Enriching E-Commerce Fraud Detection by using Machine Learning

Veena Malik, S. C. Dharmadhikari

Abstract: ***As there has been a proliferation of the internet platform, it has been increasingly getting affordable for a lot of individuals. The rise has been instrumental in achieving several services including the E-commerce platform. This has led to an unprecedented increase in the amount of fraud that is being committed on this platform. The fraud that is being committed on the E-commerce platforms is very different from the frauds committed on other platforms online. Numerous researches have been performed to combat the evils of credit card frauds and money laundering rings. But there is a severe lack of research on the fraud that is committed on the E-commerce platform. Therefore, this research paper defines an innovative approach for the identification of fraud on E-commerce platforms through the implementation of machine learning approaches. The presented technique utilizes Linear Clustering, Entropy Estimation and Frequent itemset mining in addition to the inclusion of Artificial Neural Networks, Hypergraph formation and Fuzzy classification. The implementation of this system will give more security for E-commerce platform-based transactions by identifying fraudulent activities with better efficiency. The methodology has been tested extensively through rigorous experimentation to evaluate the performance metrics which yielded significantly positive results.***

Keywords: Linear Clustering, Entropy Estimation, Frequent Itemset, Hyper graph, Artificial Neural Network, Fuzzy Classification.

I. INTRODUCTION

The E-commerce platform has been rapidly developing in the internet landscape and growing to cover a lot of territories and making the internet an economical hub. This has been largely responsible for the growing number of users on this platform. The E-commerce sale are predictable to achieved $630+ billions by the end of 2020 [1]. This can possible, only if the E-commerce platform ensures smooth functioning which keeps the experience of the users at an optimum level. Along with the various legitimate users, the E-commerce platform is prey to various criminals with malicious intent that are inclined towards criminal activities. These individuals utilize this platform to commit various fraudulent activities that are harmful to the users as well as the owners, or maintainers of the E-commerce platform. Due to the introduction of the internet platform, there has been a large influx of users to capitalize on this platform.

There was a significant increase in the number of applications that were developed to take advantage of the internet platform. Intensive competition and innovation led to the aggressive building of largescale applications and services that started changing the landscape of the internet platform. The innovations played an important part in shaping the internet platform that we see today. The novel inventions have been important in providing an increasing list of services that are enabled specifically due to the prevalence of the internet platform. Most of these innovations were established through extensive research and spending time in active development for years before becoming public. Innovations such as the cloud platform and the recent E-commerce approaches have been in development for a long time. After sound development spanning several years, it has led to the eventual release and widespread success and implementation that we see nowadays. The E-commerce platform offers increased convenience to the users as it allows the users to buy and sell items from the comfort of their homes. This allows a lot of people to be connected through commerce on the online platform. The most benefits are to the elderly and the disabled individuals as it enables ease of use and benefits of accessing items and other products easily without much effort. The prevalence of fraudsters on the E-commerce platform hurts a lot of these individuals as fraudulent transactions tend to decrease trust in the platform of the users. This leads to the E-commerce platform being largely devoid of any users and also hampers the smooth operation for the organization. Therefore, the identification of fraudulent activities and transactions on the E-commerce platforms is imperative and their prevention is dependent on the effective and timely detection. The users also bear the brunt of the frauds as the attackers or the fraudsters target the users predominantly as they tend to be easier targets. Thus, this paper deals with the identification of fraudulent transactions on the E-commerce platform through the use of machine learning frameworks.

Some of common sign to identify the potentially fraudulent transaction are as follows [2]:

a. *Larger than average orders:* Stolen payment cards usually have the short lifespan, so the fraudster main intention is to maximize the spending amount in the one or two transaction which are quite high as compare to normal transaction done by that card.
b. *Multiple transaction in a short amount of time:* This is very important sign of the fraudster. Always the fraudster will try to utilized the maximum amount of the stolen credit/debit card before the card is blocked.

Revised Manuscript Received on August 17, 2020.

Veena Malik, PG Scholar, Department of IT, PICT, Pune, India. E-mail: veenamalik123@gmail.com.
Dr. S. C. Dharmadhikari, Associate Professor, Department of IT, PICT, Pune, India. E-mail: scdharmadhikari@pict.edu

c. *Multiple cards from a single IP Address:* When a single person place a multiple order from same computer (IP address) with multiple cards with different names and their shipping address, then such transaction can be said to be the suspicious transaction.
d. *Fast shipping:* In normal cases we always prefer and select less expensive shipping options. In case of fraudster money is not an issue but will pay the overnight shipping charges for its order.
e. *Unusual location:* Some order is region specific as result transactions that come from countries one usually doesn't get order from.

