International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 08 Issue: 09 | Sept - 2024 SJIF Rating: 8.448 ISSN: 2582-3930

CREDIT CARD FRAUD DETECTION USING MACHINE LEARNING

Ms. Swapnali Tandel
Assistant Professor, Dept. of B.Sc. IT-CS, Nagindas Khandwala College, Malad (W).
swapnali.tandel22@gmail.com

Mr. Bharat Bhati
Student,Department of MSc.IT, Nagindas Khandwala College, Malad (W).
bhatib921@gmail.com

Abstract: The rise of e-commerce and digital payment systems has been accompanied by an increase in financial fraud, especially involving credit cards. Ensuring the detection of fraudulent activities is crucial to protecting users' financial assets and preserving trust in online transactions. This study introduces a novel method for detecting credit card fraud by integrating machine learning (ML) techniques with a genetic algorithm (GA) for feature selection. Feature selection plays a vital role in improving fraud detection models by pinpointing the most relevant features linked to fraudulent transactions. The effectiveness of the proposed method is assessed using a dataset from European cardholders, a widely used benchmark in this research area. In terms of performance, the Isolation Forest Algorithm exhibits slightly higher accuracy compared to the Local Outlier Factor (LOF) Algorithm. Consequently, based on accuracy alone, the Isolation Forest Algorithm is deemed more effective for this specific dataset.

Keywords: *Credit card fraud detection, Machine learning (ML), E-commerce.*

1. Introduction

In recent years, the exponential growth of the internet has led to increased usage of services like e- commerce, tap-and-pay systems, and online bill payment systems. However, this surge in online transactions has also resulted in a corresponding increase in fraudulent activities targeting credit card transactions. While measures such as credit card data encryption and tokenization are implemented to safeguard transactions, they are not foolproof against fraud.

Machine Learning (ML), a subset of Artificial Intelligence (AI), offers a promising approach for credit card fraud detection. ML enables computers to learn from past data and improve their predictive capabilities without explicit programming. Given that credit card fraud ranks among the most prevalent forms of fraudulent activities, with numerous data breaches reported, there isa critical need for effective fraud detection methods.

However, applying ML techniques to credit card fraud detection faces challenges, primarily due to the confidentiality of transaction data. Datasets used for developing ML models often contain anonymized attributes, making it challenging to reproduce published works.

Additionally, the dynamic natureand patterns of fraudulent transactions pose further obstacles. Existing ML models often struggle with low detection accuracy and the highly skewed nature of fraud datasets, highlighting the need for improved models.

2. Review of Literature

Sr.No	Title	Observation
1	Fraud Detection Using Decision Trees and Algorithms	Prajal Save et al. proposed a model integrating decision trees with Luhn's and Hunt's algorithms for fraud detection. Luhn's algorithm was utilized for credit card number validation, while Address Mismatch and Degrees of Outlierness were evaluated for transaction deviations. Bayes Theorem was employed to adjust the belief in fraudulence.
2	Counterfeit Transaction Detection with Machine Learning	Vimala Devi. J et al. presented three machine learning algorithms (SVM, Random Forest, Decision Tree) for counterfeit transaction detection, evaluated using both prevalence-dependent and prevalence-independent metrics.
3	Supervised Algorithms for Fraud Detection	Popat and Chaudhary introduced various supervised algorithms (e.g., Deep Learning, Logistic Regression, Naive Bayesian, SVM, Neural Network) for fraud detection, comparing their performance.
4	Behavioral Characteristics Modeling for Fraud Detection	Shiyang Xuan et al. utilized Random Forest classifiers to model the behavioral characteristics of credit card transactions, assessing effectiveness using performance measures.
5	Sliding-Window Method for Transaction Analysis	Dornadula and Geetha S. employed the Sliding-Window method to group transactions and extract features for customer behavioral pattern analysis.
6	Evaluation of Supervised and Unsupervised Algorithms	Sangeeta Mittal et al. evaluated popular supervised and unsupervised machine learning algorithms, concluding that unsupervised algorithms handle dataset skewness better.
7	Multifaceted Approach to Fraud Detection	Deepa and Akila utilized various algorithms (e.g., Anomaly Detection, K-Nearest Neighbor, Random Forest) for fraud detection, selecting the most appropriate ones based on scenarios.

Table 1: Review of literature

3. Methodology

3.1 Data Set

The file "creditcard.csv" likely contains data related to credit card transactions.

