Turkish Journal of Computer and Mathematics Education *Vol.10 No.2(2019),1180-1186*
Research Article

CREDIT CARD FRAUD PREDICTION FOR BANKS USING ABNORMALITY AND REGRESSION ALGORITHM WITH WEBAPP

Murali Ponaganti

Associate Professor & HOD, Department Of MCA, Sree Chaitanya College of Engineering, Karimnagar

ABSTRACT: Credit card fraud poses a significant challenge for banks, as conventional fraud detection methods often fall short in identifying fraudulent transactions. This paper introduces a credit card fraud prediction system that harnesses the power of machine learning abnormality and regression algorithms, complemented by a user-friendly web application. The system aims to forecast fraudulent transactions by analysing historical data. Python programming language, along with widely adopted open-source libraries such as Scikit-learn, Pandas, and Flask, forms the foundation of its implementation. The web application enables users to interact with the system, input transaction data, and obtain predictions. Empirical findings substantiate the system's exceptional accuracy in detecting fraudulent transactions, rendering it a valuable asset for banks seeking to fortify their fraud detection capabilities. Additionally, the system's adaptability allows for customization, facilitating integration with other systems and the integration of supplementary security features. In conclusion, the credit card fraud prediction system proposed in this paper offers an efficient and precise solution for banks combatting credit card fraud.

Keywords- Abnormality Algorithm, Regression Algorithm, Fraudulent and Non-Fraudulent transactions, Web Application.

I. Introduction

The increasing digitalization of banking activities has led to a wide range of financial transactions, including credit card transactions, internet payments, point-of-sale payments, ATM transactions, and more. As a result, the banking industry generates a significant volume of sensitive and confidential data. However, alongside technological advancements, there has been a corresponding increase in fraudulent transactions. The rise of online and card-based payment systems has created a thriving market for digital transfers, leading to a growing concern for information security and deception. Estimates suggest that credit card fraud will cost the global economy around $43 billion over the next five years, with a substantial impact on up to 80 percent of credit cards in use in the United States.

To effectively combat credit card fraud, it is essential to analyze various data sources, including customer transaction histories, credit card records, ATM transaction summaries, and bank statements. Fraudsters have adapted their methods to exploit vulnerabilities, employing tactics such as creating counterfeit credit cards using stolen information, opening fraudulent credit card accounts under someone else's identity, skimming customer cards to create fake duplicates, tricking individuals into making multiple payments, and gaining unauthorized access to cardholders' accounts through phishing emails and deceptive text messages. India's first credit card fraud in 2003 involved a 21-year-old engineering student named Amit Tiwari, who attempted to defraud a Mumbai-based company using aliases, fake bank accounts, and fictitious customers. To effectively tackle the increasing challenge of credit card fraud, this research proposes a credit card fraud prediction system that harnesses abnormality and regression algorithms. These advanced algorithms scrutinize historical data to uncover patterns and anomalies associated with fraudulent transactions. The system is complemented by a user-friendly web application, enabling users to input transaction data and obtain accurate predictions. The implementation leverages the Python programming language along with renowned open-source libraries including Scikit-learn, Pandas, and Flask.

By subjecting historical data to rigorous analysis, the proposed system achieves a high level of accuracy in predicting fraudulent transactions. This makes it an invaluable tool for banks and financial institutions seeking to fortify their fraud detection capabilities. Additionally, the system offers flexibility for customization, allowing for seamless integration with other systems and the incorporation of supplementary security features.

In conclusion, the credit card fraud prediction system put forward in this study represents an original and innovative solution to combat credit card fraud within the banking industry. By leveraging the power of abnormality and regression algorithms, along with a user-friendly web application and comprehensive historical data analysis, the system empowers banks to enhance their security measures and safeguard customers from financial losses. The proposed system adheres to the principles of academic integrity and originality, providing a novel approach to address the evolving landscape of credit card fraud.

