IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 4, April 2024

DOI: 10.17148/IJARCCE.2024.134151

A NOVEL FRAMEWORK FOR CREDIT CARD FRAUD DETECTION

G. Kavya[1], E. Bhagyasri[2], K. Jyothi[3], N. Firoz Basha[4]

Student, Computer Science and Engineering & Business Systems, RGMCET, Nandyal, India[1-4]

Abstract: Since from the last few years there is a significant increase in credit card transactions are playing a vital role. Thus, it is leading to significant financial losses everywhere in present days. It is very challenging task to process the huge amount of data and it is making data sets unbalanced and complex. There are basically two major problems while handling data. It is analysed with fraud and non-fraud transactions, and it doesn't contain relevant, appropriate, and correlated data that affects their prediction performance in a negative way. Followed by, it has involved the interest of machine learning (ML), which consists of fraud detection as a main theme. It has been involved by various ML methods such as Logistic Regression (LR), Support vector machines (SVM), Decision Trees (DT), Random Forest (RF), and K-Nearest Neighbours (KNN). However, the above methods cannot meet the excellent performance required to find and predict abnormal fraud patterns. In this project the main contribution is to provide a framework for fraud detection (FFD). Firstly, we have to overcome the unbalanced data issue, the framework uses an under sampling technique. Followed by, we have to select the relevant features by applying the feature selection (FS) mechanism. Next, Neural networks is mainly builds the ML model and it aims to handle the capability, a modified version of the Particle Swarm Optimization (PSO) algorithm, Polynomial Self Learning PSO (PSLPSO), is proposed for hyper parameters C and σ. Finally, the framework's effectiveness is depicted in the experimental results on a transaction dataset of real credit card.

I. INTRODUCTION

Technological advancements in the industry have significantly contributed to the substantial growth of credit card transactions in recent years. Several factors have fuelled this increase:

1. Digital Payment Systems: The emergence of digital payment systems, including mobile wallets, contactless payments, and online platforms, has made credit card usage more convenient and accessible for consumers both in physical stores and online.

2. E-commerce Boom: The rapid growth of e-commerce has driven the adoption of credit cards as the preferred payment method for online shopping due to their convenience, security, and flexibility.

3. Mobile Apps and Banking: Mobile banking apps provided by financial institutions allow users to manage their credit cards and conduct transactions on the go, further facilitating the increase in credit card usage.

4. Security Enhancements: Industry advancements in credit card security measures have increased consumer confidence in using credit cards for transactions, contributing to the overall growth in transactions. However, this growth has also led to significant losses, with approximately 32 billion U.S. dollars lost worldwide in 2021 due to fraudulent activities. Credit card fraud poses a serious threat, prompting the adoption of machine learning (ML) for fraud detection.

ML algorithms such as Logistic Regression (LR), Support Vector Machines (SVM), Decision Trees (DT), Random Forest (RF), K-Nearest Neighbours (KNN), and Artificial Neural Networks (ANN) are commonly used for credit card fraud detection. However, these algorithms face challenges due to the high-dimensional and unbalanced nature of collected data. To address these challenges, feature selection mechanisms are applied to select relevant features for fraud detection algorithms. Despite the use of various ML algorithms, none can consistently achieve outstanding performance. In this paper, the Support Vector Data Description (SVDD) model is proposed as a solution. SVDD, based on SVM, is designed for one-class classification and can handle outliers better than SVM. It identifies anomalies by determining a minimum-volume hypersphere surrounding positive samples in feature space. The remainder of the paper includes a literature review, the proposed Framework, experimental results, conclusions, and future research directions.

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 4, April 2024
DOI: 10.17148/IJARCCE.2024.134151

II. RELATED WORK

Various ML techniques have been used in experimental research to detect and predict fraudulent transactions. In ref. [8], Sarah et al. proposed an unsupervised feature learning method to improve the performance of various classifiers using a stacked sparse auto-encoder (SSAE) to predict fraud.

In their works [9], the authors implemented an approach to detect credit card fraud using a neural network ensemble classifier and a hybrid data resampling method. The ensemble classifier is obtained using a long short-term memory (LSTM) neural network as the base learner in the adaptive boosting (AdaBoost) technique.

