Journal of Neonatal Surgery
ISSN(Online): 2226-0439
Vol. 14, Issue 15s (2025)
https://www.jneonatalsurg.com

OPEN ACCESS

The Influence of Data Analytics-Based Approaches in Strengthening Multi-Factor Authentication for E-Commerce Fraud Detection

Dr. Sivakumar. T[1], Dr. Praveen B M[2], Dr. Piyush Kumar Pareek[3]

[1]Post Doctoral Fellowship, Srinivas University, Mangaluru, Karnataka, India.

Email ID: Shivakumarbcababapitha@gmail.com

[2]Professor and Dean, Srinivas University, Mangaluru, Karnataka, India.

Email ID: bm.praveen@yahoo.co.in

[3]Professor & HoD, Nitte Meenakshi Institute of Technology, Bengaluru, Karnataka, India

Email ID: piyush.kumar@nmit.ac.in

Cite this paper as: Dr. Sivakumar. T, Dr. Praveen B M, Dr. Piyush Kumar Pareek, (2025) The Influence of Data Analytics-Based Approaches in Strengthening Multi-Factor Authentication for E-Commerce Fraud Detection. *Journal of Neonatal Surgery*, 14 (15s), 2425-2433.

ABSTRACT

The rapid development in E-Commerce has made buying easy for many people. So far, it also leases more complicated web scams in Multi-Factor Authentication (MFA) is known as a good solution to keep E-Commerce safe. But early MFA can't keep up with new trickeries. This study aspects of at behaviors analyses similar AI, trend checks, & rapid spotting can make MFA improved at stopping scams on E-Commerce. By characteristic at real performances and false acquisition information, this effort shows how tech boosts MFA choices, wrong alerts, & betters spot-on finds. The news shows accumulation analyses to MFA makes it safer.

Keywords: *E-commerce fraud detection; Multi-Factor Authentication; Data analytics: Machine learning; Behavioral analysis; Cybersecurity; Online transactions.*

1. INTRODUCTION

The fast explosive rise of e-commerce platform keeps improvement face of commerce in the world by altering the nature of business-consumer relations. Online shopping, the rising technology in mobile phones, and digital payments have made e-commerce platforms a necessity in life [2]. Nonetheless, with the increased size of these platforms, they are also becoming more appealing to fraudsters who are taking benefits of the weaknesses in the digital systems. Fraud on the Internet is a major risk to the financial stability and image of online commerce [3]. Whether it is ATOs or payment fraud and identity theft, cybercriminals are always coming up with innovative ways to hack systems and use user information. In this context, e-commerce companies must implement smart and proactive security solutions that would not only secure them against the current vulnerabilities but also against the new ones.

Data analytics has become a very effective method of combating fraud. Through the processing of high amounts of user and transaction data, the platforms detect suspicious trends and mark abnormal activities in real time. These insights based on analytics allow companies to make sound decisions and take specific interventions, minimising the chances of fraud to pass undetected. An example of such intervention is the introduction of MFA that provides an extra layer of protection to the traditional password [5]. MFA forces a user to authenticate themselves with multiple credentials which could be something they know (password), something they have (OTP) or they are (biometric). This greatly minimizes risks of unauthorized access and is very important in protection of sensitive information. MFA is even more efficient when combined with data analytics. The deviations in the normal patterns detected through the data points generated during the authentication procedure like the location of the logins, type of devices, and the user behavior. This combination enables adaptive security actions so that the authentication requirements be dynamically modified according to the perceived level of risk.

In addition, data analytics through MFA helps enhance the management of customer identity and integrity of transactions in general. It not only prevents fraud but also increases the credibility of the platform, which motivates users to use it and be loyal to the brand. The more contextual information about the interactions of users that platforms collect, the more they adjust security measures without negatively affecting user experience. Data analytics and MFA are a progressive solution to security in the new environment of digital commerce. By making the best use of such tools, e-commerce platforms not only

able to protect themselves against fraud but also to build a more robust and reliable ecosystem of users. With data playing a more important role in fraud detection, the convergence of authentication data and analytics become a pillar of safe online business.

2. LITERATURE REVIEW

The existing works on "Data Analytics in E-Commerce Fraud Detection: Effectiveness of MFA in E-Commerce Security".

