Final Project

Enhancing Fraud Detection in Financial Transactions: A Machine Learning Approach with AutoML Optimization

Yang Hu

Campbellsville University

DS61580H224 Data Mining and Warehousing

Dr. Sean Perryman

Nov 24th, 2024

Introduction

Financial fraud poses a significant threat to the integrity of global economic systems, resulting in substantial financial losses and undermining trust in financial institutions. The widespread adoption of digital financial transactions has further exacerbated this issue, providing fraudsters with increasingly sophisticated methods to exploit vulnerabilities. Traditional fraud detection approaches, often reliant on manual oversight and predefined rule-based systems, are no longer sufficient to counteract the evolving tactics of fraudsters. As such, there is an urgent demand for advanced, adaptive, and efficient fraud detection mechanisms.

Recent advancements in machine learning (ML) and artificial intelligence (AI) provide promising solutions to enhance fraud detection capabilities (Thilagavathi et al., 2024). These technologies enable the real-time analysis of vast amounts of transactional data to identify patterns and anomalies indicative of fraudulent activities. For example, studies have shown that Support Vector Machines (SVM) and Rule-Based Systems (RBS) are effective in identifying fraudulent activities within financial transactions (Khanum et al., 2024). Similarly, the integration of Graph Neural Networks (GNNs) with anomaly detection techniques has demonstrated potential by representing transactions as graphs to capture complex relationships and behaviors (Cheng et al., 2024).

This paper analyzes the application of machine learning techniques in detecting fraud in financial transactions. By reviewing current methodologies and evaluating their effectiveness, this study aims to contribute to developing more robust and adaptive fraud detection systems. Specifically, the analysis focuses on the comparative performance of various ML models, including deep learning approaches, in adapting to the dynamic nature of fraudulent behaviors. Through this exploration, this research seeks to provide insights to inform the design of next-

generation fraud detection systems capable of safeguarding financial transactions against emerging threats.

Data Collection and Preprocessing

The data was collected from Kaggle Credit Card Transactions Dataset (Credit Card Transactions Dataset, 2024). This dataset contains credit card transactions, including both fraudulent and non-fraudulent transactions, and was last updated in August 2024, as of September 2024. The data is formatted in a CSV file with 24 columns, including a target result labeled "is_fraud." The dataset consists of 1,296,675 rows and has a size of 354.1 MB in CSV format.

The data preprocessing process began with loading the dataset, followed by removing rows with missing values using the na.omit() function. This step ensured that the dataset did not contain incomplete observations, which could negatively impact the quality of the subsequent analysis. To further enhance compatibility, column names were renamed using the make.names() function, guaranteeing they adhered to valid syntax rules for later processing. Basic exploratory steps were conducted, such as summarizing each feature's structure and statistics to understand the data better.

Feature engineering was performed to enrich the dataset with more informative variables. The first step was calculating age from the 'dob' (date of birth) column by subtracting the birth year from the current year, creating the 'age' feature. This 'age' feature was then bucketed into age categories to enable effective analysis based on different age groups. In addition, the transaction amount ('amt') was bucketed into multiple categories, providing a more convenient way to group and analyze spending behaviors. Categorical features, such as 'city' and 'job', were

also converted into factors, and their levels were renamed to maintain compatibility while ensuring they represented distinct, meaningful categories.

The transaction date and time were processed by splitting the 'trans_date_trans_time' column into separate columns for 'month' and 'date of the month.' This transformation allowed for easier extraction of time-based insights, such as identifying monthly or day-specific trends in the data. Once these new columns were created, the original columns, including 'trans_date_trans_time', 'dob', 'age', 'amt', and 'merchant', were removed to prevent redundancy and reduce the risk of data leakage. As a further data-cleaning measure, any remaining rows with missing values were removed once again.

The 'is_fraud' column was explicitly converted into a factor to prepare the dataset for modeling, ensuring it was appropriately formatted for classification purposes. Furthermore, all imported numerical features were categorized into quantiles, which helped normalize their representation and facilitated smoother input into machine learning models. This transformation was particularly useful for features requiring a more uniform representation for modeling.

Lastly, the script ensured that the 'is_fraud' column was properly identified in the dataset. It performed a validation step that stopped execution if the column could not be found, preventing potential errors in further processing. The dataset was thoroughly cleaned and structured by carrying out these steps, making it suitable for analysis and predictive modeling with minimal risk of inconsistencies or inaccuracies.

Hypothesis

Credit card fraud transactions are not random but exhibit identifiable patterns, which machine learning models can effectively leverage. The machine learning models are expected to outperform traditional rule-based systems in fraud detection. Unlike rule-based systems that rely

on static, predefined thresholds, machine learning models are designed to analyze large datasets and adapt to evolving patterns of fraudulent behavior. This adaptability is hypothesized to result in higher accuracy and reliability in identifying fraudulent transactions than traditional approaches.

