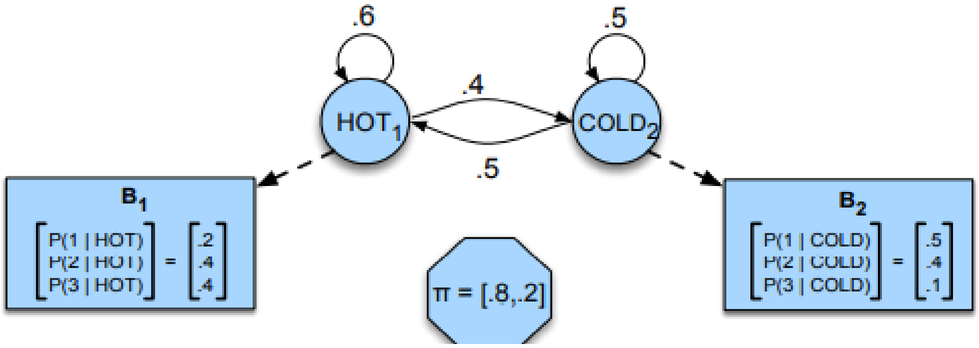

.6
.5
.4
HOT1
COLD2
.5
B1
P(1 | HOT) P(2 | HOT) P(3 | HOT) = .2 .4 .4
π = [.8,.2]
B2
P(1 | COLD) P(2 | COLD) P(3 | COLD) = .5 .4 .1

FIGURE 5

Figure 5 has two hidden states HOT(H) and COLD(C). Sequence of observations O, each an integer (1, 2, 3) shown in the matrices B1 and B2, refer to the number of ice creams eaten by a person on a given day. Hidden sequence Q of states of weather (HOT or COLD), needs to found out that made a person to have an ice cream. Table 1 is the simplified table of figure 5.

	p(...|H)	p(...|C)
p(1|...)	0.2	0.5
p(2|...)	0.4	0.4
p(3|...)	0.4	0.1

TABLE 1

The numbers in the table indicates the random probabilities. The probabilities of 1, 2 and 3 ice creams on a hot day are 20%, 40% and 40%, which totals up to 100%. Therefore, it can be guessed that on hot days, someone ate 3 ice creams, with 40% probability and on cold days ate 1 ice cream with a probability of 50%. Hence it can be guessed that, if someone ate 3 ice creams, the odds are 4 to 1 indicating that it was a hot day. But if someone ate 2 ice creams, the odds are 4 to 4, indicating that it was not clear of a hot or a cold day. Similarly eating 1 ice cream has odds of 2 to 5, which clearly indicates of a cold day.

2.2.2. Application of HMM in Fraud Detection

HMM is the perfect model to be implemented for the Fraud Detection system as there are many events which are not observable. Here in this system, can detect frauds just based on the cardholder's spending habits. HMM has the most reduced number of False Positive transaction which are identified as malicious by FDS even though they are legitimate, which makes it perfect to use for Fraud Detection.

The details of items purchased in Individual transactions are usually not known to any Fraud Detection System (FDS) running at the bank that issues credit cards to the cardholders. For each transaction taking place, it is sent to Fraud Detection System (FDS) for verification. The fraud detection system (FDS) accepts the card details such as Card number, CVV number, card type, expiry date, and amount to authenticate, if the transaction is legitimate or not.

When HMM is implemented, clusters of training sets are created to detect the spending habits of cardholder. Details regarding the purchases, like number of items bought, or description of the items are not known to FDS. FDS only works with the amount of transaction. HMM makes a prediction for three price ranges for any transaction: Low, Medium, and High. Based on the amount of various transaction, it forms a cluster in either low, medium, or high price range. According to the spending habits of the cardholder, it tries to figure out any variations occurring. Initial set of probabilities is selected based on the previous data and makes a sequence for future transactions. If FDS figures out the transaction is fraud, it creates an alarm, and a module of security information comes up. This module has various security questions such as date of birth, and some personal information, such as account number. If the security questions are answered correctly, the transaction goes through. We assume that only the cardholder is the one who knows the personal information. If the security questions are not answered correctly in 3 attempts, the transaction is rejected, and the card is blocked.

3. TECHNIQUES AND ALGORITHM USED

As discussed in the earlier section, for a full specification of HMM, we need two estimation model parameters, N and M, with a total of three probability distributions A, B and π. We can write them as a complete set as ƛ = {A, B, π}. The observation sequence O discussed in the above section can be produced through various possible state sequences.