These are some of the attributes from E-commerce data set which help in identifying the fraudulent transactions. There is no need of a physical card in the scenario of fraud that take place online, only the information of the card is enough for the transaction. Presently, there are two ways of fraud detection [3] as misuse detection and anomaly detection. In the first technique is to collect the database of fraudulent pattern and this database is used to detect the fraudulent transaction. In this type of system supervised machine learning algorithm can be used to first trained the model and then use this model to identify the incoming transaction [4][5]. Where as in second anomaly detection method begin with extracting the user profile of its normal transaction of single user separately which is legal transaction and then calculate the acceptance degree for the new transaction [6]. The central idea of behavior profiling (BP) is that each person is different so their behavior will be different of doing the online shopping. Also, their pattern of doing payment and ordering pattern will also be different. So, in this system we will be using the second misuse detection method rather than anomaly detection method. The methodology proposed in this paper utilizes Linear clustering to cluster the features extracted from the dataset. The dataset is preprocessed to condition the data and reduce the redundancies that are encountered. The preprocessing also inherently increases the accuracy of the consecutive machine learning processes as the cleaned data leads to a fewer number of collisions. The relevant features from the E-commerce transaction dataset are extracted and the resultant data is clustered using the linear clustering approach [7]. This clustering groups similar data which can be analyzed through the use of the Entropy estimation. The Entropy estimation allows the calculation of the information gain through the Shannon information gain [8], which details the entropy of the data and reveals the importance of the data that is clustered using the Linear Clustering approach. The resultant list is transferred for the frequent itemset extraction that is used for the creation of a hypergraph that can be visualized through the Neo4j. The Artificial Neural Network is utilized for the implementation of the machine learning approaches that leverage the creation of neurons and utilizing them in a network modeled after the human brain to reveal fraudulent transactions. The resultant data from the ANN module contains the most relevant and the identified frauds in the E-commerce transaction dataset. The outcomes revealed by the ANN module are further classified using the Fuzzy classification approach. The Fuzzy crisp values are capable of effectively and completely classifying the data through the use of an extensive range of values that provide accurate and effective classification [9]. The fuzzy classification approach utilizes the ANN output to effectively classify and provide E-commerce fraudulent labels. The application of machine learning greatly enhances the procedure that realizes the timely prediction of the fraudulent transactions on the E-commerce platform. This research paper dedicates section 2 for the revival of past works, whereas section 3 narrates the detailing of the proposed model. The obtained results are evaluated in section 4 and finally section 5 concludes this research work along with the scope for the future enhancement.

II. RELATED WORK

S. Surbhi explains that there has been an increased use of the internet services for the purpose of buying and selling a lot of products. These products are paid for in different forms of internet-based money. The transactions are facilitated through the use of the internet platform that utilizes digital money [10]. These transactions are highly critical as they are the source of large number of financial frauds that are committed through this platform. Therefore, the authors in this paper have defined an innovative approach for the evaluation of the various techniques that are utilized for the purpose of performing different frauds on the internet paradigm. The authors have thought upon the utilization of the time series data for this purpose in the future which is the limitation that is observed. R. Rambola states that the banking field has gone through tremendous change over the years due to the fact that this paradigm has a major impact on the day to day life of the average human being. The banking services are utilized by the majority of the individuals for the purpose of saving and other purposes [11]. Due to advancements that are options for the user for the utilization of these services, but there is always a risk of being a target of fraud or stealing that can happen. Therefore, the authors in this approach define an innovative technique that utilizes the data mining approaches to identify fraud in the banking transaction. J. Kingston introduces the paradigm of online fraud that is committed by the fraudsters that are being highly motivated to commit various types of fraud on the internet platform. There is a growing concern over the different techniques that are being utilized by the fraudsters trying to steal and manipulate credit cards and bank account details through the online platform [12]. The authors in this paper have provided an in-depth study into the various reasoning and representations that are essential for committing the fraud through the fraud plans. The authors convey that there are inconsistencies in various fraudulent transactions that are obvious and can be used for effective prediction. O. Elrajubi discusses the predominance of various online approaches for the purpose of performing any types of transactions. These transactions include online E-commerce websites, banking facilities and other transfers that are being increasingly done through the internet platform [13]. This is one of the most popular destinations for fraudsters to target unsuspecting individuals and trap them into a fraud that would extract their hard-earned money from them through nefarious means.

