International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 09 Issue: 04 | Apr 2022 www.irjet.net p-ISSN: 2395-0072

Online Transaction Fraud Detection System Using Machine Learning & E-Commerce

Shayan Wangde*, Raj Kheratkar*, Zoheb Waghu*, Prof. Suhas Lawand**

Student, Professor***
Department of Information Technology, Pillai College of Engineering, Navi Mumbai, Maharashtra, India

---

Abstract* – *We here come up with a system to develop a website which has capability to restrict and block the transaction performing by attacker from genuine user's credit card details.. We tried to detect fraudulent transaction before transaction succeed. While registration we take essential login details which is enough to catch fraudulent user activity. Therefore we have used (BLA) Behavior and Location Analysis. The items purchased are shown in cart where transactions are usually not known to any Fraud Detection System (FDS).Therefore , We used BLA to detect this problem. An advantage to use BLA approach to reduce number of positive false transactions identified as malicious by an FDS although they are genuine. An FDS runs at a credit card issuing bank. Each incoming transaction is submitted to the FDS for verification. FDS receives the card details and transaction value to verify, whether the transaction is genuine or not. The types of goods that are bought in that transaction are not known to the FDS. Bank declines the transaction if FDS confirms the transaction to be fraud. User spending patterns and geographical location is used to verify the identity. If any unusual pattern is detected, the system requires re-verification. The previous data of the user the system recognizes unusual patterns in the payment procedure.

Keywords---credit card frauds, fraud detection system(FDS), Behavior and Location Analysis (BLA), fraud detection, real-time credit card fraud detection.

1. Introduction

With the advancement of cutting-edge technology and global connectivity, fraud has risen dramatically. There are two ways to identify fraud: prevention and detection. By serving as a layer of defense, prevention helps to keep fraudsters at bay. The information utilized in this study is divided into two categories: categorical data and numerical data. Initially, the dataset contained categorical data. Data cleaning and other basic pre-processing techniques will be used to prepare the raw data. However, there has been a significant surge in fraudulent transactions, which has had a significant impact on the economy. There are a few different types of credit card fraud. Card Not Present (CNP) and Card Present (CP) frauds are the two forms of fraud that may be recognized in a collection of transactions.The security mechanism of modern computer networks, such as the Internet, has three levels: network IP level security, transport-level security, and application-level security (end-to-end security).

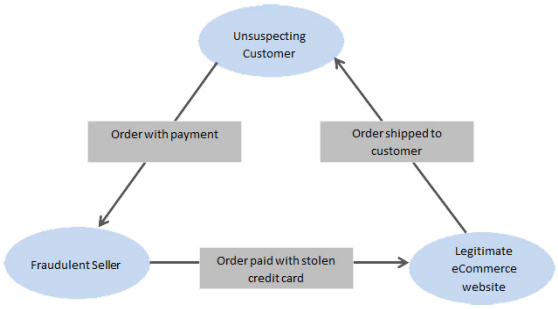
How Payment Gateways can get fraud

1. Phishing is a sort of cybercrime that includes stealing personal and business information via the use of fake emails, websites, and text messages. Phishing is a type of social engineering attack that is commonly used to obtain sensitive data from users, such as login passwords and credit card information. Phishing occurs when a hacker poses as a trustworthy entity and convinces a victim to open an email, instant message, or text message. Normal phishing, spear phishing, whaling, cloning, and email phishing are the five forms of phishing attacks.

2. Credit Card Fraud is when someone uses a revoked, canceled, reported lost, or stolen credit card to receive something of value with the aim to deceive. Credit card fraud can also be committed by using credit card numbers without actually having the card. What methods are used to commit credit card fraud?

a) Skimming: Scammers install a card skimmer on a credit card swipe machine.
b) Dumpster diving: When you throw away invoices or documents that include your complete credit card numbers, fraudsters can retrieve the information and conduct fraud.
c) Hacking: Thieves can gain access to companies with which you've done business or credit card processing companies. After that, they'll take part in data breaches.

3. Triangulation The fraudster acts as a middleman between a customer and an unwary merchant in triangulation fraud. The user will place an order for the desired item with the fraudster, who will then use stolen credit card information to make the transaction from a legal merchant. Customers' trust in small and large e-commerce stores, as well as the marketplace, is being eroded by this fraud. Customers are far less inclined to make purchases if they feel compelled to regularly check their bank account for indicators of fraud after making an online purchase.



