IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

CREDIT CARD FRAUD DETECTION

Mrs M.M. Phadatare
Dept of Computer Science and Engineering, AISSMS College of Engineering, Pune, Maharashtra, India

Gurshan Singh
Dept of Computer Science and Engineering, AISSMS College of Engineering,Pune, Maharashtra, India

Rohan Kalaskar
Dept of Computer Science and Engineering, AISSMS College of Engineering, Pune, Maharashtra, India

Rushikesh Ghuge
Dept of Computer Science and Engineering, AISSMS College of Engineering, Pune, Maharashtra, India

Harsh Tiwari
Dept of Computer Science and Engineering, AISSMS College of Engineering, Pune, Maharashtra, India

***Abstract*— Credit cards are widely used in most financial aspects due to the exponential development in online purchases, which increases the risk of fraudulent transactions. By examining different user behaviour from past transaction history databases, these fraudulent transactions can be demonstrated. Any deviation from the usual patterns of behaviour raises the risk of a fraudulent transaction. In this research, ensemble learning algorithms (XGBoost) is used. The builded system will determine whether a transaction is authentic or fraudulent using this models. Therefore, financial losses brought on by fraudulent transactions can be reduced by incorporating this methodology into fraud detection systems.**

Keywords: Machine learning ,Open CV, XGBoost, Decision Tree, Logistic Regression, Random Forest

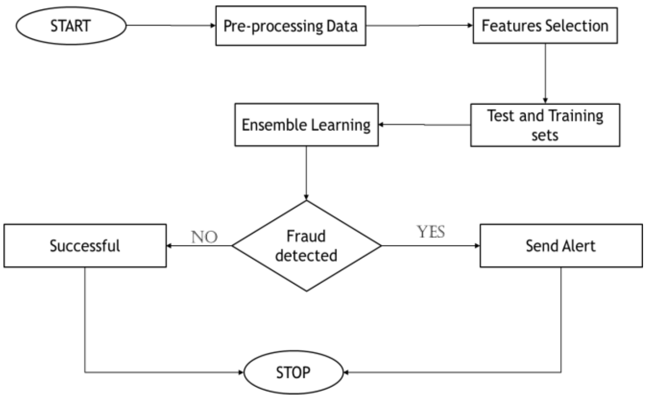
I. INTODUCTION

The age of cash lessness is upon us as the world progresses, yet every improvement has its downsides. Credit card usage has increased, which has led to an increase in fraudulent transactions. The use of a credit or debit card that has been reported lost, stolen, or cancelled in order to get something of value is known as credit card fraud. It has an impact on the entire consumer credit industry. One of the fraud categories that is expanding most quickly and that is also the hardest to stop is credit fraud. The security of the website may have been compromised, or the owner's negligence may have led to this fraud.

This research paper's main justification is to draw attention to the similarities between fraudulent credit card transactions and legitimate ones. The first step towards achieving this goal is to develop a machine-learning-based fraud detection system that can quickly and accurately identify fraudulent transactions. XGBoost and other ensemble learning algorithms are used in the system. The system can predict if a transaction is fraudulent or real by manipulating these models.

To achieve greater predictive performance that could be attained from any one of the fundamental learning algorithms alone, ensemble learning models employ many algorithms. It makes this model faster, more accurate, and more useful than the other models.