Therefore, the authors in this publication define a novel fraud detection paradigm that utilizes the speaker recognition techniques to effectively eliminate the instance of fraud committed through effective and timely detection. V. Mareeswari narrates that there are a large number of frauds that are committed on the online platforms. This is due to the fact that lot of people are not technologically sound and have not understood the various nuances that are related to the technological advances that are happening in the world. Also, the E-commerce websites are increasing the convenience of buying and selling which lures a large section of users to their platform [14]. This increased user base means a large incidence of credit card fraud being performed in the platform. Therefore, the authors have presented an innovative approach for fraud detection through the use of a Hybrid Support Vector Machine. E. Tarmazakov explains that there has been an increase in the amount of fraud that is being committed online on various E-commerce websites and mobile communication networks. The conventional approaches for providing a solution to fraud consisted of evaluation of subscriber behavior profiles and controlling the integrity of the data flows based on individual events [15]. These prevention approaches have been analyzed and the researchers in this publication have provided various patterns and rules as modern approaches for counteracting fraud in mobile communication networks effectively. X. Min states that fraud is an undesirable outcome that is performed by people with malicious intents that intend to illegitimately gain access and monetary superiority through criminal activities. This approach is highly undesirable as it causes extensive amount of damage in the form of losses to the various users and the provider of the service [16]. Therefore, the authors in this paper provide an effective and secure technique for the detection of fraud through the use of behavior characteristics and deploying principal component analysis and k-means clustering for effectively classifying and identifying fraud based on signaling data. S. Delecourt introduces the paradigm of online payment that is being highly popular and almost the default payment technique used by a lot of users in a plethora of countries and major cities. This is due to the fact that online payment is highly convenient and useful for the end user and allows for secure and fast payments to the seller online. This also increases the incidence of fraud that is being committed on online mobile payment which has led to large scale losses [17]. To provide a solution to this problem the authors in this paper have proposed a mobile payment fraud detection that implements an adversarial engine for the purpose of detecting fraud effectively. A. Kasgari discusses the various rule and techniques that are used by regulatory organizations for monitoring and regulating various prices and other systems related to their field. Regulations are necessary as they provide even ground for the purpose of doing business and providing services to the common people effectively. Therefore, the authors in this paper have analyzed intelligent visual fraud surveillance system for the purpose of identifying price manipulation that is committed by various stock markets and is a form of debilitating fraud [18]. The experimental results conclude in the favor of the intelligent visual fraud detection system as it has a really high accuracy which is observed in the results. M. Zamini narrates the increasing popularity of the E-commerce websites that offer a large variety of products and other services through the online platform. These applications are getting increasingly popular due to the increased convenience that is offered by these approaches such as E-commerce. This is usually combined with online payment to provide a seamless interface for doing business effectively through the internet [19]. These approaches also increase the incidence of fraud that is being effectively harming a lot of users on the platform. For this purpose, the authors have proposed fraud detection technique that utilizes clustering based on autoencoders for faster and effective fraud detection. K. Yang explains that there is a prevalence of financial services that are being improved by technological advancement and performing the various transactions electronically as it is highly convenient and useful for the end user. These practices have completely shifted the financial services to the electronic paradigm which has led to a considerable increase in the amount of fraud that is being committed. Fraud causes insane amount of losses that should be reduced and to reduce this for researchers in this publication have proposed a memory enhanced technique for effective detection of fraud [20]. The main limitation that is being encountered in this publication is that the system requires detailed description of the users which is not been provided by the authors in this publication. B. Omair states that there are various different actors, objects, benefits, decision points, activities and events that comprise a successful business. Fraudsters can have a negative impact on a business and can lead to the generation of obstacles and eventual failure of the business which can be highly debilitating for a market and an individual [21]. Therefore, the incidence of fraud needs to be reduced as it forms a deliberate act that can be highly deceiving and difficult for the business owners. Therefore, the authors in this approach have analyze various fraud detection techniques and produced a taxonomy of fraud detection metrics that can be applied for business processes. N. Malini introduces the concept of online financial services that have been gaining traction significantly in the recent years due to the increased usability that is being offered by this platform to the end users. The online financial services have opened up the Gateway for various E-commerce websites that can provide complete services such as payments along with the purchase of various items directly from one portal [22]. These approaches have been proven to be detrimental as it has promoted economic fraud that takes place with malicious intent on these platforms. Therefore, the authors in this Publication have defined an innovative approach that utilizes outlier detection and KNN for the purpose of effective and fast fraud detection.

