Rana Shantaram Mahajan[1],
Vinod Shantaram Mahajan[2]

Credit Card Fraud Detection Using Machine Learning

Abstract - Credit card fraud remains a significant challenge in the modern financial landscape, causing substantial economic losses to individuals and institutions. This study presents a machine learning-based approach to proactively identify fraudulent credit card transactions. To improve detection accuracy and minimize false positives, multiple supervised learning algorithms including Logistic Regression, Decision Tree, Random Forest, Support Vector Machine (SVM), and K-Nearest Neighbors (KNN) were implemented and evaluated. The study makes use of a publicly accessible credit card transaction dataset, validating its findings with preprocessing methods and performance metrics like accuracy, precision, recall, and F1-score. A comparative analysis highlights the strengths and weaknesses of each model in handling imbalanced and high-dimensional data. The results demonstrate that ensemble methods like Random Forest offer robust performance in identifying anomalous transaction patterns. The proposed system integrates these models into an admin-controlled platform for real-time detection, making it a scalable and effective solution for fraud mitigation in financial systems

Key Words: Credit Card Fraud, Machine Learning, Fraud Detection, Supervised Learning, Support Vector Machine (SVM), Logistic Regression, Random Forest

1.Introduction

As financial institutions have matured in the 21st century, they have made their business capabilities available to the public through online banking. In the current Machiavellian economic environment, electronic methods of payment are essential. They have made the process of buying products and services which are less complex. In order to make their lives easier by allowing them to safeguard themselves without carrying cash, visitors to banking institutions regularly take cards via them.In addition to assisting cards, credit cards provide visitors with protection against personal items that might get lost, stolen, or destroyed after the purchase. Guests are required to verify the sale via the retailer prior to utilising their debit or credit card to finish any transaction. Statistics show that 2287 million cardholders were issued by Visa and MasterCard in the 4th month of 2020. 1156 million Master cards were issued in an encyclopedic fashion, compared to 1131 million Visa cards.These figures illustrate the widespread acceptance and recognition of employing credit cards for transactions. To ultimate consumers. The high probability that transnational transactions will fall into this category makes this group vulnerable to fraudsters. Additionally, it may be effortless to manipulate individuals in a social setting. Despite the fact that credit cards have many advantages for customers, there are security and fraud issues with them. Fraud with credit cards is one of the difficulties that the banking industry and other financial groups are currently working to identify and address. It occurs when individuals who are not authorised use their credit cards to wrongfully acquire property or plutocrats. Credit card information theft can be possible by using unsecured websites and online services. They might also originate from identity theft schemes. Fraudsters may take drug users' credit and disbenefit card information without their knowledge or agreement. One of the main reasons the banking industry loses money is theft associated with credit and debit cards, according to "U.K. finance" [4]. Technology has advanced to the point of significant imminence. This leads to a substantial financial loss on a global scale. Determining credit card fraud is therefore essential to minimizing monetary losses. Machine literacy

[1]Assistant Professor,Department of E&TC Engineering, Sir Visvesvaraya Institute of Technology,Nashik
(rana.entc@gmail.com)
[2]Assistant Professor,Department of Computer Engineering, D.N.Patel College Of Engineering, Shahada
(vinodsm@rediffmail.com)

effectively distinguishes between authentic and fraudulent transactions. The impediment to the exchange of ideas regarding fraud discovery is one of the primary issues with discovery approaches. According to a review by "U.K. Finance," there were £574.2 million in credits or disbenefit fraud occurrences registered in the UK in 2020 [4].

2. Literature Survey

1. Credit Card Fraud Detection Using Machine Learning

Authors: R. Carcillo, Y. Belhadi, A. Bifet, M. Snoek, F. Petitjean

Summary:

An extensive review of machine learning methods for detecting credit card fraud is provided in this study. By classifying techniques into supervised, unsupervised, and semi-supervised learning, it draws attention to practical issues such idea drift, class imbalance, and data privacy.

Key Contributions:

Provided a taxonomy of ML algorithms for fraud detection.

Addressed real-time detection and adaptability to evolving fraud patterns.

Discussed various public datasets and evaluation metrics.

2. Credit Card Fraud Detection with Decision Trees and Random Forests

Authors: G., Boracchi, G., Caelen, O., Dal Pozzolo, Alippi, C., & Bontempi, G.

