Lecture Notes in Networks and Systems 1033

Alhamzah Alnoor · Mark Camilleri ·
Hadi A. Al-Abrrow ·
Marco Valeri · Gül Erkol Bayram ·
Yousif Raad Muhsen *Editors*

Explainable Artificial Intelligence in the Digital Sustainability Administration

Proceedings of the 2nd International Conference on Explainable Artificial Intelligence in the Digital Sustainability Administration (AIRDS 2024)

Springer

Lecture Notes in Networks and Systems 1033

The series "Lecture Notes in Networks and Systems" publishes the latest developments in Networks and Systems—quickly, informally and with high quality. Original research reported in proceedings and post-proceedings represents the core of LNNS.

Volumes published in LNNS embrace all aspects and subfields of, as well as new challenges in, Networks and Systems.

The series contains proceedings and edited volumes in systems and networks, spanning the areas of Cyber-Physical Systems, Autonomous Systems, Sensor Networks, Control Systems, Energy Systems, Automotive Systems, Biological Systems, Vehicular Networking and Connected Vehicles, Aerospace Systems, Automation, Manufacturing, Smart Grids, Nonlinear Systems, Power Systems, Robotics, Social Systems, Economic Systems and other. Of particular value to both the contributors and the readership are the short publication timeframe and the worldwide distribution and exposure which enable both a wide and rapid dissemination of research output.

The series covers the theory, applications, and perspectives on the state of the art and future developments relevant to systems and networks, decision making, control, complex processes and related areas, as embedded in the fields of interdisciplinary and applied sciences, engineering, computer science, physics, economics, social, and life sciences, as well as the paradigms and methodologies behind them.

Indexed by SCOPUS, INSPEC, WTI Frankfurt eG, zbMATH, SCImago.

All books published in the series are submitted for consideration in Web of Science.

For proposals from Asia please contact Aninda Bose (aninda.bose@springer.com).

Alhamzah Alnoor · Mark Camilleri ·
Hadi A. Al-Abrrow · Marco Valeri ·
Gül Erkol Bayram · Yousif Raad Muhsen
Editors

Explainable Artificial Intelligence in the Digital Sustainability Administration

Proceedings of the 2nd International Conference on Explainable Artificial Intelligence in the Digital Sustainability Administration (AIRDS 2024)

Springer

Editors

Alhamzah Alnoor
Southern Technical University
Basrah, Iraq

Hadi A. Al-Abrrow
University of Basrah
Basrah, Iraq

Gül Erkol Bayram
Sinop University
Sinop, Türkiye

Mark Camilleri
University of Malta
Msida, Malta

Marco Valeri
Niccolò Cusano University
Rome, Italy

Yousif Raad Muhsen
Wasit University
Wasit, Iraq

ISSN 2367-3370 ISSN 2367-3389 (electronic)
Lecture Notes in Networks and Systems
ISBN 978-3-031-63716-2 ISBN 978-3-031-63717-9 (eBook)
https://doi.org/10.1007/978-3-031-63717-9

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

Climate change and sustainable development goals are the most pressing global challenges facing society today, with potentially detrimental impacts on individuals, organizations, and societies. The impact of digital technologies on climate change is one of the key research priorities and is considered a researchable spot. Climate-intelligent information systems solutions and environmental, social, and governance intelligence leverage the transformative power of information systems to mitigate adverse environmental impacts. Information systems researchers can use these solutions to create practical impact. Furthermore, digital sustainability activities can advance environmental sustainability goals by creatively deploying technologies that create, use, or transmit electronic source data. The 2nd International Conference (AIRDS 2024) is held to address the theme "Explainable Artificial Intelligence in the Digital Sustainability Administration." The AIRDS 2024 brings together a wide range of researchers from different disciplines. It seeks to call for research contributions to mitigate and adapt to the effects of climate change, as it could cause far-reaching disruptions to communities and the economy worldwide. The main aim of the AIRDS 2024 is to provide a forum for academics, researchers, and developers from academia and industry to share and exchange their latest research contributions and identify practical implications of emerging technologies to advance the wheel of these solutions for global impact. In line with the Fourth Industrial Revolution goals and its impact on sustainable development, AIRDS 2024 is devoted to increasing the understanding and impact of explainable artificial intelligence on individuals, organizations, and societies and how artificial intelligence applications have recently reshaped these entities. In addition to the contribution of explainable artificial intelligence applications to achieve sustainable development goals.

The AIRDS 2024 attracted 89 submissions from different countries worldwide. Out of the 89 submissions, we accepted 26, representing an acceptance rate of 29.21%. The chapters explore the diverse implications of explainable artificial intelligence in various professional and social sectors. It opens with insights from explainable machine learning for real-time payment fraud detection and introduces an advanced machine learning model integrated with explainable artificial intelligence techniques to enhance the detection of payment fraud in real-time scenarios within the digital finance sector. It scrutinizes the potential impacts of robotic process automation on sustainable audit quality, including both opportunities and challenges. This volume also assesses the strategic potential of artificial intelligence in improving customer service and facilitating the retention of organizational human resources. Education continues to be a key theme, with discussions on the factors driving schoolteachers to adopt virtual educational resources to elevate the educational process. Furthermore, the chapters delve into the evolving landscape of digital sustainability within artificial intelligence, considering brand activity and consumer behavior. The book addresses serious considerations of intelligent agriculture decision support tools, artificial intelligence applications in advertising, energy reduction, big data analytics, the Internet of things, classifying solar radiation time series, and a unified technology acceptance model in boosting digital sustainability. The financial sector is not left out, with a case study explaining artificial intelligence's role in sustainable audit quality and financial sustainability. Finally, the volume looks at how artificial intelligence furthers the sustainability of healthcare systems by improving efficiency, disease diagnosis, optimizing resources, and developing in-person and remote care initiatives. Each chapter offers a distinct perspective, providing readers with a well-rounded understanding of the challenges and prospects of artificial intelligence in the digital sustainability administration.

Each submission is reviewed by at least two reviewers, who are considered experts in the related submitted paper. The evaluation criteria include several issues: correctness, originality, technical strength, significance, presentation quality, interest, and relevance to the conference scope. The conference proceedings are published in *Lecture Notes in Networks and Systems Series* by Springer, which has a high SJR impact. We acknowledge all those who contributed to the success of AIRDS 2024. We would also like to thank the reviewers for their valuable feedback and suggestions. Without them, it was impossible to maintain the high quality and success of AIRDS 2024.

