

# Analysis of Credit Card Fraud Detection Technique Using Machine Learning and Deep Learning

Jadhav Nikita B.<sup>1</sup> Prof.Rokade.M.D<sup>2</sup>

<sup>1</sup>Student <sup>2</sup>Professor

<sup>1,2</sup>Department of Computer Engineering

<sup>1,2</sup>Sharadchandra Pawar College of Engineering, Dumbarwadi, India

**Abstract**— nowadays, everything is happening online credit card is very useful in day to life. In recent time all payments are made online using credit cards and Debit cards, so the use of Credit cards and Debit cards has increased and because of that the number of cybercrimes related to transactions are done. In this Paper the Aim is to Analysis techniques for detecting credit card fraud. Latest Technology machine learning and deep learning are implement in this Paper. In Recent Day, Banks has become very difficult for detect the fraud in credit card system. Machine Learning Play very important role for detecting the credit card fraud in the transactions. Credit card fraud detection with machine learning is process of data investigation by a model that will provide the best result to preventing fraudulent transaction.

**Keywords:** Credit card fraud, Machine Learning, Deep Learning

## I. INTRODUCTION

We are Looking an world which is constantly growing up which not at all stable as a service say that 80% to 90% People has switch an online based transaction services. But the major Problem that we come across over is Misuse of user's credentials and Misuse of Personal information, the hacker take advantage of that now it is very common. The credit card Payment method has increased Day-to-Day life. The number of user's in India are reaches to 52 million in 2019. In India there are only 3 card's for every 100 people, When Compare 32 cards in USA.

A credit card is very useful because credit card records each purchase made through. When we do any purchasing then detail list of your monthly credit card is come. This can be used to determine and track our spending and purchasing. Unlike the cash method that is limited to cash on your wallet, But card does not has such limitation. In addition Most companies and institutions now tend to move their business toward online services due to the rapid increase of using modern technology in all fields thus online transactions (e.g. booking a hotels) require a customer to have a credit card to access the services and complete the transaction in such an efficient way that it might be hard and time consuming to perform while using cash payment. However credit cards are the cause of cybercrimes.

Credit card fraud occur when unauthorized person gain access to your information and use it to make purchase. The fraudsters perform fraudulent activities by making unauthorized access such activities cause a financial loss for both company and customer. Unfortunately, there is no fool-proof way to stop hackers from making such attempts, and fraudsters seem to always find new ways to tap into your information. But you can limit your chance of becoming a victim, or minimize the damages from fraudulent activity that's already occurred on your accounts .Credit card fraud

detection technique used to recognize fraudulent credit card transaction so that customers are not charged for items that they did not purchase. In this IEEE Paper we used Machine Learning (ML) is the scientific study of algorithm and statistical Models that computer systems use in order to perform a specific task effectively without using explicit instructions relying on patterns and inference instead.

The Paper structure is as follow: section II cover related work, Section III Problem definition and solution, section IV and V cover the Machine and Deep Learning Technique, section VI cover Dataset, section VII cover Evolution Metrics, section VIII cover Methodology, and section XI cover Result and last conclusion and reference.

## II. RELATED WORK

In this section of IEEE paper, we review some previous work related to the fraud detection. The Literature's have proposed different techniques of Machine Learning and Deep Learning.

- 1) Sam Macs, Karl Tuyls, Bram Vanschoenwinkes and Bernard ManDerick: In this paper we focus on two machine Learning Techniques to the credit card fraud detection Problems: Artificial Neural Network (ANNs) and Bayesian Belief Networks (BBNs). Also Discuss the feed forward Multi-Layer Perceptron's that are interconnected by set of weighted connection.
- 2) S Kiruthika, Sowmyarani C.N: In this paper Introduce machine learning, cloud computing and web service technology. Local outlier factor algorithm and Isolation forest algorithm are used to develop the credit card fraud detection model using unsupervised learning techniques.
- 3) M.unmul safa, R.M.Ganga: In this Paper Machine Learning algorithm are apply on dataset. In this paper investigate the Performance of Logistic regression, K-Nearest Neighbor algorithm on highly skewed data. In this paper the working is implemented in python. The result show optimal accuracy for Naïve Bayes, Logic Regression, and K – Nearest algorithm classifier are 83.00%, 97.69% and 54.86% respectively.
- 4) Apapan Pumsirirat, Liu Yan: In this study, they proposed deep learning based auto-encoder and RBM machine, also they focused on ways to build AE based on Keras, RBM, and H2O.
- 5) Thanh Thi Nguyen, Hammad Tahir, and Mohamed Abdelrazek &Ali Babar: This Paper thought study of deep learning method for credit card fraud detection problem.