To prove the hypothesis, the performance of machine learning models must show significantly higher accuracy, fewer false positives, and better adaptability to evolving fraud patterns than rule-based systems. If these criteria are met, the hypothesis is confirmed. Conversely, the hypothesis is invalidated if rule-based systems achieve similar or superior performance or if machine learning models fail to identify fraud patterns consistently.

Proposed Method

In the proposed method for financial fraud detection, multiple machine learning models are utilized, encompassing both classification and clustering approaches. This diverse range of techniques aims to effectively identify fraudulent patterns in financial transactions by leveraging different strengths of supervised and unsupervised learning models. The methods utilized include Logistic Regression, Decision Tree, Random Forest, Hierarchical Clustering, DBSCAN Clustering, and AutoML using EvalML (Krishna & Praveenchandar, 2022; Zender et al., 2023).

Methods and Justification

Logistic Regression will be used as a baseline model for fraud detection. This classification method is widely recognized for its simplicity and interpretability, making it useful for identifying the basic correlation between financial features and fraudulent outcomes (Krishna & Praveenchandar, 2022).

Random Forest, an ensemble learning method, provides higher accuracy compared to individual models such as Logistic Regression or Decision Tree. Combining multiple decision

trees for classification helps improve the detection rate by effectively reducing overfitting and increasing the robustness of predictions (Krishna & Praveenchandar, 2022).

LightGBM (Light Gradient Boosting Machine), a gradient-boosting framework, is employed for its high efficiency and accuracy in handling large datasets with imbalanced classes. Its ability to capture subtle patterns in data and its computational speed make it a powerful tool for fraud detection, particularly in scenarios requiring rapid analysis of evolving patterns.

XGBoost (Extreme Gradient Boosting), another gradient-boosting method, is utilized for its exceptional performance in handling complex and large-scale datasets (Nijwala et al., 2023). Its optimization capabilities and scalability ensure high precision and recall, making it a preferred choice for minimizing false positives and false negatives in fraud detection tasks.

To automate the model selection, hyperparameter tuning, and evaluation process, AutoML using EvalML will be applied. This enables efficient exploration of various models and provides the most suitable one for fraud detection in a given dataset, potentially enhancing overall accuracy and efficiency (Zender et al., 2023).

Software Tools and Programming Requirements

The proposed methods will be implemented using R for the machine learning models, including Logistic Regression, Random Forest, LightGBM, and XGBoost. R provides various packages, such as caret and randomForest, which offer efficient implementations for these models, making them suitable for statistical modeling and visualization. EvalML facilitates automatic testing and selecting the best models, enabling quick iteration without writing much custom code. Programming will be required for model development, involving both custom code for model training and evaluation, as well as leveraging existing tools such as EvalML for

automation. The level of programming needed is moderate, involving scripting for data preprocessing, parameter tuning, and evaluation of different models.

Validation

The validation of the proposed fraud detection methods involves splitting the dataset into training and testing sets and evaluating model performance using key metrics such as AUC (Area Under the Curve), F1-score, and overall accuracy. An 80/20 split will be used, where 80% of the data is allocated for training and 20% for testing, ensuring sufficient data for learning and reliable evaluation. The AUC metric will measure the model's ability to distinguish between fraudulent and legitimate transactions across various thresholds, with higher values indicating better discriminatory capabilities. Overall accuracy will provide a general measure of the correctly classified transactions; however, due to the class imbalance in credit card fraud datasets, accuracy may not fully reflect the model's efficacy. Therefore, the F1-score, which balances precision and recall, is emphasized as it addresses the trade-off between false positives and false negatives.

In the context of credit card fraud detection, false positives (legitimate transactions incorrectly flagged as fraudulent) are more problematic than false negatives (fraudulent transactions missed). False positives can lead to customer dissatisfaction due to blocked legitimate transactions, create operational inefficiencies through unnecessary manual reviews, and erode trust in the system. On the other hand, false negatives, while undesirable, pose less immediate disruption and are often addressed by secondary detection measures or customer reporting. As such, prioritizing the F1 score ensures that the models maintain a balanced approach, minimizing false positives while still effectively identifying fraud.

K-fold cross-validation will be applied during the evaluation process to ensure robustness further. This technique involves splitting the data into k subsets (folds) and training the model on k-1 folds while validating the remaining fold. This process is repeated k times, with each fold being the validation set once. Cross-validation mitigates overfitting and ensures that the models generalize well to unseen data. After cross-validation, the average performance metrics across all folds will be reported to provide a comprehensive assessment.

The validation process also includes error analysis to examine instances of false positives and false negatives, helping to identify trends or patterns that may require adjustments to the feature set or model parameters. Comparative analysis between traditional rule-based systems and machine learning models will be conducted, expecting that machine learning models demonstrate superior F1-scores and AUC values, thereby validating the hypothesis that machine learning approaches outperform rule-based systems. This comprehensive validation framework ensures that the proposed methods are accurate, efficient, and practical for real-world credit card fraud detection applications.