How Online Fraud Transaction Prevented

i. Address Verification Service (AVS): This is a feature that credit card processors and banks provide to users in order to detect unusual credit card transactions and prevent fraud. The AVS compares the billing address provided by the cardholder to the billing address on file with the issuing bank. This is done as part of the merchant's request for credit card transaction authorization. In a non-face-to-face transaction, the credit card processor sends a message code back to the merchant showing the degree of address matching, thereby confirming ownership of a credit or debit card. This procedure aids the merchant in determining whether to accept or reject a card transaction.

ii. Card Verification Value (CVV): The CVV is a three- to four-digit number found on every debit and credit card. It's never a good idea to reveal the code. Because it is only available on the printed card, a CVV filter adds an extra layer of security, allowing only the cardholder to use the card. If your CVV does not match while placing an order on your e-commerce site, your payment should be refused. The merchant obtains the required card information from the consumer during a card-not-present transaction (online, email, or telephone orders) to verify the transaction. Friendly fraud is a danger that can result in a chargeback when it comes to CNP transactions. Using a CVV filter to combat fraud and reduce chargebacks is beneficial to retailers.

iii. Device Identification analysis: Rather than looking at the individual who is running your e-commerce site, Device Identification looks at the system. It analyses the operating system, internet connection, and browser to determine whether or not the online transaction should be accepted, flagged, or denied. All gadgets (phones, computers, tablets, and so on) have a unique device fingerprint, which is comparable to a person's fingerprint and aids in detecting fraudulent patterns and assessing risk.

Mode Of Banking Operations

Table 2 shows that 76.1 percent of respondents have a savings account, whereas 23.9 percent have a current account. When it came to doing digital banking, 100 percent of respondents said they were doing it. 2.2, 28.3, 63, and 6.5 percent of respondents, respectively, belong to the mobile, computer with an internet connection, mobile and computer with an internet connection, and third party groups for digital banking. 8.6, 8.71, 10.89, and 71.6 percent of respondents use ATMs, credit/debit cards, NET banking, and all of the above types of digital banking services. 73.8 and 26.2 percent of respondents, respectively, belong to the personal and institutional groups when it comes to the purpose of digital banking. Similarly, money transfer via the internet, online purchase, account and balance verification via the internet and mobile banking, information services, and all of the above categories account for 8.6, 45.5, 4.29, 2.11, and 39.1 percent of respondents, respectively.

Table 1: Nature of banking operations

Variable	Category	Percentage
Types of Account in bank	Savings	76.1
	Current	23.9
Status of doing digital banking	Yes	100
Mode of doing digital banking	Mobile	2.2
	Computer with internet	28.3
	Mobile & Computer with internet	63
Types Of Digital Banking.	ATM	8.6
	Credit/Debit	8.71
	NET banking	10.89
	All of the above	71.6
The Purpose of doing digital banking	Personal	73.8
	Institutional	26.2
Types of Services	Money Transfer	8.6
	Online Purchase	45.29
	Information Services	2.11

2. Literature Survey

i. Fraudulent online transactions In this work, Dr. mooramreddy Sridevi, Teruvayi Sai Chandu, and Dr. mooramreddy Sridevi, help in comprehending fraud transactions and may be utilized to further prepare the system to build new rules and achieve higher fraud detection precision. The results of the experiments show that the proposed application is beneficial and applicable to real-world systems.

ii. Detection of Credit Card Fraud in Real-Time An exhaustive comparison of machine learning techniques via Anuruddha thennakoon, Chee Bhagyan, Sasitha Premadasa, and effective performance measures for the detection of fraudulent credit card transactions are used in this study to determine optimal algorithms for the four fraud types.

iii. Secure Internet Payment Mechanism: This study discussed the IPS (Internet Payment System) payment

system. However, because electronically transmitted data can be quickly viewed and changed, the fraud protection in internet commerce is challenging. To achieve privacy, integrity, authentic modules willS employ a variety of cryptographic algorithms and approaches.

iv. Credit Card Fraud Detection: The goal of this study is to find the best algorithms for fraud patterns by analyzing machine learning techniques while using an effective performance metric to detect fraudulent credit card transaction.

v. Online Transaction Security Customers' trust in electronic transaction is influenced by two primary factors: privacy and security. As a result, businesses, websites, and organizations that sell products or services online should make more efforts to favorably influence their customers' impressions of privacy and security. To detect fraudulent shopping, our method examines these common traits, such as apartment address, web browser cookie information, and merchandise categories.

