Credit Card Fraud Detection Using GBC Algorithm.

[1]Ms.R M Suganya, [2]J Rajasri

[1]Assistant Professor 2, [2] UG Student
[1]Information Technology,
[1]KLN College of Engineering, Pottapalayam, India

Abstract: Credit cards are now possibly the most popular method of payment for both offline and online purchases thanks to new developments in electronic commerce systems and communication technologies; as a result, there is significantly more fraud associated with such transactions. Every year, fraudulent credit card transactions cause businesses and consumers to lose a lot of money, and con artists are constantly looking for new tools and techniques to perpetrate fraud. The increased use of electronic payments is now significantly impacted by the detection of fraudulent activities. As a result, methods that are efficient and effective for identifying fraud in credit card operations are required. This study suggests a method for using an improved Gradient Boosting classifier to identify theft in credit card transactions. Using a gradient-boosting classifier that has been optimized, we suggest a clever method for spotting theft in credit card transactions. According to the suggested methodology, classification systems have become increasingly effective at detecting credit card fraud. The research makes a contribution to the use of various classification algorithms for credit card fraud detection with various kinds of data set. Gradient Boosting Classifier is one of the classification methods. The f1 score, accuracy, precision, recall, sensitivity, and specificity are used to determine the categorization outcome.

Keywords: Gradient Boosting Classifier Algorithm (GBC), Classification algorithm,Fraud detection,credit card.

I. INTRODUCTION

Credit cards became one of the most widely used payment methods as technology developed and e-commerce services grew, which led to a rise in the number of banking transactions. In addition, the significant rise in fraud necessitates expensive banking transactions. As a consequence, uncovering fraudulent activity has grown in interest. In this study, we examine how to balance the weight of fraudulent and legitimate transactions using class weight-tuning hyperparameters.

Accurate fraud detection has become increasingly important in securing such transactions as a result of company migration to the Internet and the electronic financial transactions that take place in the expanding cashless economy. When a thief makes purchases using someone else's credit card number without that person's consent, it is called credit card theft. The widespread use of credit cards combined with inadequate security measures results in billion-dollar loses due to credit card fraud. It is challenging to estimate the losses precisely because credit card companies are usually reluctant to disclose such information. There are, however, some openly available statistics about the monetary losses brought on by credit card fraud. Losses of billions of dollars are a result of using credit cards without adequate protection. Global financial losses as a result of credit card fraud totaled $22.8 billion USD in 2017 and are anticipated to keep rising; by 2020, the sum is anticipated to reach $31 billion USD.

Application fraud and behaviour theft are the two types of credit card fraud. Fraudulent credit card applications are referred to as application theft. A new credit card procedure is started fraudulently using false identity information, and the issuer approves the request. After a credit card is legitimately granted, behaviour fraud describes credit card transactions that involve dishonest behaviour. For both cardholders and banking institutions, credit card fraud detection has been a major problem. Credit card fraud has also grown to be a significant issue for researchers because stopping even a small number of fraudulent activities would safeguard large sums of money.

For a number of reasons, including the fact that a very tiny portion of all credit card transactions are fraudulent and the fact that the distribution of data changes over time as a result of new attack techniques and seasonality, fraud detection is regarded as a challenge for machine learning.

An optimised gradient classifier boosting machine modal is used in this article to suggest an intelligent method for identifying fraudulent credit card transactions. In the suggested method, this aids in lowering variance and prejudice in a machine learning ensemble. The main focus of the suggested method is separating valid from fake credit card transactions.

II. RELATED WORKS

1. M R Dileep, A V Navaneeth, M Abhishek, "A Novel Approach for Credit Card Fraud Detection using Decision Tree and Random Forest Algorithms" vol. 44, no. 5, p. 98, 2021.

Two algorithms are used viz Fraud Detection in credit card using Decision Tree and Fraud Detection using Random Forest. 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.