I. Benchaji discusses that there has been an increasing incidence of credit card transactions and other forms of online transactions that has been gaining traction in the world nowadays.

This is due to the fact that the online transactions are increasingly useful and effective as well as being highly user-friendly [23].

These transactions are highly suited for fast and immediate transfer to the seller which can be significant in getting various different products at a low rate. But this also increases the incidence of fraud being committed through these payment methods which can lead to a lot of losses. For this purpose, the researcher in this publication have utilized genetic algorithm for improving the detection and classification of fraud by a large margin.

Y. Chen narrates that there have been various difficulties for the purpose of fraud detection in different approaches that are developed for this particular purpose. Most of the fraud detection techniques utilize data mining approaches that perform analysis of the financial statements of an organization to detect any irregularities which can be a sign of fraud being committed [24]. But this is highly limited and could lead to a lot of false positives that can be debilitating for an organization. Therefore, the researchers in this paper have proposed an effective fraud detection technique that implements big data approaches on financial statements of business groups for effectively identifying fraud and reducing the false positives that are generated in conventional approaches.

III. PROPOSED METHODOLOGY

The presented methodology for E-commerce Fraud identification system is shown in the figure 1.

The System consist of steps as follows: 1) Dataset Preprocessing and Feature Extraction, 2) Linear Clustering and Entropy Estimation, 3) Frequent Itemset and Hypergraph Estimation, 4) Artificial Neural Network, 5) Fuzzy Classification. These steps are elaborated below.

Step 1: Dataset Preprocessing and Feature Extraction – This is the first step of the proposed methodology, where an E-commerce transactions dataset has been extracted from the URL-https://www.kaggle.com/vbinh002/fraudecommerce/ data.

The extracted dataset contains several attributes such as user_id, signup_time, purchase_time, purchase_value, device_id, source, browser, sex, age, ip_address and class. The dataset is converted into a workbook format which is then provided as an input to the presented model for fraud detection. The double dimension list is utilized to read the dataset and convert it into a workable format. This is then provided as an input to the preprocessing module.

For the purpose of preprocessing few important attributes are selected and the rest are eliminated or will be utilized later. In this methodology the attributes such as source, browser and class attributes are not considered and instead the rest are stored properly in the form of a list. The preprocessed list that is formed is used to select three attributes which are considered as essential are selected to perform the process of Fraud detection in E-commerce transactions. The three attributes that are required by the proposed methodology are signup_time, purchase_time and purchase_value which are then listed in a separate list.

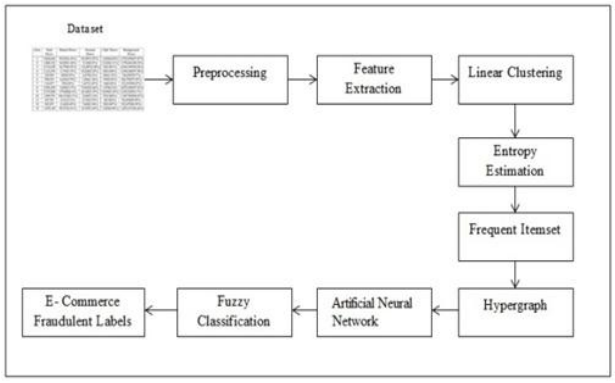

Figure 1: E-commerce fraud detection overview

Step 2: Linear Clustering and Entropy Estimation – The preprocessed list is provided as an input to this step for the purpose of Entropy estimation and cluster formation through linear clustering.

The procedure of linear clustering is performed on the preprocessed list. Firstly, the preprocessed list is divided into 5 equal ranges through the indices. The selected range is then utilized for the formation of the clusters based on the indices of the rest of the items in the list. A final cluster list is then created and the clusters are then added to this list.

The presented methodology estimates the E-commerce fraud in all the obtained clusters through the application of a protocol. Through this protocol all of the clusters are calculated for the number of rows containing the same signup and purchase time and this count is stored as P. S is used to denote the cluster size. The entropy value is subsequently calculated through the implementation of Shannon information gain through the equation given below in 1.

$$IG = -\frac{P}{S}\log\frac{P}{S} - \frac{(S-P)}{S}\log\frac{(S-P)}{S} \quad \text{_____ (1)}$$

where IG = Information Gain of the cluster

The measured value of Information gain of the cluster through the Shannon information gain equation is in the range of 0 to 1. The value closer to 1 illustrates very high importance of the cluster for the purpose of fraud detection. A double dimension list is utilized to store the index of the cluster and respective gain. The resultant list is then sorted in a descending order to achieve the best quality clusters to be sorted to the top. These clusters contain very high-quality data for the procedure of fraud detection.