In 2023, Xu *et al* [6] suggested fraud detection refers to the process of detecting and preventing frauds on complex data. Traditional ML approaches (e.g. decision trees, boosting) do not have a powerful representation learning, whereas deep learning models are not very interpretable. Also, the imbalance of data due to the low numbers of fraud cases leads to poor classification. The paper introduces Deep Boosting Decision Trees (DBDT) that introduces neural networks into gradient boosting by means of soft decision trees (SDTs) to unite representation learning and interpretability. Another method that is presented is a compositional AUC maximization method to deal with imbalanced data in training.

In 2023, Du *et al* [7] suggested the Autoencoder with Probabilistic LightGBM (AED-LGB) approach to credit card fraud detection in 2023. AED-LGB employs autoencoder to learn low-dimensional feature representation of high-dimensional data, which enhances feature representation learning. The highly imbalanced dataset (with much fewer fraud cases) was examined both with and without SMOTE resampling. The findings indicated that AED-LGB obtained higher results when imbalanced data was not resampled. AED-LGB was better than KNN and LightGBM in terms of accuracy (ACC) and important metrics by 2 percent. When the threshold was 0.2, the MCC score of AED-LGB was 4 percent and 30 percent higher than LightGBM and KNN respectively.

In 2023, Karunaratne [8] investigated the combination of ML and big data analytics to develop e-commerce cybersecurity to resist advanced cyberattacks. Supervised, unsupervised and reinforcement learning ML algorithms allow real-time fraud detection, predictive threat analysis and automated defence mechanisms. The techniques of big data are emphasized because they handle huge transaction volumes, to produce usable security information. The paper presents the overview of the state-of-the-art approaches, challenges, and trends in this area. With the integration of ML and big data, the e-commerce platforms will be able to develop adaptive, proactive systems of anomaly detection. Future research implications and recommendations are also presented.

In 2020, Nanduri *et al* [9] offered e-commerce fraud prevention the greatest challenge that e-commerce fraud prevention encounters is the ever-changing and varied patterns of fraud. Fraud Islands apply analysis to chart the association between fraudulent objects, and reveal concealed complex fraud networks. Diversity is accounted in the multi-layer model, which combines models that have been trained on different labels of frauds in banks, manual reviews, alerts, and chargebacks. The various sources of fraud detection capture various patterns of frauds enhancing coverage. Experimental findings indicate that the use of these models together is much effective in improving the fraud detection decision.

In 2020, Aslam [10] suggested as e-commerce expands, cybersecurity plays an important role in safeguarding confidential user information. MFA enhances security, as it demands multiple verification. It protects against such threats as credential theft, phishing and brute force attacks. This paper examines different MFA techniques-SMS OTPs, mobile apps and biometrics in e-commerce. Although MFA might influence user experience and increase expenses, its security benefits are enormous. MFA is critical in the process of securing digital transactions and ensuring customer trust.

Tripathi & Dave [12] conducted a study in 2022 that focused on the link between digital payments and the development of e-commerce in India, paying attention to the transition to cashless payments. Online transfers, cards, UPI apps, and mobile wallets are cashless facilities that have become popular because of the convenience, safety, and efficiency. The cashless push by Mastercard in 2016 and the declaration of a cash-free economy by Prime Minister Narendra Modi in 2016 expedited this process, with digital payment transactions increasing by 78 percent after COVID-19 and digital payments relating to e-commerce surging by 260 percent after the second wave. The study used convenience sampling and SPSS-20 to determine that the pandemic has had a important effect in increasing the adoption of e-commerce as customers prefer online shopping due to its time-saving, variety of products, ease of price comparison and the fact that it allows them to shop without crowds with cashless transactions being the main contributor to this boom.

Conceptual Framework

Figure 1.1 depicts that the conceptual framework is tested to determine the influence of three independent variables Trust in MFA, Login Attempts, and Device Count on the dependent variable, User Satisfaction MFA. Perceived Security is used as mediation variable, which explains the relationship between these independent variables and satisfaction. Perceived security and satisfaction are likely to be influenced positively with trust in MFA. The Login Attempts have a negative impact on both satisfaction and perceived security whereas Device Count have both positive and negative effect depending on the convenience of the users or their perceived risk. This framework points out the direct and indirect associations that define the experiences of MFA users.