Summary:

This study compares Decision Tree-based models, especially Random Forest, in detecting fraudulent transactions. It emphasizes handling severely imbalanced data through techniques like undersampling and SMOTE (Synthetic Minority Oversampling Technique).

Key Contributions:

Demonstrated the effectiveness of Random Forest for imbalanced classification tasks.

Highlighted the role of data sampling and feature selection in improving accuracy.

Proposed a cost-sensitive evaluation metric suitable for fraud scenarios.

3. Deep Learning for Credit Card Fraud Detection

Authors: Roy, A., Sun, J., Mahoney, W.

Summary:

The application of DNN to fraud detection is examined in this article. It looks into how deep feedforward networks and autoencoders can uncover hidden patterns in transaction data.

Key Contributions:

Applied deep learning architectures to model complex non-linear relationships.

\Emphasized unsupervised learning using autoencoders for anomaly detection.

Improved detection rates without extensive feature engineering.

4. Methods for Identifying Anomalies in Credit Card Fraud
Lee V, Smith K ,Phua C, & Gayler R. are the authors.

Summary:

This research focuses on anomaly detection approaches which are effective when fraudulent examples are rare or unlabeled. It compares traditional outlier detection methods with machine learning-based anomaly detection techniques.

Key Contributions:

examined clustering-, density-, and distance-based anomaly detection techniques.
The usefulness of these techniques in unsupervised settings was discussed.

Provided comparative analysis on performance across multiple datasets.

3. Objective

Detect Fraudulent Transactions Accurately identify fraudulent credit card transactions from large datasets.

Reduce False Positive Results To prevent user annoyance, lower the amount of valid transactions that are mistakenly reported as fraudulent (false alarms).

Increase the Accuracy of Detection To ensure accurate fraud detection, raise measures like precision, recall, F1-score, and AUC-ROC.

4.Methodology

Logistic Regression

Binary categorization using a statistical model (fraudulent or not).
It uses input features to determine the likelihood that a transaction is fraudulent.
Benefits: It works well with linearly separable data and is easy to understand.

Decision Tree

a tree-like approach for transaction classification that divides data according to feature thresholds. Simple to grasp and depict. Cons: prone to overfitting, particularly when dealing with noisy data.

Random Forest

a group approach that uses voting for the final categorization and several decision trees. decreases overfitting and increases accuracy.

Strength: Captures non-linear correlations and performs well with unbalanced datasets.

Support Vector Machine (SVM)

determines the ideal border (hyperplane) between phony and authentic transactions.

Efficient in high-dimensional environments.

Limitation: For big datasets, computationally costly.

K-Nearest Neighbors (KNN)

Classifies a transaction based on the majority label of its 'k' closest neighbors.

Strength: Simple and effective for small datasets.

Weakness: Slow prediction time and sensitive to feature scaling.

Neural Networks

A deep learning approach that mimics the human brain to detect complex patterns.

Learns non-linear relationships between features for better fraud detection.

Best suited for large datasets and continuous learning systems

5. Result And Model

The **Credit Card Fraud Detection System** is designed as a single-admin interface where the administrator handles all operations related to data upload, model training, testing, and result monitoring. The system aims to provide an efficient and automated approach for identifying fraudulent credit card transactions using machine learning techniques.

Admin Module:.

- **Upload Dataset:**

The admin can upload a training dataset (CSV format) that contains historical transaction data. This dataset includes anonymized features (V1 to V28), Amount, and a Class column indicating whether the transaction is fraudulent (1) or legitimate (0).

TrainModel:

The administrator can manually choose one machine learning algorithm from the following possibilities using the system:

- Random Forest
- K-Nearest Neighbors (KNN)
- Logistic Regression
- Support Vector Classifier (SVC)
- Decision Tree

The uploaded dataset is subsequently used to train the chosen model. Following training, performance indicators like recall, precision, F1-score, accuracy, and confusion matrix are shown to evaluate the model's efficacy.

- **Test Model:**

After training, the admin can upload a **test dataset** that does **not contain the Class column**. The system automatically evaluates **all available ML models** on this dataset to detect potential fraudulent transactions.