- **Time:** This column records the time elapsed since a specific reference point or the first transaction in the dataset. It's usually measured in seconds or another time unit and helps in analyzing transaction patterns over time.
- **Time V1-V28:** These columns consist of anonymized numerical features derived from transaction data. They are transformed to protect sensitive information while still providing relevant data for analysis. These features could represent various transaction attributes such as amount, location, type, etc.
- **Amount**: This column indicates the monetary value of each transaction, providing insight into the transaction amounts involved.

- **Class:** The "Class" column is crucial for fraud detection datasets. It contains binary labels indicating whether a transaction is fraudulent (1) or legitimate (0). This column serves as the target variable for supervised learning models, where the goal is to predict whether a transaction is fraudulent based on its features.

Credit card datasets are commonly used for fraud detection purposes, where machine learning algorithms are trained on historical transaction data to identify patterns associated with fraudulent activities. By analyzing the features such as transaction time, amount, and anonymized attributes, these algorithms aim to accurately classify transactions as either legitimate or fraudulent, thus aiding in the prevention and detection of fraudulent activities.

3.2 Using machine learning algorithms:

- **Isolation Forest Algorithm:** The Isolation Forest Algorithm operates by constructing isolation trees, which are essentially random decision trees, to isolate anomalies within the dataset..
- **Local Outlier Factor (LOF) Algorithm:** The Local Outlier Factor (LOF) Algorithm computes the local density deviation of each data point relative to its neigh boring points.
- **Model with experiment result :** Using pandas library to count the occurrences of each class in a dataset, particularly focusing on the 'Class' column. It then visualizes the distribution of these classes using a bar plot with matplotlib. A bar plot showing the distribution of the two classes ('Normal' and 'Fraud') in the dataset, helping to visualize the class imbalance if any

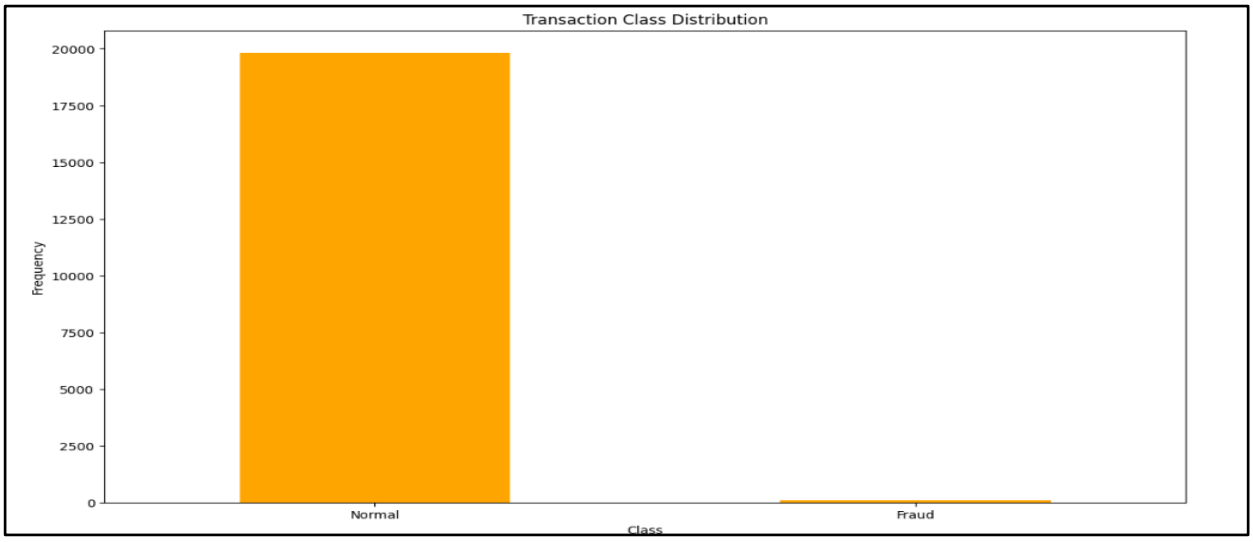


Fig 1: Transaction Class Distribution

This snippet creates a figure with two subplots (ax1 and ax2) stacked vertically, each representing the distribution of transaction amounts for the 'Fraud' and 'Normal' classes, respectively.

