IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 7, April 2023

Credit Card Fraud Detection using Machine Learning

Naga Ashwini Nayak V J[1], C. Suchika[2], N Sandhya[3], M Lakshmi[4], Roja J[5]
Assistant Professor. Department of Computer Science andEngineering[1]
Students, Department of Computer Science andEngineering[2,3,4,5]
Rao Bahadur Y Mahabaleswarappa Engineering College, Bellary, Karnataka, India

Abstract: *Credit card fraud detection is presently the most frequently occurring problem in the present world. We made an attempt for finding the frauds in the credit card business by using the algorithms which adopted machine learning techniques. We are using Decision Tree, Random Forest, and Extreme Gradient boosting algorithms. The efficiency of the model can be decided by using some public data as sample. Then, an actual world credit card facts group from a financial institution is examined. Along with this, some clatter is supplemented to the data samples to auxiliary check the sturdiness of the systems. The significance of the methods used in the paper is the first method constructs a tree against the activities performed by the user and using this tree scams will be suspected. In the second method a user activity-based forest will have constructed and using this forest an attempt will be made in identifying the suspect. The investigational outcomes absolutely show that the mainstream elective technique attains decent precision degrees in sensing scam circumstances in credit cards.*

Keywords: XG-Boost, K-Nearest Neighbor (KNN), Decision Tree, Logistic Regression, Support Vector Machine (SVM)

I. INTRODUCTION

Credit card is a small thin plastic or fiber card that contains information about the person such as picture or signature and person named on it to charge purchases and service to his linked account charges for which will be debited regularly. Now a days card information is read by ATMs, swiping machines, store readers, bank, and online transaction. Each card as a unique card number which is very important, its security is mainly relies on physical security of the card and privacy of the credit card number.

As we are moving towards the digital world cyber security is becoming a crucial part of our life. When we make any transaction while purchasing any product online a good amount of people prefers credit cards.

The credit limit in credit cards sometimes help us making purchase even if we do not have the amount during that period, but these features are misused by cyber attackers.

There exists a number of mechanisms used to protect credit cards transactions including credit card data encryption and tokenization [1].To tackle this problem, we need a system that can abort the transaction if it finds fishy.

Here, comes the need for a system that can track the pattern of all the transactions and if any pattern is abnormal then the transaction should be aborted.

Today, we have many machine learning algorithms that can help us classify abnormal transactions. Machine Learning (ML) is a sub-field of Artificial Intelligence (AI) that allows computers to learn from previous experience (data) and to improve on their predictive abilities without explicitly being programmed to do so [2]. In this work we implement Machine Learning (ML) methods for credit card fraud detection. Credit card fraud is defined as a fraudulent transaction (payment) that is made using a credit or debit card by an un authorised user [3]. One of the key issues with applying ML approaches to the credit card fraud detection problem is that most of the published work are impossible to reproduce. This is because credit card transactions are highly confidential. Furthermore, credit card fraud detection is a challenging task because of the constantly changing nature and patterns of the fraudulent transactions [4]. Additionally, existing ML models for credit card fraud detection suffer from a low detection accuracy and are not able to solve the highly skewed

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 7, April 2023

nature of credit card fraud datasets. Therefore, it is essential to develop ML models that can perform optimally and that can detect credit card fraud with a high accuracy score.

This research focuses on the application of the following supervised ML algorithms for credit card fraud detection: Decision Tree (DT) [5], Random Forest (RF) [6], Artificial Neural Network (ANN) [7], Naive Bayes (NB) [8] and Logistic Regression (LR) [9].

II. LITERATURE SURVEY

Many researches have been carried out for analyzing which method of treatment is best for the mental illness. Some of the them are as follows.

In paper [1] Y. Sayjadah, et al(2020)they have used various machine learning techniques to predict credit card fraud in bank system that based on the analysis of the results. They have proposed random forest which has prediction accuracy is more than 80%. According to them banks can use machine learning to measure credit risk of customers before surrendering them credit card. Banks main worry in to offer treasured harvests and facilities to their consumers and in order save up with their contestants they must stay advanced and creative.

