ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XIV, Issue VIII, August 2025

Adaptive Illicit Transaction Identification in Online Marketplaces via Computational Intelligence

Shaheena K V, Anila R Nambiar, Shanu Kumar

Dept. of MCA, Acharya Institute of Technology, Bangalore, India

DOI: https://doi.org/10.51583/IJLTEMAS.2025.1408000106

Abstract: Electronic trade crime has been a significant setback to the internet economy with the exploitation of weaknesses in the online business exchange. This research study evaluates the development of fraud detection methodology in online retailing and how contemporary automated learning systems have Classic exchange mechanisms based on rule-based. It offers a clear explanation of the general applicability of the methodology, both participant installation categorizer and Xgboost. This enhances the discoverability. This paper examines contemporary literature, covers the symptoms of relative performance, and covers problems, data, and research prospects. The findings explore the opportunities to adjust, evolve, and real-time ML tools in relation to internet fraud. Additionally, the paper has elucidated on the significance of the synthetic and realistic datasets in training the models reliably and there is a need to present a more explanatory solution to be implemented in real e-commerce systems. With the help of this review, this article is going to arbitrate future advances into the intellectual framework of fraud detection.

Keywords: E-Commerce, XGBoost,Electronic trade crime, Online retailing, Internet fraud detection,Machine learning, E-commerce security

I. Introduction

The evolution towards online involvement has pushed the use of e-commerce. This means that there will be a frictionless access to the market. The digital trading convenience can be offset, however, by the advanced scamming of digital fraud techniques, such as data theft, account takeover, and synthetic crimes. [1] looked at the data. Juniper Global Research Projects' online payment fraud exceeded $362 billion from 2022 to 2027.

Older forms of fraud detection systems are rule-based. This is first combat successfully to keep track of the changing aspects of fraud [6]. Auto learning (ML) is also emerging to be utilized by modern fraud detection framework since it supports the analysis of complex models and real-time detection of anomalies [7].

Of all the ML approaches, though, the overall teaching approaches including stacking and Xgboost nature of teaching shows outstanding results on the problem of fraud detection based on classification [15]. These methods are better than individual models and carry out the results of many simple students to give better generalization and reliability.

This review employs automated learning to evaluate the present state of the detection of the fraud in e-commerce, examines the general efficiency of the system and its performance, and inquires about the future where adaptive fraud detection is established.

II. Literature Survey

Earlier fraud alert systems were pre-defined rule and limit-based. Nevertheless, due to emerging fraudulent schemes, such systems have considerable weaknesses that revolve around flexibility and precision. Monteis et al. [1] were in need of immediate changes in digital security, akin to the post-country Creve of cybercrime, as well as a boost in the activities on the Internet. Kodate and Chiba [2] employed a graphic model to investigate the use of fraud detection on the basis of networks. This has been effective in not being fraudulent in decision-makings other than the manipulated user data. In our categorization of 49 studies of financial fraud, we observed that the most concentrated areas were decisions, Weisov networks and credit card fraud as based on neural networks.

Bolton and Hand [7] called attention to the necessity of the actual adaptive systems that are capable of combating class imbalances, large-scale demands, and scarce mark information. Banerjee et al. [15] made a comparative analysis of the algorithms used in fraud detection and depicted that the general model usually performs better than individual classifiers both in terms of accuracy and stability.

The presence of gaps in real-world development is still present, mostly after the description, scalability, and processing of artificial data sets.

Machine Learning Techniques

Within the context of e-commerce fraudiped modelling overall modelling models represent a high level of efficiency because they are able to process typically problematic unbalanced data whilst at the same time producing thin, non-linear fraud models. This section presents the most fundamental techniques of an auto piloted learning employed in the suggested system: the fitting classifier and XGBoost classifier, and the learning set of data applied to teach the model.

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XIV, Issue VIII, August 2025

Stacking Classifier

A means of learning about metanemicle is called packaging or generalized generalization. Metanemicle includes the predictions of a few classifiers at the basic level, including the logistical regression, choices, and support of vector machine and integrates them to the metal learning that conducts the conclusive classification [17]. This multi-layered structure allows the model to use the unique power and limitations of each basic student's solution, increasing global performance and generalization. Installation works particularly well when individual students enjoy even more predictability even when their students do not cope with their own.