IV. SYSTEM ARCHITECTURE



1.Data Source and Description

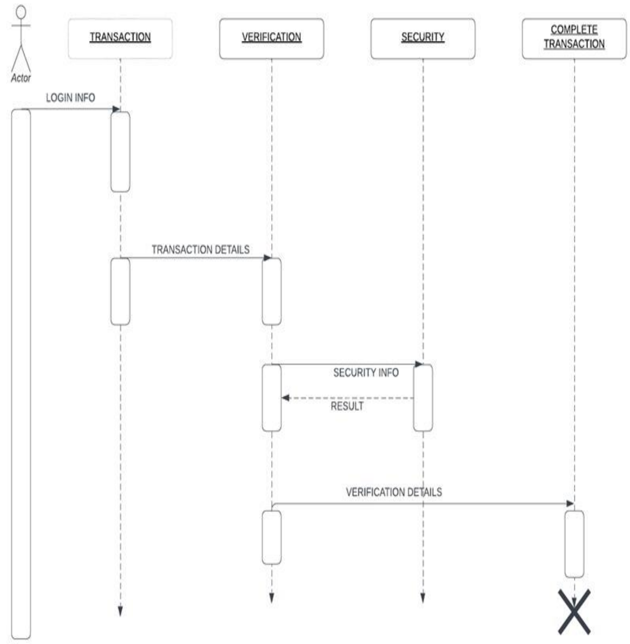
The dataset used by the system is taken from the open-source website Kaggle. There are 284,807 transactions or rows in the dataset. The dataset includes characteristics from V1 through V28 that are the PCA-transformed main components. Amount and Time are the only features that haven't undergone transformation.

2.Data Preprocessing

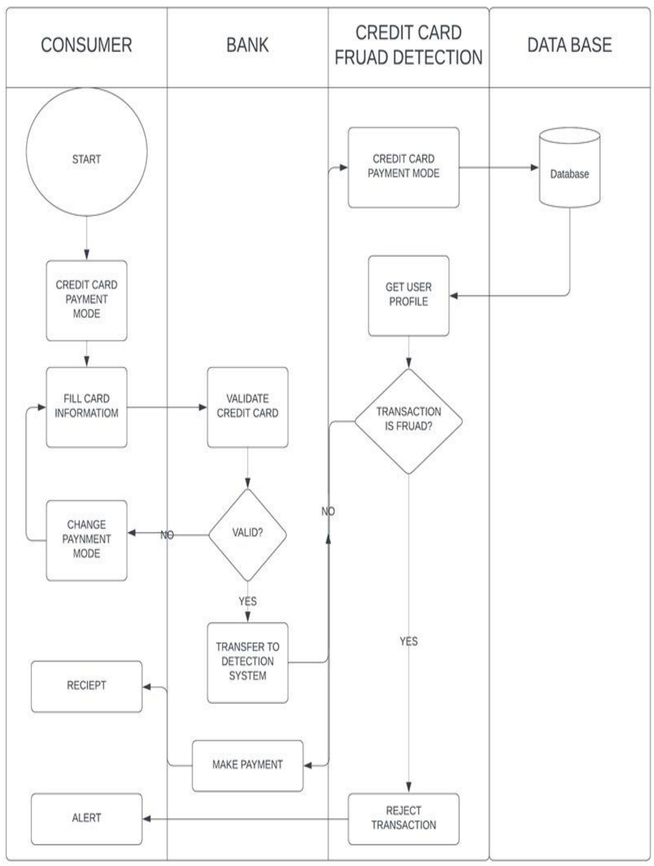
Data before passing to the model is checked for the missing values or null values as they can produce garbage results. After checking, the dataset does not contain any missing values or null values, it will be used for training and testing the model.

3. UML Diagrams

- *Sequential Diagram*



- *Activity Diagram*



4. Dividing the training set and test set

The dataset will be divided into two halves: a training dataset and a testing dataset following data pre-processing. The model is only built using the training dataset, and it is evaluated using test data. 30% of the data will be used for model testing, and 70% of the data will be used to train the model.

5. Dealing with imbalanced data

Imbalanced data is a common issue in many real-world scenarios, such as fraud detection, disease diagnosis, and anomaly detection. For example, in a fraud detection problem, the majority of transactions may be non-fraudulent (negative class), while only a small fraction of transactions are fraudulent (positive class).

SMOTE(Synthetic Minority Oversampling Technique) is an oversampling technique that generates synthetic samples from the minority class. It is used to obtain a synthetically class-balanced or nearly class-balanced training set, which is then used to train the classifier.

