Detecting Financial Fraud: Leveraging Machine Learning for Enhanced Security and Loss Prevention

Usama Ahmed

April 18, 2024

1 Introduction

The aim of this project is to develop a robust fraud detection system using machine learning algorithms. Fraudulent activities refer to deceptive or dishonest actions carried out for personal gain, often involving misrepresentation or manipulation [1]. Credit card fraud specifically pertains to unauthorized use or theft of credit card information for fraudulent transactions.We will focus on two algorithms: Support Vector Machine (SVM) and Isolation Forest. The dataset used for training and evaluation is a credit card transaction dataset containing both fraudulent and non-fraudulent transactions.

2 Description of Dataset

The dataset contains credit card transactions made by European cardholders in September 2013. It consists of numerical input variables resulting from a PCA transformation. The dataset [2] includes features such as Time, V1-V28 (principal components), Amount, and Class (0 for non-fraudulent, 1 for fraudulent). An example training data point from the dataset is shown below:

```
Time: 0
V1: -1.359807
V2: -0.072781
...
Amount: 149.62
Class: 0 (non-fraudulent)
```

3 Description of Algorithms

3.1 Support Vector Machine (SVM)

The SVM algorithm is a supervised learning algorithm that is effective for classification tasks. It works by finding the optimal hyperplane that separates different classes in the feature space [3].

Algorithm 1 Support Vector Machine (SVM)

procedure SVM(X_{train}, y_{train}, C, kernel)
    Initialize the SVM model with parameters C and kernel type
    Train the SVM model using X_{train} and y_{train}
    return SVM model
end procedure

3.2 Isolation Forest

The Isolation Forest algorithm is an unsupervised learning algorithm used for anomaly detection. It works by isolating anomalies in the dataset using binary trees [4].

Algorithm 2 Isolation Forest

procedure IsolationForest(X_{train}, $n_{\text{estimators}}$, max_samples)
    Initialize an empty list to store individual isolation trees
    for i from 1 to $n_{\text{estimators}}$ **do**
        Draw a random sample of size max_samples from X_{train}
        Train an isolation tree using the random sample
        Add the trained tree to the list
    end for
    return list of isolation trees
end procedure

4 Evaluation Procedure

The performance of the models will be evaluated using metrics such as precision, recall, F1-score, and accuracy. Cross-validation will be used to ensure robustness of the results.

5 Hyperparameter Tuning

Grid search or random search will be employed to tune the hyperparameters of the SVM and Isolation Forest algorithms, such as the regularization parameter (C) for SVM and the number of estimators for Isolation Forest.

References

[1] Sushmito Ghosh and Douglas L Reilly. Credit card fraud detection with a neural-network. In: *System Sciences, 1994. Proceedings of the Twenty-Seventh Hawaii International Conference on*. Vol. 3. IEEE. 1994, pp. 621–630.

[2] Gabriel Preda. *Credit Card Fraud Detection Predictive Models*. Kaggle. 2017.

[3] Shujun Huang, Nianguang Cai, Pedro Penzuti Pacheco, Shavira Narrandes, Yang Wang, and Wayne Xu. Applications of support vector machine (SVM) learning in cancer genomics. In: *Cancer genomics & proteomics* 15.1 (2018), pp. 41–51.

[4] Fei Tony Liu, Kai Ming Ting, and Zhi-Hua Zhou. Isolation forest. In: *2008 eighth ieee international conference on data mining*. IEEE. 2008, pp. 413–422.