In paper [2] Randhawa, Kuldeep, et al (2018)they have presented credit card fraud detection by using machine learning algorithms. Some typical models that are NB, SVM, and DL have used in the empirical study. They have proposed the best MCC score is 82% that is achieved by vote. Additional assess the hybrid mockups, noise from10% - 30% has been additional into the data models.

In paper [3] Sarah Alexandria Ebiaredoh-Mienye, et al (2020) machine learning algorithms are ineffectual for large datasets performing classification such as large credit card data set. They have proposed stacked sparse auto encoder network to gain optimal features learning. They have introduced batch normalization methods to increase the outcomes and speed of the model and further prevent over fitting. They model was optimized by using Adamax algorithm.

In paper [4] Somayeh Moradi, et al (2019) they have proposed a dynamic model for credit card fraud risk to valuation that outperformance the model used. There model has a self-motivated appliance that evaluates the behavior of corrupt clients in a once-a-month basis, credit risk that include the fuzzy factors, particularly in the financial crises. Their approach can utilize changing indeterminate issues.

In paper [5] Vaishnavi Nath Dornadula, et al (2019) they have used novel method to identify credit card fraud detection. Various classifiers are used on three altered collections advanced assessment scores are produced for each type of classifier. These self-motivated variations in strictures lead the organization to familiarize system. They have proposed that decision tree, random forest and logistic regression provided the best results and accuracy.

III. EXISTING SYSTEM

Credit card fraud recognition is a problematic issue that becomes the attention of Machine Learning researchers and scientists. Nevertheless, the issue is still challenging for credit card data which suffer from class inequity as no fraud transactions over powering succeed fraud transactions making it tough for numerous machine learning algorithms to achieve good accuracy and performance, an upright illustration can be erudite from the dataset that increase the classification performance of the machine learning techniques. Machine learning is a thinkable resolution to the challenge of credit fraud prediction because of its extraordinary feature learning aptitude in large and unstable datasets. In this work, aim of this paper to classify and categorize a well understanding between the different kinds of machine learning techniques to detection of credit card fraud/default that are continuing presently happing in this modern era. In this work the author attempts to suggestion current machine learning techniques to gain improved performance outcomes. There are the following four machine learning techniques are used in this paper to predict the credit card fraud detection accuracy namely, Logistic Regression, Naïve Bayes, and Random Forest. Dataset of credit card transactions is sourced from European cardholders containing 284,807 transactions. This paper discussed the results of the modern methods and it will predict the result for credit card fraud.

IV. PROPOSED SYSTEM

In the proposed system, Dataset is collected from the Kaggle website. The dataset is trained and tested using the following techniques: logistic regression, decision trees, svc, xgboost and adaboost. If our algorithm is applied into

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 7, April 2023

bank credit card fraud detection systems, the probability of fraud transactions can be predicted soon after credit card transaction occurs. Thereafter a series of anti-fraud strategies can be adopted to prevent banks from great losses and reduce risks.

V. OBJECTIVES

- This model is used to identify whether a transaction is fraud or not.
- Our aim here is to detect 100% of the fraud transactions while minimizing the incorrect fraud classifications.

VI. METHODOLOGY

Support Vector Machine Algorithm :

Support Vector Machine Algorithm is on of the supervised Machine Learning. It is useful for solving both regression and classification problem statement. Now let us consider the classification problem to just understand the Geometrical intuition since it is a classification problem here we can easily separable two classes points. We can separable these points with hyperplane. SVM make sures that when we are creating hyper path plane apart from that it also creates two margin lines these two margin lines will having some distance so that it will be linearly separable for both the classification points. SVM is also used for Regression problems to maintain the main features that the algorithm is characterize. Regression problem is same as the classification problem with having only having minor changes.

