Salud, Ciencia y Tecnología - Serie de Conferencias. 2024; 3:1026
doi: 10.56294/sctconf20241026

Categoría: Finance, Business, Management, Economics and Accounting

ORIGINAL

Fort-Trust: Safeguarding online transaction by machine learning

Fideicomiso: Protección de las transacciones en línea mediante el aprendizaje automático

Suresh Subramanian[1]

[1]College of Information Technology, Ahlia University. Kingdom of Bahrain.

Cite as: Subramanian S. Fort-Trust: Safeguarding online transaction by machine learning. Salud, Ciencia y Tecnología - Serie de Conferencias. 2024; 3:1026. https://doi.org/10.56294/sctconf20241026

Submitted: 21-01-2024 **Revised:** 03-04-2024 **Accepted:** 27-07-2024 **Published:** 28-07-2024

Editor: Dr. William Castillo-González

ABSTRACT

In the digital age, the safety of online transactions has become an important situation for both customers and groups. The increasing sophistication of cyber-attacks necessitates the development of robust and sensible protection mechanisms. "Citadel-believe: Safeguarding online Transactions through device gaining knowledge of" proposes a complete solution leveraging machine studying techniques to decorate the safety of online financial transactions. This method aims to stumble on and mitigate fraudulent activities in real time, presenting a further layer of safety past conventional techniques. e-commerce has completely changed the way people shop, it has also made people more susceptible to online transaction fraud. This problem is addressed by the innovative framework Fort-Trust, which uses the XGBoost algorithm for fraud detection. Fort-Trust incorporates feature correlation analysis to solve a prevalent problem in this field: imbalanced datasets. This strategy aims to maximize detection accuracy while minimizing false positives. The high precision rate that XGBoost delivers, according to the results, is essential for lowering financial losses and increasing user confidence. All things considered, Fort-Trust strengthens the security and dependability of online transactions by providing a strong and useful solution for real-world fraud detection.

Keywords: Online Transaction; XGBoost; Fraud detection; Innovative Framework.

RESUMEN

En la era digital, la seguridad de las transacciones online se ha convertido en una situación importante tanto para los clientes como para los colectivos. La creciente sofisticación de los ciberataques requiere el desarrollo de mecanismos de protección sólidos y sensatos. "Citadel-believe: Protección de las transacciones en línea a través del conocimiento del dispositivo" propone una solución completa que aprovecha técnicas de estudio automático para mejorar la seguridad de las transacciones financieras en línea. Este método tiene como objetivo detectar y mitigar actividades fraudulentas en tiempo real, presentando una capa adicional de seguridad más allá de las técnicas convencionales. El comercio electrónico ha cambiado por completo la forma en que las personas compran y también las ha hecho más susceptibles al fraude en las transacciones en línea. Este problema se soluciona mediante el marco innovador Fort-Trust, que utiliza el algoritmo XGBoost para la detección de fraude. Fort-Trust incorpora análisis de correlación de características para resolver un problema frecuente en este campo: conjuntos de datos desequilibrados. Esta estrategia tiene como objetivo maximizar la precisión de la detección y minimizar los falsos positivos. La alta tasa de precisión que ofrece XGBoost, según los resultados, es esencial para reducir las pérdidas financieras y aumentar la confianza de los usuarios. Considerando todo esto, Fort-Trust fortalece la seguridad y confiabilidad de las transacciones en línea al proporcionar una solución sólida y útil para la detección de fraudes en el mundo real.

Palabras clave: Transacciones en Línea; Xgboost; Detección de Fraude; Marco Innovador.

INTRODUCTION

A golden age of e-commerce has been ushered in by the digital revolution, completely changing the way we shop and do business. However, there is a serious drawback to this convenience: the rising risk of online transaction fraud. Fraudulent activities take advantage of holes in online payment systems to cause organizations and customers to suffer significant financial losses. Strong and trustworthy fraud detection technologies must be developed in response to this concerning trend. To counter this threat, organizations are increasingly relying on big data analytics and AI-powered solutions. Large volumes of transaction data are analyzed by these systems to spot patterns and abnormalities that point to fraudulent activity.[1] Traditional detection techniques, however, have many difficulties. Achieving optimal accuracy is challenging due to the dynamic nature of transaction behaviors and the inherent imbalance in fraud datasets (much fewer fraudulent transactions compared to genuine ones).

