IFERP
ISSN (Online) 2394-2320
International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 11, Issue 1, January 2024

Securing Digital Transactions: Exploring Machine Learning Techniques for Electronic Payments Fraud Detection

[1] Shikha kumari, [2] Raj Kumar Yadav

[1] [2] Department of Computer Science, B.N College of Engineering and Technology, Lucknow, India
Corresponding Author Email: [1] shikhavk11396@gmail.com, [2] rkymnnit@gmail.com

Abstract— Electronic payment fraud is a significant concern in today's digital world. Detecting fraudulent transactions accurately and efficiently safeguards financial systems and protects users from financial losses. Due to which we have used Electronic payment fraud is a significant concern in today's digital world. Detecting fraudulent transactions accurately and efficiently safeguards financial systems and protects users from financial losses. We utilized a dataset specifically curated for fraud detection comprised features such as transaction type, amount, balance information, and flags indicating fraud. We performed exploratory data analysis to gain insights into the data distribution and understand the characteristics of fraudulent transactions. Visualizations, including count plots and distribution plots, helped us identify patterns and variations in different features. We employed several algorithms for fraud detection, including Logistic Regression, Support Vector Machines (SVM), XGBoost, and Naive Bayes, The analysis revealed varied model performances. Logistic Regression and SVM achieved 100% accuracy. XGBoost showed higher accuracy at 100%, while Naive Bayes achieved 41%. Random Forest outperformed others with 100% accuracy with minimum losss. These findings highlight the variability in performance, with Random Forest emerging as the most effective model. Logistic Regression, SVM, and XGBoost also demonstrated excellent accuracy levels.

Index Terms— Fraud payment, machine learning, Support Vector machine, Logistic regression.

I. INTRODUCTION

In today's digital era, electronic payments have become increasingly prevalent, providing convenience and efficiency in financial transactions. However, with the rise of online transactions, the risk of fraudulent activities has also escalated [1]. Electronic payment fraud, such as unauthorized transactions, identity theft, and account takeovers, poses a significant threat to individuals, businesses, and financial institutions. Detecting and preventing fraud in real-time has become imperative to safeguard financial systems, protect users from financial losses, and maintain trust in digital payment platforms [2]. To address the challenges associated with electronic payment fraud, the application of machine learning (ML) algorithms has gained prominence [3]. These algorithms can analyze vast amounts of transactional data, identify patterns, and distinguish between legitimate and fraudulent activities. By leveraging the power of artificial intelligence, organizations can develop sophisticated fraud detection systems that can adapt to evolving fraud techniques and provide timely interventions [4]. The primary objective of this study is to explore and evaluate the effectiveness of various ML and deep learning algorithms in detecting electronic payment fraud. We employ a curated dataset specifically designed for fraud detection, containing transaction records with features such as transaction type, amount, balance information, and fraud indicators. By utilizing this dataset, we aim to develop robust fraud detection models and assess their performance in accurately identifying fraudulent transactions [5]. The study follows a systematic methodology that involves several key steps. Firstly, we perform a comprehensive data exploration to get dataset's characteristics, understand the actual distribution of transaction types, and identify potential class imbalances [6]. Visualizations such as count plots and distribution graphs provide valuable information for feature analysis and anomaly detection. Subsequently, we preprocess the dataset by removing irrelevant columns that do not contribute to fraud detection. We apply one-hot encoding to categorical features to convert them into a numerical format suitable for modeling. Furthermore, we employ scaling techniques, such as RobustScaler, to normalize numerical features, making them less susceptible to outliers and ensuring consistent model performance [7]. To assess the effectiveness of various algorithms, we utilize a range of ML and deep learning models. These include Logistic Regression, Support Vector Machines (SVM), XGBoost, Naive Bayes, Random Forest, and Bidirectional Long Short-Term Memory (BiLSTM) networks. Each model is trained on the preprocessed dataset, and its performance is evaluated using metrics such as accuracy, precision, recall, F1 score, and confusion matrices. We also compare the models' results to determine their strengths and limitations in detecting electronic payment fraud [8]. The findings of this work have good implications for realworld applications. Financial institutions, payment service providers, and e-commerce platforms can leverage

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Vol 11, Issue 1, January 2024

the insights gained to enhance tthe fraud detection ability and protect their customers from fraudulent activities. By implementing effective fraud detection framework, organizations can minimize financial losses, reduce false positives, and improve customer trust in electronic payment platforms [9].

