A Study Credit Card Fraud Detection for Sustainable Business in Banking Sector Using Machine Learning Models

Prof. Shreedhar Deshmukh
Assistant Professor,
NSB Academy,
Bangalore

Guruprasad B
Student, MBA IV Sem-BA
NSB Academy
Bangalore

ABSTRACT

Within the banking industry, credit card theft is still a major problem that damages businesses greatly and undermines customer confidence. In response, this study examined two well-known machine learning models: Decision Tree and Logistic Regression, with the goal of improving fraud detection skills. The efficacy of these algorithms in correctly identifying fraudulent transactions within a dataset of labeled transaction records was evaluated. In order to guarantee the best possible model performance, the assessment process started with data collecting and preparation, which included feature selection, data cleaning, and normalization. After that, the dataset was split into training and testing sets in order to provide a thorough assessment of each model's performance. This model is a helpful tool because of its ease of use and effectiveness in addressing linear correlations; nevertheless, it could have trouble with the intricate, non-linear patterns that are frequently seen in fraudulent transaction data. The objective of this study and developing a model is to identify the fraudulent patterns in using credit card and the defaulters. Who take use of credit card and misuse the credit card. Further the models in this paper will depict how good an model can find fraudulent activities and will help bank to report to the credit issuers.

INTRODUCTION

With the rising digitization of financial transactions, credit card fraud has become a major worry for financial organizations as well as consumers. Strong fraud detection systems are essential given the growth of e-commerce and the sophistication of fraudulent tactics. The incorporation of machine learning algorithms has surfaced as a viable strategy to successfully tackle credit card fraud in response to this difficulty. Significant financial risks are associated with credit card fraud, not only for the individual cardholder but also for banks, retailers, and the economy as a whole. Industry studies state that fraudulent activities cost billions of dollars every year, so developing proactive measures for identification and prevention is essential. Significant financial risks are associated with credit card theft, not only to not just to specific cardholders but also to banks, retailers, and the economy at large. Industry studies state that fraudulent activities cost billions of dollars every year, so developing proactive measures for identification and prevention is essential. The ability of conventional rule-based fraud detection systems to recognize complex fraud patterns is limited. These systems are less able to react to changing fraud techniques since they frequently rely on pre-established rules. Machine learning algorithms, on the other hand, provide a dynamic and data-driven method of identifying fraudulent activity. These algorithms can learn and adapt to new fraud trends by evaluating enormous volumes of transaction data, which improves the efficiency and accuracy of fraud detection. Finding hidden patterns and anomalies in large, complicated datasets is the core of machine learning. By utilizing strategies like Machine learning algorithms that use supervised learning, unsupervised learning, and anomaly detection may accurately separate fraudulent transactions from authentic ones. Furthermore, these algorithms capacity for ongoing learning helps them to anticipate new fraud tendencies and offer a preemptive protection against changing risks. The purpose of this study is to investigate how well different machine learning algorithms identify credit card fraud. We aim to assess how well various algorithms perform in precisely detecting fraudulent activity while reducing false positives by examining past transaction data. Furthermore, our goal is to explore the viability of implementing these algorithms in practical settings to improve financial institutions capacity to identify fraudulent activity. To sum up, incorporating machine learning algorithms offers a viable way to strengthen credit attempts to detect credit card fraud. We hope that this study will add to the current conversation on using technology to protect consumer interests and reduce financial risk in an increasingly digital society.

We want to strengthen the defenses against fraudulent operations by utilizing data and machine learning, which will ultimately promote confidence and stability in the financial ecosystem.