XGBoost Classifier

Extreme Gradient (xgboost) is a highly specialized general decision- based algorithm that produces the gradient. It should be quick and speedy and has some rules that one can use to obtain awesome gatherings [16]. Xgboost tolerates noisy parallel data and permits you to omit values. Detection of rare fraud, when it occurs in the case of Xgboost fraud detection, is highly beneficial in obtaining rare fraudulent blows, and it is very useful in keeping a very good classification accuracy since it is applicable when working with an important and unbalanced set of data.

Synthetic Data Generation

Datasets were produced using the Fython Faker library mainly as a result of the inability to get access to real financial transactions data, based on the existing confidentiality, and data sensitivity regulations. The dataset has 23,634 observations with 16 variables, such as transaction volume, customer identifiers, and payment method and fraud, which is a binary variable. User logic was applied to users and simulation of realistic behaviours of users and fraud models. The difference between the synthetic data and the actual was that the former made it easier to design and test models, as they were provided in controlled conditions, and were representative, and therefore reliable, so the movement and testing phases could be reduced.

Table I: Performance Comparison of Fraud Detection Techniques

Technique	Train Accuracy	Test Accuracy	Dataset Type	Strengths	Technique	Train Accuracy
Stacking Classifier	**100%**	**99%**	**Synthetic**	**Combines multiple base learners; high generalization**	**Stacking Classifier**	**100%**
XGBoost Classifier	**96%**	**95%**	**Synthetic**	**Fast, scalable; effective for structured/tabular data**	**XGBoost Classifier**	**96%**
Random Forest	**~90%**	**~85%**	**Real / imbalanced [6]**	**Interpretable; resistant to overfitting in small data**	**Random Forest**	**~90%**
Artificial Neural Net (ANN)	**~93%**	**~88%**	**Real / imbalanced [4]**	**Learns complex non-linear patterns**	**Artificial Neural Net (ANN)**	**~93%**

Comparative Analysis

Comparative studies to evaluate the effectiveness of the teaching technique in the identification of fraud were performed based on the XgBoost Classifier Classifiers and broadly used models like Random Forest Networks and Artificial Neurons (ANNs). Productivity can be determined by the training and testing accuracies, nature of data, and usability as indicated in Table I.

III. Proposed Methodology

The described solution to the issue of fraud detection will encompass the latest instruction methods and a set of artificial data, still, within the framework of detecting illicit operations on e-commerce sites, the solution is feasible. The system is designed in the manner that it delivers great amount of accuracy in the terms of classification as well as real time deployment and ease of deployment in assorted form in the terms of the digital media.

System Architecture

The application of its use can be implemented based on a network of balloons between fronts of working tasks using HTML, CSS, JavaScript and Python on the basis of BACAND, which pre-processes the data, models output, and analyses the performance of the data. The architecture takes the form of several modules.

It has an interface entry or offering static authentication. Datasets Module: It allows the user to import transactions in the form of CSV.

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XIV, Issue VIII, August 2025

Modules prediction: Training of general practice of models escalade and xgboost-classification. Initial presentation a raw and categorical data set is presented.

The Performance Analysis Module: Reports the measures such as the accuracy and accuracy, review, indicator F1. Visualization module: It provides graphic ideas (cakes and histograms) to disperse fraud and performance models.

This provides flexibility of scalability, maintenance and accessibility of user.

