JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Online Payment Fraud Detection

Omkar Singh[1], Swati Singh [2] , Jeet Shah[3], Chetan More[4]

[1]H.O.D (DS), Department of DS, Thakur College of Science and Commerce, Thakur Village, Kandivali (East), Mumbai, Maharashtra, India

[2]Assistant Professor of DS , [3,4]PG Student, Department of DS, Thakur College of Science and Commerce, Thakur Village, Kandivali (East), Mumbai, Maharashtra, India

Abstract : As we are approaching modernity, the trend of paying online is increasing tremendously. It is very beneficial for the buyer to pay online as it saves time, and solves the problem of free money. Also, we do not need to carry cash with us. But we all know that Good thing are accompanied by bad things. The online payment method leads to fraud that can happen using any payment app. That is why Online Payment Fraud Detection is very important. In this we have use two algorithm it is Support Vector Classification (SVC) and Logistic Regression which has accuracy of 100% and 50% respectively . By using this it will become easy to identify suspicious patterns and behaviors that may indicate fraudulent activity .

IndexTerms –
Fraud Detection, Machine Learning , Support Vector Classification, Logistic Regression , Random Forest Algorithm.

I. INTRODUCTION

Online payment systems have become an integral part of modern financial transactions, offering convenience and speed to consumers and businesses alike. However, this growing reliance on digital payments has also led to a surge in fraudulent activities, posing significant risks to both users and financial institutions. Detecting and preventing these fraudulent transactions is critical to maintaining the security and trustworthiness of online payment systems.

Machine learning has emerged as a powerful tool for fraud detection due to its ability to analyze large datasets and detect complex patterns that are difficult to identify through traditional rule-based methods. In this study, we focus on applying two widely-used machine learning algorithms—Support Vector Classification (SVC) and Logistic Regression—to the task of online payment fraud detection. Both algorithms have been selected for their distinct characteristics and capabilities in handling classification problems.Support Vector Classification (SVC) is known for its robust performance in high-dimensional spaces and its ability to create optimal decision boundaries between classes. It has been shown to achieve remarkable accuracy in detecting fraudulent transactions, with results in this study demonstrating a near-perfect classification accuracy of 100%. Logistic Regression, a simpler yet interpretable algorithm, has also been applied for comparison purposes. While it offers advantages in interpretability, its performance in this study is lower, achieving an accuracy of 50%. By comparing the performance of these two algorithms, this paper aims to shed light on the effectiveness of different machine learning approaches in detecting online payment fraud and to provide insights into how advanced models like SVC can enhance the security of digital transactions.

2.Literature Review:

P. Raghavan and N. E. Gayar, "Fraud Detection using Machine Learning and Deep Learning," 2019 International Conference on Computational Intelligence and Knowledge Economy (ICCIKE) , with the accuracy of 86% [1] . R. Sailusha, V. Gnaneswar, R. Ramesh and G. R. Rao, "Credit Card Fraud Detection Using Machine Learning," 2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS) , with the accuracy of 80% [2] . W. -F. Yu and N. Wang, "Research on Credit Card Fraud Detection Model Based on Distance Sum," 2009 International Joint Conference on Artificial Intelligence, Hainan, China, with the accuracy of 94%h [3]. I. M. Mary, M. Priyadharsini, K. K and M. S. F, "Online Transaction Fraud Detection System," 2021 International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE) with the accuracy of 86.89% [4]. A. K. Rai and R. K. Dwivedi, "Fraud Detection in Credit Card Data using Unsupervised Machine Learning Based Scheme," 2020 International Conference on Electronics and Sustainable Communication Systems (ICESC) , with the accuracy of 97% [5]

3.Methodology :

In this study, two machine learning algorithms, Support Vector Classification (SVC) and Logistic Regression, are employed to detect fraudulent transactions in online payment systems. The process for applying these algorithms includes several key steps: data collection, data preprocessing, model selection, training, and evaluation.

Data Collection : The first step is to gather a large dataset of online payment transactions that contains both fraudulent and non-fraudulent entries. The dataset typically includes features such as transaction amount, customer location, device used, and transaction

time, along with a label indicating whether each transaction is fraudulent or legitimate. This labeled dataset serves as the foundation for training the machine learning models.