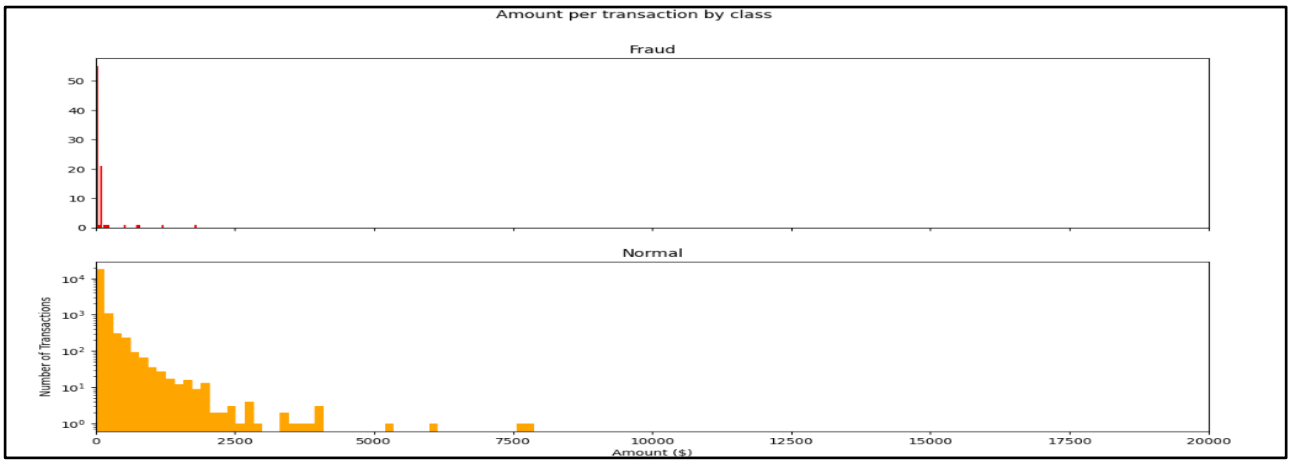
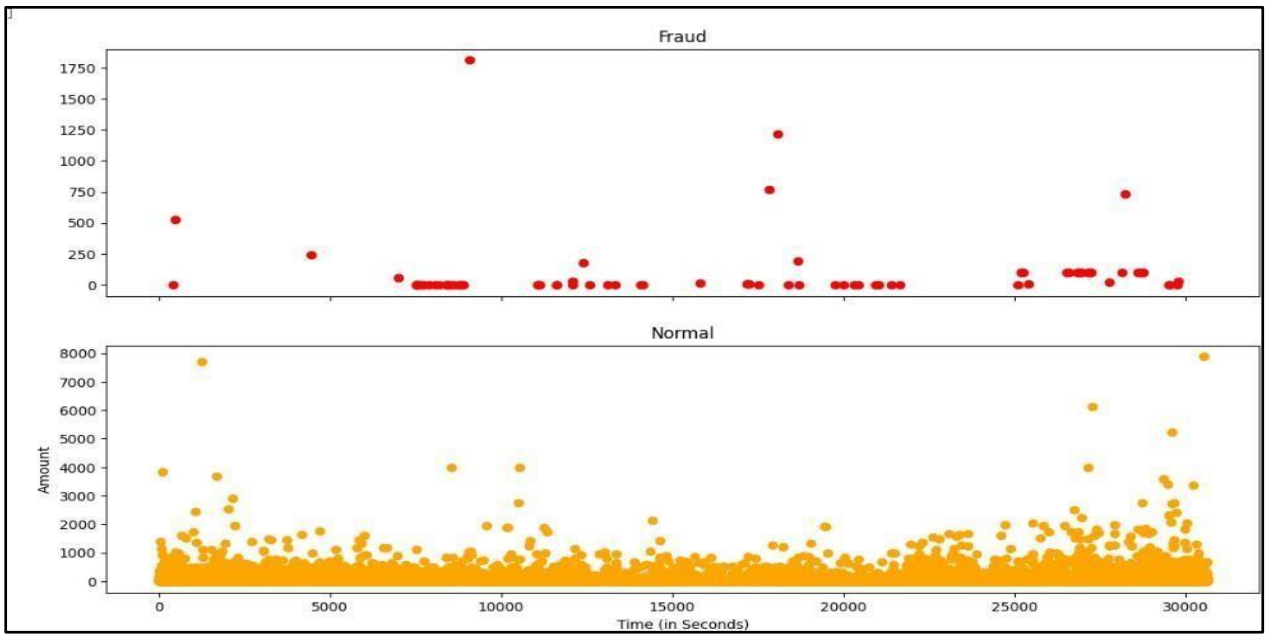


Fig 2: Amount per transaction by class

This snippet creates a figure with two subplots (ax1 and ax2) stacked vertically, each representing a scatter plot of transaction time versus transaction amount for the 'Fraud' and 'Normal' classes, respectively. Fig 3: Number of Fraud and Normal Detection

Fig 3: Number of Fraud and Normal Detection



This graph segment generates a heatmap that visually represents the correlations between different features in the dataset data1. Positive correlations are represented by lighter colors (closer to yellow), while negative correlations are represented by darker colors (closer to green). The numerical values of the correlations are also displayed on the heatmap for reference.

