International Journal Of Innovative Trends In Multi Disciplinary Scientific Research
ISSN (Online): 3139 – 2687
(www.ijtmsr.co.in) Volume 9, Issue 6, 2026

Detection of Hypocrisy in Credit Cards Using Machine Learning Algorithms and Smote Technique

T. Rajendra Prasad[1], Nalla. Jasmita Srijja[2], Paidi. Mahalakshmi[3], Nimmakuri. Ajay[4]

Department of CSE (AI&ML), NRI Institute of Technology, Agiripalli, A.P, India

Abstract

Credit card fraud has become a significant challenge due to the rapid growth of online financial transactions. Traditional rule-based fraud detection system often miss the mark to adapt to evolving fraudulent patterns. This paper proposes an ensemble-based machine learning framework for detecting credit card fraud using multiple supervised learning algorithms. The dataset is highly imbalanced, reflecting the real-world transaction scenarios, and Synthetic Minority Oversampling Technique (SMOTE) is applied to address class imbalance. Five machine learning models like – Logistic Regression, Random Forest, XGBoost, Support Vector Machine And Gaussion Naïve Bayes – are trained and evaluated using performance metrics such as accuracy, precision, recall, F1-score and ROC-AUC. The best-performing model is selected based on ROC-AUC and a majority voting mechanism is employed for financial fraud classification, Furthermore, a real- time web-based application is developed using Streamlit and deployed via ngrok for practical usability. Experimental results demonstrate ensemble decision-making improves fraud detection accuracy and reliability. Additionally, for practical usability, an on-demand internet-based program is created utilizing Streamlit and distributed using ngrok. The precision and dependability of identifying fraud are enhanced by collective decision-making, according to find from experiments.

Keywords: SMOTE, identification of credit card scam using machine learning algorithms, Streamlit, Banking security,

1. INTRODUCTION –

Globally, the use of credit cards has significantly increased as a result of the board adoption of e-commerce and online financial services. Although this expansion has made transactions easier, it has also made fraudulent activity more likely, which has led to significant losses for both clients and financial institutions.

Given to the enormous number of daily activities and the constantly habits of fraudsters, prior research highlights the need of the identification of credit card fraud (Dal Pozzolo et al., 2015). The majority of the initial systems for identifying fraud relied on expert-driven and rule-based methods. These algorithms can identify known fraud tendencies, but they aren't flexible and need to be updated manually on a regular basis. According to published research, rule-based approaches frequently fall short in

Spotting novel and intricate fraud tactics, which increases error ratio and lowers system effectiveness

(Bahnsen et al., 2013). These constraints have prompted academics to investigate computerized, and solutions based on data.

Because machine learning technologies may identify trends in past transactions, they have become helpful instruments for detecting fraud with credit cards. Fraud detection issues have been effectively addressed by a number of Greater reliability than conventional techniques by analyzing transaction attributes to categorize operations as genuine or lawful.

1. RELATED WORK –

1.1 PYTHON PROGRAMMING: -

Python's strong backing in data analysis and data mining makes it a highly popular programming language out there. The primary programming platform used for executing the proposed framework is python. It is perfect used for building apps using machine learning because of its ease of use, broad package help and robust industry growth. Within a single setup, python facilitates effective data preparation, modeling, analysis and delivery. Python's modular architecture makes it easy to integrate the website-building tools, machine learning platforms and data science resources utilized in the current endeavor.

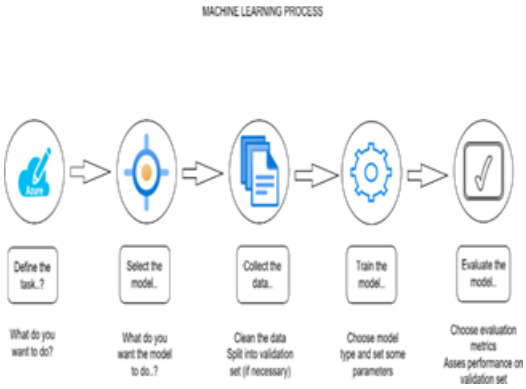
1.2 PANDAS AND NUMPY: -

Arithmetic calculation and information management are done with Pandas and NumPy. For managing massive event records, Pandas offers effective structured data like Data Framework that allow for tasks like choosing features, data cleansing and searching via event Identifiers. Highly efficient calculation and creating random numbers are supported by NumPy, which is crucial for the development of synthetic datasets and mathematical treatments used in preparation.