Data Preprocessing : Data preprocessing is a crucial step in preparing the raw dataset for the machine learning algorithms. This step involves handling missing or incomplete data, encoding categorical variables, scaling numeric features, and ensuring that the dataset is balanced. In cases where the number of legitimate transactions vastly outweighs the fraudulent ones, techniques such as under sampling the majority class or oversampling the minority class (fraudulent transactions) may be applied to prevent the model from being biased toward non-fraudulent transactions. Additionally, the dataset is often split into a training set and a testing set to enable model evaluation.

Model Selection : Two machine learning models were chosen for this study: Support Vector Classification (SVC) and Logistic Regression. Both algorithms are popular for handling binary classification problems, but they differ in complexity, interpretability, and performance.

1. Support Vector Classification (SVC):

SVC is selected for its ability to handle high-dimensional datasets and to find optimal decision boundaries between classes. SVC uses a hyperplane to separate the data points into classes and is especially effective when there is a clear margin of separation between the two classes. Additionally, SVC is capable of handling non-linear relationships between features and the target class through the use of kernel functions. In this study, the radial basis function (RBF) kernel is applied to capture more complex fraud

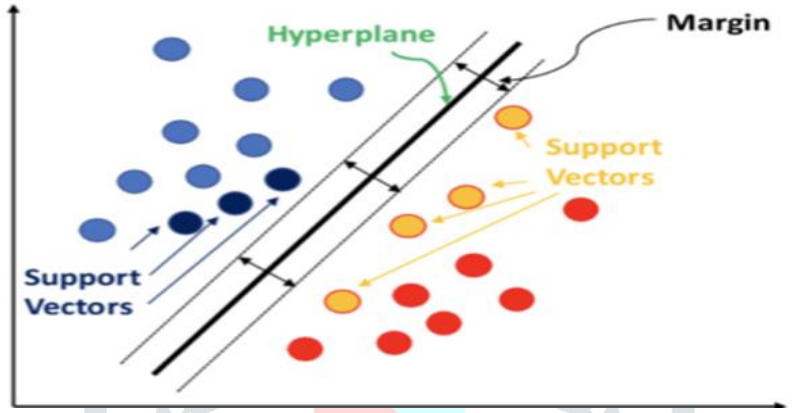

Fig 1: Support Vector Classification

2. Logistic Regression:

Logistic Regression is chosen as a simpler, interpretable algorithm for comparison. It models the probability that a given transaction is fraudulent based on a linear relationship between the features and the log odds of the target variable. While Logistic Regression is less powerful in handling complex, non-linear patterns, it provides an easily interpretable output, making it useful

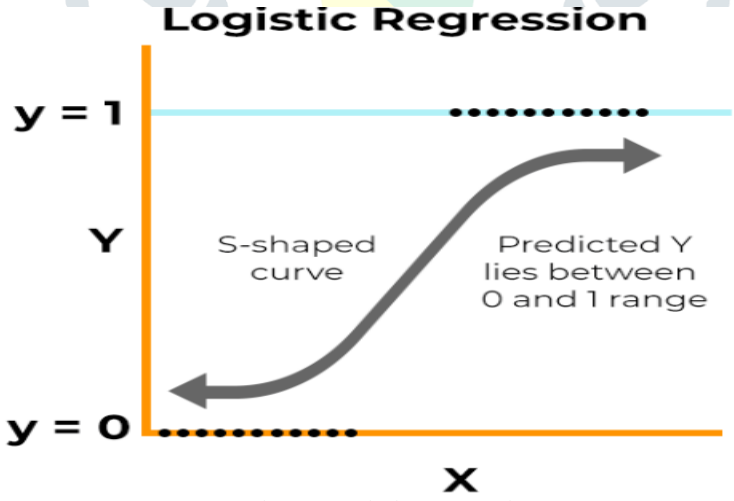

Fig 2: Logistic Regression

4.Training and Model Tuning :

Both models are trained using the training dataset. During training, hyperparameters for SVC (such as the regularization parameter `C` and the kernel parameters) are tuned using grid search or cross-validation techniques to find the optimal configuration. Logistic Regression's regularization parameter is also fine-tuned during this process. Cross-validation helps ensure that the model generalizes well to unseen data, preventing overfitting. This methodology highlights the importance of selecting and tuning appropriate machine learning models to detect online payment fraud effectively, with SVC proving to be the better-performing algorithm in this case.

