PHASE II

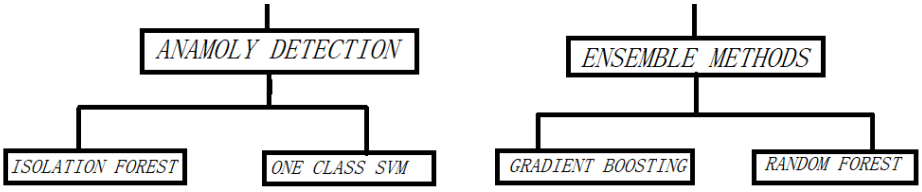
CREDIT CARD FRAUD DETECTION

Table of Contents

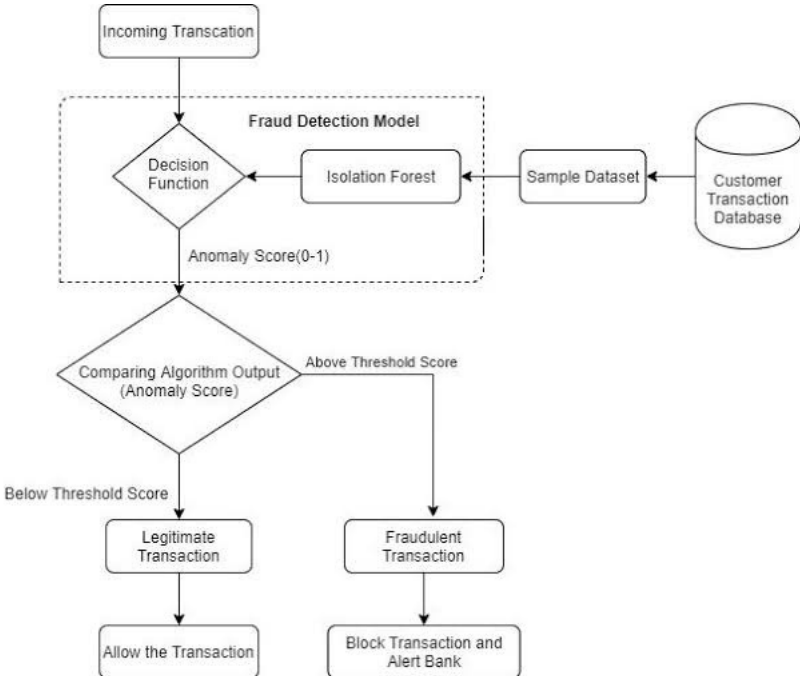
1. INTRODUCTION

Fraud detection is a critical task in various industries, including finance, healthcare, and e-commerce. Traditional methods of fraud detection often fall short in identifying complex and evolving fraudulent activities. To address this challenge, advanced techniques such as anomaly detection algorithms and ensemble methods have gained prominence in recent years. In this document, we will explore these advanced techniques and discuss how they can be leveraged to improve fraud detection accuracy.

Various Algorithm used:



2.FLOW CHART



2.1 Anomaly Detection Algorithms

Anomaly detection algorithms are designed to identify patterns or instances that deviate significantly from the norm, which may indicate fraudulent activity. Two widely used anomaly detection algorithms are Isolation Forest and One-Class SVM.

2.1.1 Isolation Forest

Isolation Forest is a tree-based anomaly detection algorithm that works by isolating anomalies rather than trying to model the normal data. It builds a binary tree structure to recursively partition the data, making it highly efficient. Anomalies are isolated in shorter paths, and their scores are higher, making it easier to detect them. Isolation Forest is particularly effective in high-dimensional datasets and is resistant to outliers.

2.1.2 One-Class SVM

One-Class SVM is a support vector machine variant used for anomaly detection. It learns a decision boundary around the normal data, aiming to separate it from any potential anomalies. It is useful when the majority of the data consists of normal instances, and anomalies are rare. One-Class SVM can capture complex data distributions and provide a measure of anomaly scores.

2.2. Ensemble Methods

Ensemble methods combine multiple models to improve predictive performance. They can be applied to fraud detection to harness the strengths of various algorithms and enhance accuracy. Common ensemble methods for fraud detection include bagging, boosting, and stacking.

Ensemble methods can:

- Reduce overfitting by combining multiple models.
- Improve generalization by capturing diverse patterns.
- Enhance robustness to noise and outliers.

2.2.1 Random forest

Random Forest is an ensemble learning method that uses multiple decision trees to make predictions. Each tree is trained on a random subset of the data, and the final prediction is based on a majority vote or weighted average of the individual tree predictions. It can improve robustness and reduce overfitting.

2.2.2 Gradient Boosting

Algorithms like Gradient Boosting build an ensemble of weak learners (typically decision trees) sequentially, where each new learner focuses on correcting the errors made by the previous ones. This can lead to better fraud detection by emphasizing difficult-to-detect cases.

3. Improving Fraud Detection Accuracy

To improve fraud detection accuracy using advanced techniques, consider the following steps:

1. Feature Engineering: Create informative features that can better represent the underlying patterns of fraud.

2. Data Preprocessing: Standardize or normalize data to ensure that different algorithms perform optimally.

3. Model Selection: Experiment with Isolation Forest, One-Class SVM, and other anomaly detection algorithms to identify which works best for your dataset.

4. Ensemble Techniques: Combine multiple models using ensemble methods like bagging or stacking. This can help in capturing diverse aspects of fraud patterns.

5. Hyperparameter Tuning: Fine-tune the hyperparameters of your models for optimal performance.

6. Cross-Validation: Use cross-validation to assess the model's performance and avoid overfitting.

7. Continuous Monitoring: Implement real-time or periodic fraud detection to adapt to evolving fraud patterns.

THANK YOU!

BY:

R. ABINAYA (abinayaraja1052004@gmail.com)

S. ANBARASI (anbarasideva14@gmail.com)

K. BHUVANA SURUTHI (sagunthalakumar16@gmail.com)

M. GAYATHRI (45gayathri@gmail.com)