Journal of Artificial Intelligence and Data Mining (JAIDM), Vol. 12, No. 1, 2024, 57-66.

Shahrood University of Technology

Journal of Artificial Intelligence and Data Mining (JAIDM)
Journal homepage: http://jad.shahroodut.ac.ir

Research paper

X-SHAoLIM: Novel Feature Selection Framework for Credit Card Fraud Detection

Sajjad Alizadeh Fard and Hossein Rahmani*

School of Computer Engineering, Iran University of Science and Technology, Tehran, Iran.

Article Info

Article History:
Received 09 October 2023
Revised 22 November 2023
Accepted 24 January 2024

DOI:10.22044/jadm.2024.13630.2480

Keywords:
Fraud Detection, Machine Learning, Feature Selection, Ensemble Learning, Explainable AI, Data Mining.

**Corresponding author: h_rahmani@iust.ac.ir (H. Rahmani).*

Abstract

Fraud in financial data is a significant concern for both businesses and individuals. Credit card transactions involve numerous features, some of which may lack relevance for classifiers and could lead to overfitting. A pivotal step in the fraud detection process is feature selection, which profoundly impacts model accuracy and execution time. In this paper, we introduce an ensemble-based, explainable feature selection framework founded on SHAP and LIME algorithms, called "X-SHAoLIM". We applied our framework to diverse combinations of the best models from previous studies, conducting both quantitative and qualitative comparisons with other feature selection methods. The quantitative evaluation of the "X-SHAoLIM" framework across various model combinations revealed consistent accuracy improvements on average including increases in Precision (+5.6), Recall (+1.5), F1-Score (+3.5), and AUC-PR (+6.75). Beyond enhanced accuracy, our proposed framework, leveraging explainable algorithms like SHAP and LIME, provides a deeper understanding of features' importance in model predictions, delivering effective explanations to system users.

1. Introduction

In the recent years, numerous studies have explored the use of machine learning methods to identify and prevent fraudulent transactions [1,2]. Some of the fundamental challenges in detecting financial fraud are outlined below [3-7]:

1. Limited access to real-world data sets
2. Imbalanced class distribution
3. Feature engineering
4. Feature selection
5. Sequence modeling
6. Explainability

Credit card transactions typically have a large number of features. Some features may not be meaningful to the classifiers or lead to overfitting (features that have many categorical values or are too sparse). Additionally, the feature selection step can enhance both the speed and performance of classifiers [7].

Explaining the operations of complex models poses a significant challenge, especially in security domains with sensitive data. Clear explanations for system users are crucial, emerging as an ethical and legal imperative in many applications [8, 9].

To address the challenges mentioned earlier, we introduce an "ensemble-based explainable feature selection framework" known as "X-SHAoLIM.". We employ an ensemble approach and thoroughly assess its effectiveness across various combinations of the best models in the state-of-the-art.

SHAP and LIME are two key explainability algorithms. The SHAP (SHapley Additive exPlanations) algorithm, developed by Lundberg and Lee in 2017, employs concepts from game theory to provide localized explanations for forecasting models. In the context of game theory, the model serves as the rules of the game, and features are akin to potential players. SHAP calculates the Shapley value by evaluating the model across various combinations of input

features, quantifying the average difference in predictions when a specific feature is present versus when it is absent—a measure that reveals the contribution of each feature to the model's prediction [10, 11].

On the other hand, the LIME (Local Interpretable Model-agnostic Explanations) algorithm serves as a local explainability method by systematically adjusting input parameters and observing resulting changes in the output. This approach enhances understanding of the model's predictions, pinpointing, which input variables significantly influenced the outcome for a specific sample [10]. From a technical perspective, LIME generates a new dataset centered around the examined sample, obtains model predictions for these perturbed instances, and subsequently trains an interpretable model, such as linear regression, on this augmented dataset. The process assigns greater weight to instances closer to the original sample, providing local explainability for the analyzed data point [12].

The structure of this paper is as what follows.

Section 2 reviews prior research across ensemble learning, feature engineering, and explainability. In Section 3, we describe our fraud detection process. The main contribution of our paper is in the feature selection stage. Section 4 evaluates our framework for quantitative and qualitative comparisons with other feature selection algorithms like ANOVA, random forest, and XGBoost. Section 5 includes a summary and conclusion, and proposes promising directions for future research.

2. Background

In this section, we review previous works in the field of ensemble learning, feature engineering, and explainability. Ensemble learning models, comprising multiple sub-models, have consistently demonstrated superior performance when compared to individual models such as logistic regression, artificial neural networks, support vector machines, and k-nearest neighbors [13]. Several studies have found that random forests are among the most effective ensemble methods [14-17].

Randhawa *et al*. [18] initially employed standalone standard models for credit card fraud detection. Subsequently, they explored the combination of models using the AdaBoost and Majority Vote techniques. To assess the algorithms' resilience against noisy data, noisy data samples were introduced. The experimental outcomes ultimately indicated that the Majority Voting method exhibited commendable accuracy in credit card fraud detection, and demonstrated robust performance even in the presence of noisy data.

Figuerola *et al*. [19] investigated the performance of tree-based ensemble learning algorithms in detecting fraudulent transactions. They specifically examined random forest, bagging, XGBoost, LightGBM, and CatBoost classifiers. They used ANOVA for the feature selection step. Their findings revealed that boosting classifiers outperformed bagging classifiers in fraud detection, with LightGBM achieving the most favorable results across multiple metrics including F1, MCC, and AUC-PR.

Since raw input features are not sufficient to detect fraudulent transactions, feature engineering strategies have been proposed. Feature selection both removes redundant features and increases learning accuracy [15, 20]. Saheed *et al*. [21] dealt with fraud detection using a genetic algorithm as a feature selection method. They first selected the top 8 features, and used those features in NB, RF, and SVM algorithms for fraud detection on the German credit card dataset. The experimental results showed that the random forest performs better than NB and SVM.

