Report on Fraud Detection with Machine Learning in Marketing

Oluchi Emmanuella Obinna

ITAI-1370: Artificial Intelligence History

Prof. Machuria Johnson

October 07, 2023

Abstract

For organizations looking to protect their financial investments and uphold client confidence, fraud detection in marketing is a critical area of concern. Utilizing machine learning to detect fraud offers the chance to automate the procedure, increasing precision and effectiveness. Large-scale labeled data sets with labels indicating its salient characteristics are used to train ML-based fraud detection algorithms. Data from both valid and fraudulent transactions that have been labeled as "fraud" or "non-fraud" may be included in this. This enables algorithms to discover which patterns and relationships connect them and use such findings to classify upcoming cases (Ahramovich, 2023.)In this document, we assess the characteristics of fraud detection using machine learning in marketing and consider why the **supervised learning paradigm** is most appropriate based on important considerations.

Availability of Labeled or Unlabeled Data:

The availability of labeled or unlabeled data is one of the crucial factors to take into account while selecting an effective machine learning paradigm. Labeled data is typically available in the context of fraud detection as past records of both fraudulent and non-fraudulent activity can be documented. Because there are so many labeled data sets, supervised learning is a great option. This data may be used to train supervised learning models to categorize fresh occurrences, which makes it an efficient and effective method for marketing fraud detection.

Nature of the Problem:

Identifying whether an activity is fraudulent or not often involves binary categorization, which is the nature of the challenge in fraud detection. A binary classifier is a mysterious device that gives some input variables one of two labels. A binary classifier can be created using a variety of machine learning techniques, including decision trees, logistic regression, and support vector machines (Rul-lan,

2020). This trait fits in nicely with the supervised learning paradigm, in which models are taught to differentiate between two different classes. The major emphasis is still on classification, while regression techniques may be used when a continuous risk score needs to be predicted. The popularity of binary classification tasks in fraud detection further supports supervised learning's appropriateness.

Desire for Interpretability or Explainability:

Understanding the reasoning behind a model's predictions is crucial to detecting fraud, especially when choices have an impact on mission-critical company operations. One of the most well-liked and frequently applied algorithms among data scientists is random forest. Supervised machine learning algorithms like random forest are frequently employed in classification and regression issues. On various samples, it constructs decision trees and uses their average for classification and majority vote for regression (Sruthi, 2023). Decision tree-based models, like Random Forest, provide some interpretability by revealing the relative importance of various features and the decision-making process. They are excellent for circumstances where transparency and explainability are valued highly because of this quality. It's crucial to remember, though, that more complicated models, like neural networks, may provide more accuracy at the expense of decreased interpretability.

Need for Feedback or Interaction with the Environment:

The demand for real-time contact or input with the environment is relatively low in traditional fraud detection jobs. A model can identify fraud autonomously after being deployed and trained on previous data. While retraining or periodic model changes may be required to adapt to changing fraud tendencies, these operations normally do not necessitate real-time interaction. This quality fits with how supervised learning operates in batches.

Conclusion:

The supervised learning paradigm is clearly the most appropriate for fraud detection in marketing when taking into account the characteristics of data accessibility, problem nature, interpretability, and interaction with the environment. It effectively uses labeled data for precise classification, fits the binary character of the issue, and strikes a compromise between accuracy and interpretability. To achieve the best results in fraud detection, the organization's specific goals and the data characteristics should be taken into consideration while selecting specific algorithms and models within the supervised learning paradigm.

References

Machine Learning for Fraud Detection: Use Cases & Guidelines. (n.d.). Www.itransition.com. Retrieved October 7, 2023, from https://www.itransition.com/machine-learning/fraud-detection#:~:text=Machine%20learning%2Dbased%20fraud%20detection,recognize%20them%20once%20they%20recur.

Rul·lan, K. (2021, February 11). *Binary Classifiers Applied to Marketing*. Datasparq Technology. https://medium.com/datasparq-technology/binary-classifiers-applied-to-marketing-f0fca6d968a6

E R, S. (2021, June 17). *Random Forest | Introduction to Random Forest Algorithm*. Analytics Vidhya. https://www.analyticsvidhya.com/blog/2021/06/understanding-random-forest/