Alhamzah Alnoor
Mark Camilleri
Hadi A. Al-Abrrow
Marco Valeri
Gül Erkol Bayram
Yousif Raad Muhsen

Organization

Conference General Chairs

Hadi AL-Abrrow	Department of Business Administration, College of Administration and Economics, University of Basrah, Basrah, Iraq
Camilleri	University of Malta, Msida, Malta
Alhamzah Alnoor	Southern Technical University, Basrah, Iraq

Honorary Conference Chair

Mohanad J. K. Al Asadi	Chancellor of University of Basrah, Basrah, Iraq

Conference Organizing Chair

Abdul Hussain Tawfiq Shibli	Dean of College of Administration and Economics, University of Basrah, Basrah, Iraq

Program Committee Chair

Marco Valeri	Faculty of Economics, Niccolo’ Cusano University in Rome, Italy

Publication Committee Chairs

Alhamzah Alnoor	Management Technical College, Southern Technical University, Basrah, Iraq
Mark Camilleri	Faculty of Media & Knowledge Sciences, University of Malta, Msida, Malta
Hadi Al-Abrrow	Department of Business Administration, College of Administration and Economics, University of Basrah, Basrah, Iraq
Marco Valeri	Niccolo’ Cusano University, Italy

Gül Erkol Bayram	School of Tourism and Hospitality Management Department, Sinop University, Sinop, Turkey
Yousif Raad Muhsen	Civil Department, College of Engineering, Wasit University, Wasit, Iraq

Conference Tracks Chairs

Sammar Abbas	Institute of Business Studies, Kohat University of Science and Technology, Pakistan
Marcos Ferasso	Economics and Business Sciences Department, Universidade Autónoma de Lisboa, 1169-023 Lisboa, Portugal
Hussam Al Halbusi	Department of Management at Ahmed Bin Mohammed Military College, Doha, Qatar
Khai Wah Khaw	School of Management, Universiti Sains Malaysia, 11800, Pulau Pinang, Malaysia
Yousif Raad Muhsen	Civil Department, College of Engineering, Wasit University, Wasit, Iraq
Gül Erkol Bayram	School of Tourism and Hospitality Management Department of Tour Guiding, Sinop University, Sinop, Turkey
Gadaf Rexhepi	South East European University, Tetovo, The Republic of Macedonia

Members of Scientific Committee

Muntader F. Saad	Head of the Dept. of Financial Sciences, University of Basrah, Basrah, Iraq
Hadi AL-Abrrow	Department of Business Administration, College of Administration and Economics, University of Basrah, Basrah, Iraq
Khai Wah Khaw	School of Management, Universiti Sains Malaysia, Malaysia
Amjad S. Abdulaali	Dept. of Economics, University of Basrah, Basrah, Iraq
Suhail A. Nasser	Dept. of Accounting, University of Basrah, Basrah, Iraq
Gadaf Rexhepi	Southern European University, The Republic of Macedonia

Wameedh A. Khdair	Department of Business Administration, College of Administration and Economics, University of Basrah, Basrah, Iraq
Sammar Abbas	Institute of Business Studies, Kohat University of Science and Technology, Pakistan
Elham J. Hamid	Dept. of Accounting, University of Basrah, Basrah, Iraq
Walid M. Rudin	Head of Admin. Info. Systems Dept, University of Basrah Basrah, Iraq
Rabee K. Thajeel	Dept. of Economics, University of Basrah, Basrah, Iraq
Baha A. Qasim	Head of Statistics Dept, University of Basrah, Basrah, Iraq
Gul Erkol Bayram	Sinop University, School of Tourism and Hospitality Management Department of Tour Guiding, Turkey
Mohammed J. Mohammed	Dept. of Financial Sciences, University of Basrah, Basrah, Iraq
Ali N. Hussein	Dept. of Statistics, University of Basrah, Basrah, Iraq
Faiza Hassan	Dept. of Financial Sciences, University of Basrah, Basrah, Iraq
Alhamzah Alnoor	Management Technical College, Southern Technical University, Basrah, Iraq
Hussan Al Halbusi	Department of Management, Ahmed Bin Mohammed Military College, Qatar
Marcos Ferasso	Economic and Business Sciences Department, Universidade Autonoma de Lisboa, Portugal

Publicity and Public Relations Committee

Ammar Y. Dhicher	Dean Assistant for Scientific Affairs, University of Basrah, Basrah, Iraq
Mohanad H. Saleh	Dean Assistant for Students Affairs, University of Basrah, Basrah, Iraq
Hadi AL-Abrrow	Department of Business Administration, College of Administration and Economics, University of Basrah, Basrah, Iraq

Finance Chair

Abdul Hussain Tawfiq Shibli — Dean of College of Administration and Economics, University of Basrah, Basrah, Iraq

Contents

Explainable Machine Learning for Real-Time Payment Fraud Detection: Building Trustworthy Models to Protect Financial Transactions

Ahmed Abbas Jasim Al-hchaimi[1], Mohammed F. Alomari[2], Yousif Raad Muhsen[3,4], Nasri Bin Sulaiman[5](✉), and Sabah Hassan Ali[6]

[1] Thiqar Technical College, Southern Technical University, Basrah, Iraq
ahmed.alhchaimi@stu.edu.iq
[2] College of Graduate Studies, Universiti Tenaga National (UNITEN), Kajang, Malaysia
[3] Faculty of Engineering, Universiti Putra Malaysia, Serdang, Selangor, Malaysia
yousif@uowasit.edu.iq
[4] Civil Department, College of Engineering, Wasit University, Kut, Wasit, Iraq
[5] Faculty of Engineering, Universiti Putra Malaysia, Serdang, Malaysia
nasri_sulaiman@upm.edu.my
[6] Iraqi Ministry of Interior, Baghdad, Al Muthanna, Iraq
sabah.hassan@mu.edu.iq

Abstract. In this study, we introduce an advanced machine learning model integrated with explainable AI techniques to enhance the detection of payment fraud in real-time scenarios within the digital finance sector. As online transactions continue to proliferate, so too do the fraudulent activities associate with them. Our approach effectively differentiates between legitimate and fraudulent transactions by meticulously analyzing key features such as transaction amount, type, and the accounts involved. Through a comprehensive evaluation of various machine learning models, the Decision Tree model emerged as the most effective, achieving an accuracy of 95.4048%, precision of 92.9461%, recall of 98.2456%, and an F1-score of 95.5224%. This study not only proposes a robust and explainable machine learning framework but also significantly enhances the transparency of fraud detection decisions. It equips financial institutions with a potent tool to safeguard their customers' assets against fraud, thereby bolstering the reliability and trustworthiness of digital payment systems.