International Journal Of Innovative Trends In Multi Disciplinary Scientific Research
ISSN (Online): 3139 – 2687
(www.ijtmsr.co.in) Volume 9, Issue 6, 2026

1.3 MACHINE LEARNING: -

Several machine learning models with supervision are used in this study to identify fraudulent charges on credit cards. Because of its easy use and comprehension in binary-categorized situations, logistic regression is utilized as an initial predictor. In order to identify intricate and irregular variations in fraud with excellent accuracy and ROC-AUC efficiency, Random Forest is used as a collaborative learning approach involving multiple trees of decisions. A stochastic algorithm called Naïve Bayes is used to make quick supervised learning algorithms, such as logarithmic regression, Support Vector Machines, Random Forests, Naïve Bayes and Gradient Boosting models (Carcillo et al.,2021). These algorithms provide
and accurate predictions via numerical data dispersion. When these approaches are used together, they enable efficient evaluation and increase the accuracy of detecting fraud.



A. LOGISTIC REGRESSION: -

Because of its ease of use and comprehension, logistic regression is utilized as an initial segmentation model. It uses a linear decision limit to assess the likelihood that an exchange is fraudulent. Logistic regression, while being simple, works well on unbalanced samples and offers a solid benchmark for contrasting more complicated algorithms.

B. RANDOM FOREST: -

The random forest is a method of ensemble learning that uses a lot of choice tree information to make better predictions and avoid excessive overfitting. It works well with data that is multidimensional yet not linear. The system that is suggested shows that Random Forest is very good at finding scams because it can find intricate relationships between deal characteristics.

International Journal Of Innovative Trends In Multi Disciplinary Scientific Research
ISSN (Online): 3139 – 2687
(www.ijtmsr.co.in) Volume 9, Issue 6, 2026

C. SUPPORT VECTOR MACHINE: -

Support Vector Machine is employed to generate an optimum choosing limit, it optimizes the gap among illegal and legal deals. SVM is very successful in large characteristic areas. The chance estimate has been made possible to allow collaborative decision-making and evaluation of different models.

D. XGBOOST: -

XGBoost is a gradient enhancement method that has been tuned for flexibility and outstanding results. It achieves excellent accuracy in prediction by gradually improving poor learners through the use of boosting techniques, XGBoost is one of the top algorithms for scam identification in this endeavor because it provides higher ROC-AUC efficiency.

E. GAUSSIAN NAÏVE BAYES: -

Naïve Gaussian a combination of the characteristics neutrality hypothesis and Bayes theorem. Bayes is a classification algorithm that uses probabilities. It works well with enormous data sets and extremely fast. It offers rapid forecasts and introduces variety to the group model despite assuming an easier collection of data.

1.4 SMOTE – (Synthetic Minority Over-sampling Technique)

The extreme inequality of classes found in the payment card transaction data is addressed by SMOTE, which is done via unbalanced learning module. The SMOTE algorithm creates synthetic deception instances by adding among actual minority class samples because scams make up a relatively small percentage of any given transaction. To avoid data leaking and guarantee objective estimation of models, SMOTE is only used on the learning sample in this research. The model's capacity to identify illegal transactions is much enhanced by this method.

1.5 SCIKIT-LEARN: -

The main ML package for creating and assessing models is called Scikit-Learn. It offers variants of several trained learning methods, as well as Gaussian Naïve Bayes, Random Forest, Support Vector Machine and Logistic Regression. Additionally, Scikit-Learn is used to separate information with training-testing separation, adjust the characteristics with Standard Scaler and assess efficiency using measure including accuracy, precision, recall, F1-score and ROC-AUC. Its uniform API architecture makes it simple to compare different approaches.

