IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Intelligent Machine For Anomalous Transaction Of Credit Card

MR.A.VENKATESWARA RAO
Dept of CSE(AI&ML)
Dadi institute of Engineering and Technology

T.VAMSI
Dept of CSE(AI&ML)
Dadi institute of Engineering and Technology

P.SAI
Dept of CSE(AI&ML)
Dadi institute of Engineering and Technology

P.LOHITH
Dept of CSE(AI&ML)
Dadi institute of Engineering and Technology

B.THARUN KUMAR
Dept of CSE(AI&ML)
Dadi institute of Engineering and Technology

ABSTRACT

fraud Detection is a major concern in that it makes both the user and financial institutions lose money. This project develops a system using machine learning to determine whether certain transactions in a credit card are fraud or valid based on transaction patterns in both Debit , payment, cash-in, cash-out, and transfer methods.With machine learning algorithms, such as Gaussian Naive Bayes, Support Vector Machines (SVM), Logistic Regression, and Random Forest, being able to identify anomalies relating to fraud.

It suggests that the model actually detects suspicious transactions in real-time, thus eliminating unauthorized use and reducing financial loss for card holders and issuers.

KEY WORDS: Gaussian Naive Bayes, Support Vector Machines (SVM), Logistic Regression.

Introduction

Detection of credit card fraud is one of the more challenging issues in the digital economy. While we look at the benefits of shopping and paying by credit card during online transactions, we must also look at the potential for malicious economic activity. This can be very expensive not just to the card holders, but to the financial institutions as well.

Project Objective

The main aim of this project is to design an appropriate system for credit card fraud detection based on machine learning tools. We can determine abnormal transaction patterns which can alert us of abnormal activities as well. Machine learning algorithms are heuristic and non-deterministic and this attribute makes the model changes the way the learn and does allow huge amounts of data, resulting in efficient fraud detection each and every time.

Approach and Methodology

To address this problem, a range of machine learning methods is applied on the transaction dataset to detect any potentially fraudulent activities. This is how we do it:

Data Preprocessing: First, we try to sanitize the data, i.e., fill the missing gaps, correct errors, and reshape the data into a

form appropriate for analysis.

Dealing with Imbalanced Data: Fraudulent transactions as a whole only comprise a minority of transactions, therefore we deal with this issue by means of oversampling or the creation of synthetic data.

Feature Selection: We seek to pick out the most significant features that assist the model in predicting fraud, including transaction amount, timing, places of occurrence, and payment methods used.

Machine Learning Algorithms: Different algorithms like Gaussian Naive Bayes, Support Vector Machines (SVM), Logistic Regression, Random Forest are being examined for selection of the model that works the most efficiently.

Evaluation Metrics: To evaluate the performance of the model, we calculate accuracy, precision, recall, and F1 score. These metrics strike a balance between detecting fraud effectively without creating too many false alarms and ensuring that undetected fraud is kept to a minimum.

System Implementation: The system is designed to conduct real- time fraud transaction analysis, thus avoiding any further loss. The system is also scalable, so it is able to support the different levels of transaction sources such as, but not limited to, debit and credit card payments, cash-ins, cash-outs, and transfers. Finally, the system is adaptive to new types of fraud, so it always improves itself as new data comes in.

Impact and Significance:
This project illustrates the ability of machine learning to transform fraudulent activities by installing effective, scalable fraud stopping mechanisms that will increase the level of security for credit card transactions.mitigating financial exposure, blocking illegal transaction and giving confidence to use electronic payment systems and in the end enhancing the safety of financial systems in the world.

System Implementation
The system is designed to perform real- time analysis, so it can identify fraud transactions in the act and react immediately to limit further losses. It's scalable, too, so it can process high volumes of transaction information from different sources such as debit, credit card payments, cash-ins, cash-outs, and transfers.

Impact and Significance
This project demonstrates how machine learning can transform fraud detection by offering an efficient, scalable solution that improves the security of credit card transactions. The system assists in minimizing financial risks, preventing unauthorized transactions, and establishing confidence in digital payment systems, which leads to a secure financial ecosystem for all stakeholders.