2. M. Devika, S. Ravi Kishan, L. Sai Manohar, N. Vijaya, "Credit Card Fraud Detection Using Logistic Regression," , 2022.

As digitalization and online transactions are getting improving day by day the usage of credit card is getting increased. Credit card frauds are the most frequently happening problems in present society. Majorly this type of frauds occurs whenever the

fraudster was using some another user credit card information. The system which can be used in order to detect the frauds in the credit card transactions is called credit card fraud detection system. In this project we will use various types of machine learning models and algorithms for identifying the fraud in the transactions. Here algorithm such as logistic regression are considered. By selecting the algorithm with best high accuracy, the fraud will be detected. The logistic regression algorithm accuracy will be nearer to 0.99% with this accuracy we can easily identify the frauds. In this web application the admin can login and the authentication will be performed, then the admin can easily upload a dataset for identifying the frauds in the given dataset.

3. Sasmita Kumari Pradhan, N V Krishna Rao, N M Deepika, Putta Harish, MA Pawan Kumar, Prem Sai Kumar, "Credit Card Fraud Detection Using Artificial Neural Networks and Random Forest Algorithms", 2021.

Credit Cards are one of the most well-built payment type which allows the customers to make the transactions easily and increases the purchasing power. It also offers various advantages or profits like cashback; reward points and it is very easy to use. But the main concern is Credit Card Fraud, it is increasing day by day rapidly. Credit card fraud continues to be the most popular form of identity theft worldwide. Now-a-days it has become too risky to use Credit Cards. Actually there are various types of Credit Card Frauds like Phishing and Vishing, Key Stroke Logging, POS Fraud, Application Fraud, Loss of Card or Theft. However safety measures and precautions should be taken by the individual in order to protect his/her money. It is a big business and there are organized crime rings that are behind the vast majority of this fraud their operations are industrialized, they're automated. So how to detect these frauds ? The answer is by using Machine Learning Algorithms. There is also an other way Conventional Fraud Detection, but Machine Learning Algorithms are far more accurate and precise. The main goal is to build a model which predicts whether a Transaction is fraud or not. In this project, several predictive models like Artificial Neural Networks, Random Forests, Support Vector Machine, K-Neighbors, Decision Tree, Gaussian Naive Bayes and Logistic Regression are used. The results of all these models are compared based on accuracy and the superior one is determined

4. Prabhat Singh, Vishesh Chauhan, Shivam Singh, Priya Agarwal, Shrey Agrawal, "Model for Credit Card Fraud Detection using Machine Learning Algorithm" , 2021

Master Card organizations can recognize fake transaction so clients do not pay for things that they didn't buy. Such issues may be handled with Machine Learning. This undertaking means to outline the demonstration of an informational collection utilizing AI with Credit Card Fraud Detection. This Fraud Detection issue incorporates displaying previous credit card transaction with information of the ones that ended up being extortion. Our model is then used to check if other purchase/order is fraudulent. The aim is, to distinguish maximum of deceitful transactions and limiting the erroneous misrepresentation arrangements. In this cycle, we have zeroed in on dissecting and pre-handling informational collections just as the sending of numerous inconsistency location calculations, for example, SVM, logistic regression, KNN and random forest calculation on the PCA changed Credit Card transaction information.

III. PROBLEM STATEMENT

To effectively predict whether a credit card transaction is fraudulent or not and implement different machine-learning algorithms for better performance and to enhance the overall performance of classifications using the Gradient Boosting Classifier algorithm.

IV. METHODOLOGY

In the proposed system, we uses Gradient Boosting Classifier algorithm for fraud detection. It repeatedly selects a function that leads in the direction of a weak hypothesis or negative gradient so that it can minimize a loss function. It attempts to reduce the chance of overfitting(gives accurate predictions for training data but not for new data) complex models.A decision-tree-based ensemble machine learning method called the gradient boosting classifier makes use of the gradient boosting framework. in unstructured data forecast issues. All other algorithms or systems typically perform worse than artificial neural networks. However, decision tree-based methods are currently thought to be best-in-class for small- to medium-sized structured/tabular data.