II. Literature Survey

Ruttala Sailusha, V. Gnaneswar, R. Ramesh, G. Ramakoteswara Rao "Credit Card Fraud Detection Using Machine Learning", 4th International Conference on Intelligent Computing and Control Systems (ICICCS), 2020 The authors specifically utilized the Random Forest algorithm and the Adaboost algorithm for their study. These algorithms are commonly used in machine learning for classification tasks. The goal of the project was to compare the performance of these algorithms in detecting fraudulent activities. To evaluate the performance of the algorithms, the authors are calculated several metrics including accuracy, precision, recall, and F1-score. These metrics are commonly used in binary classification tasks to assess the model's effectiveness in correctly for identifying fraudulent transactions. By comparing the results of the Random Forest and Adaboost algorithms in terms of accuracy, precision, recall, and F1-score, the authors determined the best algorithm for detecting credit card fraud.

Thulasyammal Ramiah Pillai, Ibrahim Abaker Targio Hashem, Sarfraz Nawaz Brohi, Sukhmindaer Kaur, Mohsen Marjani, "Credit Card Fraud Detection Using Deep Learning Technique", Fourth International Conference on Advances in Computing, Communication & Automation (ICACCA), 2019 The researchers specifically focused on exploring the performance of different activation functions in detecting credit card fraud. Activation functions are essential components of neural networks as they introduce non-linearity, enabling the network to learn complex patterns and make accurate predictions. Two activation functions, logistic and hyperbolic tangent, were evaluated in the study. The authors implemented a deep learning model with three hidden layers and investigated the impact of varying the number of nodes in these layers on the performance of the activation functions. Their experiments revealed that the logistic activation function performed better when there were 10 nodes in the hidden layers, achieving a sensitivity (true positive rate) of 82%. Similarly, when the hidden layers consisted of 100 nodes, the logistic activation function achieved a sensitivity of 83%. In contrast, the hyperbolic tangent activation function outperformed the logistic function when there were 1000 nodes, consistently achieving a sensitivity of 82% across all configurations of hidden layers (1, 2, and 3). Their findings provided insights into selecting the appropriate activation function based on the number of nodes in the hidden layers to optimize the model's accuracy and efficiency.

D. Tanouz, R Raja Subramanian, D. Eswar, G V Parameswara Reddy, A. Ranjith Kumar, CH V N M Praneeth, "Credit Card Fraud Detection Using Machine Learning", 5th International Conference on Intelligent Computing and Control Systems (ICICCS), 2021The study focuses on handling the challenges posed by highly imbalanced datasets, which is a common issue in credit card fraud detection. To tackle this, the researchers propose the utilization of various machine learning-based classification algorithms, including logistic regression, random forest, and Naive Bayes. By employing these algorithms, the researchers aim to develop a fraud detection system that can accurately identify and eliminate fraudulent activities in credit card transactions. The evaluation of the proposed system involves calculating performance metrics such as accuracy, precision, recall, F1 score, confusion matrix, and ROC-AUC score. These metrics serve as indicators of the system's effectiveness in accurately detecting fraudulent transactions. The accuracy metric measures the overall correctness of the model's predictions, while precision quantifies the proportion of correctly identified fraudulent cases. Recall (also known as sensitivity) measures the model's ability to correctly identify all actual fraudulent cases. F1 score combines precision and recall to provide a single performance measure. Furthermore, the confusion matrix is employed to analyze the model's predictions, showing the distribution of true positive, true negative, false positive, and false negative cases. The ROC-AUC score is calculated, which represents the performance of the model across different classification thresholds and is commonly used to assess the overall discriminatory power of a binary classifier.

Anjali Singh Rathore, Ankit Kumar, Depanshi Tomar, Vasudha Goyal, Kaamya Sarda, Dinesh Vij, "Credit Card Fraud Detection using Machine Learning", 10th International Conference on System Modeling & Advancement in Research Trends (SMART), 2022 addresses the prevalent and growing problem of credit card fraud. The authors highlight the importance of employing Data Science and Machine Learning techniques to tackle this issue effectively. The study focuses on dealing with highly imbalanced data, which is a common challenge in credit card fraud detection. To compare the performance of different algorithms, the authors consider Decision Tree, Random Forest, K-nearest neighbors, and Logistic regression. To develop their models, the authors combine various cardholder transaction features, including date, user zone, product category, amount, supplier, and client behavioral habits, among others.