1.6 STREAMLIT WEB FRAMEWORK: -

A current online identification of scam dashboard is created using Streamlit. Users can access fraudulent forecasts, possibility results and comparisons of models by entering their transaction ID. Streamlit offers smooth interaction using machine learning models, dynamic representation and instant development. A structure that facilitates easy comprehension and useful testing of the identification of fraud method by displaying model-specific scam rates, actual fraudulent judgements made by the majority of votes and alarm signs.

2. SYSTEM DESIGN –



3. EXISTING SYSTEM –

To find unauthorized activities, current identification of credit card fraud algorithms

mostly rely on rules-based methods and conventional statistical procedures. Previously established standards, such as large transaction minimums, anomalous buying spots or frequent transactions within a brief period of time, are used by these systems. Although these methods work well for identifying established fraud trends, they are flexible and unable to adjust to new fraud tactics.

One important shortcoming of previous approaches is they are unable to manage extremely unbalanced datasets in which illegal transactions represent a relatively small fraction of total data. This imbalance frequently results in biased forecasts of valid transactions, leading to low recognition of fraud rates. Furthermore, existing systems rely significantly on personal inspection processes, which raises operational expenses and slows fraud detection.

Existing techniques also have significant inaccurate identification rates, which means that legitimate purchases are falsely identified as fraudulent, resulting in a negative user experience. The lack of immediate training and smart decision- making hinders system growth and efficacy in contemporary digital payment platforms. As a results, these restrictions emphasize the necessity for an intelligent, flexible, and information-driven recognition of fraud platforms.

4. PROPOSED SYSTEM –

The suggested approach detects fraud with credit cards by combining supervised machine learning procedures, imbalanced class processing, and actual deployments. The software handles information regarding transactions in several steps, including preprocessing, model training, evaluation, and web-based deployment. By incorporating several classifiers and a majority voting method, the system enhances detection precision as well as resilience while also allowing for real-time fraud estimation via a dynamic user input.

4.1 DATA GENERATION AND DESCRIPTION: -

The scientific learning module's make- classification method is used to construct an artificial card transaction sample that closely simulates real-world fraudulent transaction events. The collection of data includes 10,000 Transaction files, each with 20 anonymous numerical characteristics and an extra transaction size attribute. A binary-valued variable called the target is used to identify payments as illegal or not legitimate. To better simulate real-world fraud tendencies and improve model learning, more false cases are purposefully added, which creates an incredibly unbalanced datasets reflecting real- world financial settings.

4.2 DATA PREPROCESSING: -

Data preparation is done to get the collection of data ready for efficient model training. The dataset contains no data that is missing because it was generated artificially. All numerical characteristics are converted to pixel using the Standard Scaler method to guarantee equal data dispersion and facilitate convergence between models. A specific payment ID retained for real- time event recognition and lookups. The information is then partitioned into testing and training sets using sample stratification to ensure that the pattern of class distribution is consistent over either set.

4.3 HANDLING CLASS IMBALANCE USING SMOTE: -

The credit card scam records are intrinsically skewed, with unauthorized transaction accounting for just a tiny portion of the entire data. To solve this problem, the Synthetic Minority Oversampling Technique (SMOTE) is used only on the original dataset. This method creates fake scam instances to equalize the class distribution while avoiding leakage of information. As a result, the trained models have increased ability to learn and balanced evaluations of performance.

4.4 MODEL TRAINING: -

Several classification algorithms based on ML are used to evaluate and contrast fraud detection results. The SMOTE-balancing learning set is used to train many models, including Logistic Regression, Random Forest, Support Vector

International Journal Of Innovative Trends In Multi Disciplinary Scientific Research
ISSN (Online): 3139 – 2687
(www.ijtmsr.co.in) Volume 9, Issue 6, 2026

Machine, Naïve Bayes and XGBoost. Every method is tested using metrics of performance such as accuracy, precision, recall, F1-score and ROC-AUC to ensure an accurate evaluation of detecting fraud efficiency.