Emmanuel Ileberi *et al*. [22] introduces a Genetic Algorithm (GA)-based feature selection method combined with Random Forest (RF), Decision Tree (DT), Artificial Neural Network (ANN), Naive Bayes (NB), and Linear Regression (LR) classifiers for credit card fraud detection. Results on a European cardholders dataset reveal superior performance, with GA-RF achieving 99.98% accuracy. Validation results demonstrate GA-DT's 100% accuracy and GA-ANN's AUC of 0.94, showcasing the framework's effectiveness for fraud detection. Bharat Padhi *et al*. [23] addressed challenges in credit card fraud detection by proposing a novel feature selection method, Rock Hyrax Swarm Optimization Feature Selection (RHSOFS), inspired by natural swarm behavior. Employing supervised machine learning, the approach enhances fraud identification by selecting optimal features from high-dimensional datasets generated from European cardholder dataset. In a comparative analysis, RHSOFS surpasses existing methods including Differential Evolutionary Feature Selection (DEFS), Genetic Algorithm Feature Selection (GAFS), Particle Swarm Optimization Feature Selection (PSOFS), and Ant Colony Optimization Feature Selection (ACOFS), demonstrating superior efficiency. SHAP and LIME algorithms are among the explainability algorithms utilized in various studies [24-29].

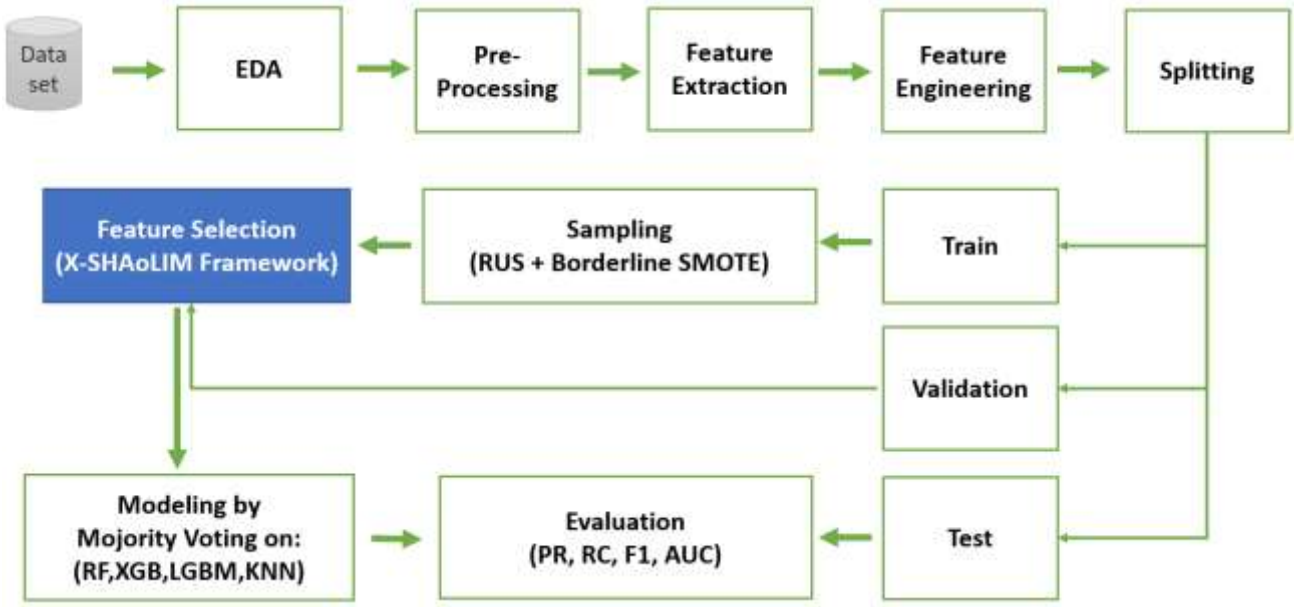

Figure 1. Steps of fraud detection process in our paper.

Sindhgatta *et al*. [26] employed SHAP and LIME algorithms to interpret predictions from diverse models using user log data, highlighting the inadequacy of relying solely on performance metrics. Their study emphasized the crucial role of model explanations in revealing feature significance and recommended incorporating explainability assessments alongside traditional performance metrics for comprehensive model evaluations.

Psychoula *et al*. [27] conducted a comparison of SHAP and LIME explainability algorithms within the realm of real-time fraud detection, spanning both supervised and unsupervised models. Their findings indicated that the SHAP algorithm yields more dependable results than the LIME algorithm. However, the LIME algorithm is recommended for real-time explainability purposes.

3. Methodology

In this section, we introduce our fraud detection process, "step by step" as shown in Figure 1. Each section describes each step (Feature Extraction and Feature Engineering steps described in one section). The main contribution of our paper is in the "feature selection step".

3.1. Dataset

As seen in the background, most studies in credit card fraud detection utilized the European credit card dataset [30], in which all the predictors are continuous and resulted after Principal Component Analysis (PCA) transformation. It is unknown the extent to which their findings can be generalized to datasets that contain a mix of continuous and categorical predictors. Thus we utilized a dataset [31] comprising credit card transactions, encompassing information from 1,000 American customers. It consists of 1,852,394 transactions and 23 columns, encompassing both numeric and categorical features. The "is_fraud" column serves as the target label, indicating the fraudulent or legal nature of each transaction. Notably, this dataset is highly imbalanced, with fraudulent transactions accounting for a mere 0.52% of the total transaction count. Table 1 shows the columns of this dataset.