Let's take a sequence Q = q_1, q_2, ……q_N,   , whereq_1 is the initial state. The probability of getting O from this given state sequence can be stated as P(O | Q, ƛ) = $\prod_{t=1}^{N} P(O_t \mid q_t , ƛ)$. This equation can also be stated as below:
P(O | Q, ƛ) = bq_1 (O_1)* bq_2 (O_2)……….. bq_N (O_N)

Also probability of state sequence Q can be stated as P(Q | ƛ) = $\pi_{q_1} a_{q_1 q_2}$* $a_{q_2 q_3}$……….. $a_{q_{N-1} q_N}$
Therefore, probability of making observation sequence O stipulated by ƛ can be stated as below:
P(Q | ƛ) = $\sum_{all\ Q} P(O \mid Q, ƛ) P(Q \mid ƛ)$ [18][19]

We get the HMM parameters for each user or cardholder. The forward-backward algorithm begins with all the initial parameters and congregates to the most nearby values.

In current system, after getting the HMM parameters, we acquire symbols from the training data of a cardholder and make an initial state series of all the symbols. Let's consider O_1, O_2, ……, O_N as one sequence. Based on the cardholder's transaction history in time t, this observation sequence is created. This input sequence is added to HMM to calculate the probability of approval. Suppose we get probability $\alpha_1$as α_1 = P (O_1, O_2, ……O_N | ƛ). If we consider O_{N+1} as a new sequence at a particular time t + 1, when a transaction is in process. Now since we have N+1 sequence, to consider only N sequence, O_1 has to be removed and hence sequences from O_2 to O_{N+1} are considered.

We get new probability as:
α_2 = P (O_2, O_3, ……O_{N+1}| ƛ).

From this we get:
Δα = α_1 -α_2

If we get results as Δα > 0, HMM contemplates O_{N+1} new sequence with lowest probability and hence the transaction is believed as a fraud transaction only if the change in the percentage in probability is more than the threshold value that is set beforehand:
Δα / α_1 ≥ threshold [20] .

Considering this system, there are three price ranges set for the spending profile of any card holder such as High (H) which ranges from ($501 to card limit), Medium(M) which ranges from ($101 to $500), and Low(L) which ranges from ($0 to $100). We can consider symbols set as V = [L, M, H] . We also define HMM parameters like State Transition Probability Matrix A, Observation Symbol Probability Matrix B, and Initial State Probability Vector π. All these three parameters are considered in the training phase of the HMM.

For the first 10 transactions, the system does not have enough information to detect a fraud based on the transactions. Therefore, for each transaction, user is asked some security questions. This is done until the user has done at least 10 transactions. Next, the HMM model gets the data for future verification based on the spending habits of the user. The algorithm and the description of the model are explained below.

3.1. Algorithm

The Algorithm section includes the training and detection phases. Training is done based on clustering scheme and it includes the following steps.

1. Based on Cardholder spending habits, identify the profile of cardholder
2. Calculation of probability relies upon the quantity of time that is past since entering current state
3. Create training sequence to train the model

Figure 6 shows the architecture of implementation of Training Algorithm.

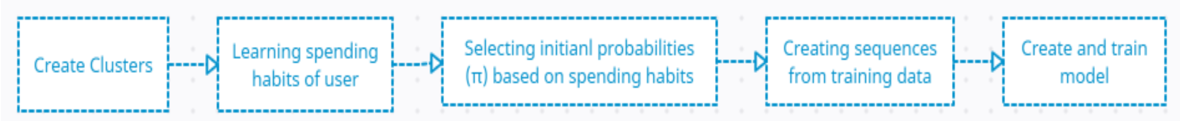

Create Clusters
Learning spending habits of user
Selecting initianl probabilities (π) based on spending habits
Creating sequences from training data
Create and train model

FIGURE 6

Detecting Fraud includes the following steps:

1. Creating observation symbol B_{N+1}
2. Creating the new sequence with addition of B_{N+1} in current sequence
3. Analyzing the difference in the probability and testing the outcome with training phase
4. If results are similar, the transaction is legitimate; otherwise, fraud alarm is created, and the system asks for verification.

Figure 7 shows the architecture of implementation of Detection Algorithm.