6. Algorithm

6.1 XGBoost

The ensemble learning technique XGBoost Classifier (Extreme Gradient Boosting) is used in the proposed system. Gradient Boosting Decision Tree, or GDBT, is the foundation of XGBoost.

5.1.1. Working of XGBoost

Progressive decision trees are produced using this algorithm. All independent variables are then given weights and put into the decision tree, which makes predictions about the outcomes. The second decision tree receives greater weights from the factors that the first one incorrectly predicted. A more accurate model is then produced by ensembleing these predictions.

6.2 Logistic Regression

Based on a set of input variables, the statistical modelling approach of logistic regression is used to forecast outcomes that are either binary or categorical. It is a special kind of regression analysis made for estimating the likelihood that an event will occur.

The dependent variable in logistic regression is binary, which means it can only have one of two potential values, such as "yes" or "no," "success" or "failure," or "0" or "1." Both continuous and categorical independent variables, usually referred to as predictors or features, are acceptable.

6.3 Naves Bayes

A well-liked classification technique called Naive Bayes is based on using the Bayes theorem while assuming that each input characteristic is independent of the others. It is referred to as "naive" because it assumes that the characteristics are conditionally independent of one another given the class, which simplifies the formulation of the probability distribution. The Naive Bayes method chooses the class with the highest probability as the projected class for a new instance by computing the probabilities of several classes given a collection of input characteristics.

6.4 Decision Tree

A decision tree is a supervised machine learning technique that bases predictions or judgements on input information in a tree-like structure. It is a model that resembles a flowchart, with each internal node standing in for a feature or attribute, each branch standing in for a decision rule, and each leaf node standing in for the result or class label.

The first step of the decision tree method is to choose the optimal feature that divides the data in the most efficient way, typically based on indicators like Gini impurity or information gain. The data is divided into subsets depending on the feature values of the chosen feature, which serves as the tree's root node. After that, the procedure is repeated recursively for every subset, adding branching and new nodes up until a stopping requirement is satisfied.

7. Training and Testing procedure

The training part of the system include training the model using random samples from the training dataset, which will have been generated by earlier steps. After a model has been successfully trained, it will be put to the test using a test dataset. To verify the model's correctness, the system would compare the output predictions to known fraud transactions. We can also get a confusion matrix, which will help us assess the model's correctness.

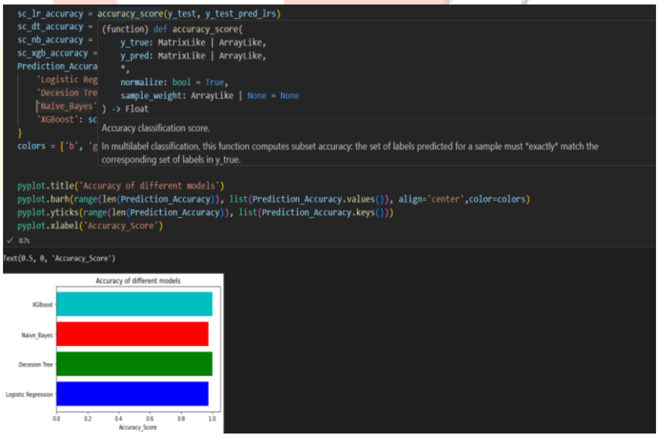
. *8. Model Evaluation*

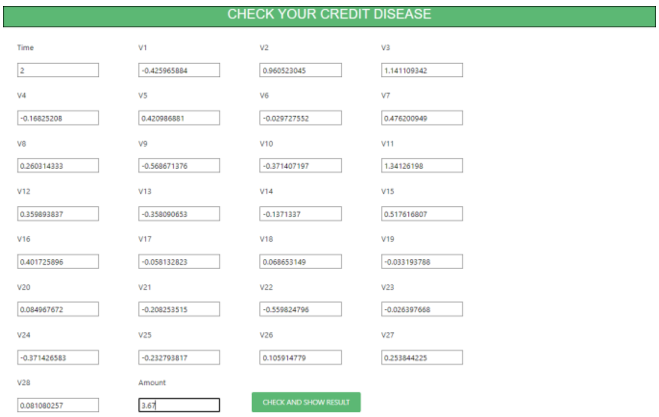
After comparing the performance metrics of the three models, it is clear that each model has its strengths and weaknesses.