vi. Secure Internet Payment System by the author Zoran Djuric explains modern computer networks, such as the Internet, have three levels of security: security at the network IP level, security at the transport level, and security at the application level (end-to-end security). An Internet payment system must meet the following basic security standards to be considered secure: secrecy (privacy), data integrity, authentication, and non-repudiation..

vii. Analysis of Internet Fraud Azhar Usmani's paper describes The dangers of using the Internet are numerous due to the numerous ways in which information posted on the internet can be modified. These records are vulnerable to security breaches that result in a variety of data breaches. Another key area of Internet-related fraud is data manipulation in electronic commerce. Users, corporations, and governments are concerned about deceit on the part of Internet enterprises and third parties. All of these types of online fraud necessitate new prevention techniques, such as legislative reforms, effective data coding, and fraud prevention strategies.

viii. Digital Banking Frauds via M.kannan, explains digital banking frauds, Different types of customers were included in the study on digital customer perceptions of digital banking fraud.It's split into three sections. Customers' age, sex, educational background, and type of employment are all part of the first part of the social-economic profile. The second section looked into the nature of their banking operations, while the third section focused on digital fraud in the form of statements, which were graded on a five-point scale. On a scale of 1-10

I.Online Shopping Fraud Detection System and Its Evaluation by the author Kenichi Yoshida, Kazuhiko Tsuda, and Hiroki Azuma describe how the online shopping fraud detection system is comprised. Although data mining technologies are used to identify fraudulent clients, they are not the most important components. The relevance of three primary components, namely the PC identification, address identifier, and goods classifier, cannot be overstated. The information provided by pre-processing improves the system's final decision accuracy.

3. Proposed Work

3.1 Overview

This section provides an overview of the system. Figure 3.1 depicts the classification of various techniques in the domain.

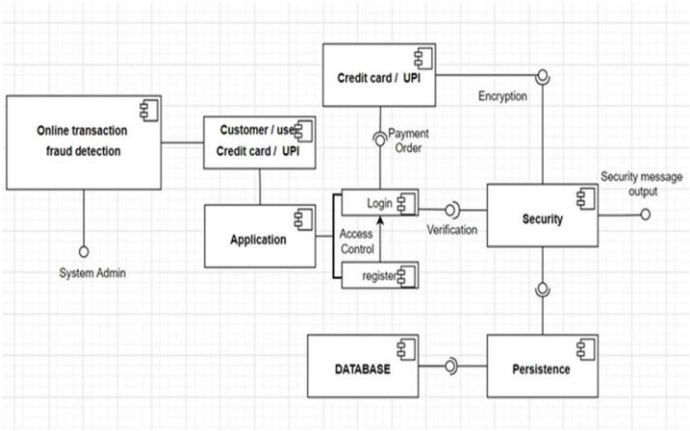
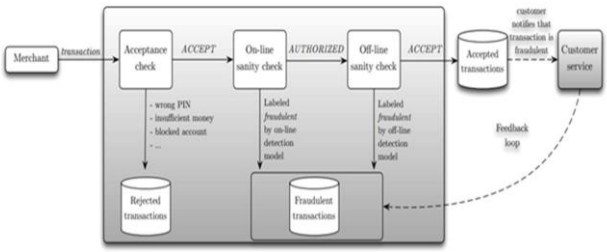


Fig. 3.1 Classification of domain techniques

Trinis and network-related characteristics Intrinsic features examine the transaction as if it were a separate entity, determining whether it falls into the typical customer profile. We derive those qualities from the credit card holder's historical transactions' RFM properties — Recency, Frequency, and Monetary Value. Network-based characteristics, on the other hand, define each transaction by constructing and analyzing a network of credit cardholders and merchants who are linked through transactions. An example network is shown in Figure 1. Using a limited set of confirmed fraudulent transactions, we use a collective inference algorithm to spread fraudulent influence throughout the network, and we decide on the suspiciousness of each network object by calculating an exposure score – that is, how much of the network object has been subjected. Previous fraudulent influences are exposed to the transaction, the connected account holder, and the merchant.

3.1.1 Existing Methodology and Systems

Content-Based Information Filtering (IF) systems require appropriate approaches for representing objects and generating user profiles, as well as strategies for comparing the user profile to the item representation.