### III. PROBLEM DEFINITION AND SOLUTION

#### A. Problem Definition:

Major Problem is that online payment does not require physical card anyone who know the credit card detail can make the fraud transaction. Card holder come to know only after fraud transaction is carried out.

#### B. Problem Solution:

Unfortunately there are no mechanism to track the information about fraud transaction. Logistic algorithm is used to find out fraud transaction. It contains only numerical input variable which are the result of PCA transformation.

### IV. MACHINE LEARNING TECHNIQUE FOR DETECTING THE CREDIT CARD FRAUD

Machine Learning (ML) is subset of Artificial Intelligence. ML algorithms build a mathematical model based on sample Data, Known as "training Data." In order to make predictions or Decisions without being explicitly programmed to perform the task. Machine Learning Model are widely used in detecting credit card fraud. We have applied five Machine Learning classifiers in the Data set which are: K-nearest neighbor, case-based reasoning, Decision Tree, Naïve Bayes, Artificial Neural Networks. On survey bases Naïve Bayes, Logistic regression, K-nearest neighbor are better than other algorithm fraud detection.

#### 1) Naïve Bayes:

Naïve is classification algorithm this algorithm depends upon Bayes thermo. This is simple and very powerful algorithm.

##### a) Bayes Theorem:

In statistics and probability theory, the Bayes theorem is a mathematical formula used to determine the conditional probability of event.

Formula-

$$P(c|x) = \frac{P(x|c)P(c)}{P(x)}$$

Likelihood
Class Prior Probability  
Posterior Probability
Predictor Prior Probability

$$P(c|X) = P(x_1|c) \times P(x_2|c) \times \dots \times P(x_n|c) \times P(c)$$

Where,

$P(c|x)$  - the probability of event  $c$  occurring, given event  $x$  has occurred.

$P(x|c)$  - the probability of event  $x$  occurring, given event  $c$  has occurred.

$P(c)$  - Probability of event  $c$ .

$P(x)$  - Probability of event  $x$ .

#### 2) Logistic Regression:

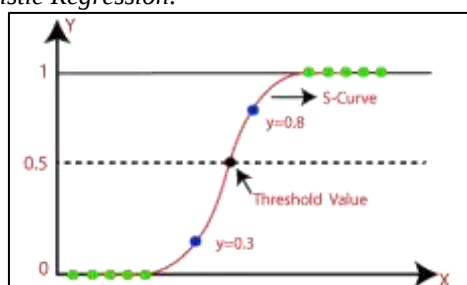


Fig. 1: Logistic Regression

This algorithm is similar to linear regression algorithm.

Logistic algorithm is used for classification of task.

Linear regression is classify as

- Binomial
- Multinomial
- Ordinal

This algorithm is easy for binary and multivariate classification task.

$1/(1 + e^{-value})$

Logistic regression equation:

$Y = e^{(b_0 + b_1 \cdot x)} / (1 + e^{(b_0 + b_1 \cdot x)})$

#### 3) K-Nearest Neighbor Algorithm:

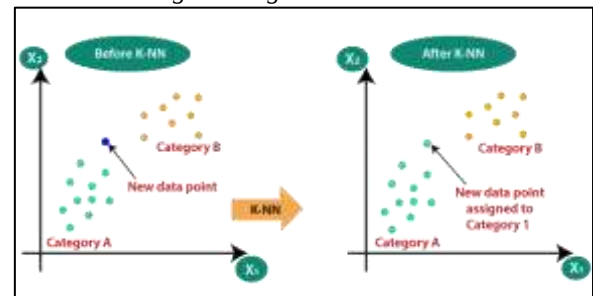


Fig. 2: K-Nearest Neighbor Algorithm

This Concept is used is several detection technique.

One of the best classifier algorithm that have been used in the credit card fraud detection is K-Nearest Neighbor. This algorithm is supervised learning algorithm where result of new instance query is classified based on majority of K-Nearest Neighbor category. The performance of KNN algorithm is categorized in three common factor:

- 1) The Distance Metrics used to locate the nearest neighbor.
- 2) The Distance Rule are used to derive a classification from nearest neighbor.
- 3) The number of neighbor used to classify the new sample

### V. DEEP LEARNING TECHNIQUE FOR DETECTING THE CREDIT CARD FRAUD

Deep learning is subset of machine learning in artificial intelligence that has network capable of learning us supervised from a data that is unstructured or unlabeled also known as deep neural learning and deep neural network.

Deep learning is artificial intelligence function which working on human brain in processing data and creating pattern used in Decision making. Based on deep learning many algorithm are implement such as neural network, convolution network, Auto-encoders.

