

Alliance College of Engineering and Design

**Department of Computer Science and Engineering and Information Technology Batch 2022 – 2026**

Data Structure & Algorithms

Topic:

Literature survey on Credit Card Fraud Detection using Decision Trees

**Student Team Members (Name and Registration No.):**

**1) Shiv Nandan Verma (2022BCSE07AED831)**

**Name of the Mentor:** Dr Rekha Nair

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Title** | **Introduction** | **Methods Used** | **Findings** | **Conclusion** |
| 1. Credit Card Fraud Detection using Decision Tree and Random Forest   ***By -*** *Dhwanir Shah and Lokesh Kumar Sharma in*  *(2023)* | During the  pandemic, online  shopping has become increasingly popular due to its convenience. Credit cards are commonly used for online payments, but they pose risks such as fraud. Credit card fraud occurs when someone  unauthorized uses another person's card. To ensure safe transactions, there's a need for effective fraud detection systems. | 1. Getting required dataset from Kaggle platform. 2. Preprocessing, Cleaning data for further process 3. Splitting Data sets into Train and Test. 4. Training ML model Using Decision Tree Classifier 5. Calculating score of ML model | 1. Genetic algorithms reduce false alerts in credit card fraud detection, predicting fraudulent transactions early. 2. Machine learning algorithms, favored by Popat and Chaudhary, offer high accuracy in detecting various types of credit card fraud. 3. Local outlier factors combined with isolation forests achieve high accuracy in fraud detection, as   demonstrated by Maniraj et al.   1. Customer behavior and location data aid in fraud detection, with the NN algorithm showing 80% accuracy, highlighted by Gulati et al. 2. Random forest with boosting technique outperforms other algorithms in credit card fraud detection, according to Bhanusri et   al. | 1. Online shopping saves time and offers convenience, but fraudulent credit card transactions are a growing concern. 2. Credit card payment is popular, but fraud causes financial losses for banks and customers. 3. Implementing secure fraud detection systems is vital to address these challenges. 4. Various machine learning algorithms, like Naïve Bayes and Random Forest, help accurately identify fraudulent   transactions.   1. These algorithms analyze data sets to detect and prevent fraudulent credit card transactions   effectively. |
| 2) Credit Card Fraud Detection Algorithm using Decision Treesbased Random Forest  Classifier | Credit card fraud is a growing concern in both online and offline transactions, where fraudsters  seek sensitive information to carry out illegal activities  without the user's | 1. Module 1 involves collecting credit card data and storing it in a database, followed by descriptive analysis. 2. Module 2 focuses on data cleaning, removing duplicate   and null values to | 1. RFA achieves high accuracy (90-95%) in credit card fraud detection, surpassing other algorithms. 2. Scikit-learn facilitates dataset classification and preprocessing, enhancing RFA   implementation. | 1. This study evaluates two types of random forest models using a real-life credit card transaction dataset. 2. Despite achieving good results,   challenges remain,  particularly with imbalanced data. |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ***By -*** *B. N. V.*  *Madhubabu,*  *T. Vyshnavi & K. Ashok in (2021)* | knowledge.  Detecting fraudulent transactions is crucial, employing various models categorized under supervised and unsupervised  learning algorithms. This paper aims to assess the accuracy of fraud detection using Random  Forest Algorithm (RFA). | obtain a refined dataset.   1. Module 3   preprocesses the cleaned dataset based on transaction amount and time.   1. Module 4   partitions the dataset into training and testing sets, applies Random Forest Algorithm (RFA), and generates a confusion matrix.   1. Module 5   evaluates the  confusion matrix  through graphical representation to enhance accuracy assessment.hms  analyze data sets to detect and prevent fraudulent credit  card transactions effectively. | 1. Confusion matrix partitioning into TP, TN, FP, and FN aids in evaluating RFA performance. 2. RFA outperforms other algorithms in accuracy assessment and graph   representation.   1. System architecture emphasizes dataset   analysis, cleaning, partitioning, RFA  application, and performance analysis for efficient fraud detection.o detect and prevent fraudulent credit card transactions effectively. | 1. Future work will address imbalanced data issues and enhance the random forest algorithm. 2. Improvements will focus on refining the voting mechanism to account for varying classifier importance. 3. These   enhancements aim to enhance the overall performance and  effectiveness of random forest in credit card fraud  detection.cient fraud detection.o detect and prevent fraudulent  credit card  transactions effectively. |
