Machine Learning Techniques Used For Detection Fraudulent Transactions

ANKITA RAI
Research Scholar, Department of Computer Application, College Of Medi-caps University, Indore

Abstract- Fraud is a big problem for banks, business, and people. It can cause many harm like losing money, trust, and respect. The old ways of detecting fraud are not so good because they are slow and can miss tricky types of fraud. But machine learning can help by catching fraud right away. This paper talks about how to use machine learning to find fraud. There are different ways to use machine learning like supervised learning and unsupervised learning, decision trees, neural networks, and finding anomalies. But there are also some problems withusing machine learning for fraud, like making sure the data is good, keeping privacy safe, and doing the right thing. In this paper there are examples of how people use machine learning to stop fraud in credit cards, insurance, and medical care. These examples show that machine learning can be very good at catching fraud right away.

Indexed Terms- Fraud detection, machine learning, Supervised learning, unsupervised learning, decision tree, neural networks, anomaly detection, data privacy ethical considerations.

I. INTRODUCTION

Fraud is a big problem for banks, businesses, and people. It can make them lose a lot of money and lose trust. Old ways of finding fraud are not good because they are slow and might not catch shifty fraud. But now machines can help by finds fraud quickly. Machines can look at lots of transactions to see if any look strange. This helps banks and businesses stop fraud quickly and avoid losing money. To find fraud, machines use math and statistics. They look at lots of transaction data to see if anything is weird. This helps them learn what fraud looks like. Then machines can find fraud right away. This helps businesses and banks stop fraud before they lose too much money. With the rise of digital payments and e-commerce, the incidence of fraudulent activities in financial transactions has increased considerably. Fraudsters continuously come up with new ways to bypass the traditional rule-based systems that are used for fraud detection. To address this challenge, machine learning (ML) techniques have emerged as a promising solution ML algorithms can detect patterns that are indicative of fraudulent activities. Supervised learning algorithms, including logistic regression, decision trees, random forests, and neural networks, can be trained on historical data to classify transactions as legitimate or fraudulent.Inaddition, unsupervised learning algorithms like clustering and anomaly detection can detect fraudulent activities without the need for labeled data. ML-based fraud detection systems have several advantages over traditional rule-based systems. For example, they are more adaptable to evolving fraud tactics and can identify fraudulent activities that rule-based systems may miss. Furthermore, ML-based systems can reduce the number of false positives, which is a major challenge for rule-based systems.

- Some common machine learning techniques used for fraud detection include:

Supervised learning: Supervised learning is like teaching a computer to learn by giving it lots of examples. Just like how a teacher teaches a student by showing them many examples supervised learning works by showing the computer lots of data with the correct answers. Then, the computer can use this data to learn how to make predictions or decisions. For example if we want to teach a computer to recognize cats, we would show it lots of pictures of cats and tell it that those pictures show cats. After seeing enough example the computer can learn to recognize cats on its own. This helps the computer make predictions or decisions about new data it hasn't seen before Models can be trained on labeled data using supervised learning algorithms. In this case fraudulent

transactions are labeled as such. This model can then be used to identify similar patterns in futures trading.

Unsupervised learning: Unsupervised learning is like exploring a new city without a map or guide. You don't know what to expect, but you start to recognize patterns and make sense of what you see. In unsupervised learning, the computer is given data without any pre-existing labels or categories. It then tries to find patterns and similarities within the data on its own. This helps identify fraudulent activities that may not be immediately apparent.

Decision trees: Decision trees can be used to build models that evaluate various characteristics of transactions and determine risk scores. Transactions that receive a high risk score can be flagged for further review.

Neural Networks: Neural networks can be used to learn complex relationships between different attributes of a transaction and identify patterns that may indicate fraudulent activity.

- Machine Learning Algorithms:-

Linear Regression:
Linear regression is a method in statistics that helps us understand the relationship between two variables. It works by drawing a straight line through a set of data points to find the line that best fits the data. This line can be used to make predictions about future data points. The formula for simple linear regression is:
y = mx + b
y is the dependent variable
x is the independent variable
m is the slope of the line
b is the intercept

Logistic Regression
While logistic regression is preferred for predicting binary outcomes such as whether a customer will buy a product or not.
The formula for logistic regression is:
p = 1 / (1 + e^(-z)) Where:
p is the probability of the positive outcome
e is the mathematical constant
z is a linear combination of the independent variables:
z = b0 + b1x1 + b2x2 + ... + bnxn

Decision Tree
This algorithm is used to make a decision based on a set of rules, where each rule leads to a different outcome. Imagine you're trying to decide whether to buy a new laptop or repair your old one. A decision tree is like a roadmap that guides you through a series of choices, each one leading to a final answer. Each decision you make is like a fork in the road that takes you down a different path. As you answer more and more questions, the decision tree narrows down your options until you arrive at the best choice for your needs. This tool can be used in many fields, such as personal finance, career planning, or even dating, to help you make better decisions based on your own unique circumstances.