4.5 ENSEMBLE DECISION MECHANISM: -

To improve forecast accuracy and eliminate each predictive bias, a method based on ensemble learning is used. A clear-cut voting procedure is used for all learned classifiers. Each model forecasts the probability of an illegal transaction, and the ultimate judgement is based on the model's consensus. The whole technique increases durability, reliability and prediction in detection.

5. EXPERIMENT RESULTS & EXPLANATIONS –

This image illustrates the way the historical data was successfully loaded as well as how 100 additional unauthorized transactions were added to improve the depiction of crime. In Figure 1, the above indications attest to the error-free finishing of the data source generation procedure. This stage aids in simulating false equilibrium scenarios found in everyday life. Moreover in Figure 2, model's performance metrices are evaluated. Differences in identifying fraud performance are shown by indices like ROC-AUC, F1-score, precision, accuracy and Recall. The low and high activity ID's as well as the sample quantities within the collected data are displayed in Figure 3. There are 10,100 distinct process ID's in all, which range from zero through 10099 in length. This verifies correct categorization and the accuracy of the resource. Figure 4 reflects an overall of 241 unauthorized activities and shows the discovered fake purchase ID's. the resultant report verifies that illegal instances were correctly identified and labelled. For the purpose of teaching and assessing models, such labels serve as ground truth. An overall of 241 forged payments and includes all discovered forged payment ID's are shown in Figure 5. The internet-based UI where consumers may provide the ID of a transaction in order to verify for criminal activity is depicted in Figure 6. Built for accurate forecasting, its design is straightforward and easy to use. The projected outcomes for a chosen deal ID is shown in Figure 6. Each one offers likelihood ratings for both fraud and non-fraud, and they all categorize the sale as secure. A majority vote is used to make the final choice. According to ROC-AUC, linear regression is the model that performs the best, as shown in Figure 8.1. Additionally, it indicates that the deal's eventual certification is SAFE. This validates the chosen system's dependability. A bar graph analyzing the ROC-AUC grades of every predictive algorithm is shown in Figure 8.2. The maximum ROC-AUC is attained by logistic regression, which is then followed by SVM, and XGBoost. The design choice process is well supported by the presentation.

FIGURE.1. DATASET GENERATION & PREPROSESSING -

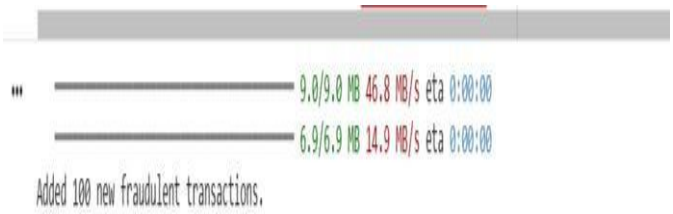


FIGURE.2. AFTER TRAINING -



FIGURE.3. DISPLAYING TRANSACTION ID COLUMN –

International Journal Of Innovative Trends In Multi Disciplinary Scientific Research
ISSN (Online): 3139 – 2687
(www.ijtmsr.co.in) Volume 9, Issue 6, 2026

```
Minimum Available Transaction IDs (first 5 samples):
0
1
2
3
4

Maximum Available Transaction IDs (last 5 samples):
10095
10096
10097
10098
10099

Total number of unique Transaction IDs: 10100
The minimum Transaction ID is: 0
The maximum Transaction ID is: 10099
```

FIGURE.4. DISPLAYING FRAUD TRANSACTIONS –

```
Total number of fraudulent transactions: 241
All fraudulent Transaction IDs:
[131, 142, 147, 233, 359, 419, 501, 521, 584, 684, 900, 1009, 1029, 1179, 1294, 1353, 1556, 1624, 1995, 2165, 2191, 2270, 2302, 2473, 2485,
```