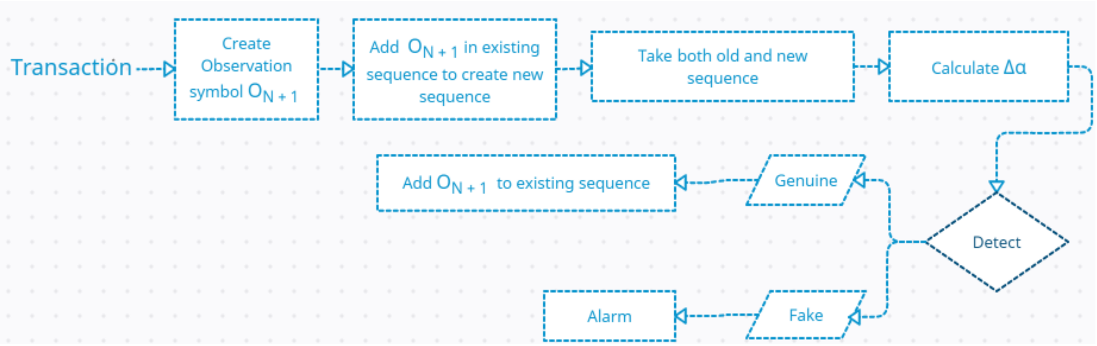

Transaction
Create Observation symbol O_{N+1}
Add O_{N+1} in existing sequence to create new sequence
Take both old and new sequence
Calculate Δα
Detect
Genuine
Add O_{N+1} to existing sequence
Fake
Alarm

FIGURE 7

3.2. Model Description: Online Shopping, Fraud Detection System

In the current existing systems, banks or other payment gateways asks for card information such as CVV, expiry month and year. But in cases where card is stolen or lost, all these details are readily available on the card itself, which makes it easier for any fraudulent to commit fraud. In addition, currently bank asks to provide a secure password for online purchases which is also not secure. This proposed model uses HMM to detect if the transaction is fraud or not at the time of transaction. This model has two modules namely: Online shopping & Fraud Detection.

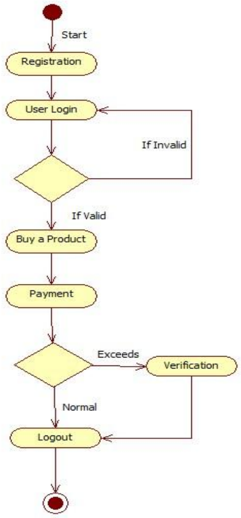
Online shopping module is the same as other online shopping websites. If the use is already registered, the user can login to the site; otherwise, the registration should be made first. The system asks a few security questions which will be used for authentications purposes. Only the users know the security questions and answers entered. After logging into the website, products can be selected and added into the cart for further transactions.

In the fraud detection section, once the user clicks for the payment, the system asked the user for card details. This includes card number, cvv, and expiration month and year. Once entered, the system checks the information with database. If the system finds that there is some variance in current transaction and the past transactions, the fraud detection module is activated.

Since this system is designed like other shopping website, the data of registered users, the order details, and the transaction amount are stored in the backend SQL database tables. If there are less than 10 transactions in the database for that user, then it asks directly for all the personal authentication questions that were set up during the registration. After the database is loaded with 10 transactions, it starts to compare the current transaction with the previous transactions before making a decision. Currently in this system, 10 transactions are taken into account to understand the behavior of the user, which can be increased as per requirements. In future, more than 10 transactions can be used from which HMM will learn the user's behavior for more accurate analysis.

The HMM model fetches the data from the SQL database tables and trains the data based on the users spending habits and how data are classified based on the threshold value set beforehand. The transaction amount and transaction category of the current purchase are compared with the past transactions. After performing the calculation for transaction probability, the system declares if the transaction is legitimate or fraud. If the system declares the transaction as fraud, it adds one more verification step that is the security questions. If the user enters correct answers transaction goes through. In case the transaction is fraud and wrong answers are submitted, after 3 attempts the card and account are blocked and no further transactions will be allowed from that account.

In cases where a user forgets the answers to the questions, the user can request to unblock his account by filling up the unblock request form. Figure 8 shows an activity diagram of the proposed system and Figure 9 shows the sequence diagram of the proposed system.