| 3) Decision Tree  Algorithm for Credit Card Fraud Detection  ***By*** *- Aditya Joshi, Anuj Singh,*  *Shikha Chauhan, & Anupama Sharma in*  *(2021)* | Fraud, an intentional act for monetary gain, is increasing daily. Prevention and detection systems are vital strategies. Credit card fraud rises with electronic payment usage, especially  online. Numerous algorithms tackle this issue. Efficient fraud detection is crucial due to the rapid nature of digital transactions and the challenge of handling concurrent  transactions. This study aims to employ the Decision Tree | 1. Decision tree technique logically expresses independent and dependent   properties in a tree structure.   1. Prominent classifiers like ID3, C4.5, and MLPC are employed for class model generation from decision trees. 2. Pruning is used to improve precision and prevent overfitting by cleaning up subtrees. 3. The Decision Tree model builds a small decision tree with   high precision for | 1. Attribute Selection Measurement (ASM) ranks features for split attribute selection, commonly using gain ratio, Gini's coefficient, and information gain. 2. Information Gain (IG) reduces entropy to measure impurities in datasets, commonly used in ID3 decision tree algorithm. 3. Gain Ratio normalizes information gain to address bias, commonly employed in C4.5 algorithm. 4. Decision tree algorithms like ID3 and J48 utilize information | 1. Supervised machine learning models were evaluated for predicting fraudulent transactions using Kaggle's credit card dataset. 2. Decision tree classification emerged as the most suitable model due to its accuracy and time sensitivity. 3. Despite Random Forest's higher sensitivity, Decision Tree's quicker prediction time favored its selection. 4. Decision trees are recommended for   negative detection due |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | model to assess an imbalanced dataset. | credit card fraud detection.  5) MLPC approach is utilized as pre- pruning to stop tree growth at a specified level, employing a top-down recursive  partitioning and conquest method. | gain and gain ratio for attribute selection.  5) Confusion matrix of decision tree prediction results demonstrates high accuracy in correctly predicting zeros and non-zeros. | to their efficiency in time-sensitive  predictions.  5) Time sensitivity alongside accuracy is crucial for selecting the most appropriate model for fraud detection. |
| 4) Detecting Credit Card Fraud by Decision  Trees and Support Vector Machines  **By -** *Y. Sahin and E.*  *Duman in (2011)* | Despite  advancements in fraud prevention mechanisms like CHIP&PIN, fraudulent activities persist, particularly in credit card transactions. This study focuses on developing and comparing  classification models based on decision trees and support vector machines (SVM) for detecting credit card fraud, aiming to provide insights using real data sets. | 1. Getting required dataset from Kaggle platform. 2. Preprocessing, Cleaning data for further process 3. Splitting Data sets into Train and Test. 4. Training ML model Using Decision Tree Classifier and   Support Vector Machines   1. Calculating score of ML model | 1. Decision tree models outperform SVM models in detecting fraudulent transactions on test datasets. 2. SVM models exhibit overfitting behavior on training datasets, leading to lower performance on test datasets. 3. As the training data size increases, SVM models' performance becomes comparable to decision tree models. 4. Decision tree models, especially C&RT, detect a larger number of fraudulent transactions compared to SVM models. 5. Accuracy is not an ideal metric for assessing model performance in fraud detection, as it does not prioritize detecting fraudulent transactions. | 1. With the increasing use of credit cards, credit card fraud has escalated significantly, necessitating enhanced security measures. 2. Developing accurate and efficient credit card fraud detection systems is imperative for financial   institutions to combat fraudulent activities effectively.   1. Advanced technologies are   essential for  automating and  strengthening the security of financial transaction systems.   1. Continuous efforts are needed to stay ahead of evolving fraud tactics and protect consumers' financial assets. 2. Collaboration between financial institutions,   technology providers, and regulatory bodies is crucial for combating credit card fraud  effectively. |
| 5) Credit Card Fraud  Detection | With the surge in credit card usage for  both online and offline | 1) Getting required dataset from Kaggle  platform. | 1) Information gain  achieved 0.9183 | 1) Efficient algorithm for decision tree |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Using Split Criteria in  Classification  ***By*** *- Manisha & Neena Madan*  *in(2017)* | transactions, instances of credit card fraud are escalating. Modern techniques,  particularly data  mining, are  instrumental in detecting fraudulent transactions due to their effectiveness. This paper focuses on the Decision Tree technique, aiming to classify transactions as fraudulent or non- fraudulent. | 1. Preprocessing, Cleaning data for further process 2. Splitting Data sets into Train and Test. 3. Training ML model Using Decision Tree Classifier and   Support Vector Machines   1. Calculating score of ML model | 1. The IP address has the highest gain ratio. 2. For Attribute customer name the information gain is 0.1567 3. Gini index of IP Address is 0.3333. Gini index value of Trans Amt is 0.4375. 4. Reduction in impurity of attribute IP Address is 0.1111 and Reduction in impurity of attribute TransAmt is 0.0069. | learning enhances fraud detection.   1. Information Gain, Gain Ratio, and Gini Index are effective attribute selection measures. 2. IP Address emerges as a crucial splitting attribute for identifying fraudulent customers or merchants. 3. Tracing fake IP addresses using software aids in fraud detection. 4. This approach improves the accuracy and effectiveness of   fraud detection systems. |

Thankyou!