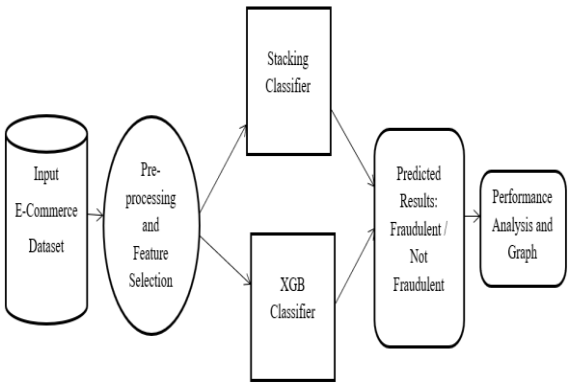


Figure: 1 System Architecture

Data Generation and preprocessing

The inaccessibility of data from real-world data generated a set of synthetic data containing 23,634 transaction records using Python's Faker library. Each observation has got 16 variables such as transaction identifier, customer identifier, transaction volume, payment method and binary labels in terms of fraud. The use of logic to simulate a balanced and realistic distribution of fraud and legal transactions is employed.

Portion of data processing instructions are contained Standardization of continuous fields of digital data. Encoding of categorical variables in digital format. Handling of zero values to maintain integrity. The choice of functions is performed on the analysis of domain relevance and correlation.

It is through these steps that it becomes assured that the model is trained using high quality and representative data and this enhances generalization and reliability.

Machine Learning Workflow

Model Training:

There were 2 general classifiers used:-

- **Complex Classifier**: Combines predictions of a number of simple models (eg. Logistic regression, solutions, etc.).
- **XGboost classifier**: Behave a gradient escalator making the use of regulation, parallelism, and trimming with assurance of evolutionary and swift learning.

Model Evaluation:

The model has been built using 70 percent of the data and was tested using the rest of the 30 percent. The evaluation measurements were test training and accuracy, confusion matrix, accuracy, inspection and indicator F1. The results that are observed are as follows:

• Installation Classifier:

Education correctness = 100 per cent,

Precision test = 99 per cent

xgboost classifier:

Educational accuracy = 96%,

Test accuracy = 95%

These findings indicate that the models are very productive with minimum rethinking and a strong detection capacity with the class imbalances.

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XIV, Issue VIII, August 2025

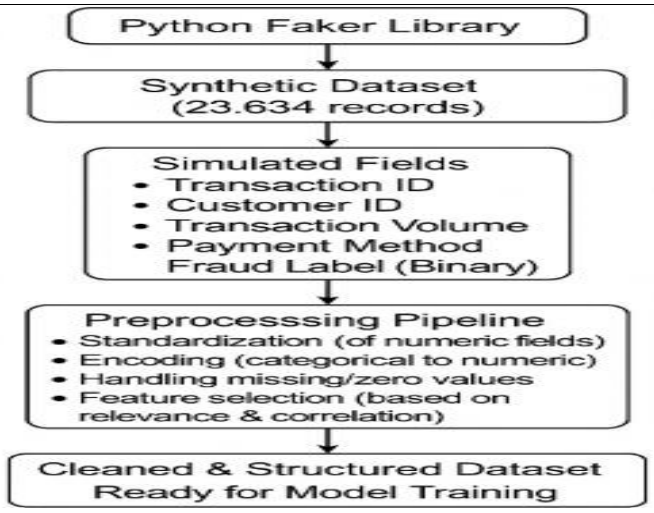


Figure 2: Data Generation and Pre processing Workflow

Model Deployment:

The model created was serialized with Joblib and into the Supported Bacand to receive real output. The Frontend interface also enables a user to download new datasets and see predictions.

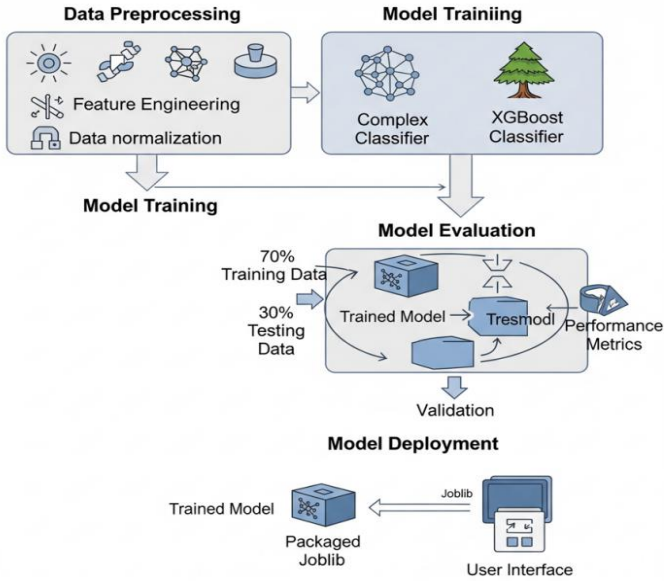


Figure 3: Proposed Machine Learning Workflow

Advantages of the Proposed Method

The suggested treatment has a number of significant benefits.