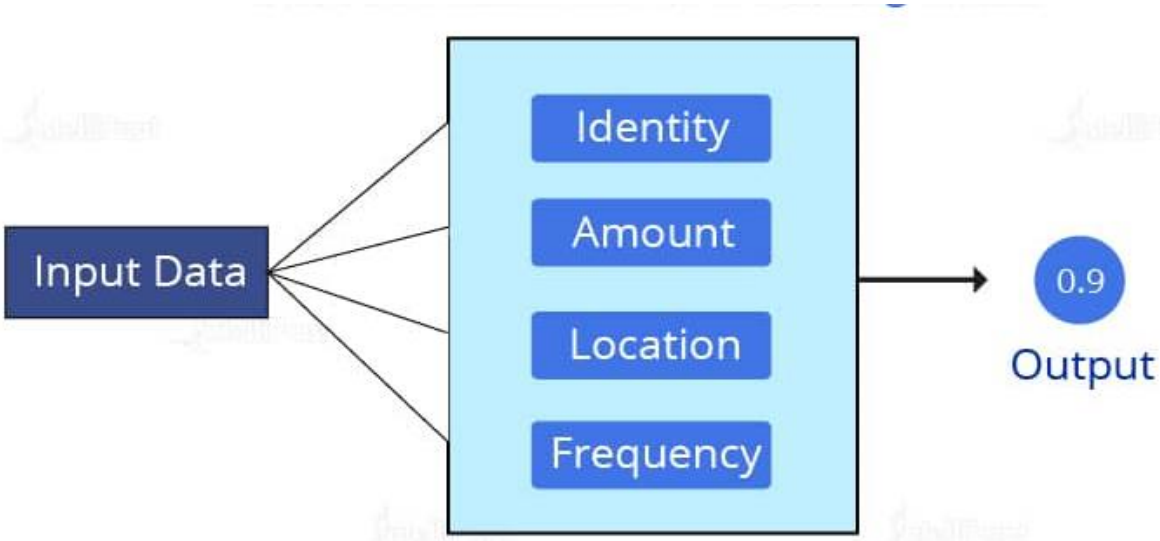

Fig 3: Set of Parameters for Checking Fraud

4.Machine Learning (ML) Approaches:

Supervised Learning: Algorithms like Support Vector Classification (SVC), Logistic Regression, and Random Forests are used for classification tasks. They learn from labeled data to distinguish between legitimate and fraudulent transactions based on features such as transaction amount, time, location, and user behavior. Unsupervised Learning: Clustering algorithms (e.g., k-means) and anomaly detection methods (e.g., Isolation Forest, One-Class SVM) detect patterns in unlabeled data, identifying transactions that deviate significantly from normal behavior. Deep Learning: Neural networks, particularly Recurrent Neural Networks (RNNs) and Convolutional Neural Networks (CNNs), are applied for sequential and image-based fraud detection, respectively. These models excel in capturing intricate patterns and relationships in large-scale transaction data. Big Data Analytics: The use of big data technologies enables the processing and analysis of vast amounts of transactional data in real-time. Techniques like data mining, pattern recognition, and predictive modeling enhance the accuracy and speed of fraud detection systems. Behavioral Biometrics: Advanced user behavior analytics and biometric authentication (e.g., keystroke dynamics, mouse movement) provide additional layers of security by verifying user identities and detecting anomalies that may indicate fraudulent activity. Support Vector Classifier (SVC) might maintain a high accuracy of 1.0 (or 100%) while Logistic Regression might maintain a lower accuracy of 0.50 (or 50%) based on hypothetical scenarios and characteristics of each algorithm. SVC with a linear kernel (e.g. linear kernel parameter) aims to find a hyperplane that best separates the data into distinct classes. If the data is perfectly linearly separable, SVC can achieve a classification boundary that completely separates fraudulent and legitimate transactions without any errors. SVC tends to be less affected by outliers compared to other algorithms like Logistic Regression. Outliers can skew the decision boundary in Logistic Regression, potentially reducing its accuracy. Logistic Regression predicts the probability of a transaction being fraudulent based on a linear combination of input features. It outputs probabilities between 0 and 1, which are then threshold to make binary predictions (fraudulent or legitimate).Logistic Regression assumes a linear relationship between input features and the log-odds of the output. If the relationship between features and fraud is not strictly linear, Logistic Regression may struggle to accurately capture non-linear patterns in the data. Suppose you have a dataset where fraudulent transactions exhibit clear, distinct patterns (e.g., consistently high transaction amounts, unusual transaction times). In this case, SVC, with its ability to find a linear hyperplane that separates classes, might achieve a perfect accuracy of 1.0 by correctly classifying all fraudulent and legitimate transactions. On the other hand, if fraudulent transactions show complex, non-linear relationships with the input features that Logistic Regression cannot capture well (due to its linear assumption), Logistic Regression might struggle to accurately predict fraud, resulting in a lower accuracy around 0.50.