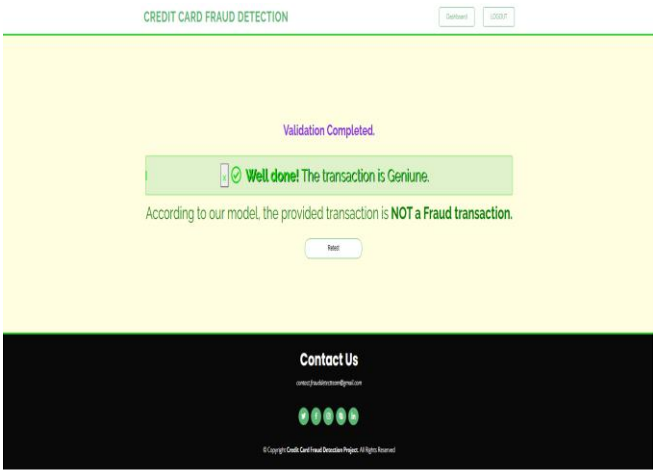
The XGBoost model demonstrates the highest accuracy among the three models, with an impressive accuracy of 99.98%. This indicates that it correctly classifies the majority of instances, making it a reliable choice for prediction tasks.

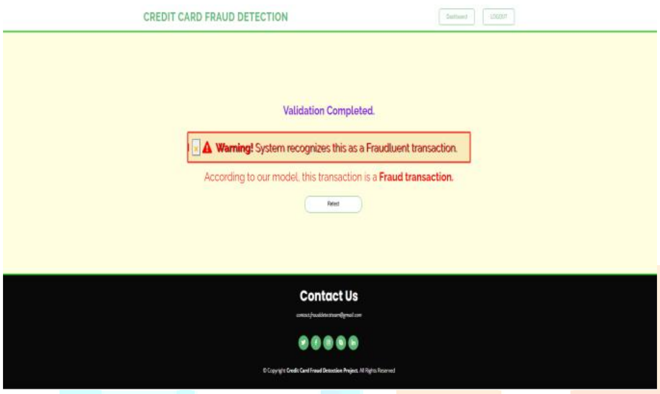
MODEL	ACCURACY	PRECISION	RECALL
XGBOOST	99.98	94.44	94.44
LR	97.54	5.85	90.27
DECISION TREE	99.93	71.50	92.36
NAÏVE BAYES	97.45	5.39	85.41

V. MODEL SNAPSHOT









VI. CONCLUSION

In conclusion, the credit card fraud detection system built using XGBoost has achieved an impressive accuracy of 99.98%. This high level of accuracy is a testament to the effectiveness of the XGBoost algorithm in accurately identifying fraudulent transactions and minimizing false positives.

By leveraging XGBoost, the system has been able to effectively capture the complex patterns and relationships within credit card transaction data. XGBoost's ability to handle imbalanced datasets and its robustness to outliers and noise have contributed to the system's exceptional performance.

VII. REFFERENCES

[1] M. Suresh Kumar, V. Soundarya, S. Kavitha, E.S. Keerthika, E. Aswini" CREDIT CARD FRAUD DETECTION USING RANDOM FOREST ALGORITHM" IEEE 2019

[2] Dr. Anju Pratap, Anu Maria Babu" Credit Card Fraud Detection Using Deep Learning" IEEE 2020

[3] Zaiyyan Khan, Siddhesh Jadhav, Hashim Malik, Munira Ansari "Credit Card Fraud Detection" IJERT 2021

[4] D. Eswar, CH V N M Praneeth, Raja Subramanian, D. Tanouz "Credit Card Fraud Detection Using Machine Learning" IEEE 2021