3.2. Exploratory data analysis

In order to gain an overview of the dataset, we perform Exploratory Data Analysis (EDA), which includes the following steps:

- Investigating the number of unique values in non-continuous features.
- Checking the range of continuous features.
- Analyzing feature values to distinguish between fraudulent and legitimate transactions.
- Identifying feature values that occur most frequently in fraudulent transactions.

According to Figure 2, the features "trans_num", "unix_time", "trans_date_trans_time", "merch_long", and "merch_lat" have the most distinct values.

Irrelevant features such as unique identifiers, features with a high number of unique values, and redundant features were removed from the dataset to prevent models from overfitting.

Table 1. Dataset features.

Feature/s	Type	Description
is_fraud	Binary	Whether the transaction is fraud or not
amt	Continuous	Amount of the transaction
city-pop	Continuous	Population of the city the customer lives
unix-time	Continuous	Time of the transaction in unix time
trans-day-trans-time	Interval-scale	Date and Time of the transaction (trxn)
dob	Interval-scale	Date of birth of the customer
first / last	Nominal	First and Last name of the customer
gender	Binary	Gender of the customer
merchant	Nominal	Merchant the customer is paying to
merch-lat / merch-long	Continuous	Merchant's Latitude and Longitude
street / city / state	Nominal	Street, City, and State where customer lives
zip	Nominal	ZIP code on credit card
lat / long	Continuous	Latitude and Longitude of the customer
cc-num	Nominal	Credit card number of the customer
trans-num	Nominal	Unique trxn num. for each and every trxn
category	Nominal	Shopping category
job	Nominal	Job of the customer

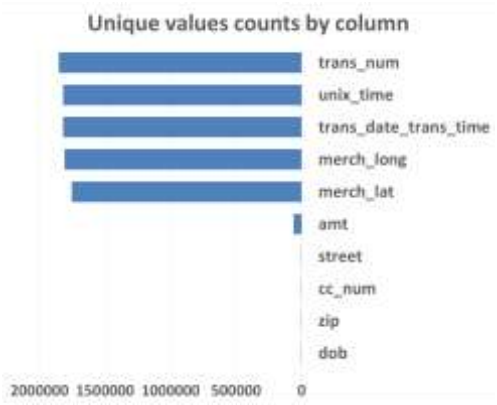

Figure 2. Features with most unique values.

The attributes "trans_num", "unix_time", "merch_lat", and "merch-long" were removed initially, and the attribute "trans_date_trans_time" was subsequently removed after extracting meaningful features from it. According to Figure 3, the average amount of fraudulent transactions was \$530, while the average amount of legal transactions was \$67.

Figure 3. The amount of fraudulent transactions is nearly 10 times that of legitimate transactions.

According to Figure 4, approximately 50% of the fraudulent transactions belong to the "grocery_pos" or "shopping_net" category.

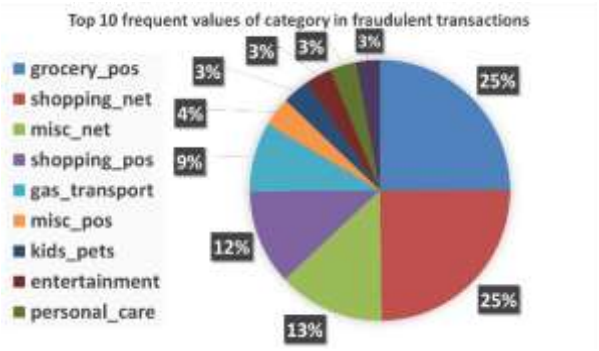

Figure 4. Top 10 frequent values of category for fraud (in percentage). Approximately 50% of the fraudulent transactions belong to the "grocery_pos" or "shopping_net" category.

3.3. Sampling

In order to train and evaluate our models, we divided the original dataset into the following three sets in a stratified manner:

- Train set (including 60% of the original set)
- Test set (including 20% of original datasets)
- Validation set (including 20% of the original set)

After dividing the original dataset into training, testing, and validation sets, we performed sampling only on the training set. In order to create the possibility of comparison with the same conditions, we follow Figuerola *et al*. [18] sampling approach,

and combine RUS and Borderline-SMOTE methods. Initially, the samples of the majority class were reduced to 20 times the number of samples in the minority class. Subsequently, the samples of the minority class were increased to 90% of the number of samples in the majority class. Figure 5 illustrates the resulting imbalance ratio (legal/fraudulent) after sampling.

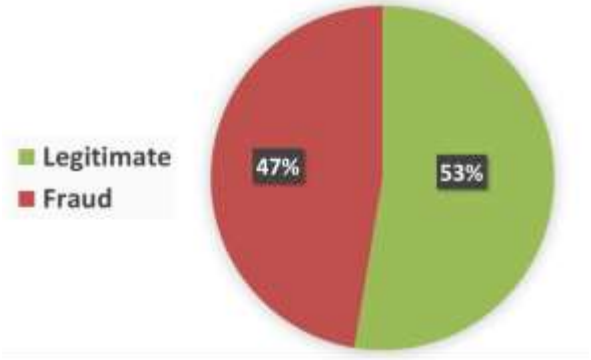

Figure 5. Class distribution after applying sampling algorithms.

3.4. Feature engineering

Even though transactions involve numerous attributes, there is a necessity to create new attributes to enhance the description of transactions. Attributes such as "trans_date_trans_time" and "dob" do not provide valuable information to the model in their raw form due to their high number of unique values and should be replaced with more informative features. Additionally, the nature of the transaction sequence is often overlooked. For these reasons, it becomes essential to extract new features that provide a more detailed description of transactions, as detailed in Table 2.

In this work, the Ordinal Encoding method was employed to encode ordinal variables, while the Target Encoding method was utilized to encode nominal variables. Subsequently, after converting the categorical features into continuous ones, the Z-Score normalization method was applied.