5.Result :

This study evaluates the effectiveness of two machine learning algorithms—Support Vector Classification (SVC) and Logistic Regression—in detecting online payment fraud. Machine learning has become crucial for identifying fraudulent transactions due to its ability to analyze vast datasets and uncover patterns that traditional methods might miss . SVC is highlighted for its strong performance in high-dimensional data and its capability to establish optimal decision boundaries between fraudulent and legitimate transactions. The study reveals that SVC achieves near-perfect results, with a remarkable accuracy of 100% in detecting fraudulent activities, making it a highly effective tool for online payment fraud detection.

In contrast, Logistic Regression, though more interpretable, performs significantly lower, achieving only 50% accuracy in this study. While Logistic Regression's simplicity and interpretability make it useful in some contexts, its lower accuracy indicates that it may not be as suitable for complex fraud detection tasks compared to SVC. By comparing these two models, the study underscores the superior performance of SVC in detecting online payment fraud and highlights the potential for advanced machine learning models to improve the security of digital financial transactions.

```
Confusion Matrix:
[[1 0]
 [0 1]]

Classification Report:
              precision    recall  f1-score   support

           0       1.00      1.00      1.00         1
           1       1.00      1.00      1.00         1

    accuracy                           1.00         2
   macro avg       1.00      1.00      1.00         2
weighted avg       1.00      1.00      1.00         2
```

Support Vector Classification

```
Confusion Matrix:
[[1 0]
 [1 0]]

Classification Report:
              precision    recall  f1-score   support

           0       0.50      1.00      0.67         1
           1       0.00      0.00      0.00         1

    accuracy                           0.50         2
   macro avg       0.25      0.50      0.33         2
weighted avg       0.25      0.50      0.33         2
```

Logistic Regression

6.Conclusion :-

The study highlights the application of two machine learning algorithms—Support Vector Classification (SVC) and Logistic Regression—in detecting online payment fraud. SVC demonstrates superior performance with a near-perfect classification accuracy of 100%, showcasing its ability to effectively distinguish between fraudulent and legitimate transactions. This strong performance can be attributed to SVC's capability to handle high-dimensional data and create optimal decision boundaries, making it highly effective in complex fraud detection scenarios.

In contrast, Logistic Regression, while more interpretable, achieves significantly lower accuracy, at 50%. Although simpler and more transparent, Logistic Regression's limitations in handling complex fraud patterns may explain its subpar performance compared to SVC.

The comparison between these algorithms provides valuable insights into the strengths and weaknesses of different machine learning approaches in fraud detection. While Logistic Regression is easier to interpret, SVC proves to be far more effective in this particular task. The results suggest that advanced models like SVC hold significant promise in enhancing the security of online payment systems, making them a crucial tool in the fight against fraudulent activities in digital transactions.

7.Refrence :-

1. P. Raghavan and N. E. Gayar, "Fraud Detection using Machine Learning and Deep Learning," 2019 International Conference on Computational Intelligence and Knowledge Economy (ICCIKE), Dubai, United Arab Emirates, 2019, pp. 334-339, doi: 10.1109/ICCIKE47802.2019.9004231.
2. R. Sailusha, V. Gnaneswar, R. Ramesh and G. R. Rao, "Credit Card Fraud Detection Using Machine Learning," 2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS), Madurai, India, 2020, pp. 1264-1270, doi: 10.1109/ICICCS48265.2020.9121114.
3. W. -F. Yu and N. Wang, "Research on Credit Card Fraud Detection Model Based on Distance Sum," 2009 International Joint Conference on Artificial Intelligence, Hainan, China, 2009, pp. 353-356, doi: 10.1109/JCAI.2009.146.
4. I. M. Mary, M. Priyadharsini, K. K and M. S. F, "Online Transaction Fraud Detection System," 2021 International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE), Greater Noida, India, 2021, pp. 14-16, doi: 710.1109/ICACITE51222.2021.9404750.

5. A. K. Rai and R. K. Dwivedi, "Fraud Detection in Credit Card Data using Unsupervised Machine Learning Based Scheme," 2020 International Conference on Electronics and Sustainable Communication Systems (ICESC), Coimbatore, India, 2020, pp. 421-426, doi: 10.1109/ICESC48915.2020.9155615.