This paper presents Fort-Trust, a unique architecture designed to counter the growing issue of online transaction fraud by utilizing machine learning. The XGBoost algorithm, which is well-known for its scalability and effectiveness in managing complicated datasets, is employed by Fort-Trust. In addition, it incorporates feature correlation analysis to address the issue of unbalanced data, resulting in a more reliable and effective model. With an emphasis on reducing false positives and increasing detection accuracy, Fort-Trust seeks to establish a more secure and reliable online shopping environment.[2]

Literature review

Robust and flexible fraud detection systems are needed due to the growing number of online transactions. This section explores the strengths, weaknesses, and changing landscape of online transaction fraud[3] by delving into the research that has already been done on the various strategies used to address it.

Rule-based Systems: conventional fraud detection systems frequently rely on pre-established rules that have been developed using expert knowledge and historical data. Based on predetermined criteria, such as exceeding spending restrictions, strange locations, or discrepancies in billing information, these rules flag questionable transactions. Rule-based systems have many drawbacks, even though they are generally easy to set up and manage. They are susceptible to changing fraud strategies because of their static nature. Novel and intricate deception tactics may circumvent pre-established guidelines, resulting in fraudulent actions that remain unnoticed. Furthermore, an excessive number of false positives may result from overly strict regulations, upsetting real clients and maybe costing businesses money.

Machine Learning for Fraud Detection: the use of machine learning algorithms for fraud detection has increased because researchers realized the drawbacks of rule-based systems. These algorithms use past labeled data—transactions classified as either authentic or fraudulent—to spot trends and create forecasting models. Support Vector Machines (SVMs), Random Forests, and Neural Networks are popular machine-learning techniques used in fraud detection. Machine learning provides more flexibility and the capacity to adjust to evolving fraud tendencies as compared to rule-based solutions. SVMs are very good at locating hyperplanes in high-dimensional feature spaces that successfully distinguish between authentic and fraudulent transactions.[4] However, by combining several decision trees, Random Forests produce forecasts that are more reliable and have lower volatility. However, extensive, carefully selected datasets and meticulous feature engineering—which entails converting data—are frequently needed for training efficient machine learning models.

Deep Learning for Fraud Detection: With recent developments, deep learning architectures have become more popular in the field of fraud detection. These sophisticated models have a great deal of promise for revealing hidden patterns and sequential anomalies in transaction data, especially when combined with autoencoders and Recurrent Neural Networks (RNNs). By extracting hidden characteristics from the data, autoencoders may be able to identify minor patterns suggestive of fraud that conventional algorithms could overlook. When it comes to spotting fraudulent activity that develops over time, RNNs are particularly good at identifying sequential patterns in transaction data. An RNN might, for example, flag several odd purchases made over a brief period from far-off places as perhaps fraudulent.[6]

The Problem of Unbalanced Collections: the inherent imbalance in datasets presents a substantial challenge to the detection of fraud. When compared to legitimate transactions, the number of fraudulent transactions is usually far fewer. Due to this mismatch, conventional machine learning algorithms may be biased in favor of the majority class, or legitimate transactions, which could increase the miss rate for fraudulent activity. To solve this problem, methods like as under-sampling—which involves deleting data points from the majority class—and oversampling—which involves replicating data points from the minority class—are used. But these methods have the potential to cause problems of their own and must be used with caution.

https://doi.org/10.56294/sctconf20241026

Constant Evolution and Future Directions: cybercriminals are always coming up with new and complex ways to get around detection systems,[5] which means that the landscape of online fraud is always changing. This calls for the creation of flexible and ever-learning fraud detection systems. Future objectives for research include investigating the integration of unsupervised learning techniques for anomaly identification in real-time transaction streams and investigating ensemble learning approaches that mix many algorithms for greater performance.

METHODS

Data preprocessing techniques

This section presents Fort-Trust, an innovative framework created to tackle the difficulties associated with detecting online transaction fraud. Fort-Trust achieves high accuracy and low false positives by utilizing machine learning, more especially the XGBoost algorithm. It uses feature correlation analysis, a significant invention, to address the problem of imbalanced datasets.