IV. ALGORITHMS

The complete process of Entropy estimation for the clusters is depicted in the algorithm 1 below.

ALGORITHM 1: Cluster Entropy Estimation

//Input : Cluster List C_L
//Output:Gain Estimation List GE_L
entropyEstimation(C_L)
1: Start
2: $GE_L = \emptyset$
3: ***for*** i=0 to Size of C_L

```
4:    TMP=∅ [Temp list List]
5:    S_G= C_L[i] [ S_G= Single Cluster]
6:    count=0
7:   for j=0 to Size of S_G
8:        R_L=S_G[j]
9:        SIGNUPTIME=R_L[0]
10:       PURTIME=R_L[1]
11:       if (SIGNUPTIME = PURTIME), then
12:       count++
13:       end if
14:  end for
15:   A=count, C=Size of S_G, B=C-A
16:   GAIN=(-A/C) log(A/C)-(B/C)log(B/C)
17:   TMP_[0]=i
18:   TMP_[1]= GAIN
19    GE_L = GE_L + TMP
20:  end for
21:   return GE_L
22: Stop
```

Step 3: Frequent itemset and Hypergraph Estimation – The list generated in the previous step with the top cluster data is then combined to form a single list. Then from this single list a purchase value list is generated through the extraction of all the transactions where signup time and purchase time are the same. A frequent item list is generated by subjecting the Purchase value list to hash set function calculation to get the unique purchase value which is stored subsequently.

The generated frequent item list is utilized to perform insight evaluation of the attribute in a single input list. Through this process the purchase time and signup time are assessed for their equality to obtain the purchase value. The obtained purchase value is correlated with the values in the frequent item list. This is then utilized to generate a hyper graph object containing the purchase value and user ID as the nodes of the graph and "purchase value" as the edge String. The neo4j graph database is utilized to store the resultant hyper graph which can be viewed through the browser.

The complete process of hyper graph formation is shown in the below algorithm 2.

ALGORITHM 2: Hyper Graph Formation

```
//Input : Frequent Item List F_L
//Output: Hyper Graph Object H_G
hyperGraphFormation(F_L)
1: Start
2: H_G =∅
3:   for i=0 to Size of F_L
4:    VAL=F_L[i]
5:    TMP=∅ [Temp list List]
6:   for j=0 to Size of S_G
7:        R_L=S_G[j]
8:        USERID=R_L[0]
9:        SIGNUPTIME=R_[1]
10:       PURTIME=R_[1]
11:       PURVAL=R_L[3]
12:       if (SIGNUPTIME= PURTIME), then
13:       if (VAL = PURVAL), then
14:             TMP[0]=USEID  [NODE 1]
15:             TMP[1]= PURVAL [NODE 2]
16:              E_D = "PURCHASE _ VLAUE" [ E_D = Edge]
17:             TMP[2]= E_D
18:       end if
19:       end if
20:   if (TMP != ∅), then
21:        H_G= H_G +TMP
22:   end if
23:  end for
24:  end for
25:   return GE_L
26: Stop
```

Step 4: Artificial Neural Network- The frequent item list obtained in the previous step is taken as an input for the ANN model. In this step the transaction rows in the frequent item input list are utilized to estimate the hidden and output layers for the attributes such as purchase value and User ID. This is performed through the utilization of the target values for the random weights W1, W2, W3, W4, W5, W6, W7, W8, B1, B2. Here B1 and B2 are considered as the bias values which are implemented for providing stability to the neurons. Through the utilization of the Equation 2 and 3 for the estimation of Hidden layer and Activation function Output layers. The measured value of the output layers are aggregated with the target values to form the new prediction list which is stored as the fraud detection probability list.

$$X = (AT1^* W1) + (AT2^*W2) + B1 \quad \text{_____ (2)}$$

$$H_{LV} = \frac{1}{(1+\exp(-X))} \quad \text{_____ (3)}$$

Where AT1 is the USERID and AT2 is the Purchase Value. Then the sigmoid function is given by Equation 3 of the neural network. H_{LV} – depicts the hidden layer value.

