ISSN: 2583-9888
DOI: doi.org/10.51219/JAIMLD/ritambhara-jha/27

URF PUBLISHERS
connect with research world

Journal of Artificial Intelligence, Machine Learning and Data Science

https://urfpublishers.com/journal/artificial-intelligence

Vol: 1 & Iss: 1

Case Report

Credit Card Fraud Detection by Implementing Supervised Machine Learning

Ritambhara Jha*

Ritambhara Jha, Senior Project Lead, Intellect Design, USA

Citation: Ritambhara Jha. Credit Card Fraud Detection by Implementing Supervised Machine Learning. *J Artif Intell Mach Learn & Data Sci 2022*, 1(1), 46-48. DOI: doi.org/10.51219/JAIMLD/ritambhara-jha/27

Received: May 06, 2022; **Accepted:** May 18, 2022; **Published:** May 20, 2022

***Corresponding author:** Ritambhara Jha, Senior Project Lead, Intellect Design, USA, E-mail: jha.ritambhara@gmail.com

ABSTRACT

Credit card fraud is a serious concern to both financial institutions and consumers. To address this issue, supervised machine learning (ML) algorithms have developed as effective methods for detecting fraudulent transactions in real time. This study digs into the deployment of supervised ML for credit card fraud detection, emphasizing the need of robust data collecting, thorough preprocessing, and smart model selection. A varied array of algorithms, including Logistic Regression, XGBoost Classifier and Convolutional Neural Network, is provided, each with distinct capabilities in discovering hidden patterns within transaction data.

Keywords: Fraud Detection, Neural Network, Supervised Machine Learning Models

1. Introduction

Credit card fraud falls into two categories. The first is physical card theft, while the second involves taking critical information from the card, such as card number, CVV code, card type, and so on. A fraudster may withdraw a big sum of money or make a major transaction by stealing credit card information before the cardholder discovers it. As a result, businesses utilize a variety of machine learning approaches to determine which transactions are fraudulent and which are legitimate. Credit card fraud (CCF) is a sort of identity theft in which someone other than the owner uses a credit card or account credentials to conduct an illegal transaction. Fraudmay occur if a credit card is stolen, lost, or counterfeited. Card-not-present fraud, or the use of your credit card number in e-commerce transactions, has also become more widespread as online shopping has grown in popularity.

The purpose of this paper is to analyze various machine learning algorithms in order to determine which algorithm is most suitable for credit card fraud detection.

2. Related Work

Increased fraud, such as CCF, has resulted from the expansion of e-banking and several online payment environments, resulting in annual losses of billions of dollars. In this era of digital payments, CCF detection has become one of the most important goals[1]. Out of about 1.4 million overall reports of identity theft in 2020, there were 393,207 incidents of CCF[2].Credit card fraud was the most prevalent sort of identity theft in 2022, with 440,666 cases. In the first three quarters of 2023, 318,087 reports of credit card fraud were filed[3].

When designing a system, the cost of fraudulent behavior and the cost of prevention should be considered. When the algorithm is exposed to new sorts of fraud patterns and routine transactions, it loses its flexibility. Because effectiveness varies depending on the issue description and its parameters, a thorough grasp of the performance measure is required[4].

Credit card data is vulnerable to skewed distribution, commonly known as class imbalance. Andrea et al. claim that their approach solves class imbalance as well as other difficulties such as idea drift and verification delay. They have also depicted the most significant performance matrix that may be employed in the identification of credit card fraud. The research also contains a formal model and a sophisticated learning method for dealing

with 'verification delay' as well as a 'warning and feedback' mechanism. Experiments have revealed that the accuracy of notifications is the most relevant measure[5].

3. Dataset

The dataset has been collected and analyzed during a research collaboration of Worldline and the Machine Learning Group (http://mlg.ulb.ac.be) of ULB (Université Libre de Bruxelles) on big data mining and fraud detection. This dataset contains 492 frauds out of 284,807 transactions that happened over the course of two days. The dataset is very uneven, with positive transactions accounting for 0.172% of all transactions.

4. Implementation

After cleaning and preprocessing the dataset, severe class imbalance is observed. As displayed in (Figure 1), dataset is heavily skewed dataset, with 99% of data classified as non-fraud and only 0.03% classified as fraud.

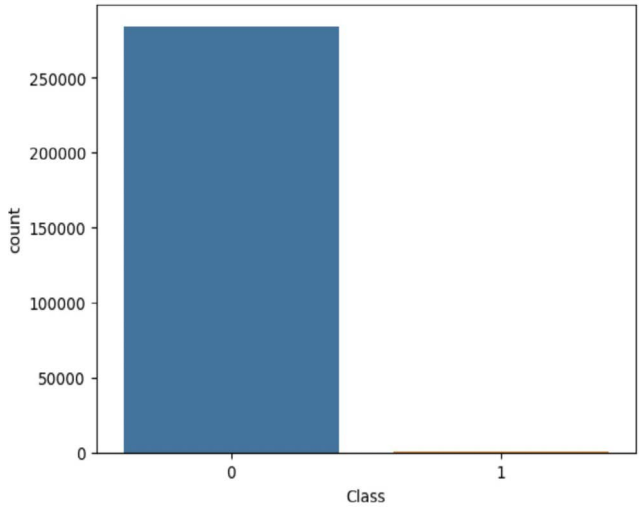

Figure 1: Class imbalance observe.