Keywords: Digital Finance Security · Real-time Fraud Analysis · Decision-Tree-Model · Transaction Analysis

1 Introduction

In the swiftly evolving landscape of digital finance, the advent of online transactions has brought unparalleled convenience and efficiency to consumers worldwide. Credit cards, once a luxury, are now a cornerstone of everyday commerce, facilitating billions of transactions across the globe (Al-Enzi et al., 2023; Husin et al., 2023; Atiyah

A. Alnoor et al. (Eds.): AIRDS 2024, LNNS 1033, pp. 1–25, 2024.
https://doi.org/10.1007/978-3-031-63717-9_1

et al., 2023, b, c; Soltani et al., 2023). The rapid ranging of these digital payments coupled with the increasing number of fraudulent activities has brought great challenges to financial systems (Alnoor et al., 2024, b, c). Another vulnerable point of financial institutions is financial fraud (Fig. 1), for instance payment fraud, which damages online payments security and many customers' faith. Statistically, The Federal Trade Commission represent-ed that more than 3.7 billion dollars was lost to credit card fraud in 2020 (Wright et al., 2020; Ahmed et al., 2024; Avila-Cano et al., 2023).

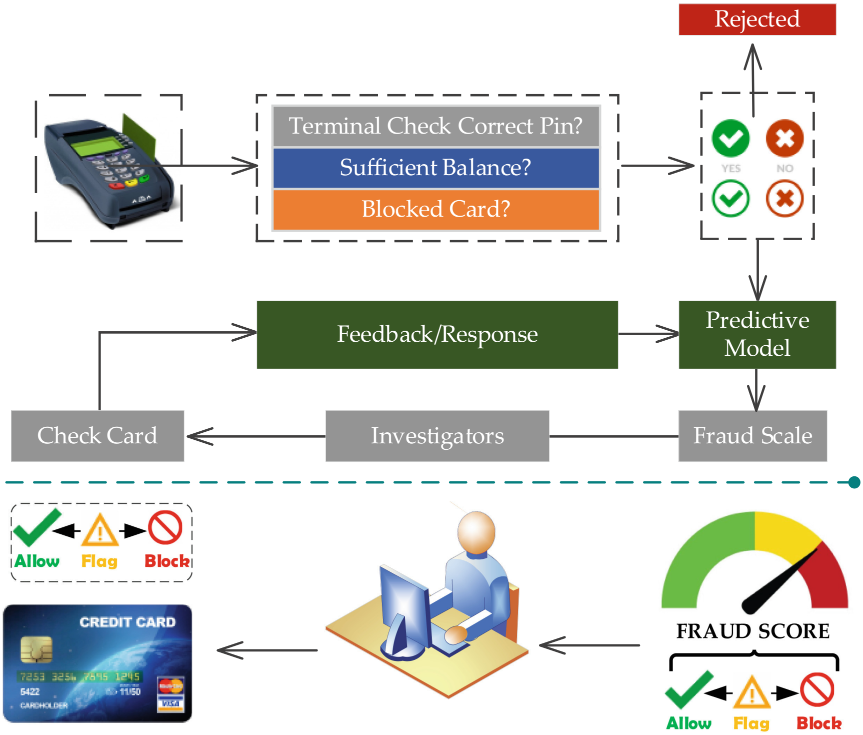

Fig. 1. Typical Online Money Transaction Process.

This demonstrates the need for efficient fraud identification technology. The technology standards that exist today find it hard to meet both the complexities and the volume of transactions while at the same time assuring seamless user experience associated with digital payment and seamless user experience (Al-Hchaimi et al., 2022, 2023; Muhsen et al., 2023; Atiyah et al., 2023, b, c; Innan et al., 2024). The fight against the payment fraud is still complicated by many difficulties. First, the development of fraud techniques is accelerating rapidly - fraudsters are always outpacing traditional controls, constantly creating new ways by which fraud could penetrate the screening mechanisms. Further, the figure of the pecuniary chases the detection process as every transaction has its particular unevenness. Traditional systems find it difficult to differentiate between legitimate and fraudulent transactions and make many true-false decisions. This negatively affects the completion of many legitimate transactions and leads to discontent of the customers. One more issue which demands urgent attention is immediate detection of threat. The fraud detection process can take a long time, which is when the fraudulent transaction happens, before the intervention can be made, financial losses occur and logistical nightmares persist in trying to explain them (Muhsen et al., 2023a; Abbas et al., 2023).

Considering these obstacles, the purpose of this project would be the creation of a model that is able to explain fraudulent payments and identifies them before they can be processed. Transactional properties including the amount, the type, and the account numbers are going to be analyzed and a model will be trained using dataset which contains both fraudulent and non-fraudulent transactions (Chen et al., 2019; Ni et al., 2023; Nijwala et al., 2023; Ali et al., 2024; Alnoor et al., 2023; Alnoor et al., 2024, b, c; Cai et al., 2024). The target of it is to provide a highly accurate solution. It is of great importance for financial sector to deliver this accuracy because it helps them to save money, and it is also trusted by their customers which is a vital component of financial markets.

The widespread phenomenon of payment fraud not only becomes more widespread but also, more technologically complex, make it an important research area to address this problem. The financial and psychological consequences that the individual victims who suffered the attack had to experience in addition to the impact of this incident on the overall public trust in online payment systems creating a consistent necessity for advanced detection techniques. This initiative is based on a firm conviction that bringing about honest AI in detecting fraud is capable of changing how financial institutions pick and trip on fraudulent transmissions. With improved transparency and comprehension of these processes, stakeholders will be able to directly participate in the detection of inconsistencies (Ahmadi, 2023; Alnoor et al., 2024, b, c; Pallathadka et al., 2023). This will make electronic payment solutions more relevant to a broader user base, resulting in a higher level of trust in the technology as well as its widespread adoption. The main contribution of this paper can be summarized as follows:

(i) Proposes a novel approach that integrates explainability into the ML model, ensuring that decisions are transparent and justifiable.
(ii) Demonstrates the model's effectiveness in discerning fraudulent transactions as they occur, markedly reducing the potential for financial loss.
(iii) Enhances the model's ability to recognize and learn from complex patterns of fraud.
(iv) Offers a comprehensive evaluation of the model's predictive accuracy, leveraging a diverse dataset to benchmark against existing fraud detection methodologies.
(v) Outlines actionable guidelines for integrating the developed model into existing fraud detection frameworks, ensuring that financial institutions can seamlessly adopt and benefit from this research.
(vi) Aims to empower financial institutions with advanced tools to safeguard against fraudulent transactions, restoring confidence in the security of digital payment platforms.

The reminder of this paper is outlined as: Sect. 2 parents the literature review. Section 3 investigates the research methodology, furthermore, Sect. 5 shows the results and discussion. Finally, Sect. 5 concludes this paper.

2 Literature Review

The prevalence of credit card fraud and its evolving complexity have made it a focal point for numerous studies aiming to harness advanced technologies for more effective detection solutions. Authors of (Cherif et al., 2023) critically evaluated research up to 2021, revealing a considerable exploration gap in deploying deep learning methods for credit card fraud detection. This gap underscores the necessity for novel approaches capable of handling the intricate patterns of fraudulent transactions. For instance, authors of (Asha et al., 2021) demonstrated the superiority of ANN models over traditional SVM and KNN models in accuracy, precision, and recall, yet highlighted the challenges of class imbalance and the need for undersampling strategies. Table 1 lists more details about the transaction fraud and utilized ML approaches for detection, mitigation, and prediction.