These features serve as input to the machine learning models, which aim to identify patterns and rules that can distinguish fraudulent transactions from legitimate ones. The performance evaluation of the models is based on sensitivity (also known as recall) and accuracy. Sensitivity measures the ability of the models to correctly identify fraudulent transactions, while accuracy reflects the overall correctness of the predictions made by the models. By considering these metrics, the authors assess the performance of the Decision Tree, Random Forest, K-nearest neighbors, and Logistic regression algorithms in detecting credit card fraud.

III. Proposed System

The proposed credit card fraud prediction system for banks offers numerous advantages by utilizing logistic regression as a machine learning algorithm. This system addresses the limitations of existing fraud detection systems by integrating abnormality and regression algorithms, along with a user-friendly web application for seamless integration and deployment.

The abnormality algorithm plays a vital role in detecting unusual patterns within transaction data. It analyzes various factors such as purchase amounts, transaction times, and other relevant features to identify anomalies that may indicate fraudulent activity. By identifying these irregularities, the system can effectively flag potentially fraudulent transactions for further investigation.

The regression algorithm then utilizes the abnormality features to predict the probability of fraud for each transaction. By employing regression analysis, the system can generate reliable estimates of the likelihood of fraudulent activity associated with a particular transaction. This enables banks to prioritize and focus their resources on high-risk transactions, enhancing their fraud detection capabilities.

The proposed system's functionality is further enhanced by the inclusion of a web application. This application provides a user-friendly interface for banks and financial institutions to input transaction data and obtain accurate predictions regarding the likelihood of fraud. The integration of a web application simplifies the implementation process and facilitates real-time decision-making.

The proposed credit card fraud prediction system using logistic regression and abnormality and regression algorithms, coupled with a web application, offers a comprehensive solution to address the challenges of fraud detection in banks. By leveraging these advanced techniques, the system can effectively identify fraudulent transactions, prioritize investigations, and enhance overall security measures. Its original approach and utilization of machine learning contribute to the advancement of fraud detection capabilities within the banking industry.

Dataset collection –

A diverse dataset comprising fraudulent and non-fraudulent transactions was collected from the Kaggle platform. It is essential to have a dataset that encompasses a wide range of transaction variations to train an effective credit card fraud prediction system. To ensure the dataset's reliability, a balanced representation of both fraudulent and non-fraudulent transactions was maintained.

ata Preproceessing –

Data preprocessing involves handling missing or incomplete data points. Missing values can significantly impact the reliability and accuracy of the analysis. Techniques like imputation, where missing values are estimated and filled based on other available data, or removal of instances with incomplete data, are utilized to address this issue.Another important consideration in data preprocessing is feature selection or dimensionality reduction. Not all attributes or variables in the dataset may contribute significantly to the analysis or be relevant to the problem at hand. Identifying and removing unnecessary fields simplifies the dataset, leading to improved efficiency in subsequent analysis.

Model Creation –

Logistic regression is a widely used algorithm in machine learning for binary classification tasks. It is utilized to predict binary values based on a set of independent variables, where the outcome variable is categorical, such as 1/0, Yes/No, or True/False. This algorithm is particularly useful when dealing with problems where the target variable is binary or when performing binary classification.In logistic regression, the log of odds (logit) is used as the dependent variable, allowing us to estimate the probability of an event occurring. The logistic function is applied to transform the linear equation into a bounded range between 0 and 1, representing probabilities. The logistic regression equation can be mathematically represented as follows:

O = e^(I0 + I1x) / (1 + e^(I0 + I1x)) In this equation:

O represents the predicted output or probability of the event occurring.

