CREDIT CARD FRAUD DETECTION USING DIFFERENT MACHINE LEARNING CLASSIFIERS

Mr. Jigul Nimavat[1], Prof. Ankur Goswami[2], Dr. Dushyantsinh Rathod[3], Dr. Vijaykumar Gadhavi[4], Dr. Jaykumar Dave[5]

[2,3,4,5]Computer Engineering Department,
Assistant Professor[2], Professor & HOD[3], Dean[4], Associate Professor[5]
Sankalchand Patel College Of Engineering, Visnagar,Gujarat[1,2]
Ahmedabad Institute of Technology,Ahmedabad,Gujarat[3]
Swaminarayan University ,Kalol,Gujarat[4]
Silver Oak University ,Ahmedabad,Gujarat[5]

ABSTRACT
With rapid advancement in the E-commerce field, fraud is spreading all over the world, causing major financial losses. In current scenario, Major cause of financial losses is credit card fraud. Credit card frauds are easy and friendly targets. E-commerce and many other online sites have increased the online payment modes, increasing the risk for online frauds. In recent years, For banks has become very difficult for detecting the fraud in credit card system. Machine Learning(ML) plays a important key role for detecting the credit card fraud in the transactions. The main address of the research is to design and develop a fraud detection method for Streaming Transaction Data, with an objective, to analyse the past transaction details of the customers and extract the behavioural patterns. the proposed system is implemented with Power Boosting Tree Classifier (PBTC) to detect the frauds. The conclusion of our study explains the best classifier by training and testing using supervised techniques that provides better solution.
Keywords: Credit Card, Machine Learning, Supervised Technique, Power Boosting Tree Classifier.

I. INTRODUCTION

In today's world the credit card fraud is the biggest issue and now there is need to fight against the credit card fraud. "credit card fraud is the process of cleaning dirty money, there by making the source of funds no longer identifiable." The purpose may be to obtain goodies without paying, or to obtain unauthorized funds from an account or to avail some kind of services. Credit card fraud is also an add on to identity theft. On daily basis, the financial transactions are made on huge amount in global market and hence detecting credit card fraud activity is challenging task. The promising way to detect the fraud is to analyze the spending behavior of the cardholder. Every day, new and new researches are performed by the researchers in the different fields. Many researchers of finance field considered this problem as a challenging and important problem. The use of machine learning is proposed by the researchers to deal with this problem. Detecting the fraud means identifying the suspicious one, If any abnormality arises in the spending behavior then it is considered as suspicious. This research is to propose a credit card fraud detection system using supervised learning algorithm. supervised algorithms are evolutionary algorithms which aim at obtaining better solutions as time progresses. To

Overcomes issues of we propose Machine learning method using 'Structural Similarity', to identify common attributes and behavior with other bank account transaction. Detection of credit card fraud transaction from large volume dataset is difficult, so we propose case reduction methods to reduces the input dataset and then find pair of transaction with other bank account with common attributes and behavior. To elude computational complexity & to provide better accuracy in fraud detection in proposed work. Support vector machine(SVM) is a method used in pattern recognition & classification. It is a classifier to predict or to classify patterns into two categories which may be fraudulent or non fraudulent.

II. METHODOLOGY

Power Boosting Tree Classifier(PBTC):-

PBTC is an optimized distributed gradient boosting library designed for efficient and scalable training of machine learning models. It is an ensemble learning method that combines the predictions of multiple weak models to produce a stronger prediction. PBTC has become one of the most popular and widely used machine learning algorithms due to its ability to handle large datasets and its ability to achieve state-of-the-art performance in many machine learning tasks such as classification and regression.

