Online Fraud Detection using Deep Learning Techniques

Rakesh R[#1], Suthan R [*2]

[#]*PG Student, CSE Department , Sridevi institute of technology and management, Visveswaraya Technological University*

* *Assistant Professor, CSE Department, Sridevi institute of technology and management, Visveswaraya Technological University*

rakheshrraki@gmail.com, sutrenuka123@gmail.com

Abstract - Fraud has no permanent patterns. They are constantly changing their behavior; therefore, we need to use unsupervised reading. Fraudsters learn new technologies that allow them to commit fraud through online transactions. Fraudsters consider the common behavior of consumers, and fraudulent methods are changing rapidly. Therefore, fraudulent schemes need to detect online transactions through unsupervised learning, as some scammers commit fraud as well as using internet users and switch to other strategies. This paper aims to 1) focus on fraud cases that are not available based on previous history or supervised learning, 2) to create a deep and restricted Auto-encoder model with Boltzmann (RBM) machine that can recreate standard transactions to detect irregularities from common patterns. The proposed in-depth instruction based on auto-encoder (AE) is an unregulated learning algorithm that works on previous applications by setting inputs equal to the results. RBM has two layers, input (visible) layer and hidden layer.

Keywords — Credit Card, Fraud, Autoencoder, Deep Learning.

I. INTRODUCTION

Fraud discovery in online shopping programs is a hot topic these days. Fraud investigators, banking systems, and electronic payment systems such as PayPal must have an effective and sophisticated fraud detection system to prevent fraudulent transactions that are rapidly changing. According to a report by CyberSource from 2017, fraudulent losses to the order channel, that is, the percentage of fraudulent losses on their web store were 74 percent and 49 percent of their mobile channels [1]. Based on this information, the study finds an accumulation of subtle patterns of behavior that have evolved over the past. A good fraud detection system should be able to accurately detect fraudulent transactions and should enable the detection to take place in real-time transactions.

Fraud detection can be divided into two groups: anomaly detection and abuse [2]. Anomaly acquisition programs bring regular transactions to training and applying novel deception techniques. On the other hand, the fraud detection system is misused by labeled operations as usual or to sell frauds to be trained in data history. Therefore, this abuse program includes a supervised learning program as well as a poorly acquired learning program program. What is the difference between supervised learning and supervised learning? The answer is a supervised course with a data label. They use labeled data sets to train and contribute precisely by changing study-level parameters. After that, they used data-level parameters in the database, techniques that initiated supervised learning such as multilayerperceptron (MLP) to create a model based on database history.

This supervised reading is worse, because in the event of a new fraudulent transaction that does not match the database records, then this transaction will be considered true. While, unattended reading gets information from new transactions and detects unfavorable patterns in new transactions. This supervised learning is much harder than supervised learning, because we have to use the right techniques to detect undesirable behavior.

Neural networks were introduced to detect credit card fraud earlier. Now, we focus on in-depth subfield learning in machine learning (ML). Based on in-depth reading in the first instance, they use in-depth learning to know by processing the image. For example, Facebook uses in-depth learning at work to tag people and know who that person is for future reference. In addition, an in-depth study of neural networks has many algorithms used to detect fraud, but in this paper, we have selected AE and RBM to determine whether conventional data transactions are as effective as novel tricks. We believe that some of the most common transactions in data sets that are labeled as crimes also reflect suspicious behavior. Therefore, in this paper we focus on unsupervised reading.

II. RELATED WORK

Ten years ago, a credit card was introduced in the financial sector. Now, credit card has become a popular way to pay for online purchases of goods and services. Since the introduction of credit cards, fraudsters have tried to manipulate users' behavior to pay for themselves. Because of these problems, much research into detecting credit card fraud has focused on pattern matching where unusual patterns are identified as different from standard practice. Many ways to detect credit card fraud have been introduced in the last few years. We will briefly review some of those methods below [3] [4].

K's closest neighbor algorithms are used to detect credit card fraud. This method is a supervised learning method. KNN is used for classification of credit card fraud by calculating the nearest point. If a new transaction comes and the point is close to fraudulent transactions, KNN identifies this transaction as fraud [5]. Many people confuse KNN with K-means integration, whether it is the same strategy or not. K-means and KNN are different. K-means is an uncontrolled learning method, used for integration. K-Means attempts to discover new patterns from data and to combine data into groups. On the other hand, KNN is a number used to compare a nearby neighbor to distinguish or predict new transactions based on past history. The KNN distance between the two data sets can be calculated using a different method, but mainly using the Euclidean range. KNN is very helpful.