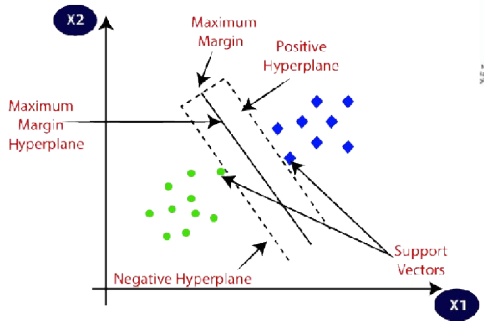

Fig. 1. Support Vector Machine Algorithm

Logistic Regression Algorithm :

Logistic Regression is one of the simple and commonly used for Machine Learning algorithms for two-classes classification. It predicts the probability of a binary event utilizing a logic function. Binary event means it will identify person is having disease or not. Logistic Regression is regression model to predict probability for given data entry belongs the category number one just like linear regression. Using sigmoid function logistic regression models the data and it provide constant output. It is used to predict the probability of dependent variable. These dependent variable has only two necessary classes. In dependent variable data is coded as either 1 or 0.

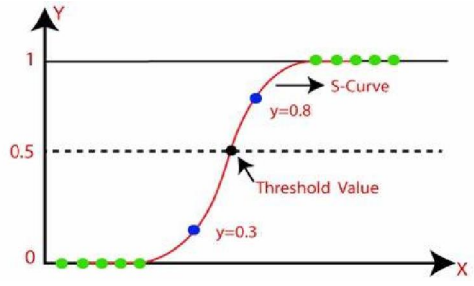

Fig. 2. Logistic Regression Algorithm

$$log\left[\frac{y}{1-y}\right] = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + \cdots + b_nx_n$$

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 7, April 2023

ADABoost:

AdaBoost is short for Adaptive Boosting. Basically, Ada Boosting was the first really successful boosting algorithm developed for binary classification. Also, it is the best starting point for understanding boosting. Moreover, modern boosting methods build on AdaBoost, most notably stochastic gradient boosting machines.Generally, AdaBoost is used with short decision trees. Further, the first tree is created, the performance of the tree on each training instance is used. Also, we use it to weight how much attention the next tree. Thus, it is created should pay attention to each training instance. Hence, training data that is hard to predict is given more weight. Although, whereas easy to predict instances are given less weight.

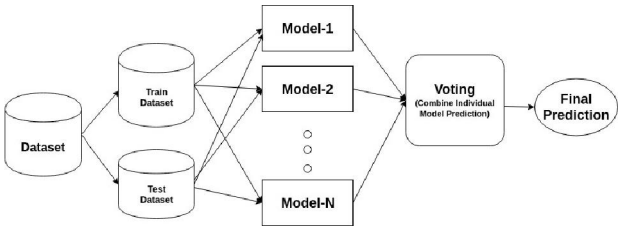

Fig. 3. ADABoost

XGBoost:

XGBoost or the Extreme Gradient boost is a machine learning algorithm that is used for the implementation of gradient boosting decision trees. When we talk about unstructured data like the images, unstructured text data, etc., the ANN models (Artificial neural network) seems to reside at the top when we try to predict. While when we talk about structured/semi-structured data, decision trees are currently the best. XGBoost was basically designed for improving the speed and performance of machine learning models greatly, and it served the purpose very well. The XGBoost is having a tree learning algorithm as well as linear model learning, and because of that, it is able to do parallel computation on a single machine. This makes it 10 times faster than any of the existing gradient boosting algorithms. The XGBoost and the GBMs (i.e. Gradient Boosting Machines) uses tree methods by using the gradient descent architecture. The area where XGBoost leaves the other GBMs behind is the area of system optimization and enhancements over the algorithms.

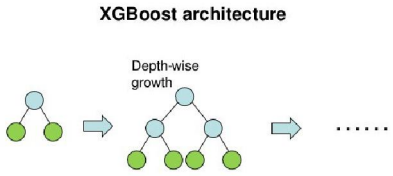

Fig. 4. XGBoost

Decision Tree Algorithm:

This represents the given data into a tree like structure in a hierarchal structure which consists of N number of nodes which is also called as attributes, there are some directed links or edges so that it establishes a relation between all the entities available. A data set divided into large number of rows and columns where last column is referred as a class and the rows or tuples are referred as instance and other columns are called as attributes or features. Decision tree has a primary node or a root node which has no incoming edges and it has two or more number of out-going edges. At the end of the tree there are leaf nodes or terminal nodes which has incoming edges but no out-going edges. And there are some nodes in between root and leaf nodes which are called as internal nodes it has exactly one incoming node and several out-going nodes. Edges consists of conditions known as attribute test conditions. This algorithm is represented in the form of a linked list in memory map. This tree is traversed in the left to right format.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 7, April 2023