One of the key features of PBTC is its efficient handling of missing values, which allows it to handle real-world data with missing values without requiring significant pre-processing. Additionally, PBTC has built-in support for parallel processing, making it possible to train models on large datasets in a reasonable amount of time.
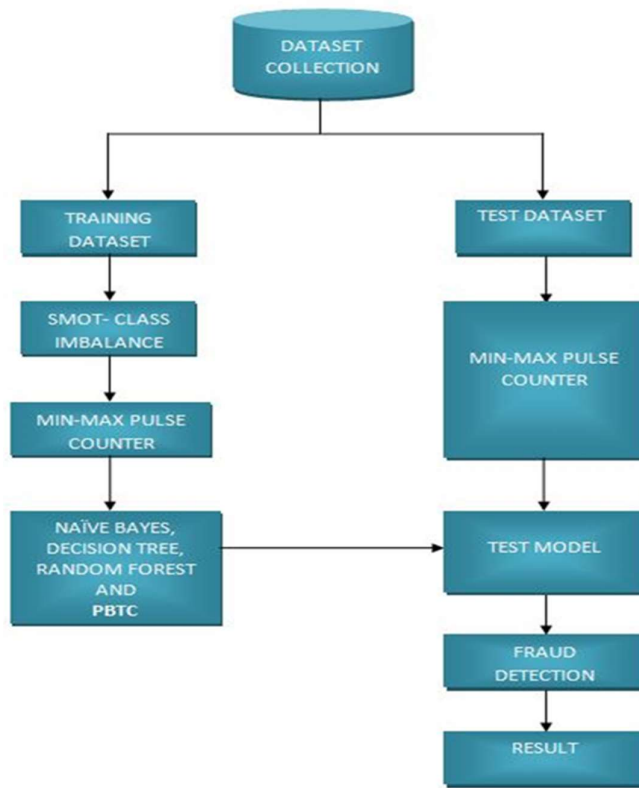

Fig 1: Architecture of Proposed System

1.   **Data processing:-**

Data Pre-processing is a way of changing original data into a clean dataset. In simple words, when the data is gathered from various sources, it is not acquired in processed form, which is not reasonable for evaluation. Pre- processing corresponds to the modifications made to the dataset before feeding the algorithm. Several algorithms of machine learning make assumptions about your data. It is often a very good idea to plan the data in such a way that the problem structure is better presented to the machine learning algorithms.

2.   **Data Classification:-**

While many software optimizations have been used for implementations, when the function dimension is high and the size of data is huge, the performance and expandability are still unsatisfactory. PBTC can be used in a variety of applications, including Kaggle competitions, recommendation systems, and click-through rate prediction, among others. It is also highly customizable and allows for fine-tuning of various model parameters to optimize performance. PBTC which was proposed by the researchers at the University of Washington. It is a library written in C++ which optimizes the training for Gradient Boosting.
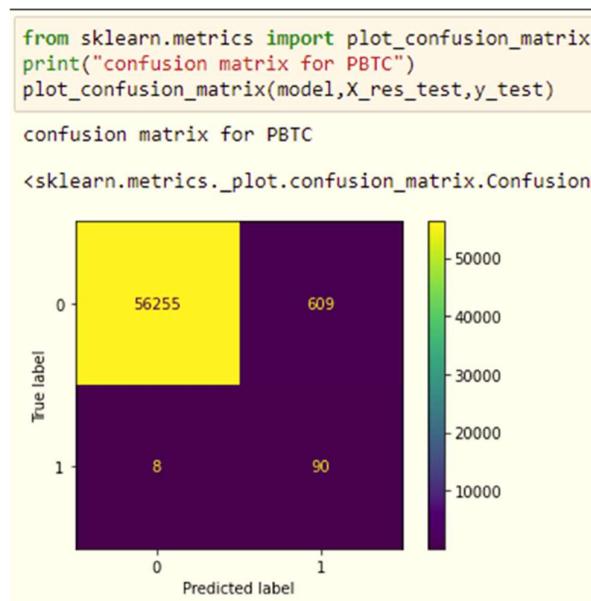
III.   RESULTS & ANALYSIS FOR PBTC