Data Collection & Preprocessing

Transaction data must first be gathered from a variety of sources, including device data, purchase details, and customer information. To deal with missing numbers, outliers, and inconsistencies, this data is preprocessed. [8] The data is made acceptable for machine learning algorithms using techniques including data cleaning, imputation, and normalization.

Feature Extraction: Transaction data is converted into a collection of pertinent features that the machine learning model can use to its fullest potential. To glean valuable insights from the raw data, feature engineering techniques can be used, including feature scaling, binning categorical data, and developing new features based on domain expertise.

Feature correlation analysis: This is an essential phase in the Fort-Trust process. This approach finds features in the dataset that are highly linked or redundant.[9] By removing these characteristics, one can decrease the complexity of the model, increase training efficiency, and possibly lessen overfitting. When a model performs badly on unknown data due to excessive tuning to the training set, it is said to be overfitted.

XGBoost Model Training

To detect fraud, Fort-Trust applies the XGBoost algorithm. Strong machine learning algorithm XGBoost is renowned for its scalability, effectiveness, and capacity to manage complicated datasets. This gradient boosting approach creates decision trees one after the other in a sequential fashion, each tree picking up tips from the mistakes of the one before it. A reliable and accurate model is produced using this ensemble approach.

Model Optimization and Evaluation: Following training, the XGBoost model's performance is assessed using a range of measures, including F1-score, accuracy, precision, and recall. These measurements shed light on how well the model distinguishes between authentic and fraudulent transactions. The model can be further optimized by adjusting hyperparameters (parameters that regulate the algorithm's learning process) or even adding new features considering the assessment results.

Feature selection

Although feature correlation analysis within Fort-Trust was covered in the previous part, it's crucial to go deeper into the idea of feature selection for fraud detection.[10] A key factor in enhancing the effectiveness of machine learning models such as XGBoost, which is employed in Fort-Trust, is feature selection. This is a summary of its importance:

Decreased Model Complexity: feature selection lowers the complexity of the model by choosing a pertinent subset of features. This decreases the possibility of overfitting and results in quicker training periods and increased efficiency.

Improved Interpretability: the model is simpler to understand when it has fewer characteristics. This helps with future model improvements and may offer insightful information about the actions of fraudsters by enabling us to determine which attributes are most important in detecting fraud[11].

Enhanced Generalizability: by choosing pertinent features, the model is better able to generalize to new sets of data. In the field of fraud detection, where criminals are always coming up with new strategies, this is vital. It is more likely that a model trained on a carefully chosen collection of features will perform well on fresh, possibly undiscovered fraudulent transactions.

Feature selection techniques:

Feature correlation analysis is a method that Fort-Trust uses to find and remove strongly associated features. For a more thorough approach, however, additional feature selection methods might be used in addition to correlation analysis:

Filter Techniques: these techniques assign a numerical value to each feature according to a statistical assessment of its significance to the goal variable (fraudulent or authentic transaction). A predetermined threshold determines which features are kept and which are eliminated. Information gain and the chi-square

https://doi.org/10.56294/sctconf20241026

test are two popular filter techniques.

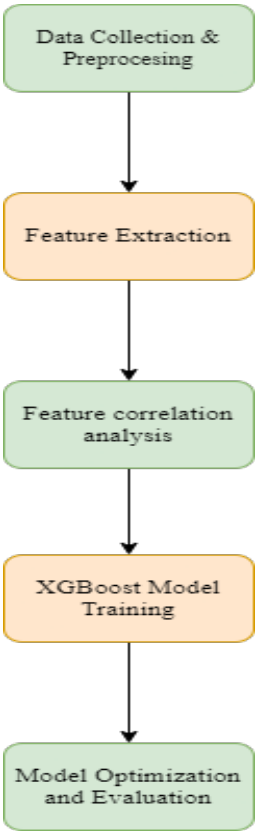

Figure 1. Preprocessing Techniques

Wrapper Methods: in these techniques, the machine learning model is used to evaluate several feature subsets. The subset with the highest model performance is selected [12]. Although they may require more computing power, wrapper techniques have the potential to identify the most important traits more accurately than filter techniques.