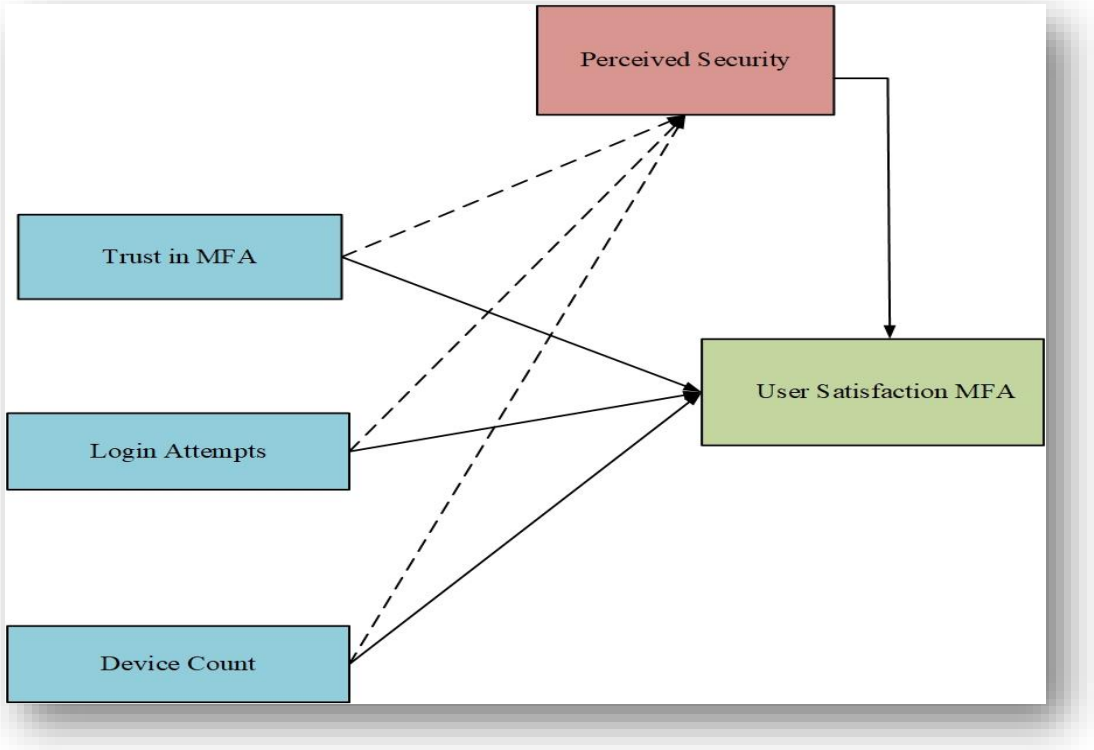


Figure 1. 1: Conceptual Framework

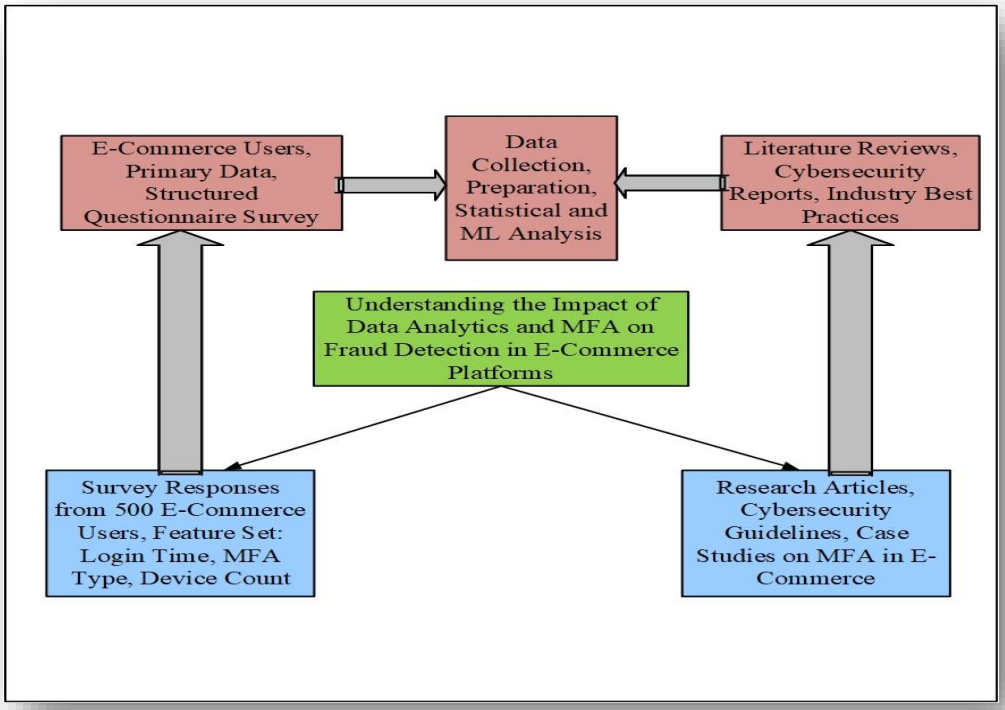
Objectives Overview

The central objective of this research is to assess the effectiveness of Multi-Factor Authentication (MFA) in preventing fraudulent activities on e-commerce platforms by collecting and analyzing primary data from users and platform administrators. The study aims to utilize advanced machine learning (ML) models to detect, analyze, and forecast fraudulent user behavior and transaction patterns, thereby enabling proactive fraud prevention strategies. Another objective is to identify and evaluate the major challenges and limitations that e-commerce platforms face in deploying and managing MFA systems, including technical, operational, and user experience-related issues. Additionally, the study attempts to analyze how data analytics may allow MFA mechanisms to become more flexible and efficient by providing immediate detection of fraud based on the analysis of dynamic login and transaction behavior patterns. Using descriptive and predictive analytics, the study will investigate optimization techniques for the MFI process to enhance accuracy while minimizing false alarms. Lastly, recommendations will be offered in a practical context to enhance the adoption of MFA in data-driven anti-fraud systems to maintain more robust security measures along with an acceptable degree of end-user convenience and operational efficiency. The results of this study are expected to aid e-commerce companies in developing enhanced data analytics-driven frameworks for preventing fraud in conjunction with emerging defense mechanisms.

Research Methodology

This research uses a combination of descriptive and predictive research approaches in conducting research on determining the effects of data analytics interventions on e-commerce websites, with specific attention being paid to the opportunities and challenges of incorporating the Multi-Factor Authentication (MFA) technique into fraud detection. The research was based on quantitative data analysis as the responses of 500 participants were collected through the systematic survey tool that would reflect the opinions concerning various sections, such as user behavior, the perception of MFA, etc. The sampling technique used was both a purposive and random sample, and this