[5] Santanu Kumar Rath, Debachudamani Prusti, Aditya Sai Kandukuri, S. S. Harshini Padmanabhuni "Detecting Default Payment Fraud in Credit Cards" IEEE 2019

[6] Donglin li "Credit card fraud identification based on unbalanced data set based on fusion model" IEEE 2019

[7] Anuruddha Thennakoon, Chee Bhagyani, Sasitha Premadasa, Shalitha Mihiranga, Nuwan Kuruwitaarachchi "Real-time Credit Card Fraud Detection Using Machine Learning" IEEE 2019

[8] Vehbi Cagri Gungor, Cengiz Gezer, Gokhan Goy "Credit Card Fraud Detection with Machine Learning Methods" IEEE 2020

[9] Hint Wint Khin, Aye Aye Khine "Credit Card Fraud Detection Using Online Boosting with Extremely Fast Decision Tree" IEEE 2020

[10] Abhishek M, Navaneeth A V,Dileep M R" A Novel Approach for Credit Card Fraud Detection using Decision Tree and Random Forest Algorithms" IEEE 2021

[11] Andrew. Y. Ng, Michael. I. Jordan, "On discriminative vs. generative classifiers: A comparison of logistic regression and naive bayes", Advances in neural information processing systems, vol. 2, 2002

[12] John Richard D. Kho, Larry A. Vea "Credit Card Fraud Detection Based on Transaction Behaviour" published by Proc. of the 2017 IEEE Region 10 Conference (TENCON), Malaysia, November 5-8, IEEE 2017. 5

[13] Yashvi Jain, Namrata Tiwari, ShripriyaDubey, Sarika Jain, "A Comparative Analysis of Various Credit Card Fraud Detection Techniques, Blue Eyes Intelligence Engineering and Sciences Publications 2019"

[14] Learning Robert A. Sowah, Moses A. Agebure, Godfrey A. Mills, Koudjo M. Kaumudi, "New Cluster Under Sampling Technique for Class Imbalance "IJMLC 2016

[15] Baraneetharan, E. "Role of Machine Learning Algorithms Intrusion Detection in WSNs: A Survey." Journal of Information Technology 2, no. 03 (2020).

[16] Mohamed Jaward Bah, Mohamed Hammad "Progress in Outlier Detection Techniques: A Survey" Hongzhi Wang, of the IEEE 2019

[17] A. Bifet and R. Kirkby Massive Online Analysis, Technical Manual, Univ. of Waikato, 2009.

[18] R. Bolton and D. Hand, "Unsupervised Profiling Methods for Fraud Detection," Statistical Science, vol. 17, no. 3, pp. 235255, 2001.

[19] P. Brockett, R. Derrig, L. Golden, A. Levine, and M. Alpert, "Fraud Classification Using Principal Component Analysis of RIDITs," The J. Risk and Insurance, vol. 69, no. 3, pp. 341-371, 2002, doi: 10.1111/15396975.00027.

[20] R. Caruana and A. Niculescu-Mizil, "Data Mining in Metric Space: An Empirical Analysis of Supervised Learning Performance Criteria," Proc. 10th ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining (KDD '04), 2004, doi: 10.1145/1014052.1014063.

[21] P. Christen and K. Goiser, "Quality and Complexity Measures for Data Linkage and Deduplication," Quality Measures in Data Mining, F. Guillet and H. Hamilton, eds., vol. 43, Springer, 2007, doi: 10.1007/978-3-54044918-8.

[22] C. Cortes, D. Pregibon, and C. Volinsky, "Computational Methods for Dynamic Graphs," J. Computational and Graphical Statistics, vol. 12, no. 4, pp. 950-970, 2003, doi: 10.1198/1061860032742.

[23] Experian. Experian Detect: Application Fraud Prevention System, Whitepaper, http://www.experian.com/products/pdf/expe rian_dete ct.pdf, 2008.