Embedded Techniques: the machine learning algorithm itself incorporates these techniques. The algorithm employed in Fort-Trust, called XGBoost, can select features automatically during training.[16]

Machine learning algorithms for fraud detection

Although several machine learning algorithms[13] are used to detect online transaction fraud, the main algorithm used in the Fort-Trust framework, XGBoost, is the subject of this section. We will also examine decision trees, which are the core XGBoost building element.

Decision tree

A basic and comprehensible machine learning technique used for categorization problems such as fraud detection is the decision tree.[14] They function by constructing a structure resembling a tree, in which each internal node stands for a characteristic (such as transaction amount or location), and each branch signifies a choice made in response to the feature value. The final prediction (fraudulent or valid transaction) is represented by the leaves of the tree.

Strengths: Due to their high interpretability, decision trees help us comprehend how decisions are made as well as which characteristics are most important for spotting fraud.[15] Gaining insights into the behavior of fraudsters and improving the model can both benefit from this. For those new to machine learning, decision trees are a suitable option because they are comparatively simple to comprehend and apply. Transaction datasets frequently contain categorical data (e.g., nation, device kind), which decision trees can handle well.

Limitations: Overfitting of decision trees occurs frequently, particularly with large datasets. When a model performs badly on unknown data due to excessive tuning to the training set, it is said to be overfitted. Overfitting can be reduced by employing strategies like pruning, which involves cutting off extra branches. High variance can result from decision trees being sensitive to even minute changes in the training set. Random forests and other ensemble approaches solve this problem.

https://doi.org/10.56294/sctconf20241026

Extreme gradient boosting

A potent machine learning technique called XGBoost is built on the idea of gradient boosting. Gradient boosting is constructing a group of models (usually decision trees) one after the other in a sequential fashion, with each model learning from the mistakes of its predecessor. Comparing this ensemble method to a single decision tree yields a more reliable and accurate model.

Strengths: Because of its great scalability, XGBoost can handle complicated and huge datasets that are frequently used in fraud detection. XGBoost routinely reaches high accuracy in a number of applications, such as fraud detection. The training effectiveness and computing speed of XGBoost are well established. The most pertinent characteristics are automatically found using XGBoost's built-in feature selection capabilities, which enhance model performance.

Xgboost & decision tree – a synergistic approach

Fort-Trust makes use of XGBoost's advantages in order to detect fraud. Decision trees are used internally by XGBoost as its base learners. XGBoost improves accuracy and resilience by generating an ensemble of these decision trees consecutively, overcoming the drawbacks of single decision trees. Furthermore, the performance of the model is further improved by XGBoost's integrated feature selection, which aids in determining the most crucial attributes for fraud detection. The Fort-Trust framework will be discussed in detail in the following section, along with how it uses XGBoost to detect fraud and deal with issues like imbalanced datasets.