To resolve the data imbalance, an oversampling technique is performed. Here, Synthetic Minority Oversampling Technique is applied. Later, Min Max Scaling technique is applied to normalize the data in the range of 0 to 1. For each value in a feature, Min Max Scaler subtracts the minimum value in the feature and then divides by the range. The range is the difference between the original maximum and original minimum. Min Max Scaler preserves the shape of the original distribution. Min Max Scaler doesn't reduce the importance of outliers thus this was the best suited scaling technique for our dataset.

5. Model Creation & Result

In this paper, Machine Learning models created are Logistic Regression, XGBoost Classifier and Convolutional Neural Network as a deep learning model.

Logistic regression is used to describe data and the relationship between one dependent variable and one or more independent variables. Here, Logistic Regression has a large proportion of False Negatives, which can cost the bank millions of dollars. At the same time, a substantial number of false positives might have an impact on customer relationships. Logistics regression is not a suitable model.

Extreme Gradient Boosting, is a scalable, distributed gradient-boosted decision tree (GBDT) machine learning library. In this scenario, XGBoost has zero False Negatives and seems to perform better than the Logistic Regression.

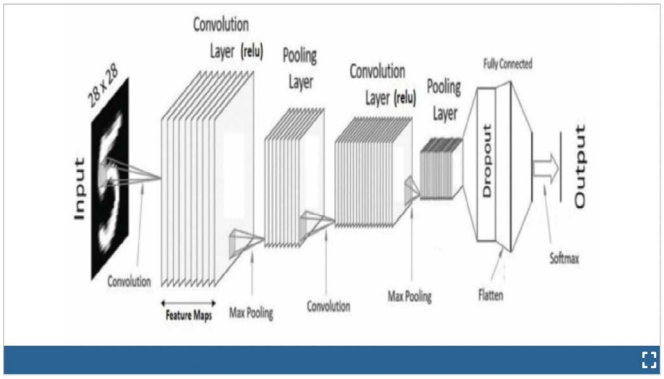

Figure 2: Convolutional neural network[1]

The CNN model proposed, consists of 14 layers, including a convolutional layer with a kernel size of 32 2 and a ReLU activation function, a batch normalization layer, and a dropout layer with a dropout rate of 0.2. Then, after a batch normalization layer and a dropout layer with a dropout rate of 0.5, we add another convolutional layer with a kernel size of 64 2 and a ReLU activation function. Then a flattening layer with a kernel size of 64 2 and a ReLU activation function is added, followed by a dense layer and a dropout layer with a dropout rate of 0.5, followed by three dense layers. The activation function of the first dense layer is 100. The activation function of the second dense is (50). The ReLU activation function of the third dense layer is (25). Finally, we include a dense classification layer with a sigmoid activation function. The accuracy at 100 epochs is 94.72% which is best suited model for our dataset.

6. Conclusion and Future Work

Implementing supervised machine learning for fraud detection has various advantages:

Improved accuracy**:** Machine learning models can outperform traditional rule-based systems in detecting more fraudulent transactions while reducing false positives.

Adaptability: Models may adapt and evolve in response to new data and emerging fraud strategies, allowing them to remain watchful against shifting threats.

Real-time detection: Real-time analysis of transactions allows for fast intervention and reduces financial losses.

Personalized profiling: A more refined approach to fraud detection is to tailor it based on individual spending habits and risk profiles.

Fraudsters are always devising new tactics of deception. A strong classifier can deal with the changing nature of fraud. A fraud detection system's top objective is accurately forecasting fraud instances and decreasing false-positive cases. Supervised ML is an effective approach for reducing credit card fraud. Financial institutions may increase their defenses against sophisticated fraudsters by exploiting its data-driven strategy, agility, and real-time analysis capabilities. Overcoming data difficulties, reducing bias, and ensuring regulatory compliance areall critical aspects in this process.

In future work, we can focus on overcoming the existing challenges such as data quality and availability, model bias, computational resources and regulatory compliance.

7. References

1. Alarfaj FK, Malik I, Khan HU, Almusallam N, Ramzan M, Ahmed M. Credit card fraud detection using state-of-the-art machine

learning and deep learning algorithms. IEEE Access, 2022;10: 39700-39715.

2. Balogun AO, Basri S, Abdulkadir SJ, Hashim AS. Performance analysis of feature selection methods in software defect prediction: A search method approach. Appl Sci, 2019;9: 2764.

3. Caporal J. Identity theft and credit card fraud statistics for 2024. Ascent, 2024.

4. West J, Bhattacharya M. An investigation on experimental issues in financial fraud mining. Procedia Comput Sci, 2016;80: 1734-1744.

5. Pozzolo AD, Boracchi G, Caelen O, Alippi C. Credit card fraud detection: A realistic modeling and a novel learning strategy. IEEE Trans Neural Networks Learn, 2017:29.