CREDIT CARD FRAUD DETECTION USING AUTOENCODERS

A Project Report
Presented to
The Faculty of the College of
Engineering

San Jose State University
In Partial Fulfillment
of the Requirements for the Degree
Master of Science in Software Engineering

By
Ashika Anand Babu
Anusha Gangasani
Ramya Mahesh
May 2023

TABLE OF CONTENTS

1. Abstract:

The usage of credit cards has dramatically increased these days. As credit cards become the most popular payment method for both online and regular purchases, related fraud cases are also on the rise. It is crucial for credit card companies to recognize fraudulent credit card transactions so that customers are not charged for these fraudulent transactions. There are two main reasons why the credit card fraud detection problem becomes challenging. First, the characteristics of normal behavior and fraud are constantly changing. Second, the credit card fraud data set is highly asymmetric. Problems like these can be solved with data science, and its importance together with machine learning cannot be overemphasized. In this project, we have used autoencoders to detect fraudulent transactions. This project demonstrates modeling a dataset using machine learning and credit card fraud detection. The problem of detecting credit card fraud involves modeling previous credit card transactions using data that turns out to be fraudulent. This model is then used to detect whether new transactions are fraudulent. Our goal is to detect 100% of fraudulent transactions while minimizing false fraud classifications. Credit card fraud detection is a good example of classification. This process focused on analyzing and preprocessing the dataset and applying autoencoders to detect fraudulent transactions.

2. Introduction:

Credit card fraud refers to the unauthorized and unwelcome use of a credit card account by someone other than the account owner. The abuse can be stopped with the use of necessary preventative measures, and the behavior of such fraudulent acts can be researched to lessen it and safeguard against recurrence. In other terms, credit card fraud is the use of another person's credit card for personal gain when neither the cardholder nor the organization responsible for providing the card are aware that the card is being used. Monitoring user populations' behavior is a key component of fraud detection since it helps identify, detect, and prevent undesirable behaviors including fraud, intrusion, and defaulting.

This is a really pertinent issue that has to be addressed by communities like machine learning and data science, where an automated solution is possible. From the standpoint of learning, this issue is particularly difficult because it is characterized by many characteristics, like class imbalance. There are significantly more legitimate transactions than fraudulent ones. Additionally, the statistical characteristics of the transaction patterns frequently vary over time.

The deployment of a fraud detection system in the real world is not without difficulties, though. In instances from the real world, the enormous volume of payment requests is swiftly reviewed by automated tools to choose which transactions to authorize. Algorithms for machine learning are used to analyze all permitted transactions and flag any that seem suspect. Professionals who are investigating these reports get in

touch with the cardholders to get their confirmation of whether the transaction was legitimate or fraudulent. The automated system receives feedback from the investigators, which is then utilized to train and update the algorithm to eventually improve the performance of fraud detection over time.

Methods for detecting fraud are always being improved to prevent criminals from altering their fraudulent tactics. These scams are categorized as: Card theft, account bankruptcy, device intrusion, application fraud, counterfeit cards, telecommunication fraud, and credit card fraud both online and offline.

3. Related Work:

In general, a crime is described as a violation when it is done with the intent to harm another person. Fraud may be perpetrated for a number of purposes, including amusement, taking advantage of a company or organization, retaliation, causing financial loss, tarnishing one's reputation, and others. There are numerous different forms of scams as well, including tax fraud, forgeries, bankruptcy fraud, identity theft, health fraud, insurance fraud, and many more. Only credit card frauds are being discussed here, and they can fall into one of two categories: either offline credit card frauds or online credit card scams.

When a person's credit card is lost or stolen, credit card fraud occurs offline. Online frauds are when an attacker or hacker steals information and uses it to carry out illicit acts. Internet usage is dramatically rising as a result of the quickly evolving technologies. Significantly, this is encouraging a lot of credit card theft.

In general, the process of determining whether transactions are legitimate or fraudulent has been referred to as credit card fraud detection. Researchers frequently employ data mining and machine learning techniques to research and identify credit card theft since they are widely used to combat cybercrime. Machine learning is founded on learning intelligence and developing its own model for the purpose of classification, grouping, and other uses, as opposed to data mining, which focuses on uncovering useful intelligence.