- **View Results:**

The system displays detailed evaluation results, including:

- Predicted labels for test data
- A comparison chart of model performance (accuracy for each model)
- Confusion matrix and classification report (for training data)

Model and Evaluation:

The detection pipeline follows these steps:

- **Preprocessing:**

Includes normalization of the Amount feature, handling of imbalanced data (e.g., using SMOTE or undersampling), and separation of training and testing sets.

- **Label Encoding and Splitting:**

For objective assessment, the data is divided into train and test sets, and the Class label is transformed into numeric form (0: Legitimate, 1: Fraud).

- **Model Training:**

Out of the five supervised learning models, the administrator chooses one to train. These models work well for classification tasks requiring datasets that are skewed.

- **Model Testing:**

During testing, the system applies **all models** to the new dataset (without Class) to identify likely frauds. This provides a comparative insight into model effectiveness.

- **Performance Metrics:**

The system computes:

- Accuracy: Out of all predictions, correct ones

F1-Score: Harmonic mean of accuracy and recall

- Confusion Matrix: Comprehensive categorization breakdown
- accuracy: Accuracy of positive (fraud) predictions
- Recall: Capacity to capture real fraud instances

This model-driven system allows for dynamic evaluation and comparison of multiple machine learning algorithms, supporting informed decisions in fraud detection with minimal manual intervention.