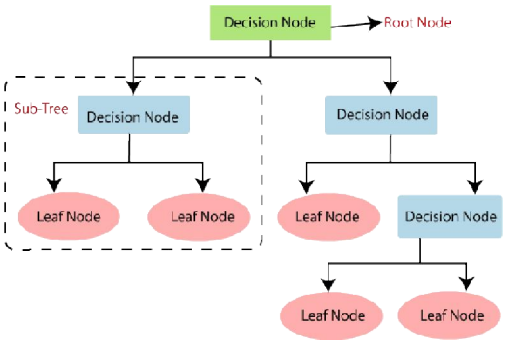

Fig. 5. Decision Tree Algorithm

VII. PROBLEM STATEMENT

- The Credit Card Frauds are increasing heavily so fraud financial loss is increasing drastically.
- Every year due to fraud Billions of amounts lost.
- As online payment does not require physical card, anyone who knows the details of card can make fraud transactions.

VIII. EXPECTED OUTPUT

The proposed system has credit card fraud detection dataset which is used for classified whether the fraud is happened or not according to their features. The overall records in the dataset are distributed into two main category training and testing datasets. The proposed system applied on this data and tries to create accurate model which predict accurate results. In this proposed system, used Logistic Regression, decision tree , SVC , adaboost and xgboost algorithms. Finally analyze the outcomes by the help of Comparing Models and Confusion Matrix. In the field of machine learning, a confusion matrix, normally known as an error matrix, is a specific table design that authorizations perception of the implementation of a calculation. Each line of the matrix expresses to the instances in a predicted class while every section speaks to the cases in a real class.After the accessibility of the data, created a predictive model that is bases on Decision tree algorithm and this classified data based on various organized features of credit card fraud detection. After the complete data accessibility predicts the all five algorithms one by one and obtains their accuracy.

IX. CONCLUSION

Machine Learning is remarkable in enhancing the perception and discovery of mental condition of people, including public health, therapy, preventive medicine and assistance, has demonstrated initial positive results.

In this project, various machine learning classification techniques and methods are used to analysed and predict the accuracy of credit card fraud detection. Anti-fraud approaches can be adopted to prevent banks from major damages and minimize threats. The objective of the study was taken differently than the typical classification problems in that we had a variable misclassification cost. Four machine learning algorithms namely Logistic Regression, Decision tree, SVC, adaboost and xgboost are compared in terms of accuracy using the credit card fraud detection dataset. By the experimental outcomes and results it's evident that Decision tree algorithm predicts the credit card fraud detection with the accuracy of 99.18 % and also with good precision rate. Banks can make the most of the machine learning techniques which can contribute in boosting their performance and image in the industry.

REFERENCES

[1] Sarah Alexandria Ebiaredoh-Mienye, Ebenezer Esenogho and (2020). Effective Feature Learning using Stacked Sparse Autoencoder for Improved prediction of Credit Card Default. Effective Feature Learning using Stacked Sparse Auto-encoder for Improved prediction of Credit Card Default.

[2] Somayeh Moradi and Farimah Mokhatab Rafiei (2019). A dynamic credit risk assessment model with data mining techniques: evidence from Iranian banks. Moradi and Mokhatab Rafiei Financial Innovation, springer. doi.org/10.1186/s40854-019-0121-9.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 7, April 2023

[3] Vaishnavi Nath Dornadula and Gheeta S (2019) Credit Card Fraud Detection using Machine Learning Algorithms. International Conference On Recent Trends In Advanced Computing. Pp 631-641. doi-10.1016/j.procs.2020.01.057.

[4] Devi Meenakshi, Janani, Gayathri (2019). Credit Card Fraud Detection Using Random Forest. International Research Journal of Engineering and Technology (IRJET). Vol 6(3).