Credit Card Fraud Detection

Teammates:

B Sai Moesha - AP21110011261

B VijayaSravya - AP21110011261

G TarunSai - AP21110010690

Introduction

In today's digital age, where financial transactions are increasingly carried out online, the risk of credit card fraud has become a significant concern for both financial institutions and consumers alike. According to recent statistics, credit card fraud continues to pose a substantial threat, with billions of dollars lost annually due to fraudulent activities. machine learning algorithms helps to detect and prevent credit card fraud effectively. Machine learning models provide a proactive defense against fraudulent transactions while minimizing false positives By analyzing vast amounts of transactional data, our models can identify patterns indicative of fraudulent activity.

Data set Description

Context :-

It is important that credit card companies are able to recognize fraudulent credit card transactions so that customers are not charged for items that they did not purchase.

Content :-

The dataset contains transactions made by credit cards in September 2013 by European cardholders. This dataset presents transactions that occurred in two days, where we have 492 frauds out of 284,807 transactions. The dataset is highly unbalanced, the positive class (frauds) account for 0.172% of all transactions.

It contains only numerical input variables which are the result of a PCA transformation. Unfortunately, due to confidentiality issues, we cannot provide the original features and more background information about the data. Features V1, V2, ... V28 are the principal components obtained with PCA, the only features which have not been transformed with PCA are 'Time' and 'Amount'. Feature 'Time' contains the seconds elapsed between each transaction and the first transaction in the dataset. The feature 'Amount' is the transaction Amount, this feature can be used for example-dependant cost-sensitive learning. Feature 'Class' is the response variable and it takes value 1 in case of fraud and 0 otherwise.

Given the class imbalance ratio, we recommend measuring the accuracy using the Area Under the Precision-Recall Curve (AUPRC). Confusion matrix accuracy is not meaningful for unbalanced classification.

Description of Related Research Papers

1. **Credit card fraud detection using machine learning: A survey**
2. **Credit Card Fraud Detection Using Machine Learning**
3. **Design and Implementation of Different Machine Learning Algorithms for Credit Card Fraud Detection**
4. **Application of Machine Learning Techniques in Credit Card Fraud Detection**
5. **Credit Card Fraud Detection Using Machine Learning Algorithms**

Credit Card Fraud Detection Using Machine Learning

The machine learning techniques used in this research paper for credit card fraud detection include K-Nearest Neighbors (KNN), Support Vector Machine (SVM), and Logistic Regression.

The most accurate model in credit card fraud detection, is the Support Vector Machine (SVM) model. It achieved an accuracy score of 99.94% with only 51 misclassified instances.

It is believed that using this model will help decrease credit card fraud and increase customer satisfaction by providing a more secure experience.

Credit card fraud detection using machine learning: A survey

The dataset used in the paper for credit card fraud detection is a sample of 280,000 credit card transactions with PCA-transformed features. The survey delves into strategies for handling dataset shift, feature engineering techniques, and sequence modeling for fraud detection.

The survey discusses the challenges faced in detecting fraudulent transactions due to the imbalance in class distributions.

The paper suggests the use of Convolutional Neural Networks (CNN) for credit card fraud detection. CNNs have been proposed as a powerful tool for detecting fraudulent transactions due to their ability to learn complex patterns and features from data.

Design and Implementation of Different Machine Learning Algorithms for Credit Card Fraud Detection

The document provides an in-depth discussion of the machine learning techniques used, such as logistic regression, decision trees, random forest, and Catboost, along with their respective features and characteristics. The dataset used for this analysis was obtained from Kaggle. The machine learning algorithm that demonstrated the highest accuracy for detecting credit card fraud in the document is Catboost, with an accuracy of 99.87%. The accuracy percentages of the different algorithms analyzed and compared in the document are as follows:

1. **Logistic Regression: 93.70%**
2. **Decision Tree: 99.40%**
3. **Random Forest: 99.60%**
4. **Catboost: 99.87%**

Application of Machine Learning Techniques in Credit Card Fraud Detection

The machine learning models used in the study on credit card fraud detection include, Logistic Regression ,Random Forest: ,XGBoost.
These models were selected for their effectiveness in handling imbalanced datasets and their potential to improve the accuracy of credit card fraud detection systems.

In the study on credit card fraud detection, the Random Forest model, when used with a combination of SMOTE and Tomek Links removal, achieved a precision score of 0.84 and a recall score of 0.84 53. This indicates that the model had a high level of accuracy in correctly identifying both fraudulent and non-fraudulent transactions.

Credit Card Fraud Detection using Machine Learning Algorithms

The techniques were applied to analyze and detect fraudulent activities in credit card transactions.,Logistic,Regression,Decision Trees,Random Forest,Local Outlier Factor,solation Forest,One-Class SVM.

accuracy values obtained after applying different machine learning techniques for credit card fraud detection. Here are some of the accuracy values reported in the paper:

1. **Logistic Regression: Accuracy of 97.18%**
2. **Decision Tree: Accuracy of 97.08%**
3. **Random Forest: Accuracy of 99.98%**
4. **Local Outlier Factor: Accuracy of 45.82%**
5. **Isolation Forest: Accuracy of 58.83%**
6. **One-Class SVM: Accuracy of 70.09%**

Random Forest algorithm achieved the highest accuracy of 99.98% 9. This indicates that Random Forest performed exceptionally well in detecting fraudulent transactions in the dataset compared to the other algorithms mentioned in the paper.

Conclusion

Based on our preliminary research, we have used the below algorithms:

1. **Random Forest**
2. **KNN**
3. **Logistic Regression**
4. **Decision Tree**
5. **SVM**

We have discovered that the Random Forest classifier outperforms the Decision Tree, Logistic Regression, KNN, and SVM.

Thank you very much!