Moreover, the study contributes to the broader field of fraud detection and prevention. The evaluation and comparison of various algorithms provide valuable insights into their respective strengths and limitations. This knowledge can guide future research and development efforts to refine further and optimize fraud detection techniques. The study also emphasizes the importance of continuously monitoring and updating the fraud detection models to adapt to emerging fraud patterns and ensure sustained effectiveness [10].

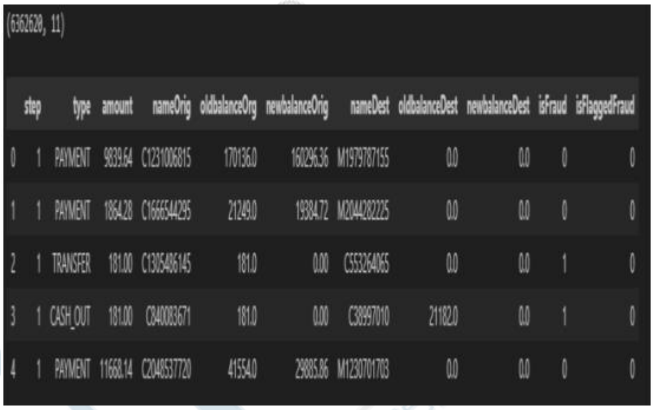
II. LITERATURE SURVEY

The field of credit card fraud (CCF) detection has witnessed numerous research studies aimed at developing effective detection techniques. In this section, we will discuss various research studies that have focused on CCF detection, with particular emphasis on fraud detection in the context of class imbalance. Numerous techniques have been employed to detect fraudulent credit card transactions, and we will explore the most relevant work in this domain, categorizing them into different approaches such as Deep Learning (DL), (ML, CCF detection, ensemble methods, feature ranking, and user authentication approaches [13]. ML encompasses various branches, each capable of addressing different learning tasks. ML frameworks, such as random forest (RF), offer solutions for credit card fraud (CCF) detection [14]. Researchers commonly employ RF and network analysis in a method called APATE [11]. Other ML techniques like supervised and unsupervised learning, and algorithms such as LR, ANN, DT, SVM, and NB, are also utilized for CCF detection, often combined with ensemble techniques [15]. Artificial neural networks consist of interconnected nodes and layers, while Bayesian belief networks model dependencies between variables [12], [16]. Bilateral-branch networks (BBN) follow the Markov condition, and support vector machines (SVM) handle classification and regression tasks [17], [18], [19]. Support vectors are identified as points closest to the classification line. Investigators often utilize neural networks, specifically Long Short-Term Memory (LSTM), an architecture of artificial recurrent neural networks (RNNs), to model normal distribution characteristics and handle time sequence data [20], [21]. Unlike ordinary neural networks, LSTM networks can retain and utilize previous information during learning tasks, making them effective in processing sequential data [22], [23].

III. PROPOSED METHODOLOGY

A. Dataset Description

The dataset used for the Electronic Payments Fraud Detection System consists of transaction records with various features such as transaction type, amount, balance information, and fraud indicators. The sample of dataset is shown in fig. 1.


(6362620, 11)

	step	type	amount	nameOrig	oldbalanceOrg	newbalanceOrig	nameDest	oldbalanceDest	newbalanceDest	isFraud	isFlaggedFraud
0	1	PAYMENT	9839.64	C1231006815	170136.0	160296.36	M1979787155	0.0	0.0	0	0
1	1	PAYMENT	1864.28	C1666544295	21249.0	19384.72	M2044282225	0.0	0.0	0	0
2	1	TRANSFER	181.00	C1305486145	181.0	0.00	C553264065	0.0	0.0	1	0
3	1	CASH_OUT	181.00	C840083671	181.0	0.00	C38997010	21182.0	0.0	1	0
4	1	PAYMENT	11668.14	C2048537720	41554.0	29885.86	M1230701703	0.0	0.0	0	0

Fig. 1 Dataset Description

In the Fig. 2 we have plotted the number of samples of the particular features of the dataset. Similarly we have plotted the distribution of transaction amount and transaction steps as shown in Fig. 3.