Step 5: Fuzzy Classification – The E-commerce fraud IDs are determined through the application of the Fuzzy Classification on the fraud detection probability list obtained from the prior step. Firstly, the minimum and maximum values are extracted from the fraud detection probability list of ANN. These minimum and maximum values are utilized and a distance is attained in between them as dividend, then this dividend is divided by divisor 5 to obtain the quotient Q. The obtained Quotient is utilized to form five fuzzy crisp values like VERY HIGH, HIGH, MEDIUM, LOW and VERY LOW.

The measured fuzzy crisp values are correlated with the fraud detection probability list according to their respective ranges to evaluate the USER ID. The obtained USER IDs are classified with respect to the fuzzy crisp ranges into different cluster. The resultant classified clusters are indicative of the different levels of fraud right in all the ranges from the VERY LOW to VERY HIGH range, which is provided as an output to the user on an interactive user interface.

V. RESULT AND DISCUSSIONS

The proposed system for E-commerce fraud detection is deployed on a laptop equipped with a Core i5 Intel Processor. The device is also powered with 6GB of primary memory installed with Windows Operating system.

The model uses Java Programming language to develop the system using Netbeans as the standard integrated development environment. An advance graph database neo4j is being used to serve the database responsibilities, which eventually stores the graph efficiently and display the same using the browser in the local host environment.

The proposed model is measured for its performance using the precision and recall parameters. These precision and Recall parameters are one of the most famous and generic way to measure the performance of any model. To understand the precision and recall four measuring entities need to be analyzed first. These measuring entities are True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN).

The True Positive (TP) are measured as the total number of correctly predicted relevant fraud transactions. True negative (TN) is the outcome of the total number of correctly predicted irrelevant fraud Transactions. False Positive (FP) can be defined as the total number of incorrectly predicted relevant fraud transactions. And finally, False Negative (FN) are the total number of incorrectly predicted irrelevant fraud transactions.

In the Proposed model True Negative cases are zero as the system not predicted any irrelevant fraud transactions. And also, the False Negative cases were also zero as there were no incorrectly predicted irrelevant fraud transactions. These two values of True Negative (TN) and False Negative (FN) are zero because the system is trained in such a way that it detects only the positive classes more efficiently. So, system did not commit any mistake of incorrectly predicting the negative class.

So, the precision actually tells what proportion of the positive fraud transactions is actually correct. Whereas Recall can be narrated as what proportion of actual positives was identified correctly. These Precision, Recall and F- Score can be given by the equation 4, 5 and 6.

$$\text{Precision} = \frac{TP}{TP+FP} \quad \text{_____ (4)}$$

$$\text{Recall} = \frac{TP}{TP+FN} \quad \text{_____ (5)}$$

$$\text{F} - \text{Score} = \frac{(2*Precision*Recall)}{(Precision*Recall)} \quad \text{_____ (6)}$$

An experiment is conducted for the various sizes of dataset to measure the accurate fraudulent transaction in the E-commerce dataset. The dataset was labelled for the fraudulent transaction class. So, by comparing with the actual labelled transactions proposed model obtained the true positive (TP) and False Positive (FP) cases in the E-commerce dataset. The obtained results are tabulated in the table 1 and the respective plot is drawn in the figure 2.

Table 1: True positive and False Positive measured cases for the E-commerce Fraud transactions

Data size	Actual	TP	FP	Precision	Recall	F-Measure
1000	50	49	1	0.98	1	0.989899
2000	108	106	2	0.981481	1	0.9906542
3000	158	155	3	0.981013	1	0.9904153
4000	209	206	3	0.985646	1	0.9927711
5000	272	269	3	0.988971	1	0.9944547

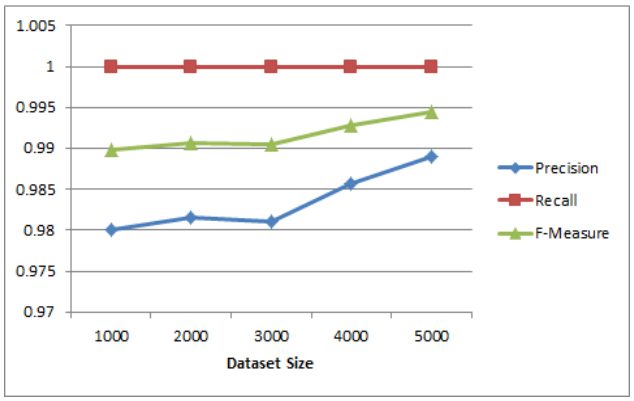

Figure 2: Comparison Precision, Recall and F-Score

The Proposed model yields an average precision of 0.9834, Recall of 1 and F-Score of 0.9916. When proposed model is compared with the other existing system for online E-commerce fraud detection as mentioned in [25]. In [25] E-commerce fraud detection is performed by using a system called ATF (Anti-Fraud System). This is designed based on some dataset by using some methodologies such as User-item bipartite graph, Traffic time series, seed identification and Fraud propagation techniques. The model in [25] also obtains a good score of Precision and recall, but lacking in short with our model of fraud detection, which is using the artificial neural network and a hyper graph technique.