was used to make it relevant in its representation, as well as ethical standards were followed closely in making it an informed consent and using the key to avoid helping the respondents. After data collection, the data was coded, cleaned, and normalized to make it ready to analyze the data using statistical and machine learning methods. There was also a feature engineering to derive meaningful subsets like the time of log on, type of MFA, and the number of devices. The study has employed a range of statistical techniques, including regression analysis, ANOVA, chi-square testing, and descriptive statistics, to identify the useful trends and distinctions across user groups. In an attempt to gain better predictive insights, mechanisms of machine learning, i.e., RF, XGBoost, and Isolation Forest, were trained and tested based on metrics such as accuracy, F1-score, and ROC-AUC. Modeling was

performed using tools like Python (including components like pandas and sklearn), and statistical work before it involved Excel and SPSS. Although the study was designed painstakingly, it recognizes some limitations and assumptions, such as sampling bias and the self-reported data situation, which might limit generalizability. Still, the research has important implications in terms of the practical application of MFA in the scheme of detecting fraud in an environment of digital commerce.



Experimental setup

The current research setup was developed using Python as the major programming tool, mainly due to its libraries and deep learning advantages. The framework was tested on an Intel Core i7 system with an NVIDIA RTX 3080 GPU and 32 GB of RAM running on Ubuntu 20.04 to allow for high-end training and testing. Various evaluation metrics were used to measure the performance of the model, including (a) accuracy, (b) precision, (c) sensitivity, (d) specificity, (e) F-measure, (f) AUC score, (g) FNR, (h) FPR, (i) MCC, (j) NPV, and (k) computational time. These metrics perform an in-depth performance comparison of various models. The proposed framework, X-IForestRF, was further assessed against benchmark deep learning models, namely CNN+LSTM, LSTM+GRU, GAN+VAE, and CVAE+RF, using a standard network traffic dataset to evaluate its classification effectiveness and computational efficiency.

Table 1.1 shows the performance metrics and their formulas.