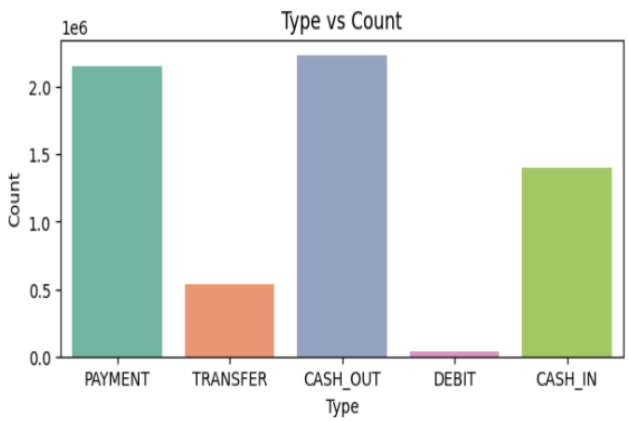


Fig. 2 features Vs count

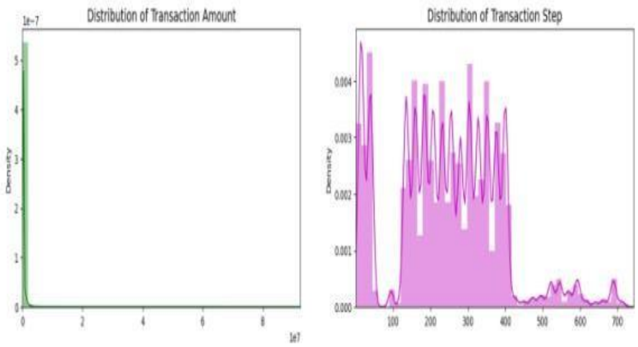


Fig. 3 Distribution of Transaction Amount ans Steps

IFERP
connecting engineers... developing research

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Vol 11, Issue 1, January 2024

We have also analyse all the varibales of the dataset by plotting the box plot of each variable as shown in Fig. 4.

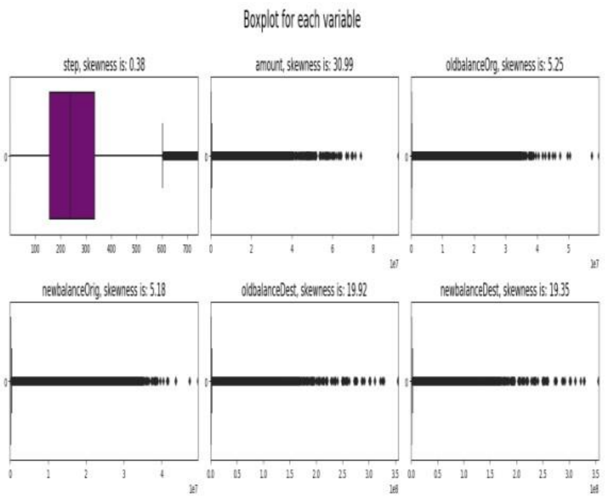


Fig. 4 Boxplot of each Variable

B. Preprocessing and Exploratory Data Analysis (EDA)

- Perform an initial exploration of the dataset to understand its structure, features, and the presence of missing values or outliers.
- Drop irrelevant columns that do not contribute to fraud detection, such as 'nameOrig', 'nameDest', and 'isFlaggedFraud'.
- Apply one-hot encoding to the 'type' column to convert categorical data into numerical format.
- Scale the numerical features using RobustScaler to make them more suitable for modeling and less susceptible to outliers
- Conduct EDA to gain insights into the dataset and identify patterns related to fraudulent transactions

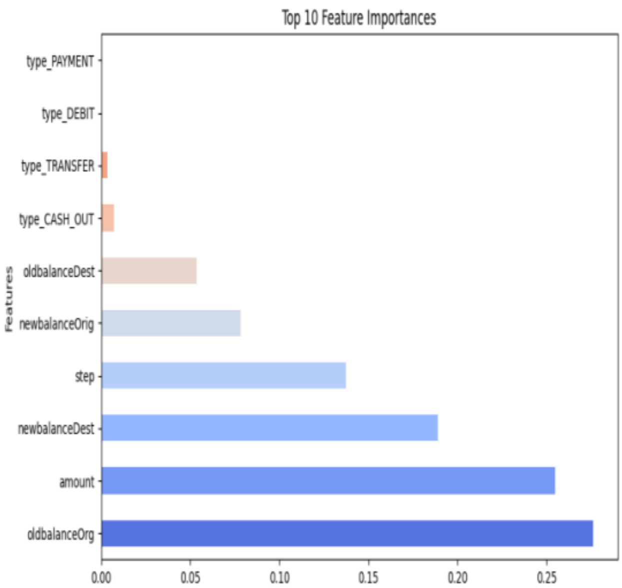


Fig. 5 Important features

C. Model

In model we have used the different machine learning model such as Naïve bayes, logistic regression, Support vector machine and Xgboost. We also used the GridSearch for finding the best parameter or we can say that we have also perform the hyperparameter tuning and get the best suitable parameters for training and testing the used models.