Support Vector Machine (SVM)
This algorithm is used for classification, regression, and detection. It works by finds the hyperplane that separates the data into different classes.If you're working on a classification problem, SVM will find the best line to separate your data into two groups. If it's a regression problem, it'll find the best line to fit your data. SVM is great for handling lots of data that might have lots of features, making it hard to tell them apart.SVM is really good at handling big, complicated datasets. It can also handle relationships between features that might not be obvious at first glance. That makes it a really useful tool for lots of things, like finding patterns in text or recognizing images.
f(x) = w^T x + b
x is the input vector
w is the weight vector that the model learns during training
b is the bias
f(x) is the predicted output for the input vector x

K-Nearest Neighbors (KNN)
This algorithm is used for classification and regression. It works by finding the k nearest neighbors to a data point and using their values to predict the value of the data point.

Step 1: Calculate the distance between the query instance and all the training instances.

KNN is a distance-based algorithm, so the first step is to calculate the distance between the query instance

and all the training instances. The most commonly used distance metric is Euclidean distance which is calculated as:
d(x, y) = sqrt(sum((xi - yi)^2))
x and y are two data instances and xi and yi are the values of the ith feature in instances x and y.
Step 2: Select the K nearest neighbors.

II. METHODOLOGY

- To find fraud gather information from various places like transaction records, user profiles, network activity logs, and other related sources. Make sure the information you get is complete and correct to the issue you are trying to solve. This is important to make sure that you have good information to identify any cheating happening.
- Once you have collected and cleaned your data, the next step is to teach your machine learning algorithm how to detect fraud using that data. You need to divide your data into two groups training set and testing set. You train the algorithm using the training set and check its performance on the testing set. This process is like teaching a child how to recognize different objects like apples, bananas, and oranges. You show the child different examples of each fruit and explain what they are. Similarly, when your algorithm correctly identifies new fraud patterns in the testing set, it means it has learned how to recognize different types of fraud.

III. FUTURE SCOPE

The future of fraud detection using machine learning (ML) techniques is promising, with potential for growth and development. ML models are becoming more complex, making it difficult to understand their decision-making process. To address this, researchers aim to develop more transparent and explainable ML models.

Data quality is crucial for the accuracy of ML-based fraud detection systems, and efforts will be made to improve data completeness, accuracy, and timeliness to reduce false positives.

Hybrid models that combine rule-based systems with ML techniques may become more common, identifying fraudulent activities more accurately and efficiently.

Real-time fraud detection systems are in high demand, and future research will focus on developing ML models that detect fraud as it occurs.

Collaboration and data sharing between organizations can lead to more effective fraud detection systems that can stay ahead of new tactics used by fraudsters. Overall, the future of ML techniques for detecting financial fraud appears promising.

CONCLUSION

ML techniques have a lot of potential when it comes to detecting fraudulent activities in financial transactions. They are able to analyze data from past transactions and identify patterns that are typical of fraudsters. This is a lot more effective than the traditional rule-based systems that can easily miss new tactics used by fraudsters.

ML algorithms can be trained to classify transactions as legitimate or fraudulent using supervised learning methods like logistic regression, decision trees, random forests, and neural networks. Alternatively unsupervised learning methods like clustering and anomaly detection can be used to detect fraudulent activities without the need for labeled data. There are a lot of advantages to using ML techniques in fraud detection. They can adapt to new tactics used by fraudsters and reduce the number of false positives. The accuracy of ML-based fraud detection systems relies heavily on the quality and relevance of the data that is used to train them. To improve the accuracy and effectiveness of these systems, researchers and practitioners need to keep developing new algorithms and techniques. As digital payments and e-commerce continue to grow it is important to have strong fraud detection systems in place to protect the security and integrity of financial transactions. It seems likely that ML techniques will play an increasingly important role in the future of fraud detection in financial transactions.

REFERENCES

[1] Randhawa, Kuldeep, et al. “Credit Card Fraud Detection Using AdaBoost and Majority Voting.” IEEE Access, vol. 6, 2018, pp. 14277–14284., doi:10.1109/access.2018.2806

[2] A. Mishra, C. Ghorpade, “Credit Card Fraud Detection on the Skewed Data Using Various Classification and Ensemble Techniques” 2018 IEEE International Students' Conference on Electrical, Electronics.

[3] Z. Kazemi, H. Zarrabi, “Using deep networks for fraud detection in the credit card transactions”, Knowledge-Based Engineering and Innovation (KBEI), 2017 IEEE 4th International Conference on pp. 630-633. IEEE.

[4] Chen, X., & Ren, Z. (2020). Application of machine learning algorithms in fraud detection of online payment systems. Journal of Electronic Commerce Research, 21(3), 262-278.

[5] Jiang, Changjun et al. “Credit Card Fraud Detection: A Novel Approach Using Aggregation Strategy and Feedback

[6] Mechanism.” IEEE Internet of Things Journal 5 (2018): 3637-3647.

[7] Randhawa, Kuldeep, et al. “Credit Card Fraud Detection Using AdaBoost and Majority Voting.” IEEE Access, vol. 6, 2018, pp. 14277–14284., doi:10.1109/access.2018.2806420