External discovery is another method used to obtain both supervised and unsupervised learning. The external vendor acquisition method is researched and categorized externally using a training database. On the other hand, uncontrolled external acquisition is similar to the aggregation of data across multiple groups based on their characteristics. N. Malini and Drs. M. Pushpa pointed out that the method of external acquisition based on unsupervised reading is preferred to detect credit card fraud over supervised external reading, because the unsupervised learning supplier does not require prior data to call data as fraudulent. Therefore, it requires training in the use of standard transactions to discriminate between formal or informal transactions [5].

Some credit card transactions contain database imbalances. Anusorn Charleonnan points out that the inequality of data sets has many facets that occur during the division. It uses RUS, a data modeling process, in an attempt to eliminate the problem of class inequality by arranging the distribution of a class of data training training. There are two major ways to address inequalities in databases, sample reduction and sampling. In his research, he also uses the MRN algorithm for the problem of credit card fraud fraud [6].

The neural artificial network (ANN) is a flexible computer framework used to solve a wide range of non-linear problems. The core concept of ANN mimics the learning algorithm of the human brain. The smallest unit of ANN is called the perceptron, represented as a node. Several perceptrons are connected as networked as the human brain. Each node has a limited connection with many other nodes in a nearby layer. Weight is just a number of floating points, and can be adjusted when the input finally comes to network training. Input is transferred from input nodes with hidden layers to output nodes. Each node can read and edit itself to make it more accurate and relevant [7].

In-depth learning creates the state of the art technology today. Most people in IT should follow this. First, ANN was introduced. After that, ML becomes the basis of ANN, and in-depth learning, the sub-region of ML. In-depth learning has been used in many fields such as image recognition on Facebook, speech recognition in Apple or Siri, and native language processing in Google Translator. Yamini Pandey used in-depth study of the framework of the H2O algorithm to identify complex patterns in the database. H2O is an open source for predictive data predictions on Big Data. Supervised learning is based on predictive analytics. The author has used multi-line H2O, which feeds the neural network to detect credit card fraud patterns. H2O performance based on a deep learning model shows less error than equal square error, root mean less mean, mean total error, and root squared log error. Therefore, these low errors increase accuracy. The accuracy of the model is also high with respect to the errors mentioned above [8]. Another concern before registering a credit card is credit card debt 'judgment.

Ayahiko Niimi uses in-depth reading to judge whether a user should be given a credit card if he or she meets certain procedures. Transactional jurisdiction refers to the validity of trademarks before making decisions. To validate the transaction, the author uses benchmark-based benchmark tests and verifies that the in-depth reading result has the same accuracy as the Gaussian kernel SVM. By comparison, the authors use five standard algorithms and change the parameters of in-depth reading five times, such as performance and drop-down parameter [9].

III. DEEP LEARNING TECHNIQUE FOR DETECT CREDIT CARD FRAUD

In-depth learning is a state of the art technology that has recently attracted the attention of the IT circle. An in-depth learning program is ANN with many hidden layers. On the other hand, the deeper feed for advanced learning in the neural network has only one hidden layer. The picture provided shows a comparison between shallow reading and in-depth reading with hidden layers. Now, we know about ANN, ML, and Deep Learning (DL). When these three terms are equated with the human body by analogy, they can be equated as follows: artificial intelligence is like a body that contains elements of intelligence, reasoning, communication, emotions and feelings. ML is like a single system that works in the body, especially the visual system. Finally, in-depth reading is compared to a visual presentation. It contains many cells, such as the retina that acts as a receptor and translates simple signals into sensory signals. Now, we will compare all three phases with the human body. In-depth reading is a common term used for a multilayer neural network. According to in-depth study, there are many application algorithms such as AE, deep convolutional network, vector support machine, and others. One problem in choosing an algorithm to solve a problem is that the engineer has to know the real problem and what each algorithm is doing in deep learning. Three in-depth learning algorithms enable unregulated learning by RBM, AE, and a small coding model. Unsupported reading automatically removes the logical features of your data, connects wireless data acquisition, and can add training-dependent customization data.

In this study using AE, we use the hyperbolic tangent function or “tanh” function encode and decode the input to the output. As a sample of the neural network, when we have already used the AE model, we have to recreate the error using regression. Backpropagation includes an "error signal", which spreads the error back and forth through a network that starts at the output units by using the condition that the error makes the difference between the actual and desired output values. Depending on the AE, we use parameter gradients to find the return.

