Credit card fraud detection using Machine Learning

Introduction

Since most people now use credit cards to pay for their necessities due to technological advancements, the number of credit card fraud cases has been steadily increasing in recent years. Nowadays, credit cards are accepted as payment methods by practically all businesses, regardless of size. Credit card fraud occurs in all types of organisations, including banks, car dealerships, and appliance manufacturers. Numerous techniques, such as data mining and machine learning, are used in conjunction with algorithmic approaches to detect credit card fraud, however the results have been largely inconclusive. Thus, it is necessary to build algorithms that are both effective and efficient and that have a big impact. We make an effort to prevent credit card fraud before it happens.

Problem statement

The majority of them now use credit cards to purchase items that they desperately need but are currently unable to afford. Credit cards are used to satisfy needs, but there is also an increase in credit card fraud, thus it is necessary to create a model that fits well and makes predictions with more accuracy.

Rationale

- Financial Loss Prevention: Avoiding significant financial losses for cardholders and institutions.
- Customer Protection: Preventing identity theft and financial harm to customers.
- Regulatory Compliance: Meeting legal requirements and avoiding fines.
- Reputation Management: Protecting the business's reputation by showing a commitment to security.
- Operational Efficiency: Reducing chargebacks and investigations, improving efficiency.
- Adapting to Fraud Tactics: Staying ahead of evolving fraud tactics.
- Data Security: Protecting transaction data from breaches and theft.

Objectives

- The main objective of the research is to find a fraudulent transactions in credit card transactions.
- When supervised learning and deep learning were compared, the deep learning algorithm performed better in terms of accuracy.

Programming language used

Python is the programming language we utilised to implement the suggested system. Python is an application-rich language suitable for beginners. Python has become the new standard in scripting languages due to its ease of use, interpretability, object-oriented, and high level of abstraction in recent years. One of the greatest languages for machine learning implementation is Python. It offers extensive libraries and packages for machine learning.

Methods

This Section describes the implementation, including the algorithm that was applied to put the suggested system into practice. The loading of the dataset is where the implementations in this paper begin. after which data pre- processing, which entails data normalization and cleansing, is completed. The dataset is divided into two subsets for the purpose of training and testing the model. In the end, the system determines if a transaction is fraudulent or not. A number of techniques have been proposed for machine learning-based card fraud credit fraud detection , including data collection, preparation, analysis, splitting, training, testing, and output analysis.

Data set

The dataset used by the suggested system was downloaded from https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud. The transactions that customers made in a European bank throughout the 2017–18 year comprisethe dataset. It has 31 columns total, of which 30 are features and 1 is the target class that determines if the transaction is fraudulent or not.

Evaluation measure

Precision, recall, and accuracy are computed and the final product is assessed using the confusion matrix as a basis. There are two classes in it: the projected class and the actual class. These characteristics determine the confusion metrics:

True Positive : Is one in which both values are positive, i.e 1.

True Negative: 0 is the situation in which both numbers are negative.

False Positive : When the true class is 0 and the non-true class is 1, this is known as a false positive.

False Negative: This occurs when the true class is zero and the actual class is one.

Precision defined as follows:

Precision = true positive / Actual result

Precision = true positive/(true positive + false positive)

Recall defined as follows:

Recall = true positive / predicted result

Recall = true positive/(true positive + false negative)

Accuracy defined as:

Accuracy = (true positive + true negative)/ total

Conclusion

In conclusion, our study demonstrates the effectiveness of machine learning algorithms in detecting credit card fraud. We achieved promising results some various machine learning models in which The XGBOOST model with Random Oversampling with Stratified K-Fold CV provided us with the best accuracy and ROC on oversampled data out of all the models we built in the oversample situations.

Following that, we adjusted the hyperparameters and obtained the metrics But out of all the models we developed, we discovered that the best outcome was obtained using Logistic Regression with L2Regularization for Stratified K-Fold cross validation (without any oversampling or under sampling).

The analysis highlighted the importance of transaction amount, time, and frequency in fraud detection. While our models performed well, future research could explore deep learning and anomaly detection methods for further improvements.

Overall, our study contributes to combating credit card fraud and shows the potential of machine learning in enhancing fraud detection systems for financial institutions

.