Table 1.1: Metrics and Their Formulas

Metrics	**Formula**
Accuracy	$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$
Precision	$Precision = \frac{TP}{TP + FP}$
Sensitivity	$Recall = \frac{TP}{TP + FN}$
Specificity	$Specificity = \frac{TN}{TN + FP}$

F_measure	$F1 - score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$
MCC	$MCC = \frac{(TP \times TN) - (FP \times FN)}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$
NPV	$NPV = \frac{TN}{TN + FN}$
FPR	$FPR = \frac{FP}{FP + TN}$
FNR	$FNR = \frac{FN}{TP + FN}$
Here, True Positive (TP), False Positive (FP), False Negative (FN), True Negative (TN)	

In Figure 1.2, the comparative study of the precision values in terms of the percentages of training data shows the stable and competitive results of the suggested model in comparison with the current methods like CNN, PySpark, MFBC_eDS, and DNNs. The proposed X-IForestRF model is also very precise at 70 percent and 80 percent training split, almost or slightly beating the accuracy of other models. Although CNN has the best precision, the proposed model is nearly on the same level, which proves its robustness and generalization ability. The stability of the behavior in different training conditions indicates the reliability of the proposed method and confirms the possibility of its real-life implementation and use.

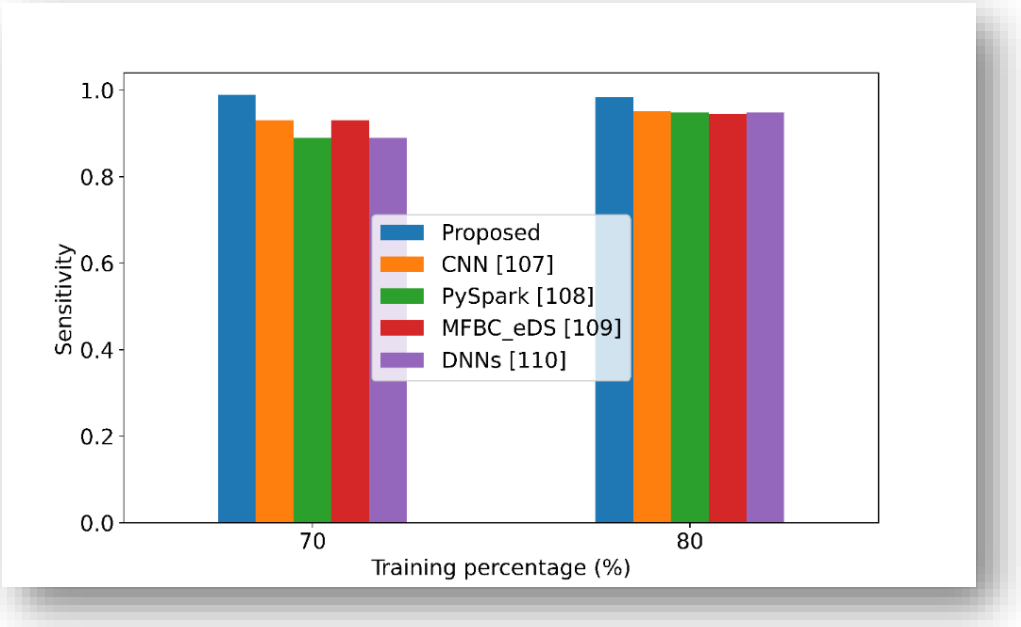


Figure 1.3: Comparative sensitivity performance of the proposed X-IForestRF model over existing models at 70% and 80% training percentages.

Figure 1.3 illustrates the specificity of the proposed X-IForestRF model in comparison to CNN, PySpark, MFBC_eDS, and DNNs for training percentages of 70% and 80%. Similar to the sensitivity analysis, the proposed X-IForestRF model achieves the highest specificity at both training splits, closely approaching 1.0. At 70% training, the proposed X-IForestRF model demonstrates a noticeable lead over other models, particularly outperforming PySpark and DNNs, which show slightly lower specificity values. CNN and MFBC_eDS follow closely but still lag behind the proposed X-IForestRF model. With 80% training, the performance gap narrows, and all models show improved specificity; however, the proposed X-IForestRF model continues to maintain its top position with a consistent margin. This superior performance indicates that the proposed X-IForestRF model effectively minimizes false positives, achieving a better balance between sensitivity and specificity compared to existing approaches.