LITERATURE SURVEY

The importance of detecting credit card fraud has gained in relevance over time due to the sophistication of schemes that compromise financial transactions. There have been a number of ghrauds that attempt to use machine learning methods for resolving challenges for fraud detection. This paper attempts to consolidate the key contributions in this topic.

1. Regular Methods of Fraud Detection: The preliminary methods for fraud detection were grounded in a rule based strategy where there was a reliance on predetermined limits both by thresholds and standards for the surfer suspicious transactions. Such systems were fairly easy to put in place, however, they were completely rigid and as such, they failed to withstand the complex and temporal fraud patterns leading to high false positives as well as false negatives.

2. Implementing AI into Fraud Detection: The inception of Tree algorithm has revolutionized fraud detection trends. Hence, the machine- learning fairly dominates as a stronger and widely suitable option, grasping patterns not seen by ordinary human behaviors within a big data set. For the categorization of fraudulent and valid transactions, mostly supervised learning methods are employed, thus far successful, mainly in the variations of Random Forests, Decision Trees, Logistic Regression, and Gradient Boosting. These models are trained and used for prediction over labeled data sets, where the previous transactions are labeled as either fraudulent or non-fraudulent.

3. **Imbalanced Dataset Problems:** Usually, the events are class-imbalanced, meaning fraud is always some much smaller class compared to the rest. To handle this, researchers such as Chawla, exist in adopting Random Under Sampling

4. **Anomaly Detection and Unsupervised Learning:** When we do not have any labeled data available, we utilize unsupervised learning methods like clustering and anomaly detection. Algorithms such as K means, Isolation Forest, Auto encoders identify outliers which can be an indication of fraud. These methods are of greatest value for finding new types of fraud that differ from previously noted fraud.

5. **Deep Learning Methods:** New breakthroughs in deep learning have been promising for fraud detection purposes. Empirical evidence has shown that deep learning models can be more responsive and more accurate than traditional methods.

6. **Evaluation Metrics:** The evaluation of fraud detection models requires metrics beyond accuracy, due to class imbalance. Measures such as Precision, Recall, F1-score, and Area Under the ROC Curve (AUC-ROC) are typically used to quantify the performance of machine learning models.

7. **Challenges and Opportunities:** Researchers have also pointed out challenges like data privacy, real-time processing requirements, and evolving fraud methods. There is also a growing popularity of hybrid approaches that combine supervised and unsupervised techniques or merge domain knowledge with machine learning for improved performance.

METHODOLOGY

1. Data Collection

Giving priority to well-established data processing related to credit card detection, data correctness and data variety is to be ensured. The historical transaction data ought to provide the model with both legal transactions and fraudulent ones. Such data can be chosen from out-of-the-box resources available to the public such as the Credit Card Fraud Detection Dataset on Kaggle or may originate from financial institutions in real life.

2. Data Preprocessing:

Cleansing: Involves handling missing values, duplicate entries, and unrelated features. For example, row
transactions with missing descriptions or inconsistent data details are either dropped or filled with imputed values.
Feature Engineering: Extract or create new features from raw data. This could include making categorical data use a numerical form like encoding categorical features and also normalizing transaction amounts. Handling Imbalanced Data: This mainly concerns fraudulent transactions that are very few in number.

3. Data Analysis and Feature Selection

1. **Exploratory Data Analysis (EDA)**: How to learn from the data? The researcher must visualize and analyze basic structure, distribution, the relationship between the features, and identify the correlation between various features.
2. **Feature Selection:** Procure the most accurate features using existing domain knowledge or automated features section approaches like Random Forest feature
The reduction of the number of features can improve the model effect

4. Model Development

Model training is the phase at which data preparation starts quickly to make sure that the datasets are balanced. This will help neutralize the class imbalance between fake records and genuine ones, innately supported by methods like random under sampling. In the process, features should be normalized or scaled; that is necessary for other machine learning algorithms, say Support Vector Machine (SVM) or Logistic Regression, the fit of which can be drastically changed by normalizing the dataset.

In the context of model training, the very first model to pick is Gaussian Naive Bayes (Gaussian NB) as a baseline; it is easy and fast to check the performance on the normalized dataset. The SVM follows next and employs kernel tricks (e.g., linear or radial basis function) for the capturing of complex relationships in the data, with appropriate hyper parameters Logistic Regression.