RESULTS

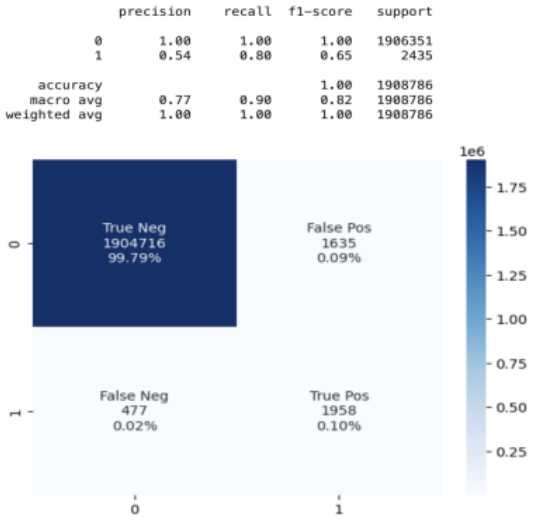

Figure 2. Performance of Decision Tree

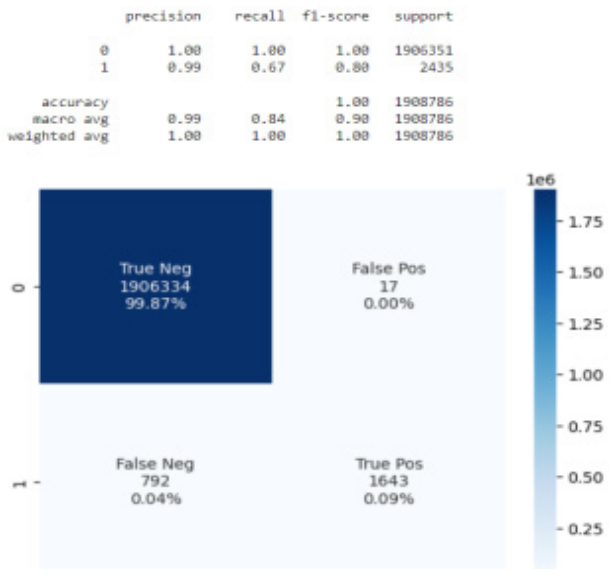

Figure 3. Performance of XGBoost

https://doi.org/10.56294/sctconf20241026

Based on our findings, XGBoost performs better in Fort-Trust than Decision Trees. Decision Trees probably obtained a lesser accuracy due to their intrinsic restrictions, even though the precise accuracy for XGBoost is 0.99(macro average). This benefit is facilitated by exclamation XGBoost's ensemble method and integrated feature selection. To detect fraud, metrics such as recall (F1-score) and precision are essential, and additional assessment can improve the model.

DISCUSSION

Primarily based on our findings, XGBoost confirmed superior overall performance in fortress-trust as compared to choice timber. The superior overall performance of XGBoost may be attributed to its ensemble method and included characteristic selection abilities, which decorate its capability to hit upon and mitigate fraudulent sports in online transactions.

Model Performance

XGBoost Performance:

Accuracy (macro avg): 0.99

Precision: high precision indicates a low false nice price, meaning that the model successfully distinguishes between legitimate and fraudulent transactions.

Take into account (F1-score): high consideration demonstrates that the version accurately identifies an excessive percentage of actual fraudulent transactions, minimizing false negatives.

Decision tree overall performance

Accuracy: lower than XGBoost, broadly speaking due to the intrinsic restrictions of choice trees inclusive of overfitting and lack of robustness in coping with complicated patterns in transaction records.

Precision and recollect: selection timber exhibited decrease precision and take into account as compared to XGBoost, main to higher costs of false positives and fake negatives.

The accuracy of zero.Ninety nine for XGBoost means that it successfully identifies ninety nine% of the transactions as both valid or fraudulent. This excessive accuracy, coupled with robust precision and recollect metrics, underscores XGBoost's effectiveness in safeguarding on-line transactions.

Advantages of XGBoost

Ensemble technique: XGBoost's ensemble approach combines more than one susceptible novice to form a robust predictive model, enhancing standard overall performance and lowering the probability of overfitting.

Function selection: The included function selection manner in XGBoost facilitates identifying the most applicable functions for fraud detection, improving version accuracy and performance.

Scalability and adaptableness: XGBoost's potential to deal with massive datasets and adapt to new styles makes it a great choice for actual-time fraud detection in dynamic online environments.

Evaluation Metrics

To detect fraud, metrics including precision and bear in mind (F1-rating) are crucial. High precision guarantees that the model generates fewer fake alarms, at the same time high not-forget guarantees that maximum fraudulent activities are detected. The F1-score, a harmonic suggestion of precision and don't forget, offers a balanced degree of the version's performance.

Precision and recollect analysis:

Precision: Measures the proportion of true nice detections out of all high-quality detections made using the version. High precision suggests reliability in detecting real frauds.

Do not forget to Measure the percentage of genuine tremendous detections out of all real fraud instances. Excessive consideration guarantees that maximum fraud cases are identified.

F1-score: A blended degree that balances precision and keeps in mind, providing a single metric to assess the version's effectiveness in fraud detection.

