IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Credit Card Fraud Detection Using Machine Learning

TALLAPUDI PREETHISRI , Mr. B.J.M. RAVI KUMAR

Master of Computer Applications, Assistant Professor

Department of IT & CA, Andhra University College of Engineering, Visakhapatnam, Andhra Pradesh, India

Department of CS & SE, Andhra University College of Engineering, Visakhapatnam, Andhra Pradesh, India

ABSTRACT

Credit card fraud detection is a system that detects and prevents fraudulent transactions before they cause financial harm. Credit card fraud refers to fraudulent activities involving payment cards, such as credit or debit cards. It can include physical theft, electronic interception, and online transactions. As credit card usage increases, financial fraud and scams increase, leading to significant losses for banks and customers. Credit card fraud detection is essential to protect users from financial loss, and these approaches help increase the security of financial transactions. Fraud detection and prevention help maintain trust in the payment system. Consumers trust credit cards for their convenience and security. Credit card fraud data is highly unbalanced, with few fraud cases. Extract relevant features from the dataset. Some common features include transaction amount, time of day, and location. Randomly remove instances from the majority class (valid transactions) to balance the dataset. Create synthetic instances of minority class (unauthorized transactions) to balance the data. Advanced algorithms and machine learning techniques analyze transaction data in real time. Credit card fraud detection based on machine learning (ML) such as Decision Tree (DT), Logistic Regression (LR), K-Nearest Neighbour (KNN) algorithms. The choice of machine learning algorithms and the performance of the evaluation metrics used are important factors that influence the accuracy of machine learning algorithms.

Keywords: Machine Learning, Decision Tree, Logistic Regression, K nearest Neighbor, Credi card fraud transactions, Valid Transactions.

I. INTRODUCTION

OVERVIEW

Credit card fraud is a common problem in the financial sector. This can lead to significant financial losses for customers and financial institutions. Therefore, there is an urgent need to develop effective fraud detection systems that can quickly identify fraudulent transactions. The advent of digital payment methods has increased the risk of credit card fraud, making it easier for criminals to conduct their business secretly. Traditional rule-based systems used to detect fraud have become less effective as fraudsters have adopted a more sophisticated approach. In response, researchers and practitioners have turned to machine learning and deep learning models as effective ways to detect fraudulent activity.

Machine learning algorithms can analyze large amounts of transaction data, which allows them to identify patterns and anomalies that could indicate fraud. Supervised learning methods Decision trees, logistic regression, and neural networks have been used to build models classify transactions as fraudulent or non-fraudulent. These models can shoot more complex patterns and relationships in transaction data, making them more efficient Fraud detection. In this review article, we examine the various approaches and techniques used in credit card fraud detection, including traditional rule-based systems and machine learning algorithms. We also discuss the challenges involved approach and present the current state of the art in the field. Credit card fraud detection is an important part of modern financial security systems detect and prevent unauthorized transactions. For example, digital and online payment methods.

The number of purchases has increased, as has the number of cases of credit card fraud, making it a major problem both consumers and financial institutions. Fraud detection involves monitoring transactions suspicious activity that deviates from the user's normal spending behavior, traditional methods. They are primarily based on rule based systems, potential fraud. However, these systems have often struggled to adapt to new fraud techniques and have been unable to generate a large number of false positives, frustrating legitimate customers. In response to these limitations, more advanced methods have been developed. In particular, machine learning becomes a powerful tool for fraud detection. By analyzing large data sets of transaction history machine learning models can identify patterns and anomalies that could indicate fraud. That's it Models can be supervised by training them on labeled data (fraudulent and non-fraudulent) transactions) or unsupervised, where they detect anomalous values without knowing in advance what constitutes fraud. Among the biggest challenges in credit card fraud detection is dealing with highly imbalanced data sets, e.g. Fraudulent transactions are rare compared to legitimate transactions, allowing for real-time detection capabilities without sacrificing accuracy. In addition, fraud detection systems should always be in place Fraudulent synthetic identity fraud or not.