This hypothetical table compares various methodologies, from traditional machine learning techniques such as Logistic Regression (LR), Decision Trees (DT), and Support

Table 1. Summary of Prior Studies in the Context of Online Transaction Fraud.

Ref	ML Type	Algorithm(s)	Domain	Dataset	Strengths	Limitations
(Asha et al., 2021)	Traditional	ANN vs. KNN and SVM	Credit card	Public (Kaggle)	High accuracy (0.9992)	Low recall; potential information loss due to under-sampling
(Singh et al., 2023)	Traditional	Shallow NN	Credit card fraud	Public (ULB)	Hybrid swarm intelligence for feature selection	Small dataset; non-deep architecture
(Ni et al., 2023)	Traditional	KNN+LDA+LR	Credit card	Public	High recall across 5 datasets	Low precision
(Jose et al., 2023)	Traditional	DT, RF, ET, GB	Credit card	synthetic (Sparkov)	Effective use of AdaBoost; large dataset	Oversampling may lead to overfitting
(Afriyie et al., 2023)	Traditional	LR, DT, RF	Credit card	synthetic (Sparkov)	Large dataset	Under sampling reduces dataset size; complexity with RF
(Liu et al., 2018)	Deep	GNN	Online payment	Private (collected from Alipay)	Robust model for large datasets	Lack of class imbalance discussion
(Liu et al., 2021)	Deep	GNN	Class imbalance for fraud	Multiple public datasets	Adaptability to various use cases and datasets	Better performance on smaller datasets; less effective for financial fraud

Vector Machines (SVM) to more contemporary approaches employing Artificial Neural Networks (ANN) and hybrid models. Each study's dataset, key strengths, and limitations are delineated, showcasing the evolution of fraud detection strategies over time.

Our paper distinguishes itself by revisiting and refining the use of traditional machine learning models—LR, DT, SVM, and K-Nearest Neighbors (KNN)—infused with the latest advancements in explainable AI to tackle credit card fraud. Unlike previous studies that predominantly leveraged either deep learning approaches or conventional models in isolation, our research synthesizes the strengths of traditional algorithms with the clarity and interpretability afforded by explainable AI. This innovative blend aims to address the critical challenges identified in prior research, including handling large and complex datasets, dealing with class imbalance effectively, and enhancing the precision and recall of fraud detection without sacrificing the model's transparency or interpretability.

Furthermore, our research uniquely combines the reliability and tested nature of traditional machine learning models with the cutting-edge advancements in explainable AI. This dual strategy is designed to not only counteract the limitations identified in previous studies but also to push the frontier in credit card fraud detection towards more transparent, interpretable, and real-time solutions.

3 Methodology

The methodology of our study (see Fig. 2) is designed to develop and evaluate explainable machine learning models for real-time payment fraud detection. This section outlines the systematic approach taken from data preprocessing to the final evaluation of the models.

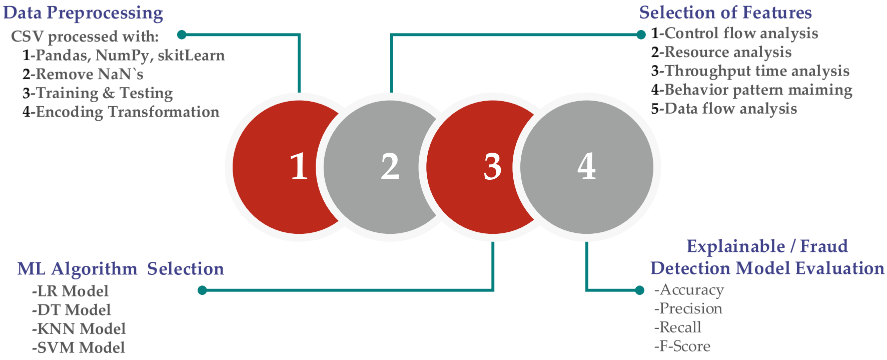

Fig. 2. Research Framework.

3.1 Data Preprocessing

Data preprocessing is the first critical step in our methodology, ensuring the quality and compatibility of our dataset with the machine learning algorithms. The following sub-steps are involved:

(i) Pandas, NumPy, and Scikit-Learn Usage: We employ Pandas for data manipulation, NumPy for numerical operations, and Scikit-Learn for applying pre-processing techniques and machine learning algorithms. These libraries facilitate the handling of CSV files and the execution of complex data transformations.
(ii) NaN Values Removal: We clean the dataset by removing or imputing NaN (Not a Number) values to maintain the integrity of our analysis and ensure that our models are trained on complete and accurate data.
(iii) Training and Testing Split: The dataset is divided into training and testing sets, allowing us to fit our models on one portion of the data and evaluate their performance on another, unseen portion, thus assessing their predictive capabilities.
(iv) Encoding Transformation: The concept of encoding is used to assign a number to all the categories thus making them interpretable to machine learning algorithms.

3.2 Selection of Features

(i) Feature selection aims to identify the most informative and relevant features for predicting fraudulent transactions. This process involves:
(ii) Control Flow Analysis: Analyzing the sequence and operations of transactions to identify suspicious patterns.
(iii) Resource Analysis: Evaluating transaction resources, such as account information and transaction amounts, to pinpoint features indicative of fraud.
(iv) Throughput Time Analysis: Investigating the duration of transaction processes to detect anomalies that may signify fraudulent activities.
(v) Behavior Pattern Mining: Mining user or account behavior over time to discover patterns that differentiate between fraudulent and legitimate transactions.
(vi) Data Flow Analysis: Assessing the movement and destination of transaction data to uncover atypical paths associated with fraud.

3.3 ML Algorithm Selection

We selected a range of machine learning algorithms to construct our fraud detection models, each chosen for its unique strengths in classification tasks:

(i) LR Model: Logistic regression becomes particularly useful for making decisions easy to understand and taking a binary classifications process.
(ii) DT Model: Decision Trees are for their intuitions and ability to model non-linearity without too much need to be undertaken in the manner of data preparation.
(iii) KNN Model: The K-Nearest Neighbors algorithm is selected because of its simplicity of implementation and the classification accuracy it provides with the training set as analogical matrix.
(iv) SVM Model: Support Vector Machines perform exceptionally well as they are famous for their ability to perform in high-dimensional spaces, therefore they are suitable for those datasets that have much number of features.