Varmedja et al. [10] proposed a credit card fraud detection method using ML algorithms such as RF, NB, and multilayer perceptron (MLP). To overcome the problem of imbalanced data, the researcher implemented the Synthetic Minority Oversampling Technique (SMOTE) technique. The experimental results demonstrated that the RF algorithm got the highest accuracy. Hence, the authors propose using a feature selection method for future works to improve the accuracy.

OMAR et al. [11] proposed an approach to select features that make data achieve the minimal overlap degree and improve classification performance. This work applies three algorithms of feature selection: Reduce Overlapping with No-sampling (RONS), Reduce Overlapping with SMOTE (ROS), and Reduce Overlapping with ADASYN (ROA), which is built through sparse feature selection to minimize the overlapping and perform binary classification. The experimental results show that the proposed algorithms as feature selection methods manage to produce good performance.

Itri, Bouzgarne, et al. [12] presented an approach that combines oversampling and feature selection methods to find the best combination for classification algorithms. The authors demonstrated that their approach drastically improved the performance of the model. The authors in [13] implemented an intelligent payment card fraud detection system. The authors implemented an approach to evaluate whether the aggregated features identified by a genetic algorithm can offer better accuracy, as compared with the original features, in fraud detection.

RTAYLI et al. [14] explored the strength of many Machine Learning methods. They provided a hybrid approach that combines three methods: The Recursive Features Elimination (RFE) method to reduce the number of features, the Hyper-Parameters Optimization (HPO) method to estimate the optimized hyperparameters to our RFC-based model, and the Synthetic Minority Oversampling Technique (SMOTE) to overcome the problem of imbalance.

III. ABOUT PROPOSED SYSTEM AND FLOW OF SYSTEM

1. Proposed Methodology:

Motivated by the contributions mentioned above, significant efforts are required to address the limitations of the proposed models, as outlined in Table 1. Each model has its unique approach to developing a fraud detection system. In light of this, the Fraud Detection Framework (FDD) adopts one of the most robust classification techniques, support vector data description. The approach is structured into three stages, where the outcomes of the initial and intermediate stages serve as inputs for subsequent ones.

The architecture of the proposed framework is depicted in Figure 4. However, prior to engaging the ML model within the FDD, the dataset undergoes normalization and scaling to ensure that all input values fall within a predefined interval. Additionally, a dataset resampling technique known as under-sampling is employed using the 'imblearn' method. This technique retains all data points in the minority class while reducing the size of the majority class. It is one of several techniques data scientists utilize to extract more precise insights from originally imbalanced datasets.