Key challenges in credit card fraud detection include managing highly imbalanced data sets, as fraudulent transactions are rare compared to legitimate transactions, and maintaining real-time detection capabilities without compromising accuracy. Additionally, fraud detection systems must constantly evolve to keep pace with fraudsters' ever-changing tactics, such as synthetic identity fraud or account takeover. The effectiveness of a lie detection system is typically measured by metrics such as accuracy, precision, recall, and the trade-off between true positives (false positives) and false positives (legitimate transactions falsely flagged). Recent developments in this area are the integration of deep learning techniques that can capture the most complex processes in data and to optimize the use of blockchain technology. Furthermore, the growing availability of big data and advanced analysis provides new opportunities to improve the accuracy and efficiency of fraud detection systems. Overall, credit card fraud detection is a dynamic and growing field, with continued research and technological advances aimed at improving the security of financial transactions and reducing their impacts on users.

PROBLEM STATEMENT

The problem for credit card fraud detection typically involves identifying fraudulent transactions within a credit card transaction record. This fraud detection systems that can accurately and quickly identify fraudulent transactions in real-time and reduce disruptions to legitimate customer transactions.

II. LITERATURE SURVEY

In 2019, Heta Naik and Prashasti Kanikar studied machine learning algorithms like Naive Bayes, Logistic Regression, J48, and Adaboost. They found that Adaboost and Logistic Regression had the highest accuracy, with time being a crucial factor in selecting the best algorithm.

S. Makki, Z. Assaghir, Y. Taher, R. Haque, M. Hacid and H. Zeineddine research describes that Credit card fraud causes significant financial loss, and researchers are working to find innovative solutions. They discovered imbalanced dataset classifications cause inaccurate results. They found LR, C5.0 decision tree algorithm, SVM, and ANN as the best algorithms based on accuracy, AUCPR, and sensitivity, using a balanced dataset for training.

Kumar, Soundarya, Kavitha, Keerthika, and Aswini developed a model for detecting credit card fraud using Random Forest techniques. The system uses decision trees for classification and a confusion matrix for performance calculation, achieving an accuracy of 90%.

In 2019, researchers analyzed credit card fraud detection techniques finding k-nearest neighbor, decision trees, and SVM have medium accuracy, while fuzzy logic and logistic regression have lowest accuracy. KNN and SVM perform better with small datasets.

C. Jiang, J. Song, G. Liu, L. Zheng and W. Luan proposed a novel process with multiple stages. First they collect the transactions made by card holder, then based on the behavioural patterns transactions are aggregated, next the dataset is classified, further the model is trained and finally the model is tested. If any abnormal behaviour arises then a feedback is provided to system about the abnormal behaviour through feedback mechanism.

III. RESEARCH METHODOLOGY

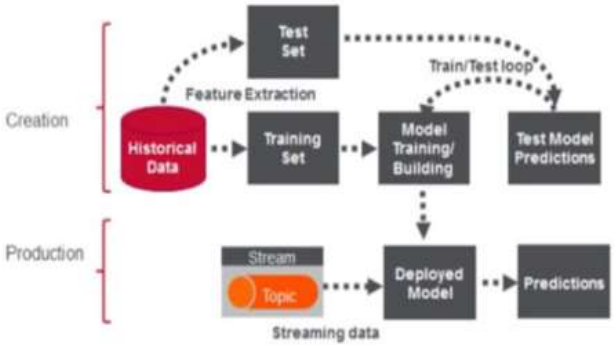


Fig: Design Approach

Problem Definition:

The problem of credit card fraud detection involves identifying unauthorized or fraudulent transactions in real-time to prevent financial loss for both cardholders and financial institutions. This requires developing a machine learning model capable of distinguishing between legitimate and fraudulent transactions based on patterns and anomalies in the data.

1. Data Collection:

Data collection is an essential step in creating an effective credit card fraud detection system. On the quality, variety and quantity of data have a direct impact on the ability of the system to accurately detect fraudulent activity. The continuation provides a general description of the key considerations and methods of collecting data in the context of credit card fraud detection.

2. Data Preprocessing

Perform preprocessing on the collected dataset to clean the text data and prepare it for analysis.

Data Clarification:

- Eliminate duplicates, correct errors and manage incorrect collateral values the integrity of the dataset. For example, incomplete transactions or records with you may need to exclude or account for incorrect values.

Normalization and Standardization:

- Normalization: Scale numerical features to a range [0,1], especially when using models that are sensitive to increasing features, such as neural networks or k-NNs.
- Standardization: Transforms the features to have a mean of 0 and a standard deviation of
  This is especially useful for algorithms like logistic regression.

Data division

- Training and evaluation split: Splits the data into training and evaluation sets to evaluate model performance.
- Tested: Ensures that the skill and probability split maintain the same proportion of fraud and no fraud in the original dataset.