High discretion: A test accuracy of up to 99 per cent of accuracy has been achieved.

Real time fraud prediction: Gives real time transactions classification. The User Interface is to make interactions easier by the use of Web Visualization.

Evolutionary and customizable Design: Fits to different contexts of trade E. Economic Implementation: Completed libraries and open source code.

Challenges and Future Scope

Current Limitations

Synthetic data set: This can be realistic though not necessarily an adequate representation of real challenges and contradictory behaviour.

Risk of Experience: Learning Accuracy is 100% and includes rethinking. There is a need to have a more generalized data set restricted fraud cases: The use of identified flights and synthetic accounts is surely not simulated.

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XIV, Issue VIII, August 2025

Future Research Directions

Applying real data: Consolidation of anonymous sector data (such as Kaggle, Paysim).

Description (XAI): An interpretation layer of stakeholders to trust the ML solution should be added. Hybrid Model: If you combine the graphics and rules based overall.

Height and Actual Treatment: Enforcement of detection at the edge devices of low latency applications

IV. Conclusion

The primary motive of even the more complex fraud programs is to conduct fraud activities through e-commerce websites. Hence, we should introduce a smarter, scalable and real detection system. Such a review is a part of the evolution of the methods of detecting fraud focusing explicitly on the styles and auto-learning ensembles like XGBoost. Such models work superior to normal algorithms. Specifically, they work with unbalanced data, engage in sophisticated frauds, and offer many classification accessories. Using comparative analysis, integration, and discussion of the final methodology of literature, using general appearance methods is a good foundation on the modern system of detecting fraud in a fairly reliable manner. Nevertheless, setbacks, like unavailability of data, explanation of the models, and the dynamic factor of the fraud cannot be overlooked as crucial barriers that should be overcome.

Future output should incorporate real transaction of data to describe downward systems and real adaptive training system systems. These issues can be converted into useful tools that can be used to enhance the security and reliability of the online market through automated solutions that are guided by training and that will identify fraud that resolves the problems.