3.1System Architecture

3.1.1 Methodology

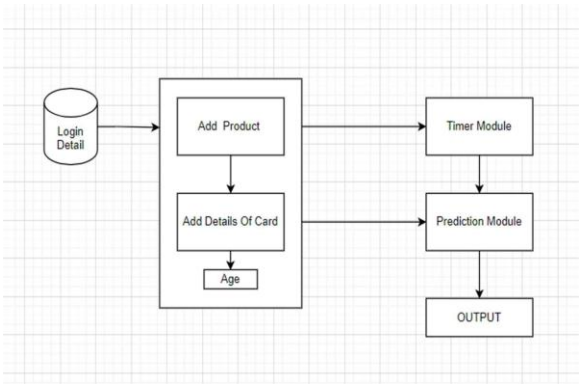


Fig3.1.1 Proposed System

3.1.2 Algorithms

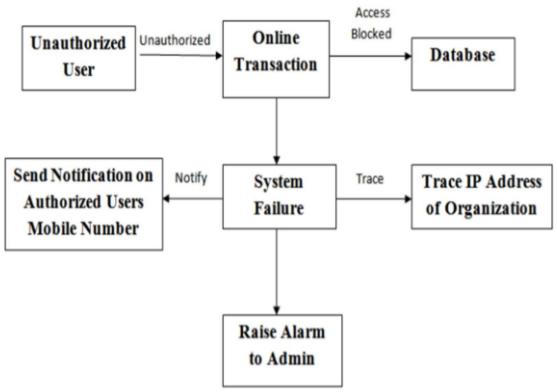


Figure 3.1.2 High-Level Architecture

In this section the working of the system will be explained: There are two modules that will predict which transaction is fraudulent: an e-commerce module and a machine learning module.

We designed a simple e-commerce website for purchasing jerseys, where users must first register, and then log in to the e-commerce website if they have previously done so.

After logging in to the website, the user can add items to their basket and then proceed to the payment gateway to complete the purchase. This card information will be submitted to a machine learning module, which will determine if the transaction is legitimate or fraudulent based on the user's age, merchant id, shopping category, and other factors.

We've included a new module called time, which gives the user two minutes to enter card information and complete the transaction. Basically, the fraudster will need more than 2 minutes to complete the fraud transaction, after which the machine learning module will anticipate that the transaction is fraudulent and deny it.

It's possible that transactions from high-risk zip codes will be blocked.

Blocking transactions from cards used too frequently (for example, in the last 30 minutes) can help discover fraud, but it can also cause false alarms (false positive)

Constraints:

Each rule has a fixed threshold that is difficult to calculate; they do not evolve over time.

- limited to yes/no results, whereas ML generates a probability o probability that can be fine-tuned (i.e. rate of occurrence of false positives and false negatives)
- ML-based systems fail to capture interactions between features, for example, the size of the transaction only important when combined with the frequency.
- Adapt to the data and, as a result, change over time.
- Instead of using a feature-by-feature threshold, this method takes into account all of the data. Rather than a binary score, it generates a likelihood.
- They usually perform better and can be paired with rules.

4. Requirement Analysis

This chapter goes over the precise details of the integration.

4.1 Software

The software's minimum requirements are:

Table 2: Software Details

Operating System	Windows 10 & Above
Programming language	JDK 1.8 Python 3.6 Asp.NET
Database	Jupyter-notebook SQL 2008

4.2 Hardware

The minimum hardware requirement are:

Table 3: Hardware Details

Processor	Dual Core
Ram	2 GB or above
Input Device	Standard Keyboard, Laptop/PC
Hard disk	50 GB or above

4.3 Dataset and Parameters

Finding large and relevant financial data sources is difficult because these data are not made available to the study community due to obvious privacy concerns.

Banksim dataset is one of the datasets utilized in this study. The information includes a credit summary of 1000 accounts with 24 traits or attributes, as well as anonymized detail data. The Banksim dataset is used to detect fraudulent transactions. This artificially generated dataset contains payments from a variety of consumers made over a variety of time periods and in a variety of quantities. This is a tagged dataset in which each account is assigned a positive or negative label (1 or 0).

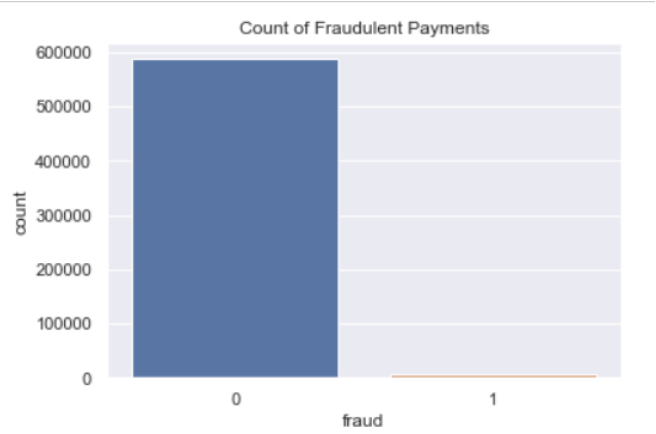
We have done the following steps in the kernel:

1. Exploring Data Analysis:

In this chapter, we'll run an EDA on the data to see what we can learn from it.

