Fraud Recognition And Uniquely Designed System (F.R.A.U.D.S)

[1]Harshal Jayant Jagtap, [2]Ruchit Kaushik Sangoi
Dept. of Information Technology Engineering, MET Inistitute Of Engineering, Nashik, Maharashtra, India

Abstract – Interestingly, most people have a different take on digital banking. There is a lot more to digital banking than just a few features that we can see on the surface. Digital banking is converting the brick and mortar banks into more secure and efficient places to operate. Online Credit Card Fraud occurs when user provide their information to the unknown persons or stolen by the unauthorized person, this information can be used for unauthorized online purchase or for accessing the credentials. Financial institutions all over the world lose billions due to credit card fraud, which necessitate the use of credit card fraud prevention. Several models have been proposed in the literature, however, the accuracy of the model is crucial. In this paper F.R.A.U.D.S (Fraud Recognition and Uniquely Designed System) is based on Machine Learning Algorithms (Hidden Markov Model and Logistic Regression) which helps to improve accuracy, security and also helps to build a scalable, fault-tolerant and high performance system..

***Key Words*: Online Credit Card Fraud Detection, Hidden Markov Model (HMM), Logistic Regression.**

1. INTRODUCTION

Credit card frauds can be defined as the use of the victim's card information without the authorization of the victim. A credit card is issued to the customers that help them to make payments that are then paid at once by the customer. As it is an easy way to make payments, people have started to use it more often. This has even attracted fraudsters to use other's cards for their own needs. The credit card can be a physical one or a virtual one. When using a physical card, the card holder has to show his card whenever a payment is to be made. The attackers only need some crucial information regarding the card such as pin code, card number, expiry date, etc., to make a payment. Such fraudulent transactions are generally done over the Internet. Frauds can be committed in such type of purchases. A fraudster only requires the knowledge of card details. Usually, the cardholder is not aware that someone else has stolen card information until a fraud transaction occurs in his account.

Machine Learning Algorithms are used for detecting fraudulent transaction which may be obtained through techniques like Neural Network, SOM, DBSCAN, HMM[1], Rule-Based Filter, Decision Tree, Naive Bayes, and Logistic Regression. The historical transactions of the user are used to identify whether the current transaction being made is genuine or not.[2]

1.1 Existing system

Global bank HSBC[4] has worked with analytics software SAS to create SAS Fraud Management, a real-time, a card fraud detection system. Using SAS Fraud Management in the US, HSBC has improved fraud detection, false positive rates and fraud case handling efficiencies. Benefits include an 87% increase in the number of data items processed, including card transactions and customer information, with a corresponding reduction of 12% in the mainframe processing overhead. This has resulted in 53% decrease in mainframe processing cost per data item, says HSBC, which is now implemented to monitor the use of more than 30 million cards in the US. It is also being used by the Halifax Bank of Scotland (HBOS). Other existing systems : Signifyd, Actimize, Similty, Agroscope, Fiserv.

1.2 Limitations with Existing system

The existing framework aims at fusing different detection algorithms to improve accuracy and using a four-layer design to handle data storage, model training, and data sharing and online detection. They had implemented the framework with latest Big Data technologies. The hybrid framework can also be applied in other similar application fields, for example, internet advertising fraud detection, telecom fraud detection and so on. However, this also is applied in other similar application fields, for example, internet advertising fraud detection, telecom fraud detection and so on. However, this work still has a lot of things to be done, e.g., better integration of more detection algorithms i.e. supervised and unsupervised algorithms.[3]

1.3 Related Work

In online trading, the credit cards are usually used as a virtual card. An attacker only needs to obtain few important information of the card (e.g. card ID, secure code)to make a fraudulent transaction on the Internet while the genuine cardholders often do not notice that his card information has been leaked, which may cause a significant financial loss both to the card-holder and credit card company. In the past decades, financial companies and researchers have developed many Credit Card Fraud Detection Systems (CCFDS)[5]. Although the main challenge for most CCFDS is how to improve detection accuracy, the computational capacity of CCFDS has become more and more important with the explosive growth of trading data. In this case, processing detection tasks and model training on so many incoming transactions with a low delay is very hard for most traditional systems.