FIGURE.5. ROC-AUC SCORES -

```
ROC-AUC scores prepared for embedding:
{'Logistic Regression': 0.7122105307640297, 'Random Forest': 0.6567148411088574, 'XGBoost': 0.693838742393
```

FIGURE.6. AFTER EVALUATING INPUT INTERFACE –

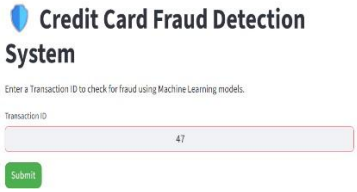


FIGURE.7.1. RESULTS WITH MODELWISE PROBABILITIES –

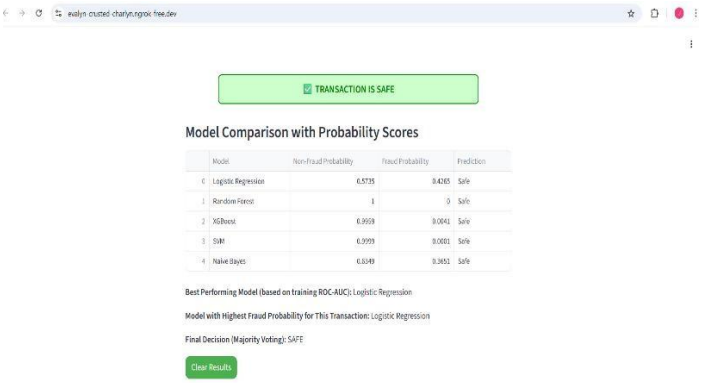


FIGURE.7.2. DETECTING FRAUD TRANSACTION –

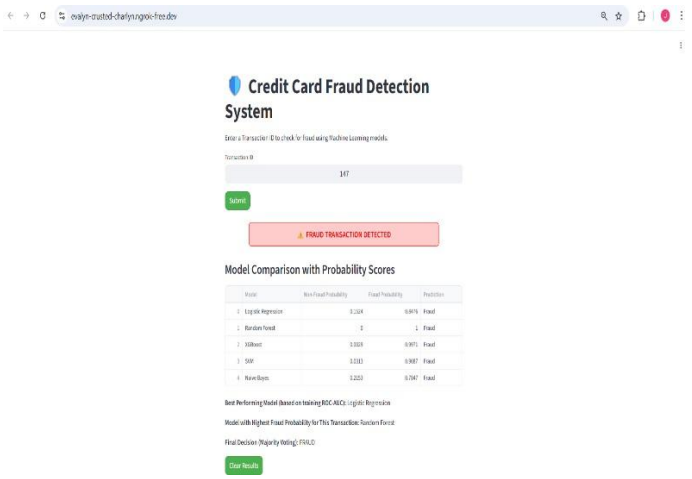
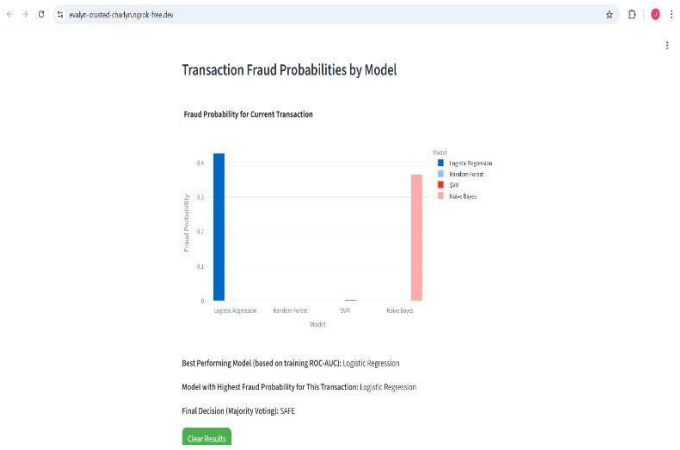