Another algorithm is RBM. There are two properties in this algorithm, the visible or input layer and the hidden layer. Each input code takes the input element into the database to be read. The design is different from other in-depth readings, because there is no extraction layer. The RBM effect returns the reconstruction of the input as shown in the image below or in Fig. 4. The point of RBM is how they learn on their own by rebuilding data; this is unregulated reading [12].

IV. PROPOSED METHOD

Autoencoder is an artificial neural network used for unsupervised learning. The purpose of the autoencoder is to study representative representation of data collection features, usually with the aim of reducing size. The simplest method of autoencoder is a feedforward, non-repetitive network similar to a multilayer perceptron [10] [11]. Since the encoder has 2 parts: one encoder and the other is a decoder that contains the input layer, one or more hidden layers and the output layer. The main difference between the autoencoder and the multiplayer perceptron is that the output layer of the autoencoder has the same number of neurons as that of the input. The purpose is to recreate its input instead of predicting the target value of the given input.

In the first phase of the proposed model the autoencoder is trained using transaction signals. The autoencoder is therefore capable of producing modified (embedded) representation of symbols, Z which can be used to retrieve original features. The features represented are smaller than the actual features that make the study of distinction in the second phase easier. In trademark conversion, only the autoencoder encoding network is used. In the second phase, the separator is trained by a labeled transaction in which each action is indicated by Z, the features generated by the autoencoder. For testing, the transaction attribute vector passes through an autoencoder (encoder only) and the corresponding variable vector is provided by a trained configuration network. The model proposed here is standard and any classification can be used in the second phase depending on the needs of the user. Our model was tested using three different dividers to prove the versatility of our model. Classifiers used are Multi-Layer Perceptron, K-Nearest Neighbor and Logistic Regression.