Feature Extraction:

Feature extraction is an essential step in building an effective credit card fraud detection system. This involves transforming raw data into a set of meaningful features that can be used by machine learning models to make predictions. The goal is to create features that help distinguish between fraudulent and non-fraudulent transactions.

Model Selection:

- Logistic Regression
- Decision Trees
- K-nearest neighbor

5. Model training and evaluation

Train the selected models using features extracted from the dataset:

- Data split: Split the dataset into training and test sets (typically an 80/20 split)

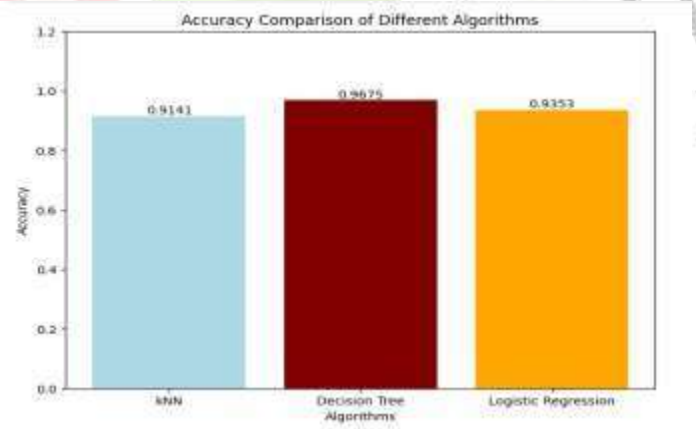
6. Deployment:

Implementing a credit card fraud detection model requires careful planning, Preparing the model for integration and monitoring. Most importantly, the model performs well in real-world conditions, is safe, and allows for future maintenance and improvements.

IV. RESULTS AND DISCUSSION

Algorithms Compared:

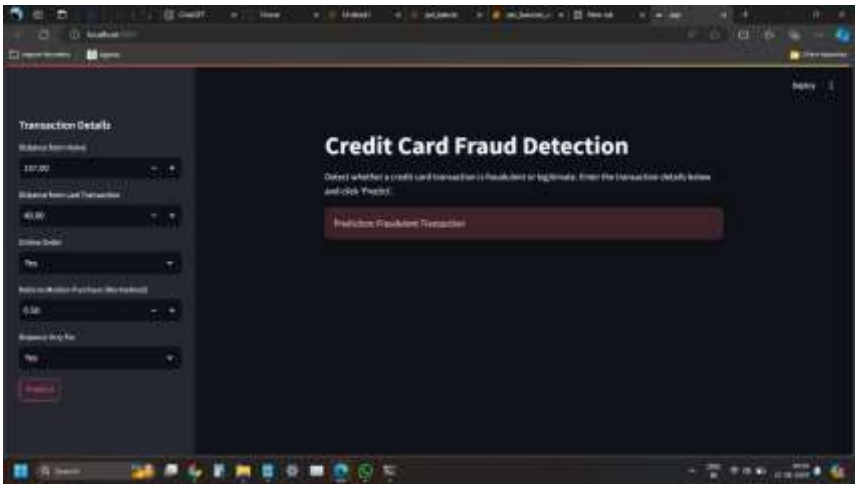
KNN: A simple, instance-based learning algorithm.
Decision Tree: A model that splits data into branches based on feature values, leading to a decision.
Logistic Regression: An analytical tool for binary classification is logistic regression.



Accuracy values:
KNN: Accuracy is about 0.9141 or 91.41%.
Decision Tree: The highest accuracy among the three algorithms with a value of 0.9675 or 96.75%.
Logistic Regression: It has an accuracy of 0.9353 or 93.53%.

The application is called "Credit Card Fraud Detection" and is intended to use input characteristics to determine whether a credit card transaction is legitimate.

This image shows the interface of a web application designed to detect credit card fraud.



The application is called "Credit Card Fraud Detection" and is designed to determine if a credit card transaction is fraudulent based on the characteristics entered.

This image shows the interface of a web application designed for credit card fraud detection.