FIGURE.8.1. GRAPHICAL REPRESENTATION –



International Journal Of Innovative Trends In Multi Disciplinary Scientific Research
ISSN (Online): 3139 – 2687
(www.ijtmsr.co.in) Volume 9, Issue 6, 2026

FIGURE.8.2



7. CONCLUSION –

An automated trained fraudulent detection device for debit card transactions was successfully developed and put into use in order to get around the challenges of spotting fraud in extremely imbalanced transaction data. Through the use of SMOTE excess sampling and efficient prior processing methods, the system enhanced learner's capacity to learn minority fraud instance. Various automated learning approaches, such as Logistic Regression, Random Forest, Support vector Machine, Naïve Bayes and XGBoost were taught and assessed to correctly recognize forged payments. According to study results, group-based systems performed better when speaking of broad forecast quality and ROC-AUC. By aggregating forecasts from several classifiers, the addition of an overall polling method improved decision resilience even more. Continuous fraud detection was made possible by the deployment of the trained models via a Streamlit web application, which offered an engaging and user-friendly interface for useful testing.

All things considered, the suggested approach provides an automatic, flexible and effective way to recognize fraud with credit cards. It is appropriate for practical banking applications since it lowers false alarms, adjust to changing scam trends and promotes safe online payment methods.

8. FUTURE SCOPE –

By incorporating immediate payment information from finance websites, the suggested fraud detection for credit card transactions method can be improved in future versions to allow for real- time crime reporting. By using the techniques of deep learning and adaptive learning processes to address changing fraud trends, the model's accuracy can be enhanced. The speed of recognition can be further improved by extra physiological and cultural factors like activity occurrence and locality. Additionally, the framework can be improved through artificial intelligence (AI) methods and deployed utilizing cloud services to enhance safety, openness and confidence in practical investments. Adopting transparent algorithms in conjunction with flexible internet-based designs can enhance compliance with regulations and confidence while facilitating effective system supervision.

9. REFERENCES –

[1]. A. Dal Pozzolo, G. Bontempi,
IEEE Computational Intelligence Magazine, vol. 10, no. 3, pp. 41-54, 2015.

[2]. F. Carcillo, Y. Boulanger, A. Dal Pozzolo, Information Fusion, vol. 41, pp.182-194, 2018.

[3]. N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic Minority Over- sampling Technique,"
Journal of Artificial Intelligence Research, vol. 16, pp. 321-357, 2002.

[4]. L. Breiman,
"Random Forest,"
Machine Learning, vol. 45, no. 1, pp. 5-32, 2001.

[5]. C. Cortes and V. Vapnik,
"Support-Vector networks,"
Machine Learning, vol.20, no. 3, pp. 273-297, 1995.

[6]. T. Chen and C. Guestrin,
"XGBoost: A scalable tree boosting system," Twenty-second IEEE SIGKDD: International symposium with learning and Information Mining Reports, pp. 785-794, 2016.

[7]. T. Mitchell,

Machine Mitchell,
McGraw-Hill, New York, 1997.

[8]. J. Brownlee, "Classification accuracy precision, recall, and F1-score,"
Machine Learning Mastery, 2019.

[9]. Steamlit Inc.,
"Streamlit: The fastest way to build data apps;"
Steamlit Documentation, 2023.

[10]. Dal Pozzolo, A., Bontempi, G.,
"Finding fraudulent activity with credit cards using flexible machine learning methods, Expert System Design with Software, vol.39, no. 15, pp. 12329- 12335, 2012.

[11]. "A Cognitive Network Composition using Knowledge Building for Better Banking Scam Recognition," Esenogho, and Mienye Swart =, Aruleba, & Obaido, IEEE Access, 2022.

[12]. "Banking ATM Fraud Prevention Using an Enhanced Variant Automatic Encoding Artificial Threat Network," Ding, kang, and Feng, Peng, & Yang, the Access Journal (or IEEE-indexed article,2023).

[13]. Alarfaj et al. "State-of-the-Art ML and Advanced Deep Learning Algorithms for Credit Card Fraud Detection," 2022; IEEE Access.