2. Proposed System

A] Architecture

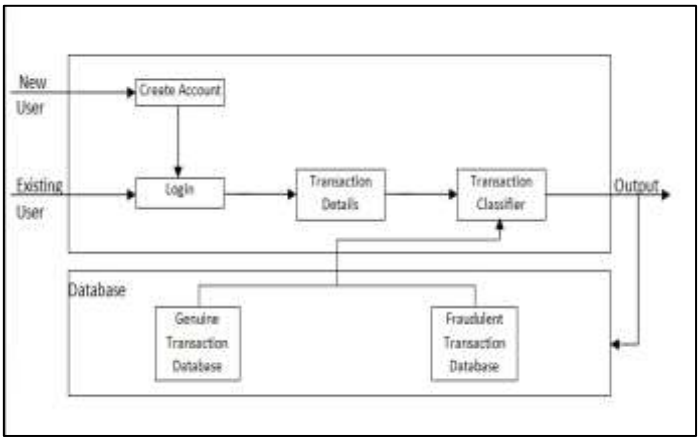


Fig. 2.1 : Architecture of System

The proposed architecture of F.R.A.U.D.S consists of 2 Layers in which the **First Layer** will be consisting of new user and existing user page respectively.

The **Second Layer** is database we are using MySql that will store the user details and transaction. details of all user. It also stores if the transaction made by the user is Genuine or Fraudulent which is handled by the **First Layer** to process the transaction of the user.