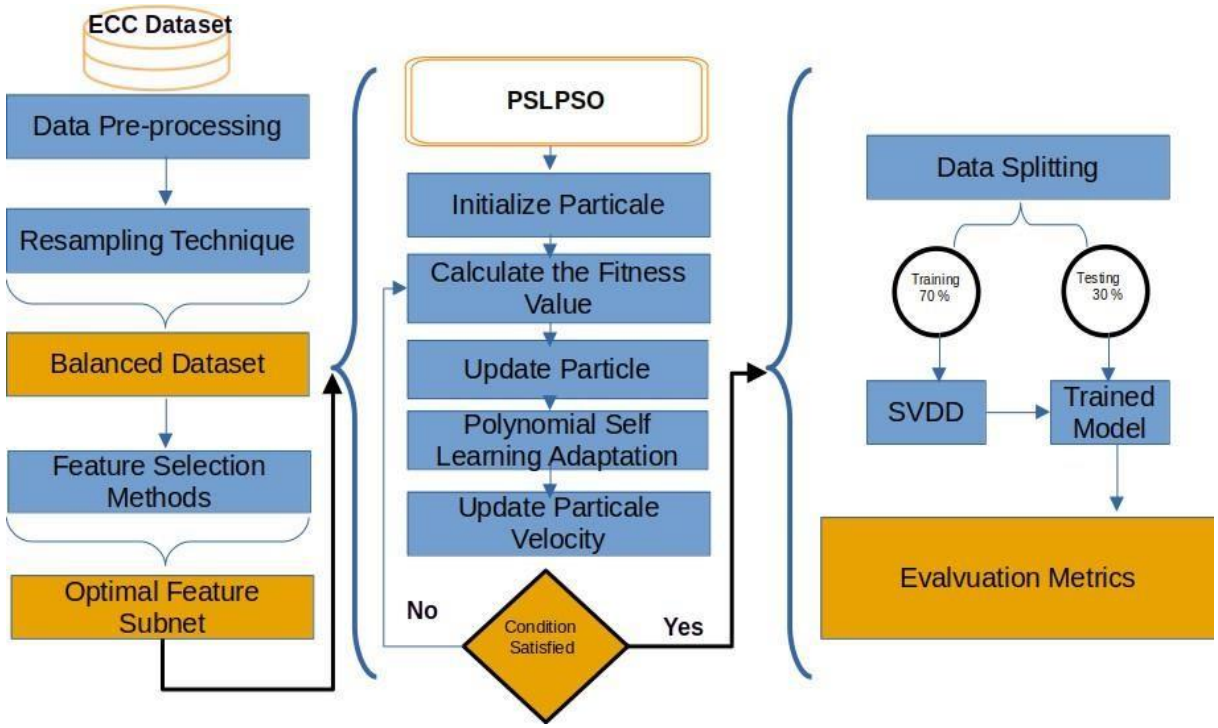
Furthermore, various feature selection techniques are implemented to enhance performance and reduce computational complexity and time for the chosen classifier model. The primary objective of feature selection is to identify the most relevant subset of features. Subsequently, a hyper-parameter optimization process is conducted using Polynomial Self Learning Particle Swarm Optimization (PSLPSO).

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 4, April 2024

DOI: 10.17148/IJARCCE.2024.134151



2. Flow of the system

Sure, here's a step-by-step process for evaluating the performance of the proposed framework, starting from the beginning:

1. Data Pre-processing:
- Load the credit card fraud dataset in CSV format.
- Perform pre-processing on the data, including checking for null values and rescaling the data.

2. Resampling Technique:
- Due to the highly unbalanced nature of the dataset, apply the 'imblearn' under-sampling technique to reduce the majority class data points to match the minority class.
- This helps prevent bias towards the majority class and improves the model's ability to predict both classes accurately.

3. Feature Selection Techniques:
- Implement feature selection methods to improve the model's accuracy and reduce computational complexity.
- Utilize three feature selection techniques: filter, wrapper, and embedded methods.
- Select the most relevant subset of features from the original dataset.

4. Polynomial Self Learning PSO (PSLPSO):
- Apply the Support Vector Data Description (SVDD) algorithm for fraud detection.
- Use PSLPSO for hyper-parameter optimization to find the optimal values for parameters such as C and Sigma.
- Initialize the swarm population of particles and set hyper-parameter values for PSLPSO.

5. Evaluation Metrics:
- Evaluate the performance of the framework using various metrics, including:
- Confusion Matrix (CM) to calculate TP, TN, FP, and FN.
- Receiver Operating Characteristic (ROC) Curve to visualize the trade-off between TPR and FPR.
- Area under the ROC Curve (AUC) to quantify the model's discriminative ability.
- Accuracy, Recall, Precision, and F1-Score to assess the model's performance comprehensively.
- Calculate these metrics based on the predictions made by the model and compare them to evaluate the effectiveness of the fraud detection framework.

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 4, April 2024
DOI: 10.17148/IJARCCE.2024.134151

By following these steps, you can systematically evaluate the performance of the proposed framework, from data pre-processing to the assessment of various evaluation metrics, providing a comprehensive analysis of its effectiveness in detecting credit card fraud.

IV. CASE STUDY OF OUR PROJECT

Here we have taken a case study based on feature selection techniques before and after.

In this case study we have used some of algorithms like Random Forest, Decision trees, Support Vector Machine, Support Vector Data Description and perform the evaluation metrics to know their performance.