3.5. Feature selection (X-SHAoLIM framework)

In this section, we describe our feature selection framework in detail.

The main application of SHAP and LIME algorithms are explainability for model prediction. As seen in the background section, most papers use these algorithms to explain the predictions of their models. Also some studies used SHAP or LIME for feature selection "alone".

Our main idea is to use an "ensemble-based approach" in the feature selection step, focusing on the SHAP and LIME explainability algorithms (both algorithms form the main cores of the proposed framework and we used them in two different stages).

We introduce our proposed framework as an "ensemble-based explainable framework" called "X-SHAoLIM". We used the word "framework" because we have actually introduced a new "structure" for feature selection. In every structure, the main components and how they interact and aggregate should be specified. Also, any structure should have the necessary flexibility to be used in any problem. Considering the previous two points, our framework includes three main components (according to Figure 6).

a) **Candidate feature selection stage:** The SHAP algorithm forms the core of this component. At this stage, different algorithms can be placed next to the SHAP algorithm and aggregated (ensemble-based approach). Also it is possible to use the SHAP algorithm with one or multiple base models (BM). For example, we used ANOVA next to the SHAP algorithm, and used the LGBM model as SHAP's base model.

b) **Voting stage:** In this stage, it is necessary to determine how to combine different features from each algorithm in the previous phase (common-based, union-based or weighted-based). For example, we give more importance to the SHAP algorithm. We make the union of the top 10 features of SHAP and the top 5 features of ANOVA.

c) **Filtering stage:** After extracting the top features from previous stage, we run the models with selected features on the "validation set", and examine the confusion matrix to indentify false predicted cases (included false positives and false negatives).

At this stage, we use the LIME algorithm to identify features that have a negative impact on the selected cases and remove them from the set of selected features.

The "street" feature was detected as the feature that has the most negative effect on both false positive and false negative cases.

Now it is clear why we called it an "ensemble-based explainable framework".

Table 2. Generated features from raw features.

trans_month	Nominal	Month of the transaction	trans_date_trans_time
trans_day	Nominal	Day of the transaction	trans_date_trans_time
trans_hour_category	Nominal	Category of transaction hour (Evening, Morning, Aftternon, Night)	trans_date_trans_time
age	Continuous	Age of customer	dob
elapsed_time_seconds	Continuous	Different between previous and current transaction time for specific card	trans_date_trans_time cc-num
diff_amt	Continuous	Different between previous and current transaction amount for specific card	trans_date_trans_time amt cc-num
sum_amt_last_7_days	Continuous	Sum of transactions amount in the last 7 days for specific card	trans_date_trans_time amt cc-num
sum_amt_last_14_days	Continuous	Sum of transactions amount in the last 14 days for specific card	trans_date_trans_time amt cc-num
sum_amt_last_30_days	Continuous	Sum of transactions amount in the last 30 days for specific card	trans_date_trans_time amt cc-num
sum_amt_last_60_days	Continuous	Sum of transactions amount in the last 60 days for specific card	trans_date_trans_time amt cc-num
cc_count	Continuous	Count of credit card for each customer	full_name cc-num
full_name	Nominal	Full name of the customer (first name + last name)	first last

We called it "X-SHAoLIM" because SHAP and LIME are the core algorithms of this framework and the first "X" character refers to the flexibility of this framework.

It is possible to combine other feature selection algorithms in the candidate selection stage and various base models for both SHAP and LIME. Figure 6 shows how we used the X-SHAoLIM framework in our fraud detection process.

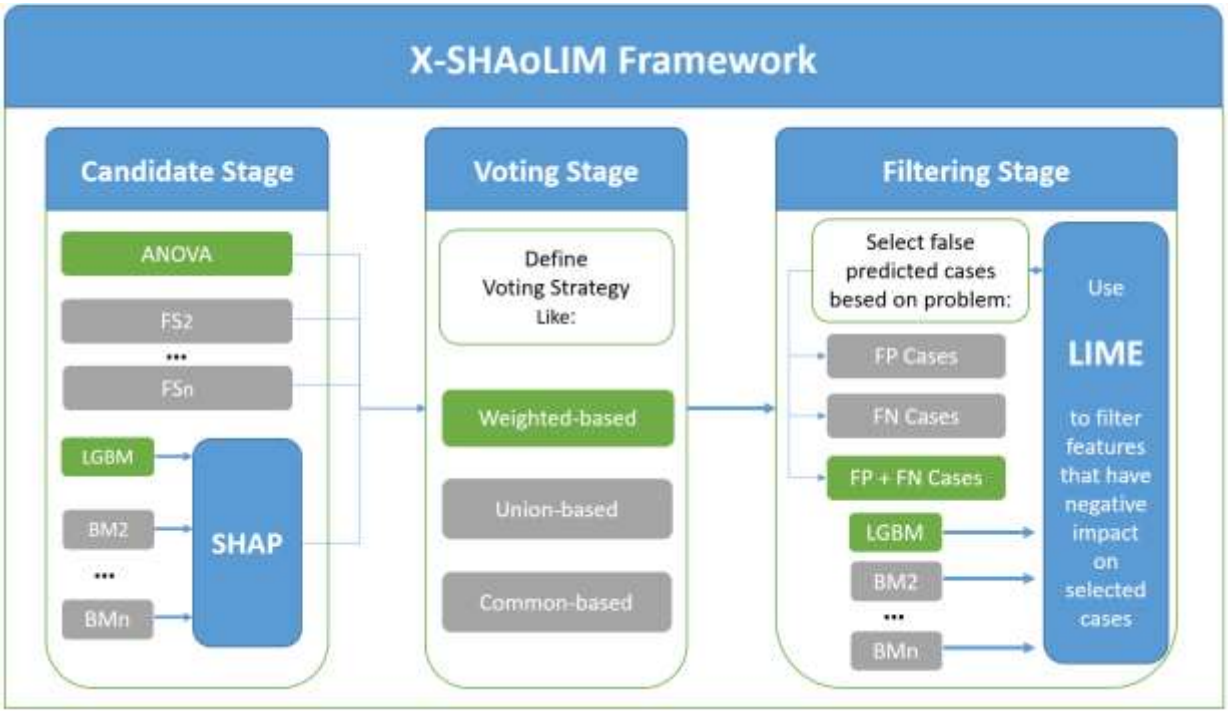

Figure 6. In our fraud detection process, we used X-SHAoLIM framework with above configuration in three stages (FS refers to Feature Selection algorithm and BM refers to Base Model).