B.] System Work-flow

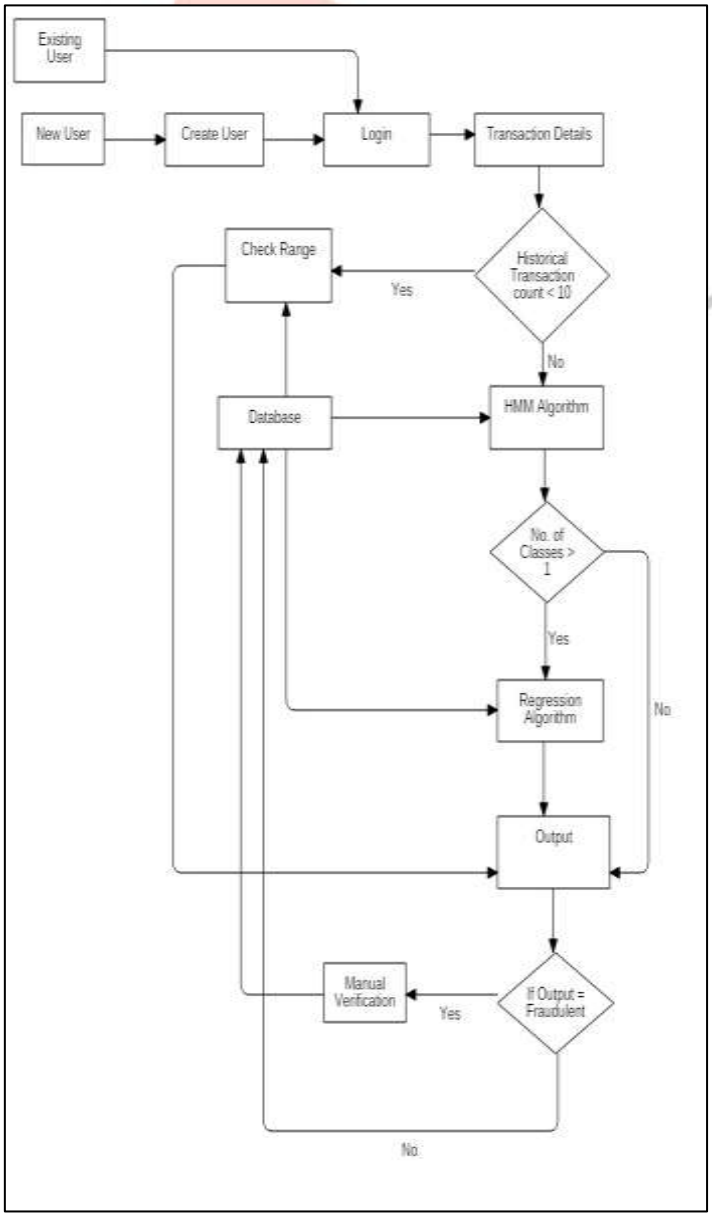


Fig. 2.2 : Work flow Diagram of System

The above work-flow diagram of F.R.A.U.D.S describe the components used which are new user, Existing user, Login Page, Transaction Page and so on. Initially, if the user is not registered he/she will have to go through the registration process. After registering the user can go to the login page and perform the transaction which the user wants to process for. In the proposed system we have used two machine learning algorithms Hidden Markov Model and Logistic Regression.
When user process the transaction it will first check if there are 10 historical transactions processed if not then it will go to the check range. If the user has historical transactions. it will check through the Hidden Markov Model algorithm which will check genuine transaction. from the database where as in Logistic Regression algorithm it will check if the transaction. done is either genuine or fraudulent.
The manual verification is done by sending OTP to the respective user through Email and it will work
only if any of the result obtained by both algorithms give a fraudulent result. Hence transaction. success notification will be displayed to the user on the screen.

3. Machine Learning Algorithms Used

a] Hidden Markov Model (HMM)

The incoming transaction is given to the FDS for a verification process. This then gives result if the card detail and verify whether the transaction generated is genuine or fraudulent. It tries to find any unauthorized transaction based on the spending of the cardholders. If the FDS confirms the transaction is fraudulent, transaction. is not processed and the message is displayed TRANSACTION UNSUCCESS, and the issuing bank declines the transaction. The concerned card-holder will be contacted and informed about the possibility that the card is being compromised. HMM never check the details of the original user as it maintains a log file. The log maintains the proof for the bank for the transaction made by the user. HMM reduces the tedious work of an employee in the bank since it maintains a log. In this process, the spending profile of customer can be categorized into three parts:

- Low Spending Profile
- Medium Spending Profile
- High Spending Profile

Every user is represented by specific patterns of the set which contains information about last 10 transactions using a credit card. The set of information contains spending profile of card holder, money spent in every transaction, the last purchase, category of transaction.

Working of HMM:

1: Initial sequence O1, O2, O3,….Or (from card-holder transaction)
2: HMM (compute the probability acceptance $\alpha_1 = P(\lambda)$ (O1, O2, O3,….Or)
3: Or + 1 (new generated symbol)
4: HMM (compute the probability acceptance $\alpha_2 = P(\lambda)$ (O1, O2, O3,….Or + 1)
5: $\Delta\alpha = \alpha_1 - \alpha_2$
6 : If $\Delta\alpha$ > threshold then the Transaction is fraudulent
7 : Else the Transcation is Genuine.

b] Regression Algorithm

Logistic regression is one of the statistical methods that analyze a dataset in which there are one or more independent variables which determines an outcome. The outcome generated is measured in only two possible outcomes. It is basically used to predict a binary outcome like 1 or 0, Yes or No, True or False. To represent binary outcome, duplicate variables are used. We also thought of logistic regression as a special case of linear regression when the outcome variable is categorical.

Working of Logistic Regression:

1: The model consists of a vector β in d-dimensional feature space.
2: For a point x in feature space, project it onto β to convert it into a real number z in the range - ∞ to + ∞

$$z=\alpha + \beta \cdot x = \alpha + \beta_1 x_1 + L + \beta_d x_d$$

3: Map z to the range 0 to 1 using the logistic function

$$p = \frac{1}{(1 + e^{-z})}$$

4: Overall, logistic regression maps a point x in d-
dimensional feature space to a value in the range 0 to 1.
5: It can interpret prediction from a logistic regression model as:
A probability of class membership
A class assignment, by applying a threshold to probability threshold represents p decision boundary in feature space.

4. Experimental Setup

The system is developed and tested using java-jdk1.7. The database is stored in MySQL database server. A web application is created using eclipse on Ubuntu- 14.04 system with 4 GB ram and i3 processor.

Dataset:
A synthetic dataset is generated for 50 users. Dataset contains account number, transaction amount and location. The location is a modest attribute. It has 3 possible values: mall, bill, and restaurant.

Results:
The system tested for transaction types based on a number of records and dataset transaction type. For new user, range check function is applied and HMM and Regression are not applicable due to lack of training dataset. As soon as the dataset size reaches the minimum threshold value (set to 10) then HMM and Regression checks are applicable. For regression checking genuine and fraudulent records must be present in a dataset. If all transactions are genuine then regression check is not applicable.

Following are results for each classifier.