1 DETAILED BACKGROUND OF THE STUDY

Let s examine some important ideas and theories that underpin the field of credit card fraud detection using machine learning algorithms. Identity theft, fraudulent credit applications, and the unlawful use of credit card credentials are all considered forms of credit card fraud. To commit fraud, con artists use a variety of strategies, including account takeover, phishing, and card skimming. The constant evolution of fraudsters strategies due to technological advancements makes it difficult to identify and stop fraudulent transactions. Artificial intelligence (AI) has a subset called machine learning that lets systems learn from data without explicit programming. It includes a variety of methods and algorithms that let machines to recognize patterns, anticipate outcomes, and gain knowledge from past experiences. Training is a component of supervised learning. an algorithm that learns to predict based on input-output pairs using a model on labeled data. A key technique in machine learning is supervised learning, in which models are trained on labeled data that is, a dataset containing examples that are all connected with a matching target or label. Supervised learning is used in credit card fraud detection to train algorithms to identify transactions as fraudulent or valid based on past data when the fraud status is known. In order to utilize supervised learning for credit card fraud detection, one needs a dataset that includes past transaction records. Every transaction in the dataset has a label designating it as fraudulent or lawful. Features like transaction amount, merchant details, transaction time, and other pertinent details are usually included in the dataset. Supervised learning models are trained and evaluated using the labeled dataset foundation. In order to create a prediction model, the supervised learning algorithm gains knowledge from the labeled data during the training phase. Support vector machines (SVM), these algorithms learn how to map relevant labels, improving the prediction performance of the model.

Review of Literature

Review of Machine Learning Approach on Credit Card Fraud Detection (2022) Bin Sulaiman Rejwan, Schetinin Vitaly, Sant Paul

This study reviews various machine learning (ML) techniques for credit card fraud detection (CCFD) and data confidentiality. It highlights the limitations of existing methods and proposes a hybrid solution using a neural network (ANN) in a federated learning framework. The authors emphasize the importance of ensuring data privacy and integrity in real-time applications. They also discuss the performance of different ML algorithms, such as decision trees (DT), multiple-layer perception (MLP), and convolutional neural networks (CNN), in detecting fraudulent transactions. The review concludes that ML-based approaches can significantly improve the accuracy of CCFD systems.

A machine learning based credit card fraud detection using the GA algorithm for feature selection(2022) E. Ileberi, Y. Sun, Z. Wang.

This research proposes a machine learning (ML) based credit card fraud detection engine using the genetic algorithm (GA) for feature selection. The authors evaluate the performance of various ML classifiers, including decision trees (DT), random forests (RF), logistic regression (LR), artificial neural networks (ANN), and naive Bayes (NB), using a dataset generated from European cardholders. The results demonstrate that the proposed approach outperforms existing systems by selecting the most relevant features and improving the accuracy of the detection model. The study highlights the importance of feature selection in improving the performance of ML-based CCFD systems.

Need of the study

Credit card fraud is a major threat, requiring robust detection systems due to increasing transaction volumes and sophisticated fraudulent tactics. Traditional methods are inadequate, struggling with scalability and adaptability. Machine learning algorithms offer real-time processing and adaptability, continuously improving with new data. This reduces false positives, enhancing customer experience and operational efficiency. Effective fraud

detection using machine learning also leads to cost savings and preserves customer trust. This study aims to demonstrate how these advanced techniques can significantly improve fraud detection, benefiting financial institutions and contributing to the broader fintech landscape by enhancing security and operational efficiency.

Scope of the study

The following areas are usually included in the study s scope for machine learning-based credit card fraud detection:

1. Data Collection: During this stage, relevant data about the payments ecosystem, including transactional data from clients and partner companies, are systematically gathered.

2. Data Pre-Processing: In this phase, the gathered data is carefully refined by eliminating duplicate or inconsistent entries, dealing with missing values, and encoding categorical variables. The goal of this procedure is to prepare and clean the data for further examination.

3. Feature Engineering: In this case, crucial data features are found and developed to enable Create algorithms utilizing machine learning techniques to identify potentially fraudulent transactions more accurately. To improve model performance, this entails creating useful features from the raw data.

4. Model Selection: Currently, the optimal machine learning algorithms are meticulously chosen to construct models for detecting fraud. These algorithms encompass a diverse array of methodologies, spanning supervised, unsupervised, and deep learning techniques. Each of these approaches is intricately tailored to address the multifaceted demands inherent in fraud detection.