V. IMPLEMENTATION

Gradient Boosting Classifier Algorithm

Gradient Boosting is a machine learning technique that boosts the effectiveness, precision, and interpretability of a model by using a group of poor learners. These students are considered to perform better than chance alone. These models, which are usually decision trees, combine their outputs to produce better final outcomes. The idea is to eliminate situations that are challenging to anticipate with precision and train new, weak learners to deal with them. Predictions are made using the entire dataset after the original model has been trained. The difference between the real value and the prediction is calculated, and the incorrect predictions are given more weight. A new model is then developed in an effort to correct the flaw in the prior model, and this process is repeated for several models. By weighing the mean of each model, we create the ultimate model.

Data Set and Data Preprocessing :

The Credit Card Fraud Detection dataset, which was published on Kaggle and includes credit card transactions from two days , was used for this project. The feature time in the dataset displays the number of seconds that passed between each transaction and the dataset's initial transaction. The feature amount, which includes the transaction amount and the feature class, informs us whether a specific transaction is legitimate or fraudulent, with 1 denoting the former and 0 the latter and etc.

A	B	C	D	E	F	G	H	I	J	K	L	M	N	
Id	over_draft	credit_usage	credit_history	purpose	current_balance	Average_Credit_Balance	employment	location	personal_status	other_parties	residence_since	property_magnitude	cc_age	othe
1	<0	6	critical/other	radio/tv	1169	no known savings	>=7	4	male single	none	4	real estate	67	none
2	0<=X<200	48	existing paid	radio/tv	5951	<100	1<=X<4	2	female div/dep/r	none	2	real estate	22	none
3	no checking	12	critical/other	education	2096	<100	4<=X<7	2	male single	none	3	real estate	49	none
4	<0	42	existing paid	furniture/	7882	<100	4<=X<7	2	male single	guarantor	4	life insurance	45	none
5	<0	24	delayed previ	new car	4870	<100	1<=X<4	3	male single	none	4	no known property	53	none
6	no checking	36	existing paid	education	9055	no known savings	1<=X<4	2	male single	none	4	no known property	35	none
7	no checking	24	existing paid	furniture/	2835	500<=X<1000	>=7	3	male single	none	4	life insurance	53	none
8	0<=X<200	36	existing paid	used car	6948	<100	1<=X<4	2	male single	none	2	car	35	none

B. Feature Selection:

When there are many characteristics, choosing significant and important ones is essential for the efficient detection of credit card fraud. The most crucial features are chosen by GBC using the sampling method, which also reduces the dimensionality of the training data. Information gain works by identifying similarities between credit card transactions, and based on the category of legal and fraudulent credit card transactions, gives the most weight to the most important features. Information gain is used as a feature selection technique in the suggested approach due to its computational effectiveness and top precision performance.

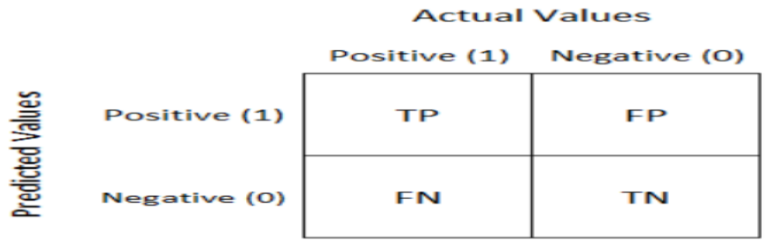
C. Model Building :

Using the Gradient Boosting classifier method to visualise a model.that build a powerful predictive model by combining numerous weak learning models. Gradient boosting frequently makes use of decision trees. It is predicated on the hunch that when previous models are combined with the best possible next model, the total prediction error is minimised.