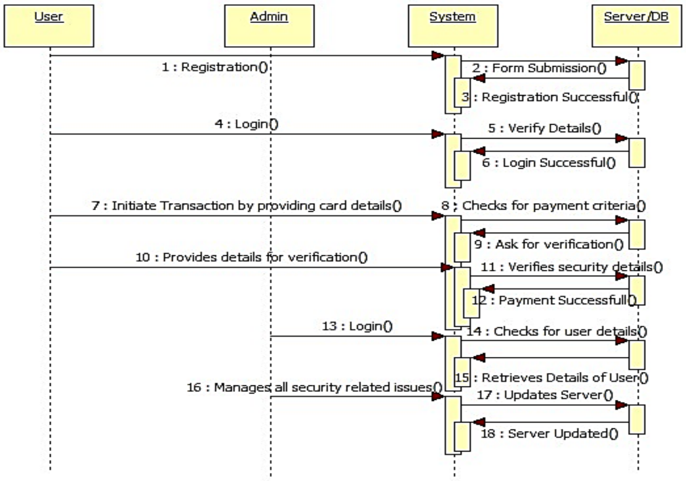

User
Admin
System
Server/DB
1 : Registration()
2 : Form Submission()
3 : Registration Successful()
4 : Login()
5 : Verify Details()
6 : Login Successful()
7 : Initiate Transaction by providing card details()
8 : Checks for payment criteria()
9 : Ask for verification()
10 : Provides details for verification()
11 : Verifies security details()
12 : Payment Successfull()
13 : Login()
14 : Checks for user details()
15 : Retrieves Details of User()
16 : Manages all security related issues()
17 : Updates Server()
18 : Server Updated()

FIGURE 8 **FIGURE 9**

Figure 10 depicts a complete diagram of working of the proposed system. The diagram starts with the user login, which if validated by the system, will allow user to go further. If the credentials entered by user are invalid, system will prompt for correct credentials. After user is logged in, products can be viewed and purchased. During that same time system will make a connection with the server. When customer buys a product system will fetch all the details related to the customer from the backend SQL tables. The incoming transaction will be mapped by the HMM algorithm based on previous transactions of the user and will make a decision whether the transaction is fraud or genuine. If detected as genuine, transaction will go through and order will be placed. If detected as a fraud transaction, additional information will be asked to user. If additional details entered correct, transaction will go through or else after 3 attempts, card will be blocked.