Table 3. Results of feature selection algorithms based on four evaluation metrics, highlighting X-SHAoLIM as the top performer.

Ensemble Model	ANOVA				Random forest				XGBoost				X-SHAoLIM			
	PR	RC	F1	AUC-PR	PR	RC	F1	AUC-PR	PR	RC	F1	AUC-PR	PR	RC	F1	AUC-PR
RF + XGB + LGBM	91	86	89	86	94	87	90	90	95	88	91	90	**97**(+6)	**88**(+2)	**92**(+3)	**93**(+7)
RF + XGB + KNN	91	85	88	83	93	86	89	86	95	86	90	88	**97**(+6)	**86**(+1)	**91**(+3)	**90**(+7)
RF + LGBM + KNN	92	84	88	84	94	85	89	88	96	86	90	89	**98**(+6)	**86**(+2)	**91**(+3)	**91**(+7)
XGB + LGBM + KNN	85	90	87	85	88	90	89	88	90	91	90	89	**93**(+8)	**91**(+1)	**92**(+5)	**91**(+6)

3.6. Modeling (Ensemble models)

In this section, we introduce the models employed in the fraud detection process. As observed in previous works, tree-based ensemble models such as random forests, XGBoost, and LightGBM have consistently demonstrated superior performance among other models. In addition to these eager models, we also included the KNN model, a lazy model. We use these four models to create various combinations. Each combination includes three base models. In each combination,
we apply majority voting to create final prediction. In the following section, we evaluate the performance of the "X-SHAoLIM" framework on these combinations.

4. Evaluation

In this section, we evaluate the performance of the "X-SHAoLIM" framework both quantitatively and qualitatively.

4.1. Quantitive evaluation

For comparative analysis, we compare our framework result with three famous feature selection algorithms (ANOVA, Random Forest, and XGBoost) on different combinations of ensemble models, based on Precision, Recall, F1, and AUC-PR. The number of features in all feature selection algorithms and our framework is 10.
Table 3 shows that the ANOVA algorithm exhibits the lowest level of accuracy in various combinations of models. In contrast, both the Random Forest and XGBoost algorithms demonstrate better performance, leading to increased accuracy. Finally, it is evident that the use of the "X-SHAoLIM" framework in the feature selection stage significantly increases the accuracy of the model, as in the improvement of accuracy (+5.6), recall (+1.5), F1 (+3.5), and AUC-PR (+6.75), compared to the least accurate algorithm (i.e. ANOVA). This framework outperforms other feature selection algorithms.
As mentioned in section 3.1, most studies used the European credit card dataset, in which all features are continuous and were transformed by PCA. As seen in the background some studies used genetic algorithm or optimization algorithms on different datasets. But we can compare our results with the work that had the same condition (same dataset and same sampling ratio). Figola *et al*. [18] compared performance of bagging and boosting models on this dataset. They used the ANOVA algorithm in the feature selection step. They showed the LGBM model had the best performance (according to Table 4).

Table 4. Compare our evaluation results with similar work on this dataset.

Method	Precision	Recall	F1-score	AUC-PR
LightGBM + ANOVA [18]	57	89	70	73
(XGB + LGBM + KNN) + X-SHAoLIM	**93**	**91**	**92**	**91**

Furthermore, beyond the accuracy enhancement, the proposed framework offers increased explainability in analyzing the impact of features on the models, which we will further examine.

4.2 Quality evaluation

Table 5 presents the top ten features identified by different algorithms.
Notably, as observed in the EDA section (Figure 3), there exists a significant disparity in the average amount of transactions between the fraudulent and legitimate labels.

According to Table 5, both the "amt" and "sum_amt_last_7_days" features emerge as the most important features across all three algorithms. Furthermore, the analysis of the "trans_hour_category" feature values in Figure 7 reveals variations between different times of the day and night in terms of the occurrence of fraud.

Table 5. Top ten features in other feature selection algorithms in the best model combination.

#	ANOVA	Random forest	XGBoost
1	sum_amt_last_7_days	amt	amt
2	amt	sum_amt_last_7_days	sum_amt_last_7_days
3	sum_amt_last_14_days	sum_amt_last_14_days	category
4	sum_amt_last_30_days	diff_amt	trans_hour_category
5	merchant	sum_amt_last_30_days	diff_amt
6	category	category	full_name
7	sum_amt_last_60_days	sum_amt_last_60_days	sum_amt_last_14_days
8	street	merchant	merchant
9	full_name	trans_hour_category	street
10	city	full_name	elapsed_time_seconds

It's worth noting that, unlike other algorithms, this feature does not appear among the top 10 features of the ANOVA algorithm (weakness of the ANOVA algorithm).