#### A. Neural network:

Neural network is series of algorithms that mimic the operations of a human brain to recognize relationships between vast amounts of data. They are used in variety of applications in financial services, from forecasting and marketing research to fraud detection.

Neural Network also be referred to as multilayer Perceptron's.

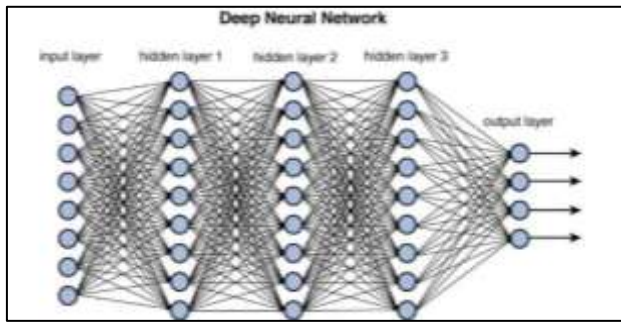


Fig. 3: Neural Network

When to use:

- 1) Tabular dataset formatted in rows and columns (CSV files).
- 2) Classification and regression problems where a set of real values is given as an input.

#### B. Convolutional neural Network:

Convolutional Neural Network (CNN) is built to handle a greater amount of complexity around pre-processing, and computation of data.

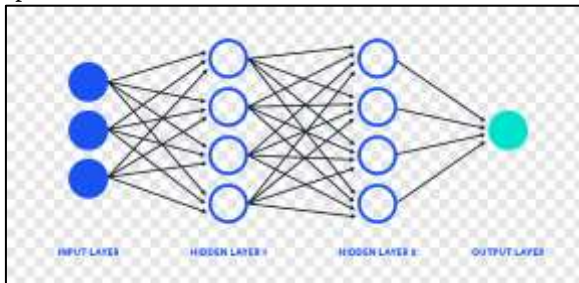


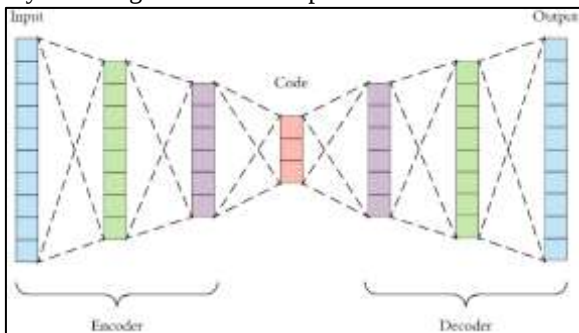
Fig. 4: Convolutional Neural Network (CNN)

When to use:

- a) Image Datasets (including OCR document analysis).
- b) Input data is a 2-dimensional field but can be converted to 1-dimensional internally for faster processing.

#### C. Auto-encoders:

Auto-encoders work by automatically encoding data based on input values, then performing an activation function, and finally decoding the data for output.



When to use:

- a) Dimensionality reduction/Feature detection
- b) Building powerful recommendation systems (more powerful than BM)

## VI. DATASET

The Dataset is used in this Paper to get hands on experience in Data mining which helps to find fraud in credit card transaction. An issue about the dataset we have here is a class

imbalance only 0.17 % of 284,807 transactions are frauds. The model is more sensitive to detect the majority class than the minority class. To solve this problem following technique are applied to the Dataset.

#### A. Random Over Sampling

Random over sampling Duplicates examples from the minority class in the training Dataset and can result in over fitting for some model.

#### B. Random Under Sampling

Random under Sampling Delete examples from a majority class and can result in losing information invaluable to a model.

#### C. Synthetic Minority Oversampling Technique (SMOTE)

SMOTE, initially presented by Nitesh Chawla in 2002, works by selecting Data that are close or similar in the feature space and drawing a line between Data are close to the minority class in the feature space.

## VII. EVALUATION METRICS

To measure, how the proposed model perform we used different Metrics. Since the Dataset used in this paper was highly imbalanced, using the accuracy metrics alone will not be accurate to measure the performance of the model.

Model Evaluation Metrics:

We Discuss Different Evaluation Metrics in the dataset to know the Performance of Metrics.

#### A. Confusion Matrix

A Confusion Matrix is an  $N * N$  matrix. Where  $N$  is the number of classes being predicted. Here some definition you need to remember for confusion Matrix:

##### 1) Precision:

It is simplest terms, Precision is the ratio between the True Positives and all the Positives.

##### 2) Recall:

Recall is the number of True Positives divided by the negatives.