4. Data

The data includes 492 fraudulent transactions out of 284,807 total transactions using European credit cards over the course of two days. A Principal Component Analysis (PCA) has been used to reduce everything except the time and amount due to privacy concerns. We can also consider that there are only 0.17% of fraudulent transactions in our given dataset and we are able to obtain over ~70% accuracy using our model for predicting credit card fraud occurrence.

5. Method

The method this study suggests makes use of the most recent machine learning techniques to identify unusual behaviors, also known as Autoencoders.

Autoencoder:

Autoencoders initially may look rather strange. These models' task is to forecast the input given the same input. Let's examine autoencoder neural networks in more detail. The following identity function is attempted to be approximated by this autoencoder:

$$f_{W,b}(x) \approx x$$

While attempting to do so might appear simple at first, it is crucial to remember that our goal is to discover structure by learning a compressed representation of the data. This can be accomplished by reducing the model's concealed unit count. Undercomplete autoencoders are those that fall into this category.

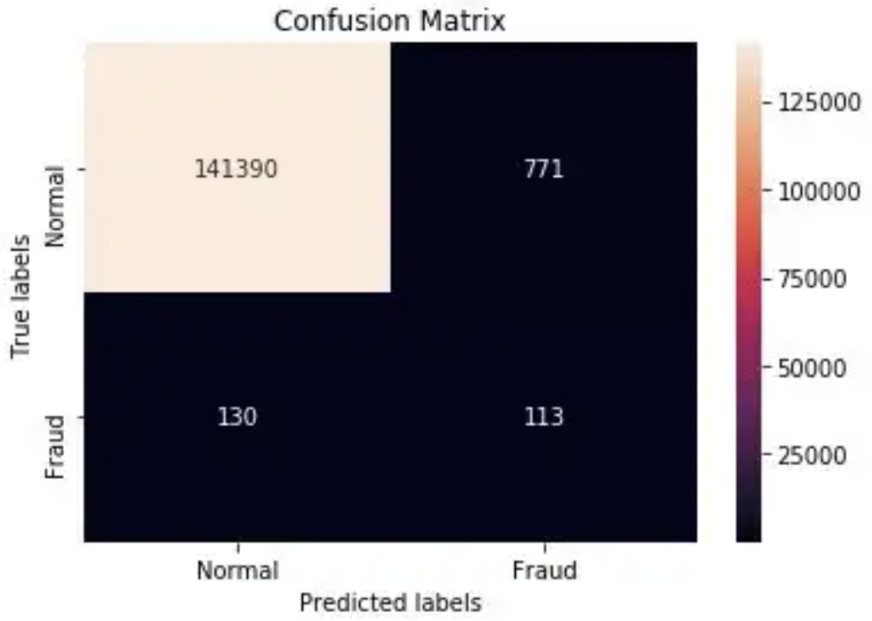
Using Reconstruction Error:

The first approach we take is to build an autoencoder and teach it just on legitimate transactions. It makes sense to assume that this autoencoder's reconstruction error will be higher for the fraudulent case than the non-fraudulent case. I learned about this method via Venelin Valkok's outstanding Medium essay. I advise you to read it because it explains the process in great detail and is simple to understand.

We'll construct a straightforward three-layer autoencoder as an example. output layer, input layer, and one hidden layer. There will be 12 neurons in the hidden layer. As previously stated, the train set's non-fraudulent examples are the only ones used to train the net. We get a network that we feed with all the train set cases after 100 epochs. The error (reconstruction error) between the input and the output can then be determined.

Based on the findings, we will label an instance in the test dataset as fraudulent if its reconstruction error exceeds three times the standard deviation from the mean, or more than 0.767519+3*3.439808=11.078922. Naturally, choosing the threshold is one of our model's hyperparameters, and in a practical application, it should be adjusted.

We can see that 113 out of 243 (46.5%) of the test dataset's fraudulent cases are caught by our algorithm. Additionally, 771 of 142161 (0.5%) non-fraudulent instances are labeled as such.