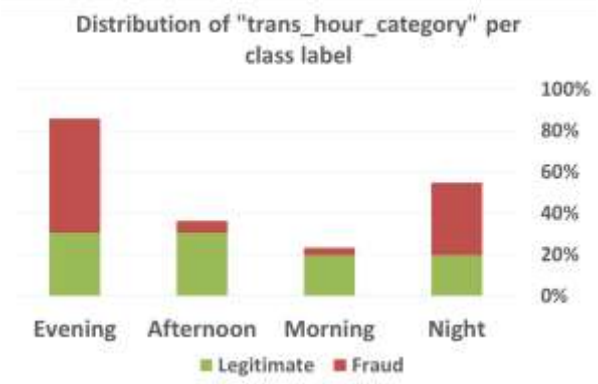

Figure 7. Distribution of "trans_our_category" feature across labels, highlighting higher fraud occurrence during Night and Evening hours.

Also sum_amt_last_60_days feature is seen among the selected features of ANOVA and random forest algorithms, but this feature is not among the selected features of XGBoost algorithm. The analysis of sum_amt_last_60_days attribute values shows that, unlike the amt attribute, there is no high difference between the average values of this attribute in fraudulent and legitimate transactions (Figure 8).

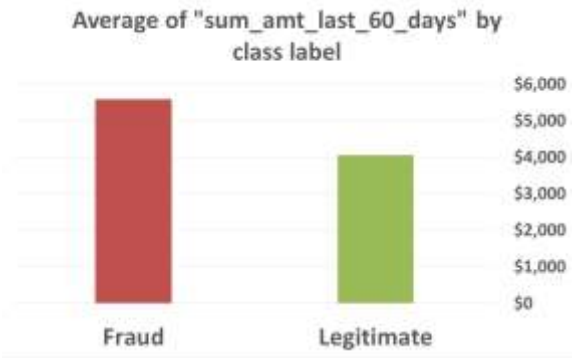

Figure 8. Average of "sum_amt_last_60_days" per class label. Note the smaller gap between fraudulent and legitimate values compared to the "amt" feature.

In the next phase, we examine the performance of the "X-SHAoLIM" framework. As depicted in Figure 9, similar to the XGBoost algorithm, the "trans_hour_category" feature is among the top 5 features of "X-SHAoLIM", while the "sum_amt_last_60_days" feature does not make it into the top 10 features. Additionally, unlike the previous three algorithms, the "full_name" feature is not included among the final features.

According to Figure 9, analysis of the Filtering component revealed that, in both false positive and false negative samples, the "street" feature has the most negative effect on these cases. Consequently, it was removed from the set of selected features. Interestingly, this feature was among the top 10 features in the ANOVA and XGBoost algorithms.

Furthermore, as shown in Figure 10, both the "street" and "full_name" features have nearly 1000 unique values, and these attributes are not included in the "X-SHAoLIM" attribute set.

5. Conclusion and Future Works

The feature selection step stands as a pivotal component in the fraud detection process, significantly influencing model accuracy and execution time.

In this paper, we have introduced an "explainable feature selection framework" based on an ensemble approach. Our research work involved the application of the proposed framework to various combinations of the best models identified in previous works, followed by a comprehensive quantitative and qualitative comparison with other feature selection algorithms.

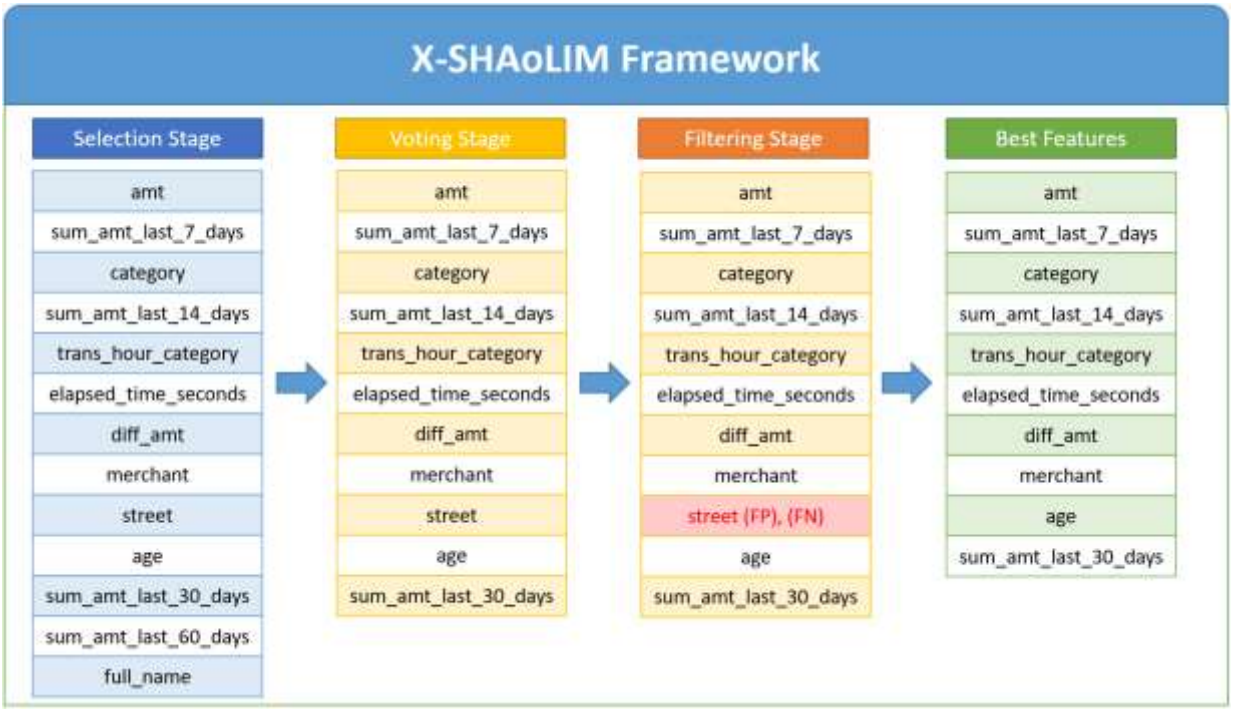

Figure 9. Traversal of features in the stages of the X-SHAoLIM framework. The street feature was identified in the filtering stage and removed from the final feature set.