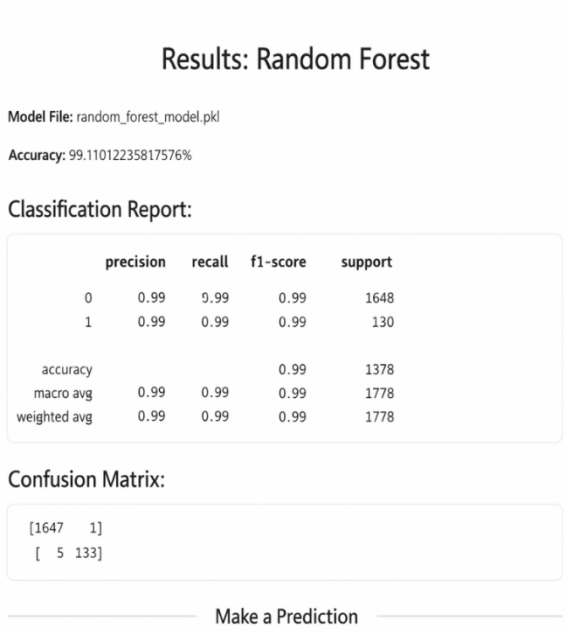

Fig.5.1 Result –Random Forest

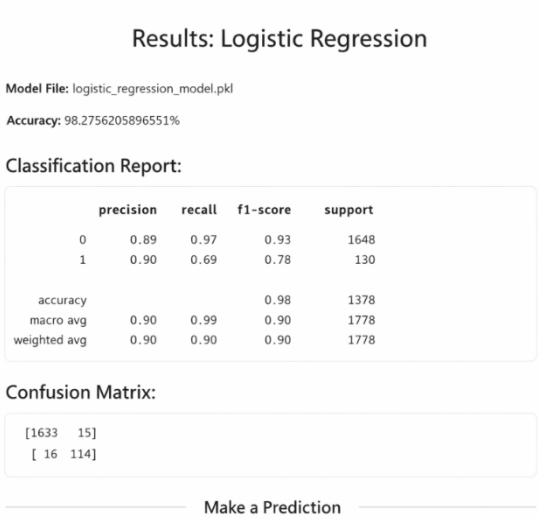

Fig.5.2 Result –Logistic Regression

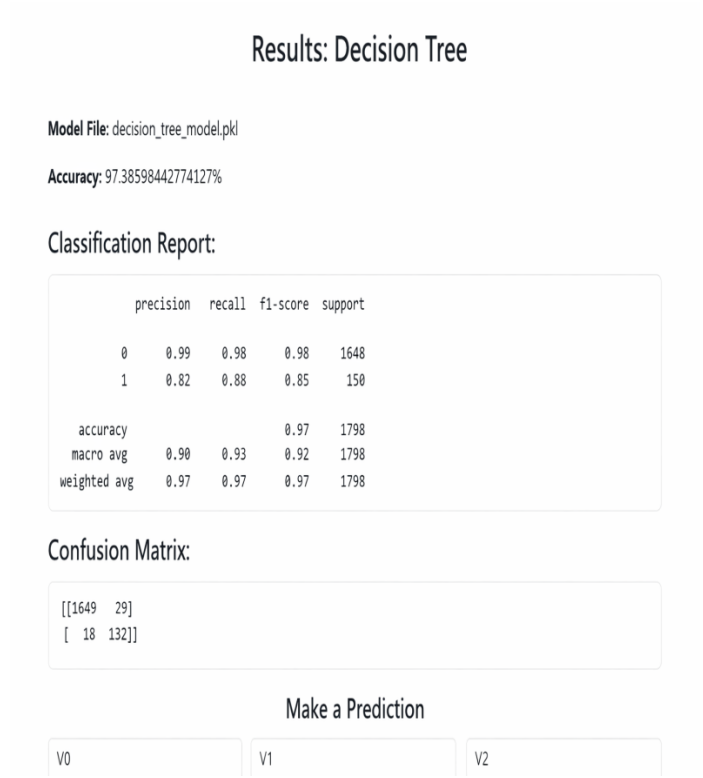

Fig.5.3 Result –Decision Tree

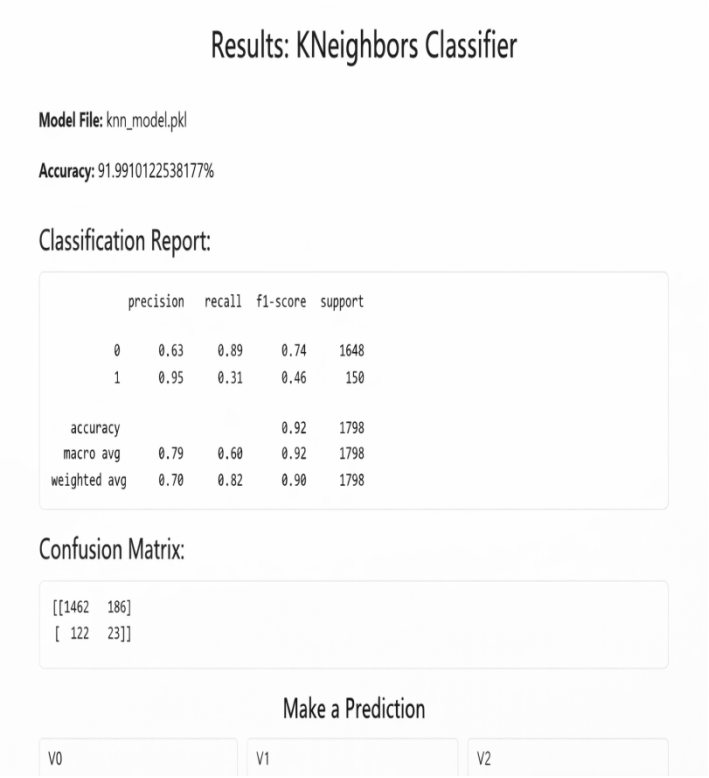

Fig.5.4 Result –KNeighbors Classifier

6. Conclusion

This paper demonstrates that machine learning algorithms can effectively detect credit card fraud with high accuracy. Random Forest performed best, achieving 99.11 % accuracy, followed

closely by Logistic Regression (98.28 %) and Decision Tree (97.33 %). Overall, the system seems good at spotting fraud in financial transactions.It can be a strong starting point for making a real-world fraud detection tool.

Reference

[1] Shu, K., Sliva, A., Wang, S., Tang, J., & Liu, H. (2017). Fake news detection on social media: A data mining perspective..

[2] Ruchansky, N., Seo, S., & Liu, Y. (2017). CSI: A hybrid deep model for fake news detection. arXiv.

[3] Kaliyar, R. K., Goswami, A., & Narang, P. (2021). FakeBERT: Fake news detection in social media with a BERT-based deep learning approach. Multimedia Tools and Applications, 80, 11765–11788.

[4] Alam, F., Cresci, S., & Chakraborty, T. (2020). Fake news detection using machine learning algorithms: A systematic review. IEEE.

[5] Ahmed, H., Traore, I., & Saad, S. (2018). Detecting fake news using machine learning: An ensemble approach. In Proceedings of the 12th International Conference on Data Mining (p. 279–285).

[6] U.K. Finance. (2020). Annual fraud report 2020.

[7] Visa. (2020). Facts & figures: Visa card statistics.

[8] MasterCard. (2020). Mastercard facts & figures.