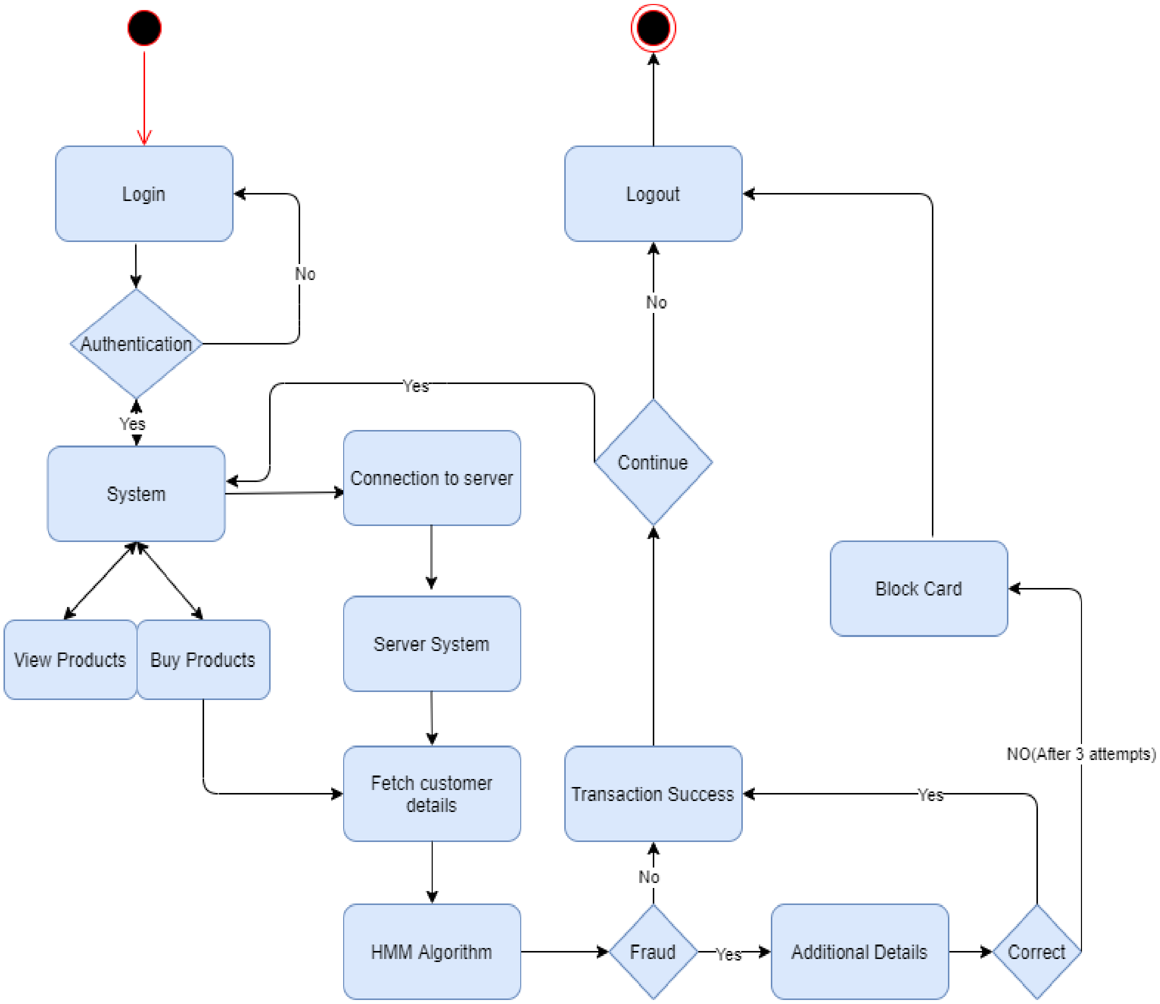


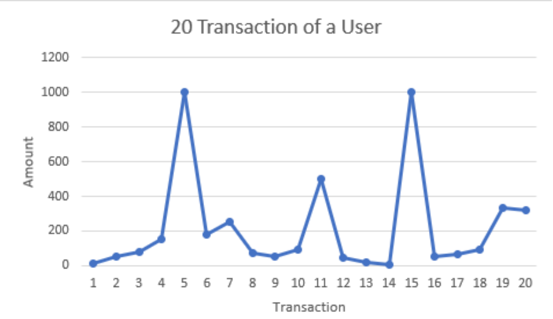
FIGURE 10

4. RESULTS AND DISCUSSION

Due to the security and privacy reasons, we are not able to perform the simulation the real time data as they are not provided by any Card issuing banks. Thus, for our proposed system a random set of total 20 transactions of a user are taken. Each transaction is classified into the purchased category and price range set before that is, High, Medium and Low. The detection system works after learning from first 10 transactions. After 10th transaction, the Fraud Detection starts detecting fraud on every incoming transaction. Following is the random data entered while making purchase on the online shopping module of the system with category of purchase and transaction amount. This data is fetched from the SQL database tables where details of all the users and transactions are stored. Table 2 shows the dataset taken for this system, showing all the transactions and categories.

No. of Trans	Amount	Category	No. of Trans	Amount	Category
1	10	Art & crafts	**11**	500	Electronics
2	50	Shoes	**12**	45	Purse
3	80	Shoes	**13**	20	Art & crafts
4	150	Purse	**14**	5	Art & crafts
5	1000	Rent	**15**	1000	Rent
6	180	Furniture	**16**	50	Purse
7	250	Furniture	**17**	68	Shoes
8	70	Purse	**18**	94	Shoes
9	55	Purse	**19**	330	Electronics
10	95	Shoes	**20**	320	Electronics

TABLE 2



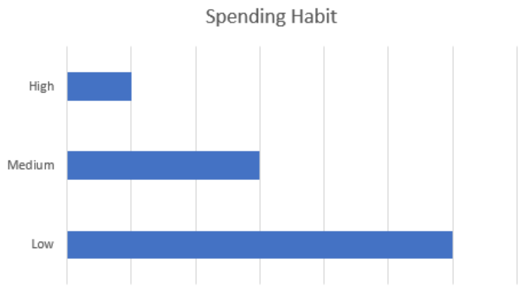
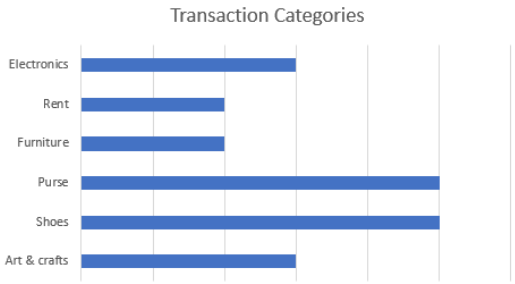