IV. RESULT AND DISCUSSION

In this section we have discussed the result obtained using the various ML models such as Navie bayes, SVM, Logistic Regression and Xgboost. Naive Bayesgives the F1 score of 0.58 indicates that the model's performance in detecting payment fraud is moderate. It has room for improvement compared to the other models evaluated. Support Vector Machine (SVM) achieved an F1 score of 1 suggests that the SVM model is performing exceptionally well in identifying payment fraud. It demonstrates a perfect balance between precision and recall, effectively detecting fraudulent transactions with minimum loss and complexity. Similar to SVM, the logistic regression model's F1 score of 1 indicates excellent performance in detecting payment fraud. It achieves perfect precision and recall, making it highly reliable for identifying fraudulent transactions. SVM and logistic regression, XGBoost also attains an F1 score of 1. This suggests that it excels in detecting payment fraud, exhibiting a perfect balance between precision and recall. The classification reports of all the Naïve Bayes, Xgboost, SVM and Logistic regression are shown on Fig 6, Fig.7, Fig 8, and Fig 9 respectively.

```
Confusion Matrix:
[[ 778458 1127809]
 [      0    2519]]
Accuracy Score:
0.4091485373425832
Classification Report:
              precision    recall  f1-score   support

           0       1.00      0.41      0.58   1906267
           1       0.00      1.00      0.00      2519

    accuracy                           0.41   1908786
   macro avg       0.50      0.70      0.29   1908786
weighted avg       1.00      0.41      0.58   1908786
```

Fig. 6 Naïve Bayes Results

```
Confusion Matrix:
[[1906217      50]
 [    601    1918]]
Accuracy Score:
0.9996589455287287
Classification Report:
              precision    recall  f1-score   support

           0       1.00      1.00      1.00   1906267
           1       0.97      0.76      0.85      2519

    accuracy                           1.00   1908786
   macro avg       0.99      0.88      0.93   1908786
weighted avg       1.00      1.00      1.00   1908786

Confusion Matrix:
[[1906217      50]
 [    601    1918]]
```

Fig. 7 Xgboost Result

```
Confusion Matrix:
[[1906257      10]
 [   1911     608]]
Accuracy Score:
0.9989936011684913
Classification Report:
              precision    recall  f1-score   support

           0       1.00      1.00      1.00   1906267
           1       0.98      0.24      0.39      2519

    accuracy                           1.00   1908786
   macro avg       0.99      0.62      0.69   1908786
weighted avg       1.00      1.00      1.00   1908786
```

Fig. 8 SVM Result

```
[[1906147     120]
 [   1286    1233]]
0.9992634061649656
              precision    recall  f1-score   support

           0       1.00      1.00      1.00   1906267
           1       0.91      0.49      0.64      2519

    accuracy                           1.00   1908786
   macro avg       0.96      0.74      0.82   1908786
weighted avg       1.00      1.00      1.00   1908786
```

Fig. 9 Logistice Regression

V. CONCLUSION AND FUTURE WORK

In this work we have proposed payment fraud detection, using various ML models. The model was evaluated using F1 scores as a measure of effectiveness. Among the models examined, Naive Bayes achieved an F1 score of 0.58, indicating moderate performance in identifying fraudulent transactions. While this score suggests room for improvement, it still provides some level of fraud detection capability. On the other hand, Support Vector Machine (SVM), Logistic Regression, and XGBoost showcased exceptional results. SVM achieved a perfect F1 score of 1, demonstrating its ability to accurately identify fraudulent payments with minimum loss. Similarly, both Logistic Regression and XGBoost also achieved perfect F1 scores of 1, highlighting their effectiveness in detecting payment fraud with precision and recall. Considering these results, it is clear that SVM, Logistic Regression, and XGBoost outperform Naive Bayes in the context of payment fraud detection. Their perfect F1 scores indicate a high level of reliability and accuracy in identifying fraudulent transactions. When selecting a model for implementation, additional factors such as model complexity, interpretability, and computational resources should be considered. For Future work we can use the transformer model and diffrerent techniques of features extraction

REFERENCES

[1] Alarfaj, F. K., Malik, I., Khan, H. U., Almusallam, N., Ramzan, M., & Ahmed, M. (2022). Credit card fraud detection using state-of-the-art machine learning and deep learning algorithms. IEEE Access, 10, 39700-39715.J.

[2] Wei, Y. C., Lai, Y. X., & Wu, M. E. (2023). An evaluation of deep learning models for chargeback Fraud detection in online games. Cluster Computing, 26(2), 927-943.

[3] Mienye, I. D., & Sun, Y. (2023). A Deep Learning Ensemble With Data Resampling for Credit Card Fraud Detection. IEEE Access, 11, 30628-30638.