CLASSIFICATION REPORT

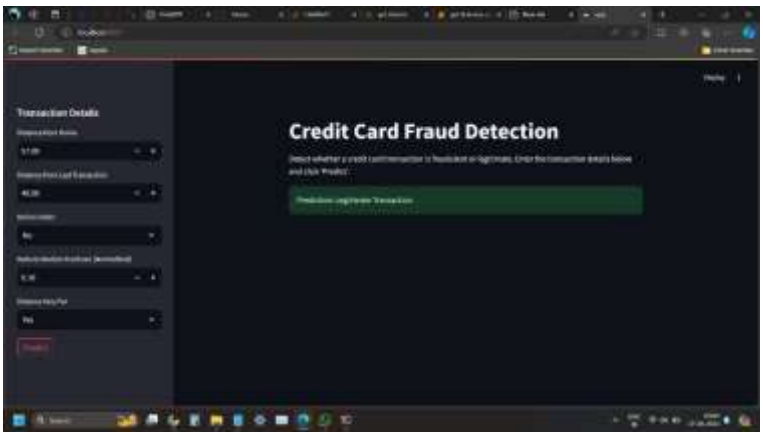


Fig: classification Report of Decision Tree

Classification Report:

	precision	recall	f1-score	support
0.0	0.94	0.99	0.97	365248
1.0	0.79	0.35	0.48	34752
accuracy			0.94	400000
macro avg	0.87	0.67	0.72	400000
weighted avg	0.93	0.94	0.92	400000

V. CONCLUSION

The project developed a credit card fraud detection model that effectively identifies fraudulent transactions by analyzing transaction distance, purchase behavior, and usage patterns, achieving high accuracy through careful feature engineering and model selection.

Machine learning significantly enhances fraud detection, enabling faster and more accurate detection than traditional methods. Real-time application can reduce fraud risk and protect customers. However, continuous monitoring and model updates are crucial due to active fraudsters. Regular retraining with new data and additional security measures can further improve the system's fraud detection capability. This project underscores the importance of data-driven strategies in combating credit card fraud.

FUTURE SCOPE

Advancements in technology necessitate the continuous adaptation and innovation of credit card fraud detection systems. Key strategies include new algorithms, real-time analytics, blockchain technology, and consumer education. This will create a more secure transaction environment, enhancing confidence and safety for businesses and consumers in digital payment landscapes.

VI. REFERENCES

[1] A. Srivastava, A. Kundu, S. Sural, and A. Majumdar, "Credit Card Fraud Detection Using a Hidden Markov Model," in: IEEE Transactions on Dependable and Secure Computing, vol. 5, no. 1, January-March 2008, pp. 37-48.

[2] Statista, the statistical portal, https://www.statista.com/topics/871/online-shopping/,

March 14, 2017.

[3] S. Yusuf, E. Duman, "Credit Card Fraud Detection Using Decision Trees and Support Vector Machines," IMECS 2011, International Conference of Computer Engineers and Scientists,

2011, 1, 442-447, 2011.

[4] Y. Wang, S. Adams, P. Beling, S. Greenspan, S. Rajagopalan, M. Vélez-Rojas, S.

Mankovski, S. Boker, D. Brown, "Privacy-preserving distributed deep learning and its

applications in credit card fraud detection", 1070-1078, 2018.

[5] S. Bhattacharyya, S. Jha, K. Tharakunnel, J. Westland, "Data Mining for Credit Card Fraud: A Comparative Study," Decision Support Systems, 50, 602-613, 2011.

[6] Y. Sahin, S. Bulkan, E. Duman, „Ein kostensensitiver Entscheidungsbaumansatz zur Betrugserkennung" Expertensystem. Appl., 40, 5916-5923, 2013.

[7] S. Panigrahi, A. Kundu, S. Sural, A. Majumdar, "Credit card fraud detection: A fusion approach using Dempster-Shafer theory and Bayesian learning", Information Fusion, 10, 354. 363, 2009.

[8] S. Maes, K. Tuyls, B. Vanschoenwinkel, B. Manderick, „Erkennung von Kreditkartenbetrug mithilfe von Netzwerken Bayesianisch und neuronal, 2002.

[9] M. Zareapoor, P . Shamsolmoali, „Anwendung zur Erkennung von Kreditkartenbetrug: basierend auf dem Bagging-Ensemble-Klassifikator", Procedia Computer. Science, 48, 679-686, 2015.

[10] L. Zheng et al., „Ein neuer Ansatz zur Erkennung von Kreditkartenbetrug auf der Grundlage eines Verhaltenszertifikats." 15. IEEE International Conference on Networking, Sensing and Control (ICNSC), Zhuhai, S. 1-6, 2018.