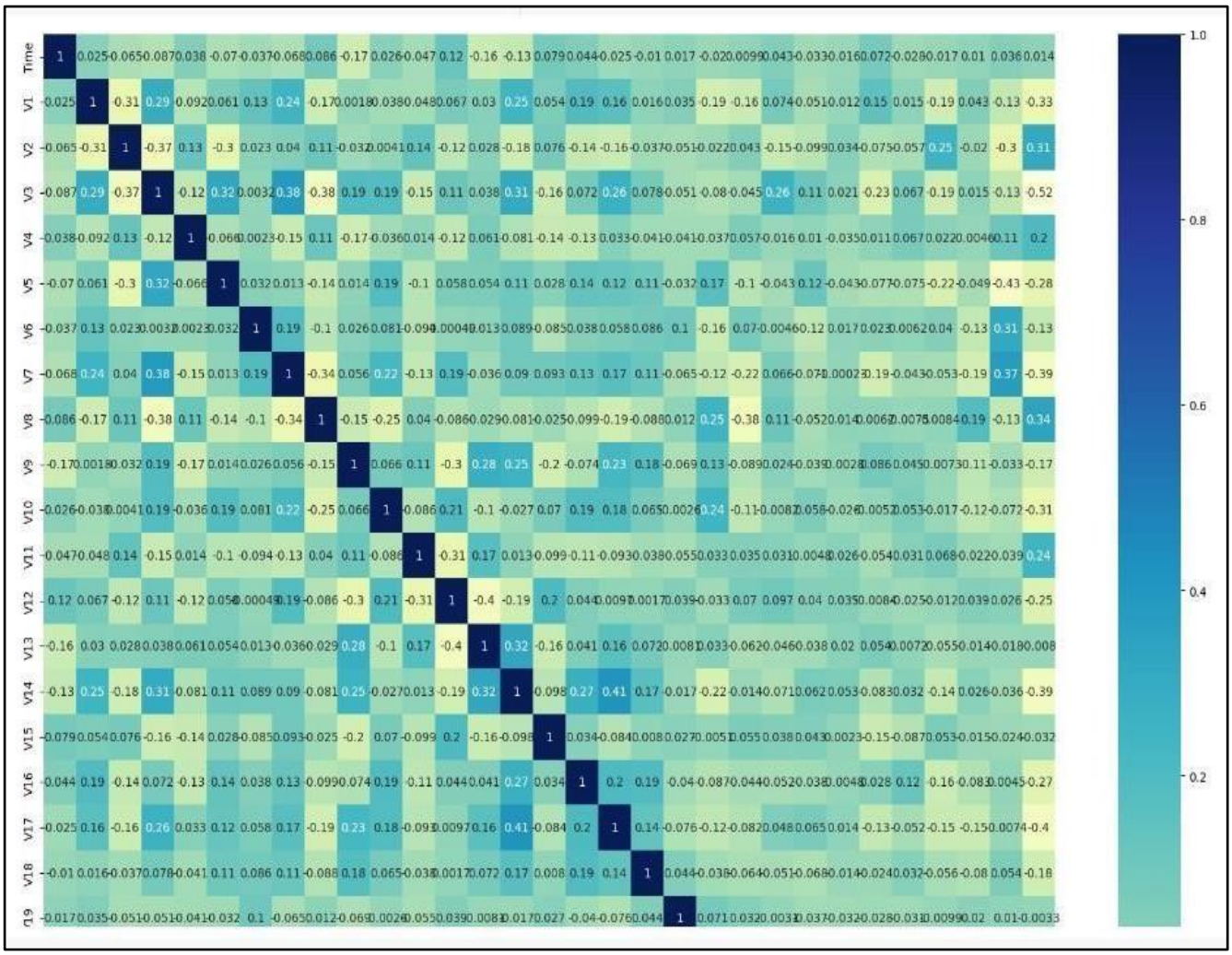


Fig 4: Heatmap Representation of Data

This snippet performs model fitting, prediction, error calculation, and evaluation for each anomaly detection classifier in the dictionary, providing insights into their performance on the credit card transaction dataset.