1.   **Classification Report of PBTC :**

```
                 precision    recall  f1-score   support

             0       1.00      0.99      0.99     56864
             1       0.13      0.92      0.23        98

      accuracy                           0.99     56962
     macro avg       0.56      0.95      0.61     56962
  weighted avg       1.00      0.99      0.99     56962
```
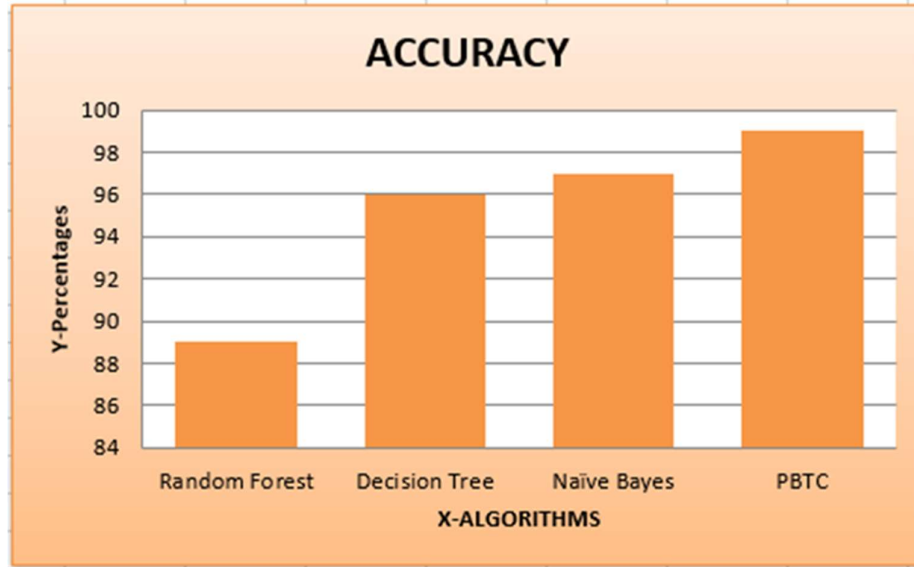
Fig-1. Classification Report of PBTC

2.   **Confusion Matrix of PBTC :**

```
from sklearn.metrics import plot_confusion_matrix
print("confusion matrix for PBTC")
plot_confusion_matrix(model,X_res_test,y_test)
```

confusion matrix for PBTC

<sklearn.metrics._plot.confusion_matrix.ConfusionM

Fig-2. Confusion Matrix of PBTC

3.   **Comparison Chart :**

Fig-3. Comparison Chart.

IV.   CONCLUSION

The proposed method using PBTC gives effective performance in fraud detection. The model is compared with existing classifiers like Naive Bayes, Decision Tree and Random Forest. As these models provides varied performance in terms of detection of fraud cases related to credit card transactions, they are used as baseline algorithms. The Results shows that PBTC methods shows great performance for all three performance metrics like precision, Recall, Accuracy over traditional techniques. In future, the cost based PBTC with effective function will be used to find the fraud detection with lower error rates.

V.   REFERENCES

[1]   R. Dhanpal & P. Gayathiri, "Credit card fraud detection using decision tree for tracing email & ip," International Journal of Computer Science Issues, vol. 9, no. 2, 2012.

[2]   Y. Sahin & E. Duman, "Detecting credit card fraud by decision trees & support vector machines," Proceeding of the International MultiConfrence of Engineers & Computer Scientist, vol. I, 2011.

[3]   C. Bahnsen, A. Stojanovic, D. Aouada, and B. Ottersten. Improving credit card fraud detection with calibrated probabilities. In SDM, 2014.

[4]   M. Gupta, J. Gao, C. C. Aggarwal, and J. Han. Outlier Detection for Tempo- ral Data. Synthesis Lectures on Data Mining and Knowledge Discovery, Morgan Claypool Publishers, 2014.

[5]   Awoyemi, J. O., Adetunmbi, A. O., &Oluwadare, S. A. (2017). Credit card fraud detection using machine learning techniques: A comparative analysis. 2017 International Conference on Computing Networking and Informatics (ICCNI). P1-9.

[6]   V. Dheepa, R. Dhanapal and D. Remigious, "A Novel Approach to Credit Card Fraud Detection Model",

Journal of Computing, Vol. 2, No. 12, pp. 96, 2010.

[7] Adewumi, A. O., &Akinyelu, A. A. (2016). A survey of machine-learning and nature-inspired based credit card fraud detection techniques. International Journal of System Assurance Engineering and Management, 8(S2), 937–953

[8] Roy, Abhimanyu, et al. "Deep Learning Detecting Fraud in Credit Card Transactions." 2018 Systems and Information Engineering Design Symposium (SIEDS), 2018, doi:10.1109/sieds.2018.8374722.

[9] Lu, Q., & Ju, C. J. J. o. C. I. T. (2011). Research on credit card fraud detection model based on class weighted support vector machine. 6(1).

[10] Akter Hossain, M., & Uddin, M. N. (2018). A Differentiate Analysis for Credit Card Fraud Detection. 2018