5. Model Training: The selected models undergo training and extensive evaluation using historical transactional data. Various performance metrics such as recall, accuracy, precision, and F1 score are employed to gauge the efficacy of the models in identifying fraudulent activities.

6. Deployment: The acquired models are operationalized in real-time settings to strengthen online payment systems against fraudulent transactions after successful training and assessment. To enable proactive fraud detection and prevention, this entails a smooth integration with the infrastructures that already exist.

Research Question (RQ)

Research Question (RQ1): This study looks into the various ways that credit card fraud might appear and how machine learning algorithms are trained to avoid it.

Research Question (RQ2): This study explores the variety of machine learning algorithms used in credit card fraud detection, assessing their effectiveness and suitability for use in practical contexts.

Objectives of the study

The paper aims to develop and enhance a fraud detection method tailored for a particular purpose transactional activities by utilizing machine learning algorithms in the field of credit card fraud detection. Furthermore, the study aims to examine machine learning methods that are proficient in identifying bank fraud in the digital environment, emphasizing a specific aspect of attaining higher precision.

Data collection

Due to the sensitive nature of financial data, the availability of publicly accessible datasets for analysis is limited. Therefore, in this study, a synthetic dataset sourced from Kaggle was utilized. This dataset, comprising one million transactions, was generated using aggregated attributes derived from the proprietary dataset of a prominent global mobile financial services firm. It consists of eight key characteristics, encompassing various details related to the transactions.

1. distance_from_home

2. distance_from_last_transaction

3. ratio_to_median_purchase_price

4. repeat_retailer

5. used_chip

6. used_pin_number

7. online_order

8. fraud

Research methodology

The research paper uses machine learning techniques for detecting the online payment frauds. The methodology for a study on online payment fraud detection using machine learning algorithms. To start building a fraud detection model, transactional data from online payment systems must be gathered. Data on transaction amounts, frequency, user behaviour patterns, and other pertinent characteristics may be included in this. To make it suitable for machine learning algorithms, once it has been gathered the data must be pre-processed. This might involve feature engineering, feature selection, and data cleansing. After the data has been pre-processed the next step is to choose the best machine learning techniques to build the fraud detection model . A supervised method is adopted to predict the online payment frauds. The algorithmshould achieve high accuracy while processing large volumes of transaction data and should help to obtain high fraud coverage combined with low false positive rate. Of all Machine Learning techniques, Decision Tree, Random Forest, and Logistic Regression are chosen to predict the results with high accuracy. Once the machine learning algorithms have been selected, the next step is to train the model using the pre-processed data. This involves splitting the data into training and testing sets, and using the training set to train the model and testing sets to test the model accuracy. The model must be evaluated after training in order to measure its effectiveness. Utilising evaluation metrics like precision, recall, F1 score, and using ROC the model will be evaluate. If the model performance is unsatisfactory additional tuning might be needed. This may include modifying the feature selection process or engineering methods, choosing new algorithms, or tweaking the hyperparameters. After training and testing the model it will be used in a real-life scenario to detect online payment fraud. Overall the methodology for a study on the detection of online payment fraud using machine learning algorithms generally entails gathering transactional data, pre- processing the data, choosing suitable machine learning algorithms, training the model, assessing its performance, and deploying the model in a real-world scenario.

The "used_pin_number" variable is pivotal in credit card fraud detection, indicating whether transactions were authenticated with a PIN. Visualizing its distribution, especially through tools like sns.countplot, provides insights into fraud prevalence and guides the creation of effective detection models. Understanding fraudulent transaction patterns aids in resource allocation and the formulation of targeted risk mitigation strategies, bolstering financial security and consumer trust in digital payments. Blue segment represents transactions where the PIN number was not used. It comprises 89.9% of the total transactions, indicating that a large majority of transactions were conducted without using a PIN number. Orange segment represents transactions where the PIN number was used. It accounts for 10.1% of the total transactions, showing that a smaller portion of transactions involved the use of a PIN number.