FIGURE 11

FIGURE 12

Figure 11 displays a graph of Transaction of the user or card holder according to the amount spend. Figure 12 depicts the graph of the spending habit of the user based on the price range set before. The graph shows that users spend almost 60% in the Low range which is from $0 to $100 followed by Medium range with 30% that is between $101 to $500 and the least is spent in High Range that comprises of just 10% ranging between $501 to $1000.



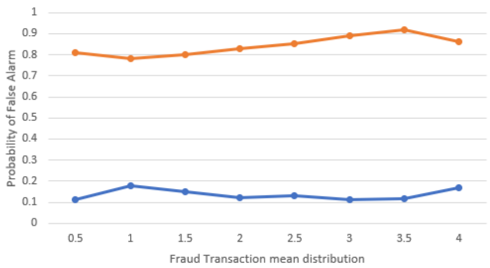


FIGURE 13

FIGURE 14

Figure 13 is a graphical representation of the Categories in which a user spends. In this system, we have taken Categories such as Art & Crafts, Shoes, Purse, Furniture, Rent and Electronics. It is shown that a user mostly spends on two categories that is Shoes and Purse with around 25%, followed by Arts & Crafts and Electronics with 15% and least is 10% spend on Furniture and Rent.

Figure 14 shows the Mean Distribution of Fraud Detection. In this graph Probability of Fraud Transaction is compared with Genuine Transaction. As illustrated, when probability of fraud transaction goes down, the probability of genuine transaction goes up vice-versa. This vice versa has helped in finding out the false alarm for the fraud transaction detection. Therefore, the instances where the probability for false alarm is more compared to threshold probability, a fraudulent alarm takes place, and a transaction is declined.

Unlike other fraud detection systems, proposed system is capable of detecting the fraud transaction the same time the transaction is made. The proposed system can detect the fraud on real time data without taking enormous amount of time contrasting the other systems. When compared to other approaches, the proposed system is approximately in the middle for the computation time. However, using this system in a practical environment, the computation time can be low which can be further enhanced using upgraded HMM algorithms. Both less and more number of transactions can be computed in satisfactory amount of time, in milliseconds and seconds.

Theoretically when considering probability matrix B, the processing time rises due to encoding of probability matrix B. Also, more processing time and memory can be used up when dealing with larger matrices. This can be fixed by allocating memory to specified matrix size which can give a better linear execution.

In addition to the processing time, verification procedure has minimal steps which in turn does not negotiate with the customer experience.

Proposed system can be most beneficial to the users who can save their hard-earned money from being used up by fraudulent. All e-commerce businesses can massively stop the losses that incurs to them due to the fraud transactions taking place. And banks can reduce their time expended in the wake of uncovering the frauds and finance that time in some other matter.

5. CONCLUSION

In the proposed system, we have discussed the usefulness of Hidden Markov Model for Detecting Fraud in Online Transaction. The steps that are used in the transaction process are considered as stochastic process of Hidden Markov Model, while the price ranges of transactions are considered as observation symbols and the items purchased are taken as states of Hidden Markov Model (HMM). The proposed system is also ascendable for handling huge amount of transactions. This system gives fast results unlike the existing system. In this system, any incoming transaction will be checked for legitimate, or fake based on the user’s spending habits. This system will also make decisions of fake or legitimate transactions based on the set threshold values. Accuracy of this Fraud Detection system is nearly 80%, and has high processing speed, according to the comparative studies made. Proposed system can be used in almost all the online shopping websites for various merchandises. It can also be implemented as an extra layer for payments gatewaying Banks and other e-commerce platforms. It is most suitable for Online Transaction Fraud Detection since there is no need to check the original user as it maintains the log of users.

6. REFERENCES

[1] Nilson Company. Connected Commerce. Connectivity is Enabling Lifestyle Evolution. November 2018.