V. RESULTS

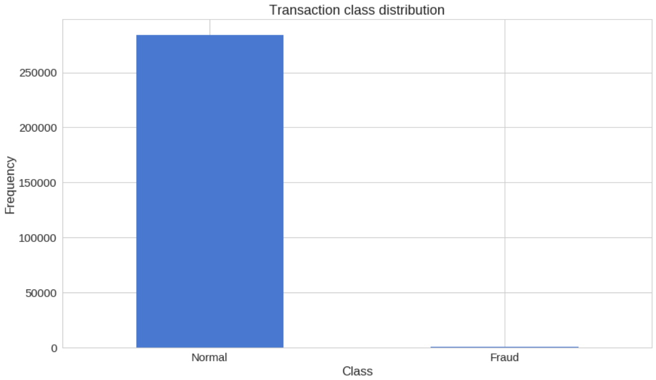

Fig1. Transaction class Distribution

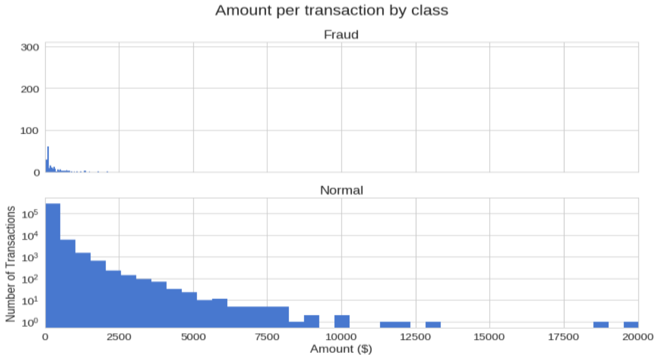

Fig 2: Amount per transaction by class

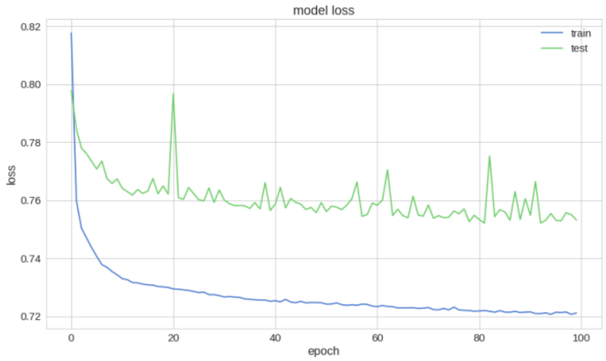

Fig 3: Evaluation of credit card fraud detection

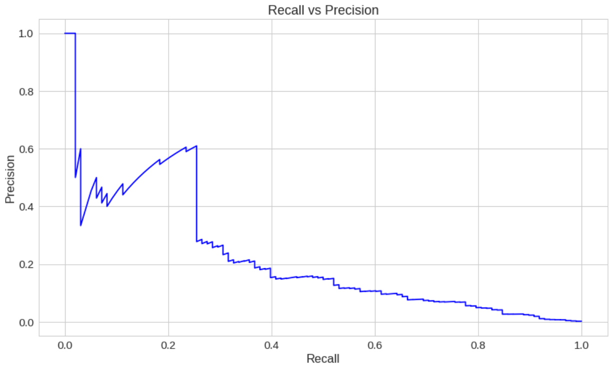

Fig 4: Recall vs Precision

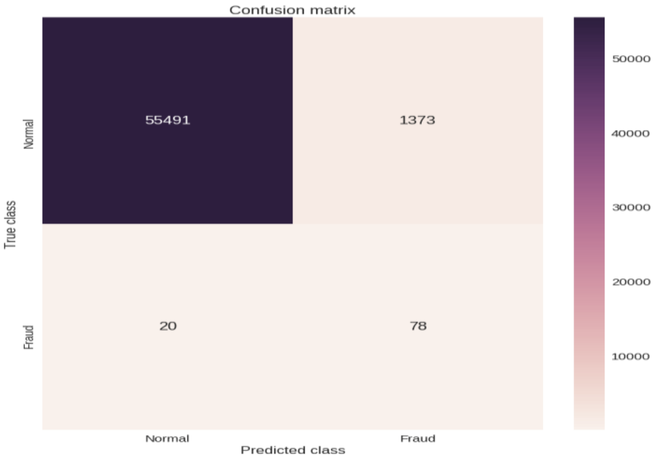

Fig 5: Confusion Matrix

III. CONCLUSIONS

Nowadays, in the case of a global computer, online payments are important, because online payments only use credit card verification information to complete the process and withdraw money. For this reason, it is important to find the best solution for finding the highest amount of fraud on online systems. AE and RBM are two types of in-depth learning that use standard transactions to detect fraud in real time. In this study, we focused on ways to build AE based on Camera, RBM, and H2O. To validate our proposed methods, we have used benchmark testing and other tools to ensure that AE and RBM in deep learning can accurately access credit card acquisition and big data such as the European Dataset. Or, in these tests, it would be better to use a real credit card transaction with a larger amount of data. We confirm that AE and RBM can make the AUC more accurate with the signals of the recipient than reflected in the results from the European Database.

REFERENCES

[1] Neal Leavitt, " Is cloud computing really ready for prime time?," in IEEE Computer Socciety, 2009.

[2] Ogwueleka, F. N., (2011). Data Mining Application in Credit Card Fraud Detection System, Journal of Engineering Science and Technology, Vol. 6, No. 3, pp. 311 – 322.

[3] Rashmi, T. V., and Keshava Prasanna. "Load Balancing As A Service In Openstack-Liberty." International Journal of Scientific & Technology Research 4.8 (2015): 70-73

[4] Maes, S., Tuyls, K., Vanschoenwinkel, B., & Manderick, B. (2002). Credit card fraud detection using Bayesian and neural networks. In Proceedings of the 1st international naiso congress on neuro fuzzy technologies (pp. 261-270).

[5] Shen, A., Tong, R., & Deng, Y. (2007). Application of classification models on credit card fraud detection. In Service Systems and Service Management, 2007 International Conference on (pp. 1-4). IEEE.

[6] Khanum, Salma, and L. Girish. "Meta Heuristic Approach for Task Scheduling In Cloud Datacenter for Optimum Performance." International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) Volume 4

[7] Sahin, Y. and Duman, E., (2011). Detecting credit card fraud by ANN and logistic regression. In Innovations in Intelligent Systems and Applications (INISTA), 2011 International Symposium on (pp. 315-319). IEEE.

[8] Autoencoder for Words, Liou, C.-Y., Cheng, C.-W., Liou, J.-W., and Liou, D.-R., Neurocomputing, Volume 139, 84–96 (2014), doi:10.1016/j.neucom.2013.09.055.

[9] M. Koziarski and M. Woźniak, "CCR: A combined cleaning and resampling algorithm for imbalanced data classification", International Journal of Applied Mathematics and Computer Science, vol. 27, no. 4, 2017.

[10] Pranav T P, Charan S, Darshan M R. (2021). Devops Methods for Automation of Server Management using Ansible . International Journal of Advanced Scientific Innovation, 1(2), 7-13. https://doi.org/10.5281/zenodo.4782271.

[11] Prajwal, S., M. Siddhartha, and S. Charan. "DDos Detection and Mitigation SDN using support vector machine." International Journal of Advanced Scientific Innovation 1.2 (2021): 26-31.