A slight good performance in the proposed model compared with that of [25] is just because the proposed model is equipped with the efficient machine learning model. Whereas the ATF doesn't use any machine learning technique rather than that it used data mining techniques and efficient graph techniques. The readings and the comparison of the proposed model and model in [25] is shown in the table 2 and subsequent plot in figure 3.

Table2: Performace comparison reading between ATF and Proposed system

Parameters	ATF	Proposed Model
Precision	0.9764	0.9834
Recall	0.9785	1
F-Score	0.9774	0.9916

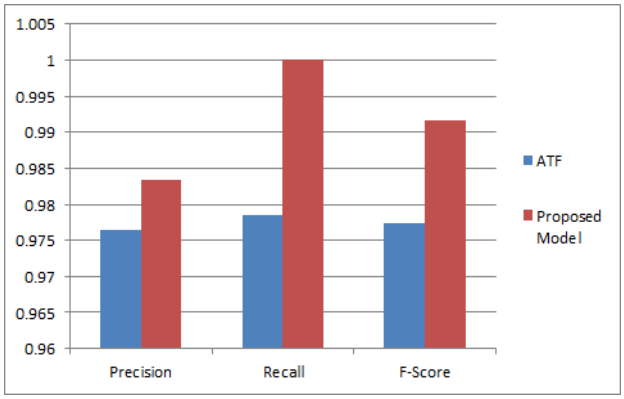

Figure 3: Comparison of Precision, Recall and F-Score

VI. CONCLUSION AND FUTURE SCOPE

The proposed methodology for the purpose of E-commerce fraud detection has been detailed in this paper. Due to technological advances the internet platform has been enhanced with various services that are offered on this platform. E-commerce is one of these services which offer increased convenience to the user, such as for elderly persons. But there have been increased cases of fraud being committed on this platform. Therefore, this methodology utilizes an E-commerce transaction dataset to effectively categorize the fraudulent transactions on this platform. The pre-processed dataset is provided as an input to perform linear clustering, Entropy estimation and frequent item set generation. The resultant clusters are provided as an input to the Artificial Neural Network module which performs the hidden layer estimation to extract the fraudulent transactions. The output obtained from the ANN module is provided to the Fuzzy classification approach to effectively classify the output using fuzzy crisp values. The methodology has been tested extensively through the implementation of rigorous testing parameters. This fraud detection system is stronger towards reducing the rate of false alarms. The outcomes of the experimentation reveal a highly satisfactory realization of a first-time implementation of such a system for E-commerce fraud detection. For future research, the proposed methodology can be improved further by increasing the accuracy considerably through the integration of much more extensive algorithms in the real time E-commerce applications.