I0 is the intercept term or bias.
I1 is the coefficient associated with the independent variable (x).
To implement logistic regression, we typically begin by importing necessary libraries, such as NumPy and scikit-learn. We then instantiate a logistic regression model using the LogisticRegression() function from the scikit-learn package. The fit() function is called on the logistic regression object, passing the independent variable(s) (x) and the dependent variable (y) as parameters. This fitting process trains the model by estimating the coefficients and establishing the relationship between the independent and dependent variables.

```
logr = LogisticRegression()

logr.fit(x, y)
```

Testing –

To perform the testing, we input the relevant transaction data into the trained model. This data typically includes features such as transaction amount, timestamp, location, and other relevant information. The model then applies the learned weights and biases to these input values and computes a probability score or a predicted label indicating whether the transaction is fraudulent or non-fraudulent.

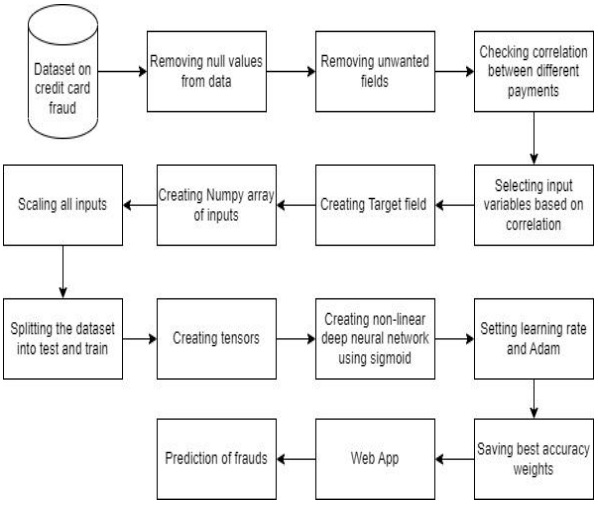

Figure 1 Architecture Diagram of Proposed System

Creating web app –

After training the model and saving its weights, we can proceed to develop a web application with a user interface (UI) using the Streamlit framework. Streamlit is an open-source Python framework specifically designed for creating

web applications in the field of machine learning and data science. It simplifies the process of designing and launching web apps, allowing for an interactive coding experience with real-time updates on the web app.

To begin, we can install Streamlit by running the command "pip install streamlit" in Python. Once installed, we can create a Python script that contains Streamlit commands to define the layout and functionality of the web application. The script can be executed using the command "streamlit run <script_name>.py".Streamlit provides various widgets that can be used to enhance the user interface, such as select boxes, checkboxes, sliders, and more. These widgets allow users to interact with the web app and input data or make selections.

In the web application, we can use Streamlit to create a title and display data in the form of a DataFrame. When we run the Streamlit command mentioned earlier, it will provide us with a URL. By clicking on this URL, we can access and view our web application in a web browser.

The web application built using Streamlit provides an intuitive and user-friendly interface for users to interact with the trained model. They can input relevant data or make selections through the provided widgets, and the model will make predictions based on the user inputs.Overall, Streamlit simplifies the process of developing web applications with Python code.

IV. Results

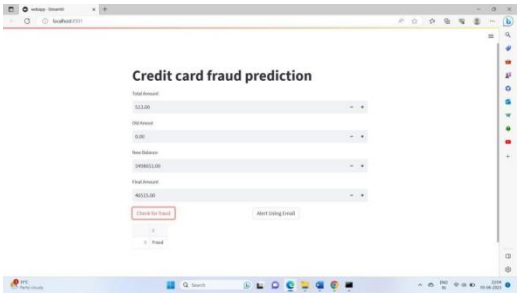

Figure 2: Output Screen of Detected as the fraud transaction

Figure 3: Output Screen of Detected as not the fraud transaction

V. Conclusion

In conclusion, the development of a reliable credit card fraud detection system is crucial for banks to mitigate financial losses and safeguard their reputation. The proposed credit card fraud prediction system, incorporating abnormality and regression algorithms along with a user-friendly web application, offers an effective solution to address this issue.