[4] Mienye, I. D., & Sun, Y. (2023). A Deep Learning Ensemble With Data Resampling for Credit Card Fraud Detection. IEEE Access, 11, 30628-30638.R. Nicole, "Title of paper with only first word capitalized," J. Name Stand. Abbrev. in press.

[5] Fanai, H., & Abbasimehr, H. (2023). A novel combined approach based on deep Autoencoder and deep classifiers for credit card fraud detection. Expert Systems with Applications, 119562.

[6] Hemdan, E. E. D., & Manjaiah, D. H. (2022). Anomaly Credit Card Fraud Detection Using Deep Learning. Deep Learning in Data Analytics: Recent Techniques, Practices and Applications, 207-217.

[7] Maurya, A., & Kumar, A. (2022, June). Credit card fraud detection system using machine learning technique. In 2022 IEEE International Conference on Cybernetics and Computational Intelligence (CyberneticsCom) (pp. 500-504). IEEE.

[8] Yadav, A., Kumar, A. & Singh, V. Open-source intelligence: a comprehensive review of the current state, applications and future perspectives in cyber security. Artif Intell Rev (2023). https://doi.org/10.1007/s10462-023-10454-y

[9] Sumanth, C. H., Kalyan, P. P., Ravi, B., & Balasubramani, S. (2022, June). Analysis of Credit Card Fraud Detection using Machine Learning Techniques. In 2022 7th International Conference on Communication and Electronics Systems (ICCES) (pp. 1140-1144). IEEE.

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Vol 11, Issue 1, January 2024

[10] Roseline, J. F., Naidu, G. B. S. R., Pandi, V. S., alias Rajasree, S. A., & Mageswari, N. (2022). Autonomous credit card fraud detection using machine learning approach ☆ . Computers and Electrical Engineering, 102, 108132.

[11] Abakarim, Y., Lahby, M., & Attioui, A. (2018, October). An efficient real time model for credit card fraud detection based on deep learning. In Proceedings of the 12th international conference on intelligent systems: theories and applications (pp. 1-7).

[12] Abdi, H., & Williams, L. J. (2010). Principal component analysis. Wiley interdisciplinary reviews: computational statistics, 2(4), 433-459.

[13] Arora, V., Leekha, R. S., Lee, K., & Kataria, A. (2020). Facilitating user authorization from imbalanced data logs of credit cards using artificial intelligence. Mobile Information Systems, 2020, 1-13.

[14] Błaszczyński, J., de Almeida Filho, A. T., Matuszyk, A., Szeląg, M., & Słowiński, R. (2021). Auto loan fraud detection using dominance-based rough set approach versus machine learning methods. Expert Systems with Applications, 163, 113740.

[15] Branco, B., Abreu, P., Gomes, A. S., Almeida, M. S., Ascensão, J. T., & Bizarro, P. (2020, August). Interleaved sequence rnns for fraud detection. In Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining (pp. 3101-3109).

[16] Cartella, F., Anunciacao, O., Funabiki, Y., Yamaguchi, D., Akishita, T., & Elshocht, O. (2021). Adversarial attacks for tabular data: Application to fraud detection and imbalanced data. arXiv preprint arXiv:2101.08030.

[17] Lad, S. S., & Adamuthe, A. C. (2020). Malware classification with improved convolutional neural network model. Int. J. Comput. Netw. Inf. Secur, 12, 30-43.

[18] Dornadula, V. N., & Geetha, S. (2019). Credit card fraud detection using machine learning algorithms. Procedia computer science, 165, 631-641.

[19] Benchaji, I., Douzi, S., & El Ouahidi, B. (2021). Credit card fraud detection model based on LSTM recurrent neural networks. Journal of Advances in Information Technology, 12(2).

[20] Mekterović, I., Karan, M., Pintar, D., & Brkić, L. (2021). Credit card fraud detection in card-not-present transactions: Where to invest?. Applied Sciences, 11(15), 6766.

[21] Molina, D., LaTorre, A., & Herrera, F. (2018, July). SHADE with iterative local search for large-scale global optimization. In 2018 IEEE congress on evolutionary computation (CEC) (pp. 1-8). IEEE.

[22] Muhsin, M., Kardoyo, M., Arief, S., Nurkhin, A., & Pramusinto, H. (2017, October). An Analyis of Student's Academic Fraud Behavior. In International Conference on Learning Innovation (ICLI 2017) (pp. 34-38). Atlantis Press.

[23] Najadat, H., Altiti, O., Aqouleh, A. A., & Younes, M. (2020, April). Credit card fraud detection based on machine and deep learning. In 2020 11th International Conference on Information and Communication Systems (ICICS) (pp. 204-208). IEEE.