3.4 Explainable/Fraud Detection Model Evaluation

Our final step involves a thorough evaluation of the developed models using key metrics to measure their performance in detecting fraudulent transactions. These metrics include:

(i) Accuracy: The proportion of total predictions that were correct.
(ii) Precision: The ratio of true positive predictions to the total positive predictions, indicating the quality of positive class predictions.
(iii) Recall: The ratio of true positive predictions to the actual number of positive class samples, reflecting the model's ability to capture positive class instances.
(iv) F-Score: The harmonic mean of precision and recall, providing a balance between the precision and recall metrics.
(v) This comprehensive methodology ensures the development of accurate, reliable, and explainable models for real-time payment fraud detection, addressing both the technical and practical requirements of such systems.

4 Dataset Exploration and Description

The "Online Payments Fraud Detection Dataset" is designed to aid in the identification and analysis of fraudulent transactions in online payment systems. Each record in this dataset encapsulates a transaction's details, allowing for a comprehensive exploration of transaction patterns and potential fraud indicators (Dornadula et al., 2019; Vanini et al., 2023). Below is a breakdown of the dataset's key features:

- step: This attribute signifies a unit of time, with one step equating to one hour. This temporal aspect can be critical in understanding the timing and frequency of transactions, which may help in detecting suspicious patterns often seen in fraudulent activities.
- type: This field describes the type of online transaction executed. Different transaction types may exhibit varying risks and patterns of fraud.
- amount: Represents the monetary value of the transaction. Analysis of transaction amounts can provide insights into typical user behavior and identify outliers that might suggest fraudulent activities.
- nameOrig: Indicates the customer initiating the transaction. By tracking the originator, one can analyze individual behavior and detect potentially fraudulent operations based on historical data.
- oldbalanceOrg: Shows the balance of the originating customer's account before the transaction took place. This can be useful in verifying the authenticity of the transaction amount and the account's typical activity level.
- newbalanceOrig: The balance in the originating customer's account after the transaction. Changes in this balance, when compared with the amount and previous balance, can help in identifying discrepancies that might indicate fraud.
- nameDest: The recipient of the transaction. Monitoring recipients can help in identifying suspicious accounts that frequently receive funds from different sources or are involved in split transactions, a common method in money laundering.
- oldbalanceDest: The initial balance of the recipient's account before receiving the transaction. This feature can aid in understanding the flow of money and whether the credited amounts align with typical account profiles.
- newbalanceDest: Reflects the balance in the recipient's account after the transaction. This is crucial for detecting whether the credited money is being quickly moved out, which is typical in layering stages of money laundering.

- isFraud: A binary indicator where '1' represents a fraudulent transaction and '0' represents a non-fraudulent transaction. This is the target variable used to train models to detect and predict fraudulent transactions.

Understanding the relationships and trends among these attributes can provide crucial insights into typical and atypical transaction behaviors, aiding in the development of robust fraud detection algorithms.

4.1 Exploratory Data Analysis

- **Univariate Analysis** (Caixeta et al., 2023): In this phase, each variable in the dataset is analyzed individually to summarize and find patterns in the data. The first aim is to explore each of the transaction types, their amount and balance distributions. For instance, by looking at the 'type' column we can see that "CASH_OUT" and "PAYMENT" are two of the most frequently used transaction types, and hence, are the prevalent activities inside the dataset. By recognizing these distributions, it is possible to detect those transaction types that have the highest likelihood of being associated with fraud.
- **Bivariate Analysis** (Shi et al., 2023): Bivariate analysis consists of analyzing correlations and interactions between two variables, which helps us determine relationships between data points. It may entail studying the patterns of transaction amounts that vary depending on the type of transactions, or comparing the old and new balances in an account to observe any anomalies that are likely fraudulent behaviors.
- **Multivariate Analysis** (Kadhuim et al., 2023): This analysis is not restricted to two variables but it looks at a multi-factorial environment to unfold the relationships within the dataset in a more complex manner. This maybe realizing how amounts of transactions, types of transaction, and account balances fluctuate over time or different steps. The fact that multivariate analysis helps in finding the patterns and glitches that would not seem obvious when we look at the variables individually cannot be overemphasized.

4.2 Bivariate Analysis: Transaction Type vs. Fraud Occurrence

This part of a bivariate analysis reveals the association between the type and whether it is classified as a fraudulent transaction. Findings that link various types of transactions to the frequency of fraud incidents can lead to complexities for producing fraud prevention measures that aim to specialized goals. Analysis of Transaction Types and Fraud Rates: Analysis of Transaction Types and Fraud Rates:

- CASH_IN: Total: 227130 transactions; Fraudulent: 0 transactions (NaN indicates nothing of this kind happened); Observation: Transactions involving "CASH_IN" do not have any committed fraud recorded. This underscores the fact that digital currency transactions are more naturally safe and tracking exists more clearly.
- CASH_OUT: The figure: 373,063 transactions; Among them: 578 transactions which are fraudulent; Summary: This amount represents the total number of "CASH_OUT" transactions where fraudulent transactions are a small yet remarkable number.
- This type may be more susceptible to fraud due to the nature of withdrawing funds, requiring more stringent monitoring and detection mechanisms.
- DEBIT: Total: 7,178 transactions; Fraudulent: 0 transactions (NaN indicates no occurrences); Analysis: Similar to "CASH_IN", "DEBIT" transactions show no fraudulent activities. This might suggest that direct debit transactions are less likely targets for fraudsters or are well-protected.
- PAYMENT: Total: 353,873 transactions; Fraudulent: 0 transactions (NaN indicates no occurrences); Analysis: Transactions classified under "PAYMENT" do not report any fraudulent instances. This might be due to the types of payments processed, which may generally involve smaller amounts or transactions within controlled environments.
- TRANSFER: Total: 86,189 transactions; Fraudulent: 564 transactions: Analysis: "TRANSFER" transactions, although fewer in total number compared to "CASH_OUT", also exhibit a notable amount of fraud. This type involves moving funds between accounts, which can be particularly attractive to fraudsters attempting to misappropriate large sums.

4.3 Transaction Type Distribution

The distribution of transaction types within the dataset is a critical factor to consider:

- CASH_OUT: 373,641 transactions
- PAYMENT: 353,873 transactions
- CASH_IN: 227,130 transactions
- TRANSFER: 86,753 transactions
- DEBIT: 7,178 transactions

Figure 3 shows that "CASH_OUT" and "PAYMENT" are the most commonly recorded transaction types. Such high frequencies indicate these are routine transactions for many users, potentially making them prime targets for fraudsters due to their frequency and volume. Conversely, "TRANSFER" and "DEBIT" transactions, while less frequent, may involve higher amounts or be linked to more complex transaction networks, potentially increasing their risk for fraudulent activities. Understanding these patterns helps in tailoring fraud detection systems to be more responsive to the nature of each transaction type.