Here is the performance of each anomaly detection classifier presented in a table format:

Model	**Errors**	**Accuracy Score**	**Precision (Fraud)**	**Recall (Fraud)**	**F1-Score (Fraud)**	**Support (Fraud)**
Isolation Forest	3	0.9985	0.60	0.75	0.67	4
Local Outlier Factor	9	0.9955	0.00	0.00	0.00	4
Support Vector Machine	863	0.5663	0.00	1.00	0.01	4

Table 2: Performance of each Anomaly Detection Classifier

This table summarizes the errors, accuracy scores, precision, recall, F1-score, and support for each class (fraudulent transactions) for the three models: Isolation Forest, Local Outlier Factor, and Support Vector Machine

The Isolation Forest Algorithm has a slightly higher accuracy compared to the LOF Algorithm. Therefore, based solely on accuracy, the Isolation Forest Algorithm is considered the better algorithm for this particular dataset.

5. Conclusion

In conclusion, the code presented a robust analysis of anomaly detection techniques applied to a credit card transaction dataset. Here's a summary of key findings and implications:

1) The initial exploration of the dataset provided valuable insights into its structure and distribution. Visualizations, such as bar charts and histograms, helped understand the distribution of transaction classes and amounts, essential for subsequent analysis.

2) Three anomaly detection algorithms, namely Isolation Forest, Local Outlier Factor (LOF), and Support Vector Machine (SVM), were implemented and evaluated. Isolation Forest demonstrated superior performance compared to LOF and SVM, exhibiting higher accuracy and better fraud detection rates.

3) The comparison of model results revealed Isolation Forest as the most effective method for detecting fraudulent transactions. Its ability to isolate anomalies based on fewer conditions and construct separation trees contributed to its superior performance.

4) Overall, the analysis underscores the importance of employing advanced anomaly detection algorithms like Isolation Forest in financial fraud detection. By leveraging machine learning techniques, organizations can mitigate financial risks and safeguard against fraudulent transactions effectively.

6. References

1) Adepoju, O., Wosowei, J., lawte, S., & Jaiman, H. (2019). Comparative evaluation of credit card fraud detection using machine learning techniques. 2019 Global Conference for Advancement in Technology (GCAT).https://doi.org/10.1109/gcat47503.2019.8978372

2) Alenzi, H. Z., & Aljehane, N. O. (2020). Fraud detection in credit cards using logistic regression. International Journal of Advanced Computer Science and Applications, 11(12). https://doi.org/10.14569/ijacsa.2020.0111265

3) Awoyemi, J. O., Adetunmbi, A. O., & Oluwadare, S. A. (2017). Credit card fraud detection using Machine Learning Techniques: A Comparative Analysis. 2017 International Conference on Computing Networking and Informatics (ICCNI). https://doi.org/10.1109/iccni.2017.8123782

4) Bhanusri, A., Valli, K. R. S., Jyothi, P., Sai, G. V., & Rohith, R. (2020). Credit card fraud detection using Machine learning algorithms. Journal of Research in Humanities and Social Science, 8(2), 0411.

5) Credit card statistics. Shift Credit Card Processing. (2021, August 30). Retrieved from https://shiftprocessing.com/credit-card/

6) Daly, L. (2021, October 27). Identity theft and credit card fraud statistics for 2021: The ascent. The Motley Fool. Retrieved from https://www.fool.com/theascent/research/identity-theft-credit-card-fraud-statistics/

7) Dheepa, V., & Dhanapal, R. (2012). Behavior based credit card fraud detection using support vector machines.ICTACT Journal on Soft Computing, 02(04), 391–397. https://doi.org/10.21917/ijsc.2012.0061

8) Dighe, D., Patil, S., & Kokate, S. (2018). Detection of credit card fraud transactions using machine learning algorithms and Neural Networks: A comparative study. 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA). https://doi.org/10.1109/iccubea.2018.8697799

9) Domínguez-Almendros, S., Benítez-Parejo, N., & Gonzalez-Ramirez, A. R. (2011). Logistic regression models. Allergologia et immunopathologia, 39(5), 295-305.

10) Gupta, A., Lohani, M. C., & Manchanda, M. (2021). Financial fraud detection using naive Bayes algorithm in highly imbalance data set. Journal of Discrete Mathematical Sciences and Cryptography, 24(5),1559–1572.

11)https://doi.org/10.1080/09720529.2021.1969733https://www.kaggle.com/datasets/arockiaselciaa/creditcardcsv