5: Model Evaluation and Optimization

Metrics elucidating class balance are particularly needed while evaluating trained models. High precision means a model correctly identifies a fraction of fraudulent transactions out of all flagged fraudulent transactions, which gives a clue about the model's correctness at predicting fraudulent.The model's full-time fraud detection capacity is indicated by the Recall ratio between fraudulent transactions correctly identified and overall transactions actually marked as fraudulent. “F1-score" refers to the harmonic average of precision and recall.

6: Real-Time Detection System

Once the model has been trained and validated, a system is built that can take the transaction data as input in real time. In this example, we build a continuous streaming data pipeline that consumes and processes transactions and predicts their probability of being fraudulent. As a result, potential fraud transactions can be flagged in real time and banks can take appropriate action (whether to block a transaction or notify the card owner)

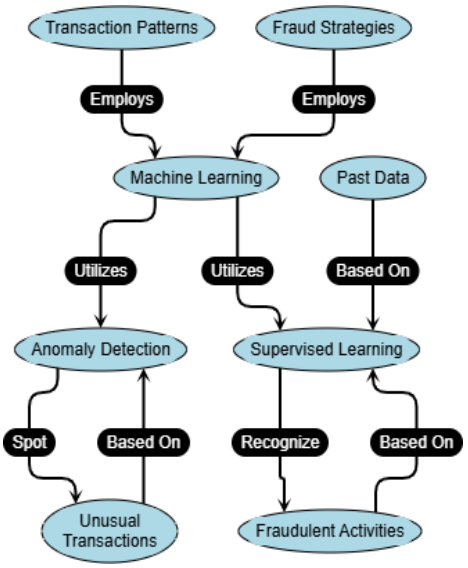
7 The Gradio Interface for Review and Moderation

Flask-based Review and Moderation Interface with Gradio Now that we can see both of our interfaces, let us look into a more user-friendly approach using Gradio, a python library for generating stunning web interfaces. Gradio allows you to build web interfaces for your models/data in a simple manner.

System Architecture

RESULT AND ANALYSIS

In the credit card fraud detection project,



we found that Random forest model outperformed logistic regression, support vector machines (SVMs), and other baseline approaches. Random Forest has its ensemble learning approach, in which predictions from several decision trees constructed using random subsets of features are combined. This not only decreases overfitting but also increases the model's capacity to generalize well, resulting in better performance on unseen transaction data.

A major advantage of this model, is its understanding of the context of transactions. the model examines more complex patterns in user behavior and transaction data through the Random Forest. It can detect subtle signs of fraud, like unusual spending habits or discrepancies between the origin and destination balances that would otherwise go unnoticed by more straightforward, rule-based approaches. it delivers a very huge result that ensures by success in getting to false alarms and getting to a best detection accuracy.

	PRECISION	RECALL	FISCORE	SUPPORT
Non - Fraudulent	0.68	0.94	0.79	2484
Fraud	0.90	0.56	0.69	2444
ACCURACY			0.75	4928
MACROAVG	0.79	0.75	0.74	4928
WEIGHTEDAVG	0.79	0.75	0.74	4928

Accuracy Table for GaussianNB model

	PRECISION	RECALL	FISCORE	SUPPORT
Non - Fraudulent	0.77	0.99	0.86	2484
Fraud	0.98	0.70	0.81	2444
ACCURACY			0.84	4928
MACROAVG	0.87	0.84	0.84	4928
WEIGHTEDAVG	0.87	0.84	0.84	4928