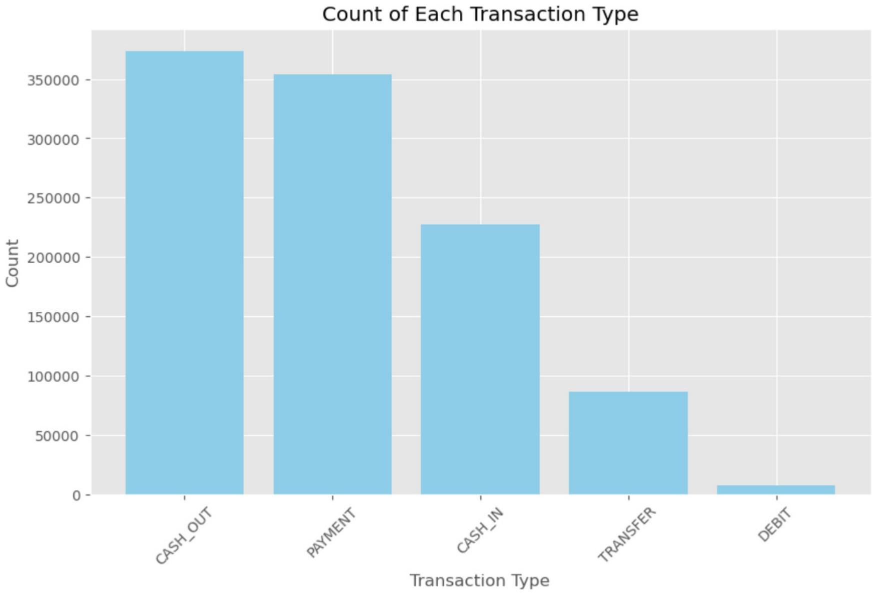

Fig. 3. Transaction Categories per Utilized Dataset.

- The dataset contains five types of transactions: CASH_OUT, PAYMENT, CASH_IN, TRANSFER, and DEBIT
- Among these types, the most common transactions are CASH_OUT and PAYMENT, with counts of 373,641 and 353,873, respectively.
- CASH_IN transactions are also quite common, but less frequent compared to CASH_OUT and PAYMENT, with a count of 227,130.
- TRANSFER and DEBIT transactions are relatively less common, with counts of 86,753 and 7,178, respectively as deployed in Fig. 4.

4.4 Distribution of Fraudulent Transactions

In the exploration of fraudulent transactions within the dataset, it is crucial to understand the proportion of transactions classified as fraud compared to legitimate transactions as shown in Fig. 5. This distribution is a fundamental aspect of the dataset that influences how fraud detection models are developed and evaluated. The dataset shows a significant imbalance in the classification of transactions:

- Not Fraud: 99.9% of the transactions are legitimate. This high percentage reflects the typical nature of transaction datasets where fraudulent activities are rare but potentially very harmful.
- Fraud: Only 0.1% of transactions are fraudulent. Although this represents a small fraction of the total transaction volume, the absolute numbers can still be significant given the large scale of data. This small percentage highlights the challenges in detecting fraud, as the models must identify these rare events without mistakenly classifying legitimate transactions as fraud.

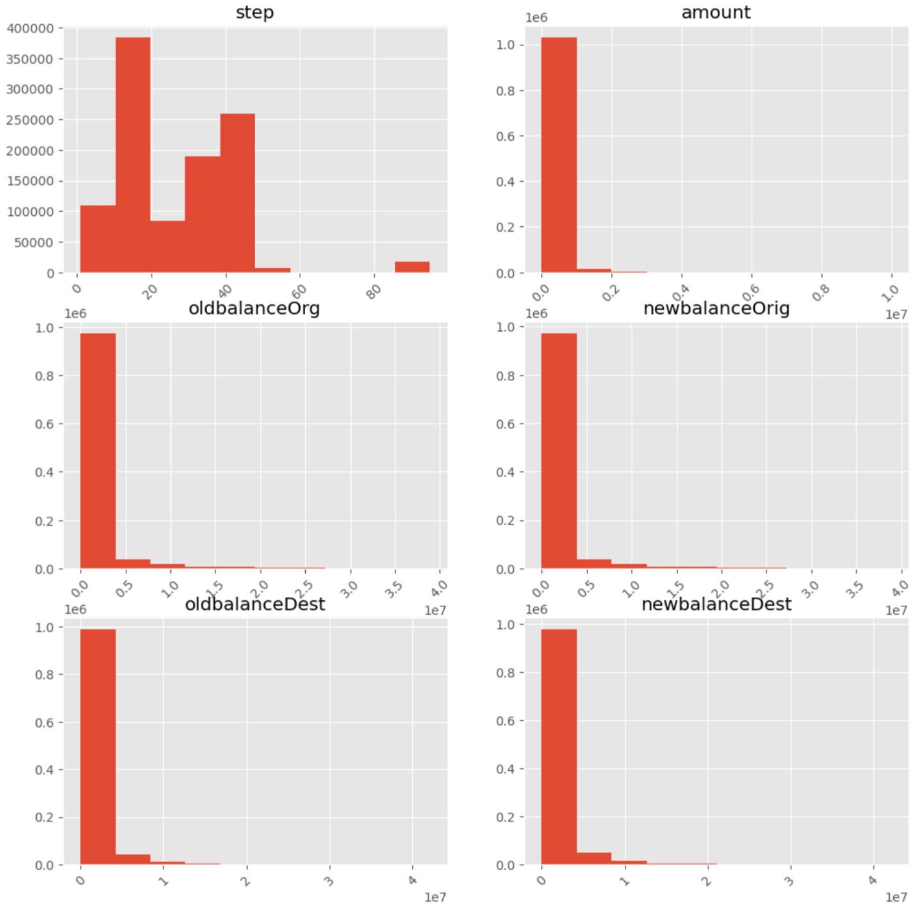

Fig. 4. Distribution of transaction attributes in Online Payment Data.

The stark disparity in the distribution between fraudulent and non-fraudulent transactions poses challenges for data scientists and analysts. Models trained on such data may have a tendency to overwhelmingly predict transactions as non-fraudulent due to the imbalance. This can lead to a high accuracy rate that is misleading because the rare fraudulent cases are the ones of most interest and are more costly when missed. To address this, techniques such as resampling the data, using anomaly detection algorithms, or applying advanced machine learning methods that focus on precision and recall balance, rather than just accuracy, are often employed. Such distribution is the very essence of the tuning process for the sake of ensuring that the model is sensitive enough to catch those critical, fraudulent transactions while causing less redundancy.