REFERENCES

1. Elina Bumbiere, "The basic of Ecommerce Fraud-What it is and how to manage," [Online] Available:http//www.printful.com/blog/the-basic-of-ecommerce-fraud-what-is-it-and-how-to-manage-it/.
2. V. Malik and Dr. S. C. Dharmadhikari, "Analysis of fraudulent Transaction Detection Techniques based on Customers Behavioural Patterns," CIIT International Journal of Artificial Intelligent System and Machine Learning, Vol. 11, No. 12, pp. 219-224, December2019.
3. W.H. Ju and Y. Vardi, "A hybrid high-order Markov chain model for computer intrusion detection," Journal of Computational and Graphical Statistics, vol.10, pp.277-295, 2004.
4. T. Guo, and G. Li, "Neural data mining for credit card fraud detection", International Conference on Machine Learning and Cybernetics, pp.3630-3634, July 2008.
5. R. Chen, S. Luo, X. Liang and V.C.S. Lee, "Personalized Approach Based on SVM and ANN for Detection Credit Card Fraud", International Conference on Neural Networks and brain, pp. 810-815, April 2006.
6. G. Mota, J. Fernandes and O. Belo, "Usage signature and applications analysis an alternative method for preventing fraud in ecommerce applications," Proc. IEEE International Conference on data science and Advance Analytics, pp.203-208, October 2014.
7. N. Kimoto and Y. Endo, "On Linear Clustering with Constraints on Cluster Size," Joint International Conference on Soft Computing and Intelligent Systems and International Symposium on Advanced Intelligent Systems, pp. 832-836, December 2018.
8. G. Ciuperca, V. Girardin and L. Lhote, "Computation and Estimation of Generalized Entropy rates for Denumerable Markov Chains," IEEE Transactions on Information theory vol. 57, July 2011.
9. T. Murata and H. Ishibuchi, "Adjusting Membership Functions of Fuzzy Classification Rules by Genetics Algorithms," Proc. of IEEE International Conference on Fuzzy Systems, 1995.
10. S. Surbhi and Sanjeev Kumar, "Fraud Detection during Money Transaction and Prevention," International Conference on Issues and Challenges in Intelligent Computing Techniques (ICICT 2019).
11. R. Rambola, P. Varshney and P. Vishwakarma, "Data Mining Techniques for Fraud Detection in Banking Sector," International Conference on Computing Communication and Automation (ICCCA 2018).
12. J. Kingston, "Representing, Reasoning and Predicting Fraud using Fraud Plans," International Conference on Research Challenges in Information Science (RCIS 2017).
13. O. Elrajubi, A. Elshawesh and M. Abuzaraida, "Detection of Bypass Fraud based on Speaker Recognition", International Conference on Information Technology (ICIT 2017).
14. V. Mareeswari and G. Gunasekaran, "Prevention of Credit Card Fraud Detection based on HSVM," International Conference on Information Communication and Embedded System (ICICES 2016).
15. E. Tarmazakov and D. Silnov, "Modern Approaches to Prevent Fraud in Mobile Communications Networks," IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (EIConRus 2018).
16. X. Min and R Lin, "K-Means Algorithm: Fraud Detection Based on Signaling Data," IEEE World Congress on Service, 2018.
17. S. Delecourt and Li Guo, "Building a robust mobile payment fraud detection system with adversarial examples," IEEE Second International Conference on Artificial Intelligence and Knowledge Engineering (AIKE 2019).
18. A. Kasgari, M. Taghavifard, and S. Kharazi "Price manipulation fraud detection by Intelligent Visual Fraud surveillance system," International Conference on Control, Decision and Information Technologies, 2019.
19. M. Zamini and G Montazer, "Credit Card Fraud Detection using autoencoder based clustering," International Symposium on Telecommunications, IST, 2018.
20. K. Yang, "A Memory-Enhanced Framework for Financial Fraud Detection," IEEE International Conference on Machine Learning and Applications, 2018.
21. B. Omair and A. Alturki, "Taxonomy of Fraud Detection Metrics for Business Processes," IEEE Access, 2020.
22. N. Malini and M. Pushpa, "Analysis on Credit Card Fraud Identification Techniques based on KNN and Outlier Detection," International Conference on Advances in Electrical, Electronics, Information, Communication and Bio-Informatics (AEEEICB 2017).
23. I. Benchaji, S. Douzi and B. Ouahidi, "Using Genetic Algorithm to Improve Classification of Imbalanced Datasets for credit card fraud detection," Cyber Security in Networking Conference (CSNet), 2018.
24. Y. Chen and C. Wu, "On Big Data-based Fraud Detection Method for Financial Statements of Business Groups," IIAI International Congress on Advanced Applied Informatics, 2017.
25. H. Weng, Z. Li, S. Ji, C. Chu, H. Lu, T. Du, and Q. He, "Online E-Commerce Fraud:A Large-scale Detection and Analysis," International Conference on Data Engineering, IEEE ,2018.

AUTHORS PROFILE

Veena V. Malik, received the B.E. degree from RIEIT College, Shiroda GOA, INDIA. She is currently pursuing the Master in Engineering degree in Information Technology from Pune Institute of Computer Technology, Pune.

Dr. Shweta C. Dharmadhikari did her Ph.D in Computer Science from DAVV, Indore (M.P.) INDIA, M.E. (CSE) from BVDU, Pune, Maharashtra. She is presently working as Associate Professor in Department of Information Technology, PICT, Pune. She has over 35 research papers published in various International/ National Journals and Conferences. Her areas of interest include Machine Learning, Text Mining, Data Science, Systems Software, Natural Language Processing, Object Oriented Programming.