1. **HMM**

Amount	Location	Type
2000	Restaurant	Genuine
9000	Mall	Genuine
450	Bill	Genuine
4500	Restaurant	Genuine
6500	Mall	Genuine
650	Bill	Genuine
2500	Restaurant	Genuine
7500	Mall	Genuine
600	Bill	Genuine
5000	Restaurant	Genuine

Test Transaction : < pending, 15000, bill >
Threshold alpha : 2.0E7
Alpha value :

Hence Transaction is Genuine.

2. **Regression**

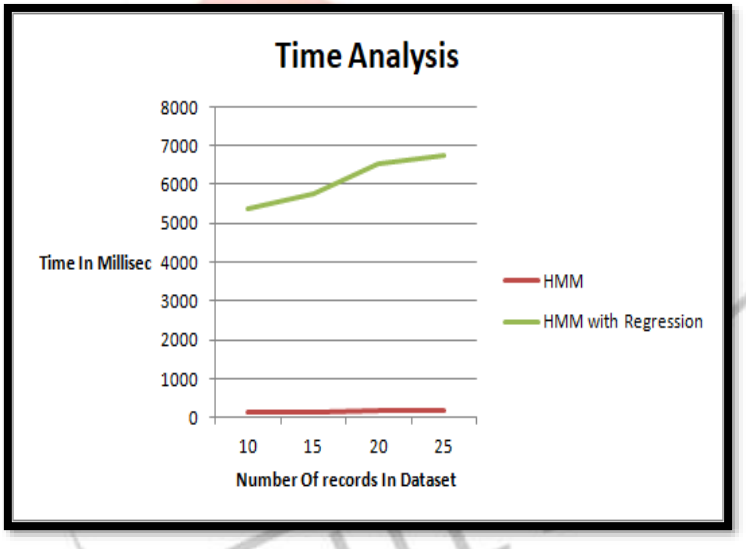
Amount	Location	Type
25000	Mall	Genuine
10000	Mall	Genuine
100000	Restaurant	Fraudulent
4000	Restaurant	Genuine
15000	Bill	Fraudulent
12000	Mall	Genuine
11000	Mall	Genuine
520	Bill	Genuine
15000	Mall	Genuine

The **Threshold** set to: 0.7 for the genuine transaction.
The **Test Transaction** <pending, 140000, bill> is tested over above dataset generates the **fdistribution** as: 0.508152
Hence Transaction is Fraudulent.

3. **Total Time Estimation**

Sr. No	Dataset Size	Applied Checks	Time in milliseconds

1	5	Range Check	63
2	10	HMM	115
3	10	HMM, Regression	5376
4	15	HMM	137
5	15	HMM, Regression	5754
6	20	HMM	178
7	20	HMM, Regression	6545
8	25	HMM	193
9	25	HMM, Regression	6752



5. CONCLUSION

Online fraud detection is challenging due to the burst amount of trading transactions that are happening every day. The Proposed System design of framework is made to solve this problem. This framework aims at fusing different detection algorithms to improve accuracy and using a two-layer design to handle data storage, model training, and data sharing and online detection. The Implementation of the framework can be done with Machine Learning Technologies, which helps to build a scalable, fault-tolerant and high-performance system.

The proposed framework can also be applied in other similar applications for example internet
Advertising fraud detection, telecom fraud detection and so on.

6. Acknowledgment

Firstly we gladly thank our project guide, Prof K. Metre, for her valuable guidance for implementation of the proposed system. We will remain thankful for excellent as well as polite guidance for the preparation of this report. Also, we would sincerely like to thank HOD of Our department Prof. N. R. Kale and other staff for their helpful coordination and support in project work.

7. REFERENCES

[1] D. Iyer, A. Mohanpurkar, S. Janardhan, D. Rathod, and A. Sardeshmukh, "Credit card fraud detection using hidden markov model," in *Information and Communication Technologies (WICT), 2011 World Congress on*, Dec 2011, pp. 1062–1066.

[2] E. Ngai, Y. Hu, Y. Wong, Y. Chen, and X. Sun, "The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature," Decision Support Systems, vol. 50, no. 3, pp. 559 – 569, 2011.

[3] M. Krivko, "A hybrid model for plastic card fraud detection systems," Expert Systems with Applications, vol. 37, no. 8, pp. 6070 – 6076, 2010.

[4] http://www.hsbc.com/about-hsbc/structure-and-network/global-banking-and-markets.

[5] Computational Intelligence in Data Mining: Proceedings of the International Conference on CIDM, 10-11 December 2016 Himansu Sekhar Behera, Durga Prasad Mohapatra