4.5 Dataset Analysis Implications

The result of fraud-related analysis implies that “CASH_OUT” and “TRANSFER” exhibit higher fraud prevalence when looked at the data shown on Fig. 6 and Fig. 7. The

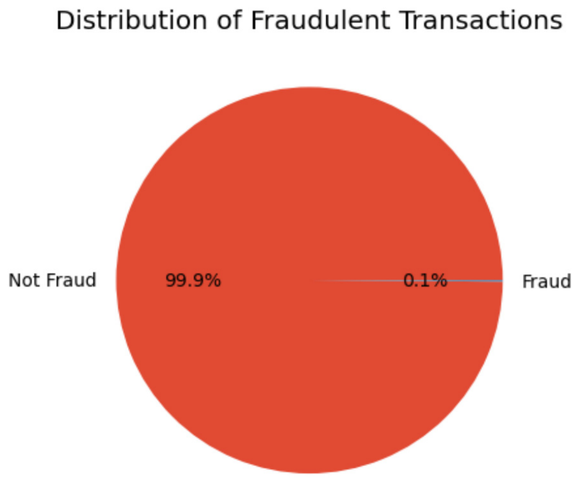

Fig. 5. Transaction Distribution per Utilized Dataset.

dissimilarity in the fraud occurrence types reveals requirement for the transaction-type-specific deterrent forecasting techniques. Institution of financial nature is necessarily required to have efficient fraud detection systems that are properly tuned for the risks categorized in each kind of transaction.

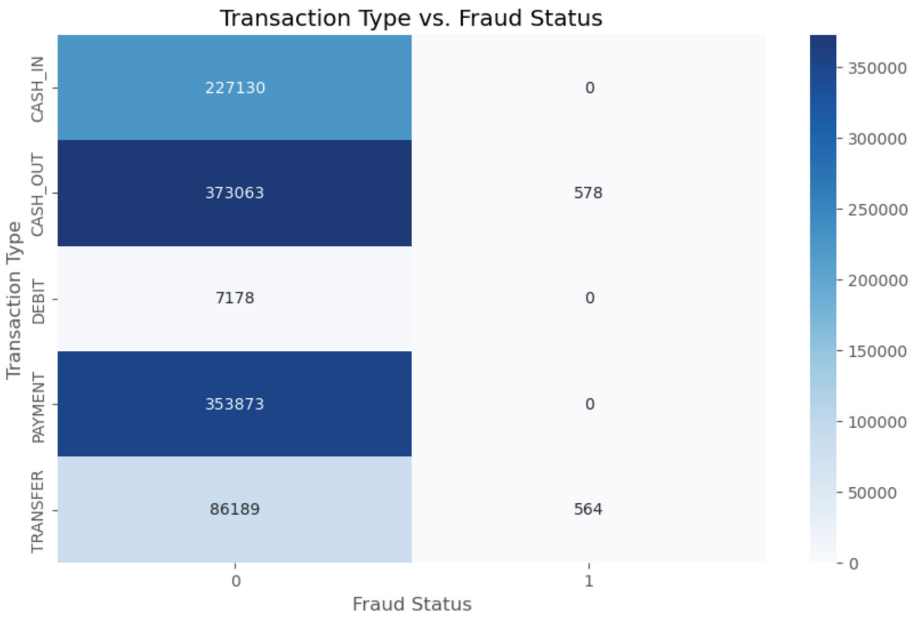

Fig. 6. Dataset Transactions' Mapping Analysis (From the Perspective of Fraud Status).

It can involve the setting of different thresholds of anomaly detections, or machine learning models that learn pattern types of fraud which are distinct from "CASH_OUT" and "TRANSFER". In the given situation, a fraud model should be carefully prepared to ascertain false positives, which could ruin a legitimate transaction, by mining useful information from a large number of legitimate transactions to distinguish between typical and atypical behavior in relatively higher-risk categories.

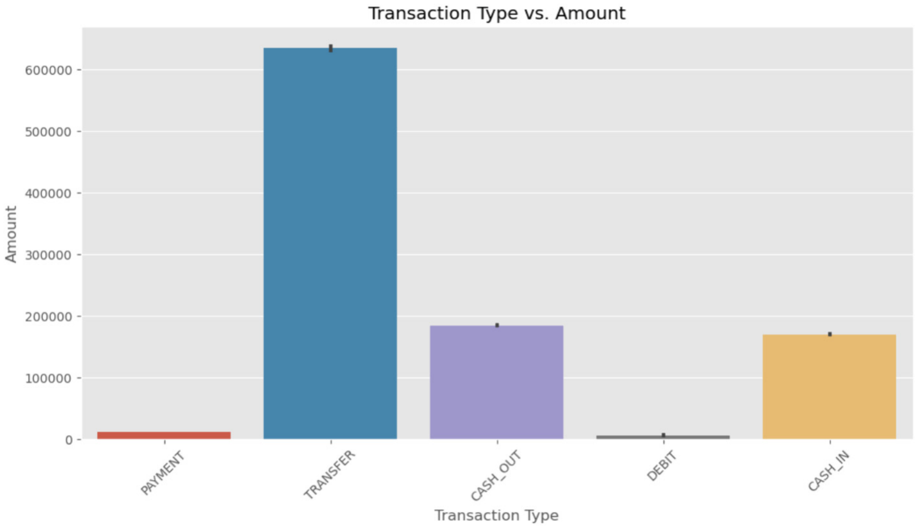

Fig. 7. Dataset Transactions' Mapping Analysis (From the Perspective of Amount).

4.6 Multicollinearity Analysis

Multicollinearity refers to the occurrence of high intercorrelations among independent variables in a dataset, which can lead to statistical issues that affect the performance and interpretability of a regression model (Derraz et al., 2023). In the context of the Online Payments Fraud Detection Dataset, examining multicollinearity helps in understanding how different features relate to each other and whether they might be providing redundant information. For examining features for Multicollinearity, in the provided snapshot of the dataset, the features include:

- step (time of transaction in hours)
- type (type of transaction)
- amount (transaction amount)
- nameOrig (identifier for the customer initiating the transaction)
- oldbalanceOrg (originating account balance before transaction)
- newbalanceOrig (originating account balance after transaction)
- nameDest (identifier for the recipient of the transaction)
- oldbalanceDest (destination account balance before transaction)

- newbalanceDest (destination account balance after transaction)
- isFraud (indicates if the transaction is fraudulent)
- Potential Areas of Multicollinearity

oldbalanceOrg and newbalanceOrig: These two features might be closely related as they both represent the state of the originating account's balance before and after the transaction. High correlation here could be due to the direct impact of the transaction amount on these balances. oldbalanceDest and newbalanceDest: Similar to the originating account balances, these features for the recipient's account might also exhibit high correlation. They represent the state of the account balance before and after the transaction affects the account. in addition, amount, oldbalanceOrg, and newbalanceOrig: The transaction amount is expected to directly influence changes in the originating account balances. Analyzing the correlation between these variables would help determine how independent each is in the context of predicting fraud.

4.7 Methods to Detect and Address Multicollinearity

Correlation Matrix: A simple and effective method to visually inspect potential multicollinearity is by generating a correlation matrix for the numeric variables. High correlation coefficients (close to −1 or 1) indicate potential multicollinearity as shown in Fig. 8.