[2] Stojanovic A., Aouada D., Ottersten B Bahnsen A.C., "Cost-sensitive credit card fraud detection using Bayes minimum risk," in *12th International Conference on Machine Learning and Applications (ICMLA)*, pp. 333-338, 2013.

[3] Nilson Company. Connected Commerce. Issue 1187. Dec 2020 Card Fraud Worldwide.

[4] About US English. (2020, November 09). Retrieved March, 2021, from https://nilsonreport.com/publication_chart_and_graphs_archive.php?1=1&year=2020.

[5] Anderson, Keith & Durbin, Erik & Salinger, Michael. (2008). Identity Theft. Journal of Economic Perspectives. 22. 171-192. 10.1257/jep.22.2.171.

[6] A. Dal Pozzolo, O. Caelen, Yann-A¨el Le Borgne, Serge Waterschoot, and Gianluca Bontempi. Learned lessons in credit card fraud detection from a practitioner perspective. Expert Syst. Appl., 41:4915–4928, 2014.

[7] Andrea Dal Pozzolo, Giacomo Boracchi, Olivier Caelen, Cesare Alippi, and Gianluca Bontempi. Credit card fraud detection and concept-drift adaptation with delayed supervised information. 2015 International Joint Conference on Neural Networks (IJCNN).

[8] Kuldeep Randhawa, Chu Kiong Loo, Manjeevan Seera, Chee Peng Lim and Asoke K. Nandi, "Credit card fraud detection using AdaBoost and majority voting," *IEEE Access*, vol. 6, pp. 14277-14284, 2018.

[9] John O. Awoyemi, Adebayo Olusola Adetunmbi, and Samuel Adebayo Oluwadare. Credit card fraud detection using machine learning techniques: A comparative analysis. 2017 International Conference on Computing Networking and Informatics (ICCNI), pages 1–9, 2017.

[10] You Dai, Jin Yan, Xiaoxin Tang, Han Zhao and Minyi Guo, "Online Credit Card Fraud Detection: A Hybrid Framework with Big Data Technologies", IEEE TrustCom/BigDataSE/ISPA, pp 1644 -1651, 2016.

[11] A. Roy and J. Sun and R. Mahoney and L. Alonzi and S. Adams and P. Beling, "Deep learning detecting fraud in credit card transactions," in *Systems and Information Engineering Design Symposium (SIEDS)*, pp. 129-134, 2018.

[12] S. Xuan, G. Liu, Z. Li, L. Zheng, S. Wang and C. Jiang, "Random forest for credit card fraud detection," 2018 IEEE 15th International Conference on Networking, Sensing and Control (ICNSC), 2018, pp. 1-6, doi: 10.1109/ICNSC.2018.8361343.

[13] William N. Robinson, Andrea Aria, Sequential fraud detection for prepaid cards using hidden Markov model divergence, Expert Systems with Applications, Volume 91,2018, Pages 235-251, ISSN 0957-4174.

[14] Fraud detection: How machine learning systems help Reveal scams in Fintech, healthcare, and ecommerce. (2020, February 27).

[15] K. Bennett. Legacy systems: coping with success. Published in: IEEE Software (Volume: 12, Issue: 1, Jan. 1995). Pages 19-23. DOI: 10.1109/52.363157.

[16] Daniel Jurafsky & James H. Martin. Speech and Language Processing. December 30, 2020.

[17] Stamp, M. A Revealing Introduction to Hidden Markov Models.

[18] V.Bhusari & S.Patil. “Application of hidden Markov Model in Credit Card Fraud Detection” November 2011. International Journal of Distributed and Parallel Systems (IJDPS) Vil.2, No.6.

[19] Rabiner, L. R. (1989, February). A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition. February 1989. Proceedings of the IEEE, Vol.77, No.2.

[20] A.Prakash (December 2012)A Novel Hidden Markov Model for Credit Card Fraud Detection. International Journal of Computer Applications, Vol.59, No.3.

[21] T.Chetcuti & A. Dingli,(2008) Using Hidden Markov Models in Credit Card Transaction Fraud Detection.

[22] Bunke, H., & Caelli, T. (2001). *Hidden Markov models: Applications in computer vision*. Singapore: World Scientific.

[23] H.Zhou, G.Sun . (2019). *A Scalable Approach for Fraud Detection in Online E-Commerce Transactions with Big Data Analytics*. Tech Science Press, Vol.60, No.1.