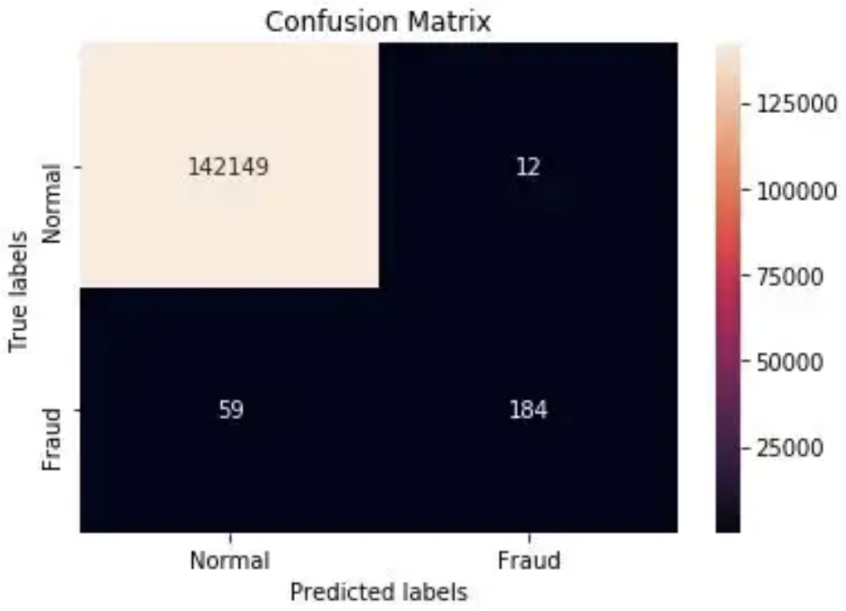


Encoder and KNN:

The encoder component of an autoencoder will be used in our second technique. The instances will be mapped into a low-dimensional space by the encoder, and the classification will be done using k-Nearest Neighbors (k-NN). The encoder will be trained using both legitimate and fraudulent transactions in this approach. One may say

that the encoder will be employed for dimension reduction, hastening the execution of the k-NN in this way.

We'll employ the same model as the first approach. The encoder portion will consist of the input layer and the inner, hidden layer with 12 neurons. All instances (both from the train and test sets) will be mapped using the encoder to a 12-dimensional space for the classification portion. The three closest neighboring examples of the train set will determine whether or not each instance in the test set is fake. The second technique recognizes 184 out of 243 (75.7%) test dataset fake cases. Additionally, 12 out of 142161 (0.008%) cases that were not fraudulent were labeled as such.



6. Experiments and Results

In our project we wanted to identify the records that are similar to a fraudulent transaction from anywhere. We are provided with very few fraudulent records and we need to train a model to identify it from any given dataset. From the above lines, it is clear that the problem in hand can be solved using autoencoders. “The aim of an auto-encoder is to learn a compressed, distributed representation (encoding) for a set of data.” And since the data we have has undergone PCA for privacy reasons, autoencoder is the best model for our situation.

One of the most important choices to make is the number of epochs. We started our model with 150 and reduced it down bit by bit until 50. This is because our model did not show any signs of reduction in loss or increase in accuracy when we set it down to 50. This significantly improved the time taken for our model to execute.

The other decision we made was to use relu as the activation functions since the accuracy was very low, almost 50% with sigmoid. Loss was very high and the model did not seem stable with sigmoid. The current network showed better convergence performance with Relu than sigmoid.”

“Also we set the batch size to 15 records, since it proved to be enough for training our model and increasing the speed of training.

7. Conclusion

Credit card fraud is definitely a dishonest criminal act. In this project, we have tried to identify fraudulent and non-fraudulent transactions. The simple deep autoencoder in Keras that has been developed can reconstruct what non-fraudulent transactions look like. The large area under the Recall vs Precision curve represents both high recall and high precision, with high precision associated with low false positive rates and high recall associated with low false negative rates. High scores for both indicate that the classifier returns correct results (high accuracy) and a majority of all positive results (high recognition).

The model can be further improved by varying the predefined threshold used to classify fraudulent and non-fraudulent transactions and by including a larger dataset. This model can also be used in day-to-day life by credit card companies to prevent fraudulent transactions.

8. References:

[1] "Credit Card Fraud Detection: A Realistic Modeling and a Novel Learning Strategy" published by IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS, VOL. 29, NO. 8, AUGUST 2018.

[2] CLIFTON PHUA1, VINCENT LEE1, KATE SMITH1 & ROSS GAYLER2 " A Comprehensive Survey of Data Mining-based Fraud Detection Research" published by School of Business Systems, Faculty of Information Technology, Monash University, Wellington Road, Clayton, Victoria 3800, Australia.

[3] David J.Wetson,David J.Hand,M Adams,Whitrow and Piotr Jusczak "Plastic Card Fraud Detection using Peer Group Analysis" Springer, Issue 2008.

[4] Building Autoencoders in Keras

[5] Autoencoders - Tutorial