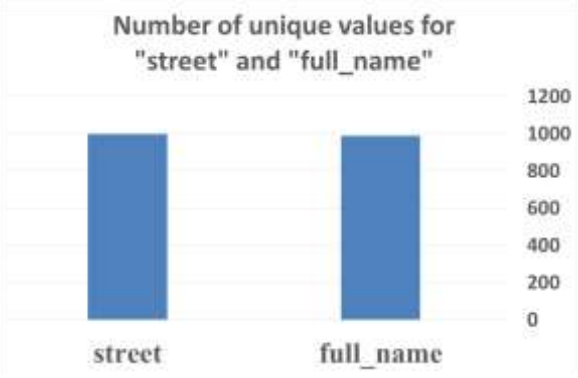

Figure 10. Number of unique values for "street" and "full_name" features. They have an excessive number of unique values and are not included in the X-SHAoLIM feature set.

The quantitative evaluation of the "X-SHAoLIM" framework across diverse model combinations has demonstrated substantial enhancements in model accuracy. Notably, we have observed significant improvements in Precision (+5.6), Recall (+1.5), F1-Score (+3.5), and AUC-PR (+6.75) compared to other feature selection algorithms, establishing its superiority. Furthermore, this paper has underscored the value of incorporating SHAP and LIME explainability algorithms into the feature selection process. Beyond enhancing model performance, these algorithms offer effective explanations of model behavior, adding an invaluable layer of transparency and interpretability to the fraud detection process. Looking ahead, future works can explore the application of our proposed framework to different datasets, diverse models, and varied combinations of models. Additionally, comparative studies with other feature selection algorithms can provide deeper insights into its performance and versatility.

References

[1] S. Mittal and S. Tyagi, "Performance Evaluation of Machine Learning Algorithms for Credit Card Fraud Detection," *IEEE Xplore*, Jan. 01, 2019.

[2] S. Beigi and M. R. Amin Naseri, "Credit Card Fraud Detection using Data mining and Statistical Methods," *Journal of AI and Data Mining*, vol. 8, no. 2, pp. 149-160, Apr. 2020.

[3] R. Yan, Y. Liu, R. Jin, and A. Hauptmann, "On predicting rare classes with SVM ensembles in scene classification," *IEEE Xplore*, Apr. 01, 2003.

[4] N. V. Chawla, N. Japkowicz, and A. Kotcz, "Editorial," *ACM SIGKDD Explorations Newsletter*, vol. 6, no. 1, p. 1, Jun. 2004.

[5] Y. Russac, O. Caelen, and L. He-Guelton, "Embeddings of Categorical Variables for Sequential Data in Fraud Context," *The International Conference on Advanced Machine Learning Technologies and Applications (AMLTA2018)*, pp. 542-552, 2018.

[6] C. Guo and F. Berkhahn, "Entity Embeddings of Categorical Variables," *arXiv.org*, Apr. 22, 2016.

[7] S. Bhattacharyya, S. Jha, K. Tharakunnel, and J. C. Westland, "Data mining for credit card fraud: A comparative study," *Decision Support Systems*, vol. 50, no. 3, pp. 602-613, Feb. 2011.

[8] A. Adadi and M. Berrada, "Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI)," *IEEE Access*, vol. 6, pp. 52138-52160, 2018.

[9] Małgorzata Magdziarczyk, "RIGHT TO BE FORGOTTEN IN LIGHT OF REGULATION (EU) 2016/679 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL OF 27 APRIL 2016 ON THE PROTECTION OF NATURAL PERSONS WITH REGARD TO THE PROCESSING OF PERSONAL DATA AND ON THE FREE MOVEMENT OF SUCH DATA, AND REPEALING DIRECTIVE 95/46/EC," *SGEM International Multidisciplinary Scientific Conferences on Social Sciences and Arts*, Apr. 2019.

[10] R. Omobolaji Alabi, A. Almangush, M. Elmusrati, I. Leivo, and A. A. Mäkitie, "An interpretable machine learning prognostic system for risk stratification in oropharyngeal cancer," *International Journal of Medical Informatics*, vol. 168, p. 104896, Dec. 2022.

[11] Rasheed Omobolaji Alabi, M. Elmusrati, Ilmo Leivo, Alhadi Almangush, and Antti Mäkitie, "Machine learning explainability in nasopharyngeal cancer survival using LIME and SHAP," *Scientific Reports*, vol. 13, no. 1, Jun. 2023.

[12] A. Gramegna and P. Giudici, "SHAP and LIME: An Evaluation of Discriminative Power in Credit Risk," *Frontiers in Artificial Intelligence*, vol. 4, Sep. 2021.

[13] I. Sohony, R. Pratap, and U. Nambiar, "Ensemble learning for credit card fraud detection," *Proceedings of the ACM India Joint International Conference on Data Science and Management of Data - CoDS-COMAD '18*, 2018.

[14] A. Dal Pozzolo, O. Caelen, Y.-A. Le Borgne, S. Waterschoot, and G. Bontempi, "Learned lessons in credit card fraud detection from a practitioner perspective," *Expert Systems with Applications*, vol. 41, no. 10, pp. 4915-4928, Aug. 2014.

[15] V. Van Vlasselaer *et al.*, "APATE: A novel approach for automated credit card transaction fraud detection using network-based extensions," *Decision Support Systems*, vol. 75, pp. 38-48, Jul. 2015.