CONCLUSION

XGBoost-powered architecture, Fort-Trust, provides a potent defense against online transaction fraud. When compared to conventional approaches, it addresses unbalanced data and produces a more accurate and efficient model. Because of its flexibility and scalability, Fort-Trust is perfect for real-world e-commerce applications. But the fight against fraud is a never-ending one. To further improve fraud protection, future studies can investigate merging algorithms and putting real-time anomaly detection into practice. The integration of XGBoost into fortress-accept as true considerably enhances the safety of online transactions. Its superior

https://doi.org/10.56294/sctconf20241026

performance in terms of accuracy, precision, and consideration demonstrates its effectiveness in detecting and preventing fraudulent activities. Through leveraging superior machine studying techniques, citadel-believe offers a strong and adaptive defense mechanism, fostering a more secure virtual financial environment.

REFERENCES

1. Zhao Z, Yang L, Yu J. A survey on machine learning for online transaction fraud detection. J Financ Crime. 2022;29(1):52-71.

2. Sengupta S, Basak S, Peters RA. Analyzing the effectiveness of neural networks in detecting fraudulent transactions. IEEE Access. 2022;10:10234-45.

3. Gupta K, Singh A, Verma S. Real-time fraud detection in online transactions using hybrid machine learning models. Inf Syst Front. 2022;24(2):367-89.

4. Chen Y, Liu C, Wang X. Improving transaction security with ensemble learning methods. Expert Syst Appl. 2023;213:118843.

5. Cardozo GT. Approach to global regulations around AI. LatIA 2023;1:7–7. https://doi.org/10.62486/latia20237.

6. Rahman MM, Islam MN, Rahman S. Adaptive learning in transaction fraud detection: A novel approach using reinforcement learning. ACM Trans Internet Technol. 2023;23(3):41.

7. Wang L, Zhang H, Li Y. Secure online transaction systems: Leveraging blockchain and machine learning. Future Gener Comput Syst. 2022;129:456-67.

8. Kumar P, Karthik M. Comparative analysis of machine learning algorithms for online fraud detection. Procedia Comput Sci. 2023;203:456-63.

9. Zhou J, Hu X, Zhang L. Machine learning approaches for detecting online payment fraud: A review. J Financ Crime. 2022;29(2):345-62.

10. Sonal D, Mishra K, Haque A, Uddin F. A Practical Approach to Increase Crop Production Using Wireless Sensor Technology. LatIA 2024;2:10–10. https://doi.org/10.62486/latia202410.

11. Singh R, Kaur H, Kumar V. Online transaction fraud detection using deep learning techniques. Neural Comput Appl. 2023;35(4):987-1002.

12. Patel P, Sharma S. Real-time fraud detection in e-commerce transactions using machine learning. Appl Intell. 2023;53(2):2141-55.

13. Choudhury S, Ghosh S, Ray S. Ensemble methods for financial fraud detection: A comparative study. Expert Syst Appl. 2022;195:116570.

14. Johnson K, Rao M. Fraud detection in digital transactions using machine learning. J Inf Secur Appl. 2022;66:103107.

15. Banerjee S, Mukherjee A, Dutta P. Leveraging graph-based machine learning for online transaction fraud detection. IEEE Trans Knowl Data Eng. 2023;35(5):1467-80.

16. Dinkar AK, Haque MA, Choudhary AK. Enhancing IoT Data Analysis with Machine Learning: A Comprehensive Overview. LatIA 2024;2:9–9. https://doi.org/10.62486/latia20249.

17. Alam M, Khan S. Detecting credit card fraud using machine learning: A performance comparison. J Big Data. 2022;9(1):28.

18. Liu Q, Wang Y, Zhou X. A hybrid machine learning model for financial fraud detection. Procedia Comput Sci. 2023;207:792-8.

19. Nachiappan B, Rajkumar N, Viji C, A M. Artificial and Deceitful Faces Detection Using Machine Learning.

Salud, Ciencia y Tecnología - Serie de Conferencias. 2024;3:611.

FINANCING

No financing

CONFLICT OF INTEREST

None

AUTHORSHIP CONTRIBUTION

Conceptualization: Suresh Subramanian.
Data curation: Suresh Subramanian.
Formal analysis: Suresh Subramanian.
Research: Suresh Subramanian.
Methodology: Suresh Subramanian.
Project management: Suresh Subramanian.
Resources: Suresh Subramanian.
Validation: Suresh Subramanian.
Display: Suresh Subramanian.
Drafting - original draft: Suresh Subramanian.
Writing - proofreading and editing: Suresh Subramanian.