References

1. S. Monteith, M. Bauer, M. Aida, J. Geddes, P. C. Whybrow, and T. Glenn, "Increasing cybercrime since the pandemic: Concerns for psychiatry," *Curr. Psychiatry Rep.*, vol. 23, no. 4, p. 18, 2021.
2. S. Kodate, R. Chiba, S. Kimura, and N. Masuda, "Detecting problematic transactions in a consumer-to-consumer e-commerce network," *Appl. Netw. Sci.*, vol. 5, no. 1, p. 90, 2020.
3. R. Samani and G. Davis, "McAfee mobile threat report," 2019. [Online]. Available: https://www.mcafee.com
4. E. W. T. Ngai, Y. Hu, Y. H. Wong, Y. Chen, and X. Sun, "The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature," *Decis. Support Syst.*, vol. 50, no. 3, pp. 559–569, 2011.
5. S. Smith and Juniper Research, "Online payment fraud: Market forecasts, emerging threats & segment analysis 2022–2027," 2024. [Online]. Available: https://www.juniperresearch.com/press/losses-online-payment-fraud-exceed-362-billion/
6. A. Abdallah, M. A. Maarof, and A. Zainal, "Fraud detection system: A survey," *J. Netw. Comput. Appl.*, vol. 68, pp. 90–113, 2016.
7. R. J. Bolton and D. J. Hand, "Statistical fraud detection: A review," *Stat. Sci.*, vol. 17, no. 3, pp. 235–255, 2002.
8. C. Phua, V. Lee, K. Smith, and R. Gayler, "A comprehensive survey of data mining-based fraud detection research," *arXiv preprint arXiv:1009.6119*, 2010.
9. L. Akoglu, H. Tong, and D. Koutra, "Graph-based anomaly detection and description: A survey," *Data Min. Knowl. Discov.*, vol. 29, no. 3, pp. 626–688, 2015.
10. D. Irani, S. Webb, and C. Pu, "Study of static classification of social spam profiles in MySpace," in *Proc. Int. AAAI Conf. Web Soc. Media*, vol. 4, no. 1, pp. 82–89, 2010.
11. A. Bhowmick and S. M. Hazarika, "Machine learning for e-mail spam filtering: Review, techniques and trends," *arXiv preprint arXiv:1606.01042*, 2016.
12. D. Savage, X. Zhang, X. Yu, P. Chou, and Q. Wang, "Anomaly detection in online social networks," *Soc. Netw.*, vol. 39, pp. 62–70, 2014.
13. Nambiar, Anila Raveendran, Ranjitha K. M., and Shaheena K. V."An Intelligent Deep Learning System for Identifying Bird Species." Convergence of Machine Learning and IoT for Enabling the Future of Intelligent Systems, Qt Analytics Publication, 2024, pp. 60–74.
14. Shaheena K. V., and Dhanalakshmi S. "Evaluating Image Segmentation Techniques: A Comparative Approach." ShodhKosh: Journal of Visual and Performing Arts, vol. 5, no. 6, 2024, pp. 854–863. Granthaalayah Publications.
15. Gangadhar, Ch, Nama Ajay Nagendra, Syed Mutahar Aaqib, C. M. Sulaikha, Shaheena K. V., and Karanam Santoshachandra Rao."COVID-19 Image Transmission Using Convolutional Neural Networks Based Algorithms for Medical Applications." World Journal of Engineering, vol. 19, no. 2, 2022, pp. 183 188. Emerald Publishing Limited.
16. Maheswari, G., and Shaheena K. V.
"MRI Brain Tumor Segmentation Algorithms and Approaches — A Survey." International Journal of Latest Engineering and Management Research (IJLEMR), 2017. [Cited by 2].
17. Guhan, T., Margaret Mary T., and Shaheena K. V.From Algorithms to Intelligence: Navigating the Frontiers of AI, Security, and Modern Computing. Qt Analytics Publication, 2025, pp. 1–128.

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XIV, Issue VIII, August 2025

18. A. Ali et al., "Financial fraud detection based on machine learning: A systematic literature review," *Appl. Sci.*, vol. 12, no. 19, p. 9637, 2022.
19. T. Pourhabibi, K. L. Ong, B. H. Kam, and Y. L. Boo, "Fraud detection: A systematic literature review of graph-based anomaly detection approaches," *Decis. Support Syst.*, vol. 133, p. 113303, 2020.
20. R. Banerjee, G. Bourla, S. Chen, M. Kashyap, and S. Purohit, "Comparative analysis of machine learning algorithms through credit card fraud detection," in *Proc. IEEE MIT Undergraduate Research Technology Conf.*, Cambridge, MA, USA, 2018, pp. 1–4.
21. N. Carneiro, G. Figueira, and M. Costa, "A data mining-based system for credit-card fraud detection in e-tail," *Decis. Support Syst.*, vol. 95, pp. 91–101, 2017.
22. A. O. Adewumi and A. A. Akinyelu, "A survey of machine-learning and nature-inspired based credit card fraud detection techniques," *Int. J. Syst. Assur. Eng. Manag.*, vol. 8, no. S2, pp. 937–953, 2017.
23. M. Ahmed, A. N. Mahmood, and M. R. Islam, "A survey of anomaly detection techniques in the financial domain," *Future Gener. Comput. Syst.*, vol. 55, pp. 278–288, 2016.
24. V. Rodrigues, L. Policarpo, and D. E. da Silveira, "Fraud detection and prevention in e-commerce: A systematic literature review," *J. Retail. Consum. Serv.*, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1567422322000904
25. J. West and M. Bhattacharya, "Intelligent financial fraud detection: A comprehensive review," *Comput. Secur.*, vol. 57, pp. 47–66, 2016.