D. Evaluation for Classification Model.

There are some evaluation measures that we can use when working with classification models to gauge the effectiveness of our models. The confusion matrix, which summarises predicted results in comparison to the actual values of our data set, is one of those assessment metrics. A confusion matrix for a binary classification problem appears like this.

→ TP stands for True Positive and displays a model's accurate forecasts for a positive class.
→ False Positive, or FP, denotes a model's inaccurate forecasts for a positive class.
→ False Negative, or FN, denotes a model's inaccurate forecasts for a negative class.
→ TN stands for True Negative and displays a model's accurate forecasts for a negative class.



Accuracy :

$$\text{Accuracy} = \frac{(TP + TN)}{(TP + FP + TN + FN)}$$

Recall :

$$\text{Sensitivity} = \frac{TP}{(TP + FN)}$$

Precision :

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

F1 Score :

$$\text{F1-score} = \frac{2 \times (\text{precision} \times \text{recall})}{(\text{precision} + \text{recall})}$$

VII. Result And Conclusion

We must always be clear about what we are attempting to achieve when working with a machine learning model. Our project's objective is to identify fraudulent transactions as they happen, and the model that did it the best was the Ada Boost Classifier, which accurately identified 147 out of 160 fraudulent transactions with a recall of 91.87%. The Ada Boost classifier, however, had the highest percentage of false negatives; 1321 legitimate transactions—or 1.54% of all legitimate transactions—were incorrectly classified as fraudulent.

VIII. Future Enhancement

Even though it is not simple, it is possible to stop known and undiscovered fraud in real time. The proposed architecture's initial goal was to identify credit card fraud in online transactions and place a strong emphasis on offering a fraud prevention mechanism to confirm whether a transaction is genuine or fraudulent.. Future technology will allow us to more quickly and accurately determine whether a credit card purchase is fraudulent or not.

IX. References

1. Y. Abakarim, M. Lahby and A. Attioui, "An efficient real time model for credit card fraud detection based on deep learning", Proc. 12th Int. Conf. Intell. Systems: Theories Appl., pp. 1-7, Oct. 2018.
2. H. Abdi and L. J. Williams, "Principal component analysis", Wiley Interdiscipl. Rev. Comput. Statist., vol. 2, no. 4, pp. 433-459, Jul. 2010.

3. I. Benchaji, S. Douzi and B. E. Ouahidi, "Credit card fraud detection model based on LSTM recurrent neural networks", J. Adv. Inf. Technol., vol. 12, no. 2, pp. 113-118, 2021.
4. J. Forough and S. Momtazi, "Ensemble of deep sequential models for credit card fraud detection", Appl. Soft Comput., vol. 99, Feb. 2021.
5. X. Hu, H. Chen and R. Zhang, "Short paper: Credit card fraud detection using LightGBM with asymmetric error control", Proc. 2nd Int. Conf. Artif. Intell. for Industries (AII), pp. 91-94, Sep. 2019.
6. R. F. Lima and A. Pereira, "Feature selection approaches to fraud detection in e-payment systems" in E-Commerce and Web Technologies, Springer, vol. 278, pp. 111-126, 2017.
7. S. Makki, Z. Assaghir, Y. Taher, R. Haque, M.-S. Hacid and H. Zeineddine, "An experimental study with imbalanced classification approaches for credit card fraud detection", IEEE Access, vol. 7, pp. 93010-93022, 2019.
8. I. Mekterović, M. Karan, D. Pintar and L. Brkić, "Credit card fraud detection in card-not-present transactions: Where to invest?", Appl. Sci., vol. 11, no. 15, pp. 6766, Jul. 2021.
9. A. Rb and S. K. Kr, "Credit card fraud detection using artificial neural network", Global Transitions Proc., vol. 2, no. 1, pp. 35-41, Jun. 2021.
10. I. Sadgali, N. Sael and F. Benabbou, "Adaptive model for credit card fraud detection", Int. J. Interact. Mobile Technol., vol. 14, no. 3, pp. 54, Feb. 2020.