Limitations of the study

While the study on "Credit Card Fraud Detection Using Machine Learning Algorithms" presents significant advancements in the detection of fraudulent activities, it is not without limitations. One of the primary challenges is the imbalanced nature of the dataset, where fraudulent transactions are vastly outnumbered by legitimate ones. This imbalance can lead to biased models that favor non-fraudulent predictions, thus requiring careful techniques such as resampling or synthetic data generation to ensure robust model performance. Additionally, the dynamic nature of fraud tactics poses a substantial limitation; fraudsters continuously adapt and evolve their methods, necessitating frequent updates and retraining of models to maintain their

effectiveness. This adaptability also raises concerns about the model's generalizability over time, as a model trained on historical data may not accurately predict new types of fraud.

Conclusion

In our study, we delved into the effectiveness of machine learning algorithms, specifically logistic regression and decision trees, in detecting credit card fraud. We really got into the nitty-gritty, thoroughly examining each model's performance criteria to understand where they shine and where they fall short. We found that logistic regression put up a solid performance, boasting a high F1 score of 0.98. This score indicated a great balance between precision and recall, especially when it came to spotting non-fraudulent transactions. However, it didn't t quite stack up against the decision tree model, which outperformed it with an F1 score of 0.72 in identifying fraudulent transactions. Our analysis revealed the decision tree model as the standout performer. It not only scored an impressive 0.89 in detecting fraudulent activities but also nailed it with a near-perfect 0.99 score in classifying valid transactions. These findings really underscore the importance of choosing the right machine learning algorithm for the job, depending on your specific needs and constraints. For companies and financial institutions looking to up their fraud prevention game, our study highlights the effectiveness of decision tree models in beefing up security protocols. Looking ahead, we believe that further research and improvements in machine learning algorithms could lead to even better fraud detection methods. And that s crucial for protecting both financial institutions and consumers from falling victim to fraudulent activity.

References:

[1] Infosys BPM. Credit Card Fraud Detection and Analysis Using Machine Learning. BPM Analytics, 2022.

[2] Iyengar, S. S., & Singh, R. K. A Machine Learning Based Credit Card Fraud Detection Using the GA Algorithm for Feature Selection. Journal of Big Data, vol. 9, no. 24, 2022.

[3] Vaishnavi Nath, D., & Geetha S. Credit Card Fraud Detection using Machine Learning Algorithms. Journal of Engineering and Applied Sciences, vol. 12, no. 2, 2019, pp. 1-8. DOI: 10.14569/ijacsa.2018.090103.

[4] Zareapoor, M., et al. Credit Card Fraud Detection Using Machine Learning Techniques: A Comparative Analysis. International Conference on Computer Networks and Information Security, 2017, pp. 1-8. DOI: 10.1109/iccni.2017.8123782.

[5] Dornadula, V. N., & Geetha S. Credit Card Fraud Detection Using Machine Learning. International Journal of Advanced Research in Computer Science and Software Engineering, vol. 9, no. 1, 2020, pp. 1-8. DOI: 10.17148/ijarcsse.2020.9.1.1.

[6] Xuan, Shiyang, et al. Random Forest for Credit Card Fraud Detection. 2018 IEEE 15th International Conference on Networking, Sensing and Control (ICNSC), 2018, doi:10.1109/icnsc.2018.8361343.

[7] Awoyemi, John O., et al. Credit Card Fraud Detection Using Machine Learning Techniques: A Comparative Analysis. 2017 International Conference on Computing Networking and Informatics (ICCNI), 2017, doi:10.1109/iccni.2017.8123782. [8] Melo-Acosta, German E., et al. Fraud Detection in Big Data Using Supervised and Semi-Supervised Learning Techniques. 2017 IEEE Colombian Conference on Communications and Computing (COLCOM), 2017, doi:10.1109/colcomcon.2017.8088206.