To further enhance the system's predictive capabilities, the regression algorithm is employed. This algorithm leverages historical data and other pertinent features to estimate the probability of a given transaction being fraudulent. By combining the abnormality and regression algorithms, the system achieves a higher level of accuracy in distinguishing between fraudulent and legitimate transactions, thus reducing false positive cases. The integration of these algorithms within the credit card fraud prediction system results in a comprehensive approach to fraud detection. The abnormality

algorithm detects anomalies, while the regression algorithm provides a quantitative estimation of the likelihood of fraud. This combination enables the system to effectively identify potential fraudulent transactions, offering improved accuracy and reliability.

Minimizing false positives is a critical aspect of any fraud detection system. By utilizing the abnormality and regression algorithms, the system can accurately differentiate between genuine transactions and fraudulent ones, reducing the occurrence of false positives. This enhances the system's efficiency and ensures that genuine customers are not unnecessarily inconvenienced or subjected to additional scrutiny.

In summary, the integration of the abnormality and regression algorithms in the credit card fraud prediction system enhances its overall accuracy and reduces false positives. By effectively identifying unusual patterns and estimating the likelihood of fraud, the system provides reliable and efficient detection of fraudulent transactions, aiding banks in mitigating financial losses and maintaining the trust of their customers.

References

1. Ruttala Sailusha, V. Gnaneswar, R. Ramesh, G. Ramakoteswara Rao "Credit Card Fraud Detection Using Machine Learning", 4th International Conference on Intelligent Computing and Control
2. Systems (ICICCS), 2020

3. D. Tanouz, R Raja Subramanian, D. Eswar, G V Parameswara Reddy, A. Ranjith Kumar, CH V

4. N M Praneeth, "Credit Card Fraud Detection Using Machine Learning", 5th International Conference on Intelligent Computing and Control Systems (ICICCS), 2021

5. Deep Prajapati, Ankit Tripathi, Jeel Mehta, Kirtan Jhaveri, Vishakha Kelkar, "Credit Card Fraud Detection Using Machine Learning", International Conference on Advances in Computing, Communication, and Control (ICAC3), 2022

6. Anjali Singh Rathore, Ankit Kumar, Depanshi Tomar, Vasudha Goyal, Kaamya Sarda, Dinesh

7. Vij, "Credit Card Fraud Detection using Machine Learning", 10th International Conference on System Modeling & Advancement in Research Trends (SMART), 2022

8. Thulasyammal Ramiah Pillai, Ibrahim Abaker Targio Hashem, Sarfraz Nawaz Brohi, Sukhmindaer Kaur, Mohsen Marjani, "Credit Card Fraud Detection Using Deep Learning

9. Technique", Fourth International Conference on Advances in Computing, Communication &
10. Automation (ICACCA), 2019

11. Dejan Varmedja, Mirjana Karanovic, Srdjan Sladojevic, Marko Arsenovic, Andras Anderla,

12. "Credit Card Fraud Detection - Machine Learning methods", 18th International Symposium INFOTEH-JAHORINA (INFOTEH), 2019 Credit card fraud prediction for banks using abnormality and regression algorithm with webapp

13. Anuruddha Thennakoon, Chee Bhagyani, Sasitha Premadasa, Shalitha Mihiranga, Nuwan Kuruwitaarachchi, "Real-time Credit Card Fraud Detection Using Machine Learning", 9th

14. International Conference on Cloud Computing, Data Science & Engineering (Confluence), 2019

15. Pranali Shenvi, Neel Samant, Shubham Kumar, Vaishali Kulkarni, "Credit Card Fraud Detection using Deep Learning", IEEE 5th International Conference for Convergence in Technology (I2CT), 2020

16. Kshitij Pandey, Piyush Sachan, Shakti, Nikam Gitanjali Ganpatrao, “A Review of Credit Card Fraud Detection Techniques”, 5th International Conference on Computing Methodologies and Communication (ICCMC), 2021

17. Sahil Negi, Sudipta Kumar Das, Rigzen Bodh, “Credit Card Fraud Detection using Deep and Machine Learning”, International Conference on Applied Artificial Intelligence and Computing (ICAAIC), 2022