#### B. Area Under the receiver operator characteristics (AUC - ROC)

The Receiver Operator Characteristic (ROC) curve is an evaluation metrics for binary classification Problem. It is Probability curve that plots the TPR against FPR at various threshold values and essentially separates the 'signal from noise'

The AUC is measure of ability of classifier to distinguish between classes and is used as a summary of the ROC curve.

#### C. F1 -Score

F1-Score is the harmonic mean of Precision and recall values for a classification problem. It can calculated as follows:

$$F1\text{ score} = 2 * (\text{precision} * \text{recall}) / (\text{precision} + \text{recall})$$

## VIII. METHODOLOGY

In this paper we implement different machine learning classifier including K-Nearest neighbor, case based reasoning Decision Tree, Naïve Bayes Artificial Neural

Network in the dataset. To overcome the highly imbalanced dataset. We used Random oversampling, Random under sampling, SMOTE. We also used Deep learning technique like neural network, Convolutional neural Network, Auto-encoders, BiLSTM and BiGUE. We used all this methodology which make the fraud detection easy.

#### IX. RESULT

| Name                 | Accuracy | Method                | Inbuilt Packages         |
|----------------------|----------|-----------------------|--------------------------|
| Naïve Bayes          | 83.00%   | Classification method | Gaussian NB              |
| Logistic Regression  | 97.69%   | Machine Learning      | Supervised Learning      |
| K – nearest neighbor | 54.86%   | Logistic Regression   | Decision Tree Classifier |

Table 1: Result of Machine Learning classifiers

| Modelling Methods | BiLSTM | BiGRU |
|-------------------|--------|-------|
| Oversampling      | 88%    | 87.4% |
| Under Sampling    | 89.07% | 90.8% |
| SMOTE Sampling    | 89.7%  | 89.7% |

Table 2: Result From Deep Learning Model Using The Three Sampling Technique.

#### X. CONCLUSION

In the global computing environment, online payments are important, because online payments use only the credential information from the credit card to fulfill an application and then deduct money. Due to this reason it is very important to find the solution to detect the fraud in online system. Accurately predicting the fraud cases and reducing the number of false positives is the foremost priority of a fraud detection systems. Here we used machine learning classifier Naïve Bayes, Logistic Regression, K-Nearest Neighbor Algorithm are better than other algorithm fraud detection. Based on Deep Learning, there are neural network, Convolutional neural Network, Auto-encoders algorithm implement. An issue about the dataset we have here is a class imbalance only 0.17 % of 284,807 transactions are frauds. To overcome that issue three method are used.

#### REFERENCES

- [1] Sam Macs, Karl Tuyls, Bram vanschoenwinkes, Bernard Manderick” credit card fraud detection using Bayesian and Neural Network”.
- [2] S Kiruthika, Sowmyarani C N “credit card fraud detection using machine Learning and Deployment of model in public cloud as a we service” International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-9, Issue-2, July 2020.
- [3] Apapan Pumsirirat, Liu Yan” Credit Card Fraud Detection using Deep Learning based on Auto-Encoder and Restricted Boltzmann Machine” (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 9, No. 1, 2018.
- [4] M. Ummul Safa, R. M. Ganga” Credit Card Fraud Detection Using Machine Learning” International

- Journal of Research in Engineering, Science and Management Volume-2, Issue-11, November-2019.
- [5] Thanh Thi Nguyen, Hammad Tahir, Mohamed Abdelrazek and Ali Babar “Deep Learning Methods for Credit Card Fraud Detection”
- [6] Rishi Banerjee, Gabriela Bourla, Steven Chen, Mehal Kashyap, Sonia Purohit, Jacob Battipaglia “Comparative Analysis of Machine Learning Algorithms through Credit Card Fraud Detection”
- [7] Sunil S.Khatal, Dr. Yogesh Kumar Sharma “Analyzing the role of Heart Disease Prediction System using IoT and Machine Learning” International Journal of Advanced Science and Technology Vol. 29, No. 9s, (2020), pp. 2340-2346
- [8] Divyarani P. Babar and Dr. Pankaj Agarkar “ Identification of Malicious Activity for Network Packet using Deep Learning” International Journal of Current Engineering and Technology E-ISSN 2277 – 4106, P-ISSN 2347 – 5161
- [9] <https://towardsdatascience.com/6-deep-learning-models-10d20afec175>
- [10] <https://www.javatpoint.com/k-nearest-neighbor-algorithm-for-machine-learning>
- [11] <https://towardsdatascience.com/credit-card-fraud-detection-9bc8db79b956>
- [12] <https://spd.group/machine-learning/credit-card-fraud-detection/>