Data The dataset comprises 9 feature columns and a goal column, as shown in the first rows below. Customer, ZipCodeOrigin, Merchant, ZipMerchant, Age, Gender, Category, Amount, Fraud are the feature columns.

Data on fraud will be unbalanced, as shown in the plot below and in the count of incidents. Oversample or undersample strategies can be used to balance the dataset. SMOTE is an oversampled technique that we will use (Synthetic Minority Over-sampling Technique). SMOTE will use the neighbor instances to generate new data points from the minority class, thus the created samples are not exact replicas, but they are comparable to the instances we have.




```
Number of normal examples:  587443
Number of fradulent examples:  7200
```

2.Data Processing

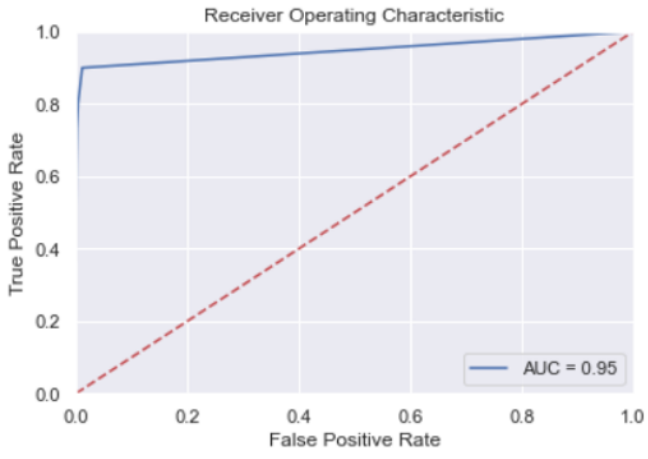
We'll pre-process the data and get ready for the training in this section.

```
Classification Report for K-Nearest Neighbours:
              precision    recall  f1-score   support

           0       1.00      1.00      1.00    176277
           1       0.85      0.73      0.79      2116

    accuracy                           1.00    178393
   macro avg       0.92      0.87      0.89    178393
weighted avg       1.00      1.00      1.00    178393

Confusion Matrix of K-Nearest Neigbours:
 [[176003    274]
 [   565   1551]]
```



3.Oversampling with SMOTE

Using SMOTE(Synthetic Minority Oversampling Technique)

[2] for balancing the dataset. Resulted counts show that now we have exact number of class instances (1 and 0).

```
sm = SMOTE(random_state=42)
X_res, y_res = sm.fit_resample(X, y)
y_res = pd.DataFrame(y_res)
print(y_res.iloc[:,0].value_counts())
```

Now we'll execute a split train test to see how well we did. I haven't done cross validation because we have a lot of cases and don't want to wait too long for training, but most of the time it would be better to do so.

As previously stated, fraud datasets will be unbalanced, with the majority of non-fraudulent incidents. Assume we have the dataset and are always correctly forecasting non-fraudulent transactions. For this dataset, and most others, our accuracy would be nearly 99 percent, owing to the low rate of fraud. Our accuracy is excellent, but we don't find any frauds, therefore it's a pointless classifier. As a result, the overall accuracy should be higher.

As a result, the base accuracy score for completing a detection should be better than forecasting always non-fraudulent.

4. K-neighbours Classifier

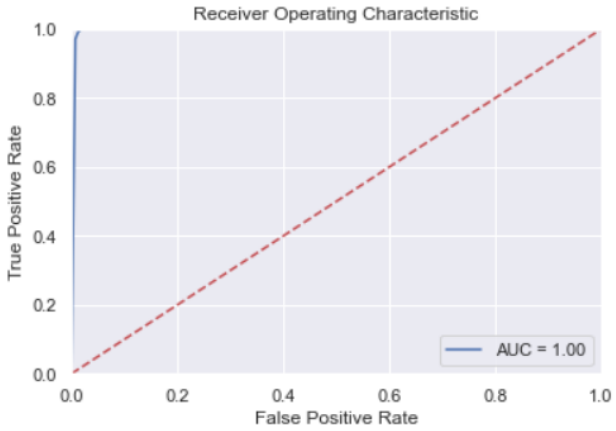
K-nearest neighbor is a term used to describe a person's closest The KNN algorithm is a supervised machine learning technique that can address classification and regression problems.

```
Classification Report for K-Nearest Neighbours:
              precision    recall  f1-score   support

           0       1.00      0.98      0.99    176233
           1       0.98      1.00      0.99    176233

    accuracy                           0.99    352466
   macro avg       0.99      0.99      0.99    352466
weighted avg       0.99      0.99      0.99    352466

Confusion Matrix of K-Nearest Neigbours:
 [[173234   2999]
 [   457 175776]]
```