[16] M. Zareapoor and P. Shamsolmoali, "Application of Credit Card Fraud Detection: based on Bagging Ensemble Classifier," *Procedia Computer Science*, vol. 48, pp. 679-685, 2015.

[17] S. D. Penmetsa and S. Mohammed, "Ensemble Techniques for Credit Card Fraud Detection," *International Journal of Smart Business and Technology*, vol. 9, no. 2, pp. 33-48, Sep. 2021.

[18] K. Randhawa, C. K. Loo, M. Seera, C. P. Lim, and A. K. Nandi, "Credit Card Fraud Detection using AdaBoost and Majority Voting," *IEEE Access*, vol. 6, pp. 14277-14284, 2018.

[19] E. Figuerola Ullastres, "Credit Card Fraud Detection using Ensemble Learning Algorithms," *norma.ncirl.ie*. May 30, 2022.

[20] A. Correa Bahnsen, D. Aouada, A. Stojanovic, and B. Ottersten, "Feature engineering strategies for credit card fraud detection," *Expert Systems with Applications*, vol. 51, pp. 134-142, Jun. 2016.

[21] Y. K. Saheed, M. A. Hambali, M. O. Arowolo, and Y. A. Olasupo, "Application of GA Feature Selection on Naive Bayes, Random Forest and SVM for Credit Card Fraud Detection," *2020 International Conference on Decision Aid Sciences and Application (DASA)*, Nov. 2020.

[22]. E. Ileberi, Y. Sun, and Z. Wang, "A machine learning based credit card fraud detection using the GA algorithm for feature selection," *Journal of Big Data*, vol. 9, no. 1, Feb. 2022.

[23] Bharat Kumar Padhi, S. Chakravarty, B. Naik, Radha Mohan Pattanayak, and H. Das, "RHSOFS: Feature Selection Using the Rock Hyrax Swarm Optimization Algorithm for Credit Card Fraud Detection System," vol. 22, no. 23, pp. 9321-9321, Nov. 2022.

[24] van, *Process Mining Handbook*. Springer Nature.

[25] W. Rizzi, C. Di Francescomarino, and F. M. Maggi, "Explainability in Predictive Process Monitoring: When Understanding Helps Improving," *Lecture Notes in Business Information Processing*, pp. 141-158, 2020.

[26] R. Sindhgatta, C. Ouyang, and C. Moreira, "Exploring Interpretability for Predictive Process Analytics," *Service-Oriented Computing*, pp. 439-447, 2020.

[27] I. Psychoula, A. Gutmann, P. Mainali, S. H. Lee, P. Dunphy, and F. Petitcolas, "Explainable Machine Learning for Fraud Detection," *Computer*, vol. 54, no. 10, pp. 49-59, Oct. 2021.

[28] W. E. Marcilio and D. M. Eler, "From explanations to feature selection: assessing SHAP values as feature selection mechanism," *2020 33rd SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*, Nov. 2020.

[29] T.-Y. Wu and Y.-T. Wang, "Locally Interpretable One-Class Anomaly Detection for Credit Card Fraud Detection," *IEEE Xplore*, Nov. 01, 2021.

[30] Kaggle, "Credit Card Fraud Detection," *www.kaggle.com*, 2018. https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud.

[31] K. Shenoy, "Credit Card Transactions Fraud Detection Dataset," Kaggle.com, 2019. https://www.kaggle.com/datasets/kartik2112/fraud-detection.

X-SHAoLIM: یک چارچوب انتخاب ویژگی جدید به‌منظور کشف تقلب در تراکنش‌های کارت‌های اعتباری

سجاد علیزاده فرد و حسین رحمانی *

دانشکده مهندسی کامپیوتر، دانشگاه علم و صنعت، تهران، ایران.

ارسال ۲۰۲۳/۱۰/۰۹؛ بازنگری ۲۰۲۳/۱۱/۲۲؛ پذیرش ۲۰۲۴/۰۱/۲۴

چکیده:

تقلب در داده‌های مالی یک نگرانی جدی برای سازمان‌های تجاری و افراد است. تراکنش‌های کارت‌های اعتباری ویژگی‌های متعددی دارند که برخی از آن‌ها ممکن است برای رده‌بندها، نامرتبط باشند و منجر به بیش برازش شوند. یک مرحله اساسی در فرآیند کشف تقلب، مرحله انتخاب ویژگی‌ها است که به‌شدت بر دقت و زمان اجرای مدل‌ها مؤثر است. ما در این مقاله، به ارائه یک چارچوب انتخاب ویژگی جمعی و توضیح‌پذیر مبتنی بر الگوریتم‌های SHAP و LIME می‌پردازیم (تحت عنوان X-SHAoLIM). ما چارچوب خود را بر روی ترکیبات متنوع از بهترین مدل‌ها در کارهای پیشین اعمال کرده و به مقایسه کمی و کیفی آن با الگوریتم‌های انتخاب ویژگی رایج پرداختیم. ارزیابی کمی چارچوب «X-SHAoLIM» بر روی ترکیبات مختلف از مدل‌های منتخب نشان داد، چارچوب پیشنهادی به‌طور میانگین باعث افزایش دقت مدل‌ها بر اساس معیارهای صحت (+۵٫۶)، فراخوانی (+۱٫۵)، معیار F1 (+۳٫۵) و AUC-PR (+۶٫۷۵) شده و همچنین به دلیل به‌کارگیری الگوریتم‌های توضیح‌پذیر SHAP و LIME، درک عمیق‌تری از اهمیت ویژگی‌ها فراهم کرده و توضیحات مؤثری به کاربران سیستم ارائه می‌دهد.

کلمات کلیدی: تشخیص تقلب، یادگیری ماشینی، انتخاب ویژگی، یادگیری گروهی، هوش مصنوعی قابل توضیح، داده‌کاوی.