Accuracy Table for SVM model

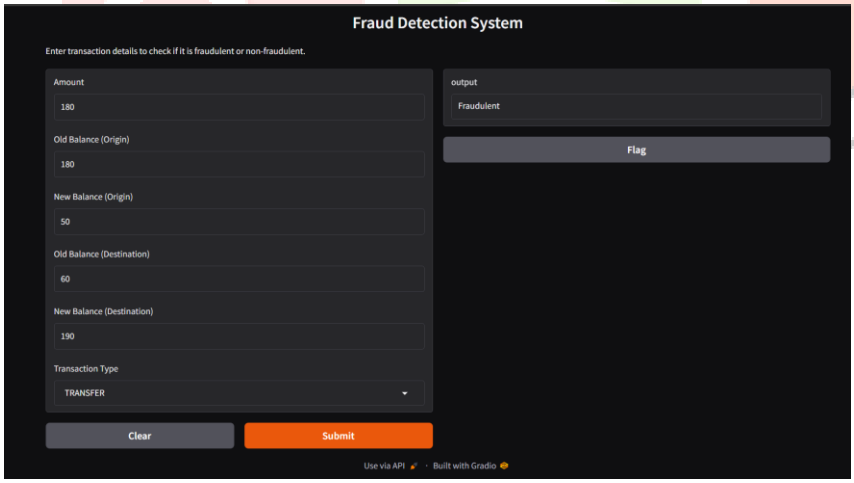
	PRECISION	RECALL	FISCORE	SUPPORT
Non - Fraudulent	0.92	0.90	0.91	2484
Fraud	0.90	0.92	0.91	2444
ACCURACY			0.91	4928
MACROAVG	0.91	0.91	0.91	4928
WEIGHTEDAVG	0.91	0.91	0.91	4928

Accuracy Table for Logistic Regression model

	PRECISION	RECALL	FISCORE	SUPPORT
Non - Fraudulent	1.00	0.99	0.99	2484
Fraud	0.99	1.00	0.99	2444
ACCURACY			0.99	4928
MACROAVG	0.99	0.99	0.99	4928
WEIGHTEDAVG	0.99	0.99	0.99	4928

Accuracy Table for Random Forest model

FINAL OUTPUT:



REFERENCES

[1] A. Ghosh, A. Reilly, Credit card fraud detection with a neural-network, Proc. Twenty- Seventh Hawaii Int. Conf. Syst. Sci., 3 (1994), 621–630.

[2] R. Jha, P. Sharma, S. Dutta, Machine learning algorithms for credit card fraud detection: A comparative study, J. Comput. Sci., 36 (2020), 101-111.

[3] P. Sharma, S. Shrivastava, K. Tiwari, Credit card fraud detection using machine learning: A review, Int. J. Comput. Sci. Eng., 8 (2021), 133–142.

[4] A. Dal Pozzolo, O. Caelen, Y. Le Borgne, et al., Learned lessons in credit card fraud detection from a practitioner perspective, Expert Syst. Appl., 41 (2014), 4915–4928.

[5] K. Pathak, D. Rawat, An ensemble learning approach for credit card fraud detection, Int. J. Data Mining Appl., 8 (2020), 99–110.

[6] S. Singh, M. Singh, Credit card fraud detection using machine learning models: Performance evaluation, Int. J. Recent Technol. Eng., 8 (2020), 573–578.

[7] A. Roy, S. Halder, Machine learning-based fraud detection in real-time transactions, J. Comput. Appl. Math., 12 (2022), 202–210.

[8] A. Patil, B. Kulkarni, Credit card fraud detection using random forest algorithm, Int. J. Eng. Res. Technol., 10 (2021), 1115–1121.

[9] A. Althoff, M. Maier, Enhancing fraud detection accuracy with deep learning models, Int. J. Soft Comput., 6 (2021), 25–35.

[10] J. Wang, S. Liu, S. Zhang, Improving fraud detection with data imbalance handling techniques, J. Inf. Technol., 43 (2021), 12–20.

[11] Y. Liu, J. Yang, J. Fang, Fraud detection using decision trees: An empirical study, J. Big Data, 4 (2022), 113–123.

[12] T. Fawcett, F. Provost, Adaptive fraud detection, Data Mining Knowl. Discov., 11 (1997), 291–316.

[13] R. Maheshwari, S. Mehta, Deep learning techniques for credit card fraud detection, Int. J. Comput. Sci. Inf. Syst., 14 (2020), 99–108.

[14] D. Patidar, M. Pandya, Credit card fraud detection using hybrid algorithms, J. Comput. Sci. Technol., 25 (2021), 77–85.

[15] A. Srivastava, A. Kundu, S. Sural, et al., Credit card fraud detection using hidden Markov model, IEEE Trans. Dependable Secure Comput., 5 (2008), 37–48.