5. Random Forest Classifier

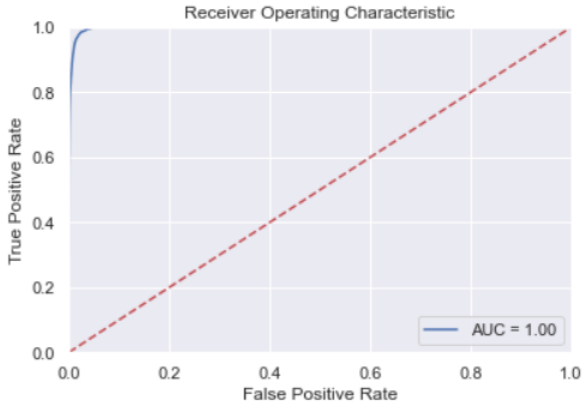
A random forest is a meta estimator that employs averaging to increase predicted accuracy and control over-fitting by fitting a number of decision tree classifiers on various sub-samples of the dataset

```
Classification Report for Random Forest Classifier:
              precision    recall  f1-score   support

           0       0.99      0.97      0.98    176233
           1       0.97      0.99      0.98    176233

    accuracy                           0.98    352466
   macro avg       0.98      0.98      0.98    352466
weighted avg       0.98      0.98      0.98    352466

Confusion Matrix of Random Forest Classifier:
 [[170296   5937]
 [  1510 174723]]
```



6.Conclusion:

We attempted to detect fraud using bank payment data in this kernel, and our classifiers produced impressive results. We used the SMOTE oversampling technique to produce additional minority class cases since fraud datasets have an imbalance class problem.

ACKNOWLEDGEMENT

It is our pleasure to offer our heartfelt gratitude to our supervisor, Prof. Suhas Lawand, for his invaluable contributions, capable leadership, encouragement, wholehearted cooperation, and constructive criticism throughout this project. We are grateful to our Department Head, Dr. Satish Kumar Verma, and our Principal, Dr. Sandeep M. Joshi, for their encouragement and for enabling us to present our study.

REFERENCES

[1] https://community.nasscom.in/wp-content/uploads/attachment/kpmg-fintech-report.pdf

[2] Teruvayi Sai Chandu, Dr. Mooramreddy sreedevi ,Online Transaction Fraud Detections , Tirupati 2020

[3] Azhar Usmani, Internet Fraud Analysis , Western Governor University , Feb-2019

[4] Z. Djuric, Securing money transactions on the Internet, RoEduNet 2004, Timisoara, Romania, may 2004

[5] Nikhil Khandare, Dr. B.B Meshram , Security Of Online Electronic Transactions ,October-2013

[6]https://www.researchgate.net/publication/354937786_Online_Transaction_Fraud_Detection_System_Based_on_Machine_Learning

[7] A. Levi, and C.K. Koc, CONSEPP: convenient and secure electronic payment protocol based on X9.59,Computer Security

[8] S. Buchegger and A. Datta, —A case for P2P infrastructure for social networks— Opportunities and challenges,‖ in Proc. WONS, 2009,pp. 161–168.

[9] C.-H. Park and Y.-G. Kim, "Identifying key factors affecting consumer purchase behavior in an online shopping context," International Journal of Retail & Distribution Management, 2003

[10] https://bfsi.eletsonline.com/18-young-indians-are-bank-fraud-victims-fis-study/

[11] https://www.newsbarons.com/banking-and-finance/financial-cybercrime-increasing-in-india-fis-pace-report/

[12] M.Kannan, "The Face of Digital Frauds in Digital Banking Scenario – A Literature-Based Study", International Journal of Science and Research, e-ISSN:2319-7064, Volume 8, Issue 4, 2019, pp.1645-1647.

[13] ACL, "Fraud Detection using Data Analytics in the Banking Industry", 2014