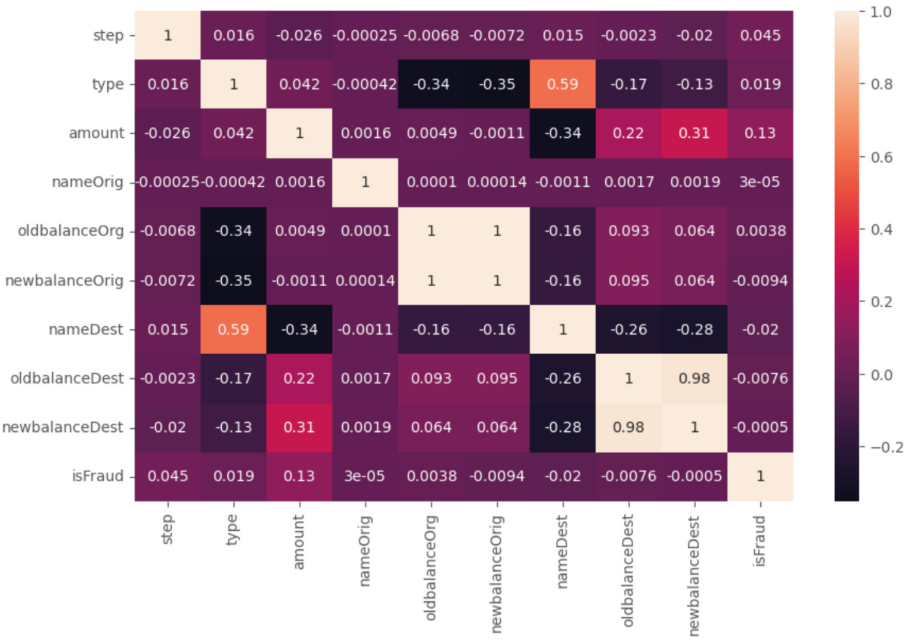

Fig. 8. Dataset Correlation Analysis Heatmap.

The below methods are specifically delas with multicollinearity related issues:

- Variance Inflation Factor (VIF): Calculating the VIF for each variable provides a quantifiable measure of how much the variance of an estimated regression coefficient increases if predictors are correlated. A VIF value greater than 10 is typically considered an indication of multicollinearity.
- Principal Component Analysis (PCA): PCA can be used to reduce dimensionality by transforming the original variables into a new set of variables (principal components) that are linear combinations of the original variables. The new components are orthogonal (independent), effectively removing multicollinearity.

4.8 Analysis of Variance Inflation Factor (VIF) Results

The Variance Inflation Factor (VIF) is utilized to identify the presence and severity of multicollinearity among the independent variables in a dataset, which is crucial for ensuring the reliability of the coefficients in regression models. Below is an interpretation of the VIF results you provided, structured to help understand each variable's impact and possible actions to address any identified issues.

- Step (VIF = 2.877871): This shows moderate multicollinearity. The variance of the regression coefficient is slightly inflated but generally not problematic.
- Type (VIF = 5.055132): The borderline high VIF indicates that this variable has substantial multicollinearity with other variables, suggesting its role might overlap with other features.
- Amount (VIF = 2.126342): Indicates low to moderate multicollinearity, which suggests that the amount of the transaction is fairly independent of the other variables.
- NameOrig (VIF = 2.857206): Similar to 'step', shows moderate multicollinearity and is not a concern.
- OldbalanceOrg (VIF = 709.443407) and NewbalanceOrig (VIF = 716.622550): Both of these variables exhibit extremely high VIF values, indicating significant multicollinearity. This suggests that these variables are providing overlapping information, possibly about the state of the originating account before and after a transaction.
- NameDest (VIF = 3.823784): Displays moderate multicollinearity, which should not typically pose a problem for modeling.
- OldbalanceDest (VIF = 38.440149) and NewbalanceDest (VIF = 41.277884): Both show high VIF values, indicating significant multicollinearity and suggesting overlapping information concerning the recipient's account balances.
- IsFraud (VIF = 1.134756): Shows minimal multicollinearity, indicating that this variable is quite independent of the other variables, which is expected for a dependent variable.

4.9 Handling Imbalanced Data

Given the nature of fraud detection, where fraudulent transactions are much less frequent than non-fraudulent ones, techniques such as Synthetic Minority Over-sampling Technique (SMOTE) can be applied to the training data to balance the dataset. Preparing data for ML involves several essential steps from loading data, handling different data types, splitting the data into training and testing sets, to feature engineering. Each step ensures that the data fed into the model is well-suited for learning patterns and making accurate predictions, crucial for tasks like fraud detection.

4.10 Feature Importance in Fraud Detection Models

Feature importance is a crucial aspect of machine learning model interpretation, helping to identify which variables significantly influence the prediction outcomes. In the context of fraud detection, understanding which features contribute most to the model's decision-making process can provide insights into the nature of fraudulent transactions and inform strategies for prevention and detection. Analysis of Feature Importance (Wadday et al., 2020; Atiyah, 2023; Atiyah et al., 2023, b, c; Husin et al., 2024). Based on the provided importance scores, the features contribute to the model's predictions as follows:

- Amount_Orig (0.40): This feature, representing the amount originally involved in the transaction, has the highest importance score. A high score indicates that the initial transaction amount is a strong predictor of fraud. Large or unusual amounts, especially in the context of the customer's usual transaction patterns, can be a significant red flag for potential fraudulent activity.
- Step (0.20): The time step of the transaction, measured in hours, also plays a substantial role in predicting fraud. This suggests that the timing of a transaction, perhaps in relation to typical customer activity patterns or during unusual hours, might indicate suspicious behavior.
- Type (0.11) and Amount (0.11): Both transaction type and amount have equal importance, indicating that the nature of the transaction and the specific transaction amount are pertinent to detecting fraud. Certain types of transactions might inherently be more susceptible to fraud, or there might be specific transaction amounts that are commonly used by fraudsters to avoid detection.
- Amount_Dest (0.10): The amount received by the destination account is also a significant factor, slightly less so than the type and transaction amount but still notable. This might reflect situations where funds are transferred to accounts that typically do not receive large sums, suggesting a possible redirection of funds for illicit purposes.
- NewbalanceDest (0.07): The new balance of the destination account after the transaction is the least influential but still relevant. Changes in this balance might help identify whether the funds are being quickly moved in a manner consistent with layering stages of money laundering or other fraudulent schemes (Al-Hchaimi et al., 2021; Muhsen et al., 2023b).

in addition, the Implications of Feature Importance.

- Prioritization of Monitoring: The identified feature importance can help financial institutions prioritize certain types of monitoring and controls. For example, transactions involving large amounts or those that occur during abnormal hours should be scrutinized more closely.
- Model Tuning: Categorizing those features that are most opportune to model further refinement helps in phase the model further. Something that has lower effect may probably be the first one that should be eliminated in the basic models. Perhaps the objectivity of accuracy won't change too much.