Result

The evaluation highlights a clear progression in performance, with advanced machine learning models significantly outperforming the baseline and simpler models. The baseline model, serving as a random classifier, achieved an F1-score of 0.667 and an AUC of 0.500, underscoring its inability to identify meaningful patterns in the data. Logistic regression improved upon the baseline with an F1-score of 0.886 and an AUC of 0.952, demonstrating its utility as a foundational model. However, ensemble methods like Random Forest performed better, achieving an F1-score of 0.933 and an AUC of 0.982, owing to their ability to capture non-linear patterns. LightGBM further advanced performance with an F1-score of 0.964 and an

AUC of 0.997, leveraging gradient boosting to handle imbalanced data effectively. XGBoost surpassed all models, achieving an F1-score of 0.971 and an AUC of 0.997, showcasing its superior capability to balance false positives and false negatives, which is critical for fraud detection. The following table shows a clear comparison between the performance of different models on the testing dataset:

Model	Accuracy	F1-Score	AUC
Baseline	0.500786782061369	0.6673656618610747	0.5
Logistic Regression	0.890243902439024	0.8860759493670886	0.952432685397202
Random Forest	0.9315499606608969	0.932973805855161	0.9816631660659004
LightGBM	0.963414634146341	0.9644902634593356	0.99714814010079
XGBoost	0.970889063729347	0.9710711493354183	0.996709249571478

Table 1: Results on Models in terms of accuracy, f1-score, and AUC.

These results demonstrate the model's ability to effectively balance precision and recall, which is critical in fraud detection scenarios where false positives and false negatives carry significant operational and customer satisfaction implications. The pipeline used by the AutoML framework that createsa a model with XGBoost included various preprocessing steps such as label encoding, feature selection, date-time featurization, and imputation, ensuring robust handling of both categorical and numerical data. The final XGBoost Classifier utilized hyperparameters optimized for binary classification, achieving high accuracy while maintaining computational efficiency. Cross-validation results further confirmed the model's reliability. The

consistency across folds, evidenced by low standard deviation and coefficient of variation, highlights the model's generalizability to unseen data.

Conclusion

The results validate the hypothesis that machine learning models, particularly those leveraging advanced techniques like AutoML and XGBoost, outperform traditional rule-based systems in detecting credit card fraud. With an AUC of 0.9967 and an F1-score of 0.9711, the XGBoost model demonstrated a superior ability to identify fraudulent transactions while minimizing false positives and false negatives. This balance is crucial in fraud detection, where false positives can lead to customer dissatisfaction and operational inefficiencies, and false negatives can result in financial losses.

The robust pipeline employed by the AutoML framework ensured effective preprocessing, feature engineering, and model optimization, significantly reducing the need for manual intervention. Furthermore, the inclusion of cross-validation provided strong evidence of the model's stability and reliability.

In conclusion, the proposed approach using XGBoost offers a scalable, efficient, and accurate solution for credit card fraud detection. Future work could focus on integrating this solution into real-time fraud detection systems and exploring its adaptability to emerging fraud patterns. These advancements could further enhance its practical application in financial systems, ensuring a secure and trustworthy transactional environment.

References

Cheng, D., Zou, Y., Xiang, S., & Jiang, C. (2024). Graph Neural Networks for Financial Fraud Detection: A review. arXiv (Cornell University). https://doi.org/10.1007/s11704-024-40474-y

Credit Card Transactions Dataset. (2024, July 23). Kaggle. https://www.kaggle.com/datasets/priyamchoksi/credit-card-transactions-dataset

Khanum, A., S, C. K., Singh, B., & Gomathi, C. (2024). Fraud Detection in Financial Transactions: A Machine Learning approach vs. Rule-Based Systems. 2024 International Conference on Intelligent and Innovative Technologies in Computing, Electrical and Electronics (IITCEE), 1–5. https://doi.org/10.1109/iitcee59897.2024.10467759

Nijwala, D. S., Maurya, S., Thapliyal, M. P., & Verma, R. (2023). Extreme Gradient Boost Classifier based Credit card Fraud Detection model. *2023 International Conference on Device Intelligence, Computing and Communication Technologies, (DICCT)*, 500–504. https://doi.org/10.1109/dicct56244.2023.10110188

Thilagavathi, M., Saranyadevi, R., Vijayakumar, N., Selvi, K., Anitha, L., & Sudharson, K. (2024). AI-Driven fraud detection in financial transactions with graph neural networks and anomaly detection. 2024 International Conference on Science Technology Engineering and Management (ICSTEM), 1–6. https://doi.org/10.1109/icstem61137.2024.10560838