Model Performances without Feature Selection Methods:

- The experiment trained the proposed approach using the full dataset's attributes.
- SVDD outperformed other classifiers with an accuracy of 90%, while SVM achieved the lowest accuracy of 83%.

Model Performances with Feature Selection Methods:

- This step aimed to demonstrate the benefits of feature selection (FS) methods in enhancing model performance.
- Table 5 presents the most important input features selected by different FS methods.
- Among all classifiers, SVDD achieved the highest classification accuracy of 93%, precision of 90%, sensitivity of 97%, and f-measure of 93%.
- LR, DT, SVM, and KNN classifiers showed no improvement with FS methods, maintaining the same accuracy as before.
- SVDD and RF models constructed from selected feature subsets using various FS approaches outperformed the original dataset feature subsets.
- The feature subset selected by the embedded feature selection technique achieved the highest classification accuracy of 93% with SVDD.

In summary, the results demonstrate the significant improvement in model performance, particularly for SVDD and RF classifiers, when utilizing feature selection techniques. These methods effectively enhance accuracy, precision, sensitivity, and f-measure, indicating their importance in optimizing fraud detection models.

V. CONCLUSION & FUTURE WORK

Conclusion:

In conclusion, this study presents a fraud detection framework that leverages the Support Vector Data Description (SVDD) model with optimized hyperparameters and incorporates resampling techniques and feature selection methods to enhance accuracy. Through experimentation, it was found that without feature selection, the model achieved 90% accuracy, which increased to 93% after implementing feature selection techniques. This underscores the importance of focusing on relevant features rather than the sheer number of attributes for improved predictive performance. The study highlights the significance of feature selection in enhancing machine learning model efficiency and emphasizes the need for exploring different combinations of resampling and feature selection strategies to optimize model performance. Future research may involve integrating fuzzy logic into hyper-parameter optimization to further enhance the framework's adaptability and efficiency in fraud detection.

Future Work:

Ahead of that, I would say that the future of credit card fraud detection using machine learning (ML) remains very active and full of opportunities for future discoveries and innovation. The application of Artificial Intelligence takes another evolutionary step as new approaches of applying advanced ML techniques, e.g., deep learning, and reinforcement learning, can be used to increase the detection abilities of fraud detection systems.

Deep learning models, which can beat human intelligence at feature extraction and hierarchy formulation from raw market data, show promise at detecting intricate behavioural patterns and interactions in credit card transactions. Furthermore, reinforcement learning algorithms may lead to the development of intelligence systems which can learn from feedback and keep on improving the detection of fraud activities by interacting in the environment.

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 4, April 2024

DOI: 10.17148/IJARCCE.2024.134151

Through these technologies application of ML technics to credit card fraud detection in the future upcoming frameworks can reach to run more accurate and scalable mechanisms, as well as overcome fraudsters' ever-evolving techniques. Furthermore, the ultimate response of credit card fraud detection in future is based on integrating active monitoring and having the capability to learn.

Along with boosting credit card transactions numbers, credibility and velocity of payments, a rising number of frauds appear, therefore fraud detection systems that operate in real-time and respond to the emergence of threats rapidly are in demand nowadays. The future models can encompass new techniques of streamed data processing and seaming computing architectures to spot any fraudulent activity quickly.

REFERENCES

[1]. Statista,"Card fraud in U.S. versus rest of the world 2014- 2021". Statista, Accessed on: Aug 28, 2023. [Online]. Available: https://www.statista.com/statistics/1264329/value-fraudulent-cardtransactions-worldwide.

[2]. Sayank Paul ,"Beginner's Guide to Feature Selection in Python.", Data camp. Accessed at 2021. [Online]. Available : https://www.datacamp.com/community/tutorials/featureselection-python, Accessed at 2021/

[3]. Khedmati, M., Erfani, M. and GhasemiGol, M., 2020. Applying support vector data description for fraud detection. arXiv preprint arXiv:2006.00618.

[4]. Mqadi, N., Naicker, N., Adeliyi, T.(2021). A SMOTE based oversampling data-point approach to solving the credit card data imbalance problem in financial fraud detection. International Journal of Computing and Digital Systems, 10(1), 277-286.