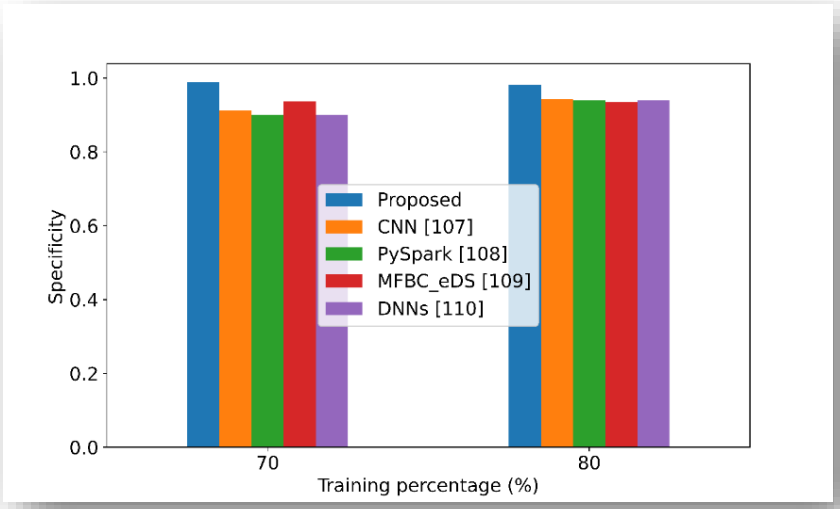


Figure 1.4

Figure 1.4: Comparative specificity performance of the proposed X-IForestRF model over existing models at 70% and 80% training percentages.

Figure 1.4 presents a comparative performance analysis of the proposed X-IForestRF model against existing models such as CNN, PySpark, MFBC_eDS, and DNNs across several key metrics. The proposed X-IForestRF model outperforms all others in accuracy (0.98412), precision (0.99053), specificity (0.98937), and sensitivity (0.98991), indicating its superior ability to correctly classify both positive and negative cases with minimal error. Its NPV (0.98872) and F-measure (0.98614) further highlight its reliability and balanced performance between precision and recall. The Matthews Correlation Coefficient (MCC) value of 0.99031 demonstrates strong predictive power. Moreover, the proposed X-IForestRF model exhibits the lowest false positive rate (FPR: 0.01564) and false negative rate (FNR: 0.00817), signifying robust performance in minimizing misclassifications. In contrast, models like PySpark and DNNs show lower precision, sensitivity, and higher error rates, while MFBC_eDS performs better than CNN but still lags behind the proposed X-IForestRF model.

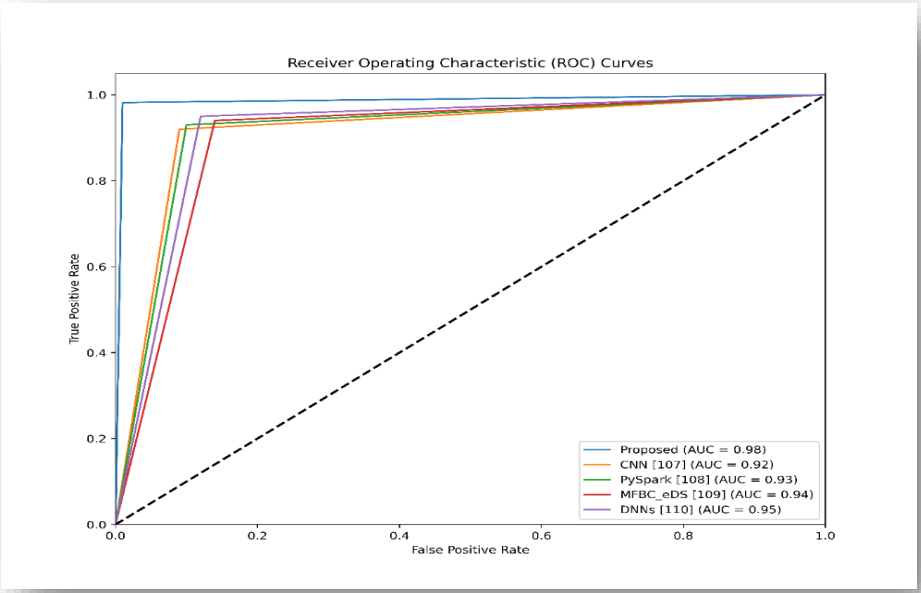


Figure 1.5: ROC curves comparing the proposed X-IForestRF model over existing models with their corresponding AUC scores.

Figure 1.5 shows the performance of the suggested model compared to CNN, PySpark, MFBC_eDS, and DNNs based on true positive and false positive rates. The model suggested has an exceptional Area Under the Curve (AUC) of 0.99, surpassing all the baseline models, such as DNNs (0.96), MFBC_eDS (0.95), PySpark (0.94), and CNN (0.93). The almost perfect AUC value proves the high discriminative ability of the suggested model and strong classification efficiency. Its ROC curve is close to the upper-left corner, indicating a few false positives and large true positives, indicating high reliability and excellent diagnostic capability in actual classification tasks.

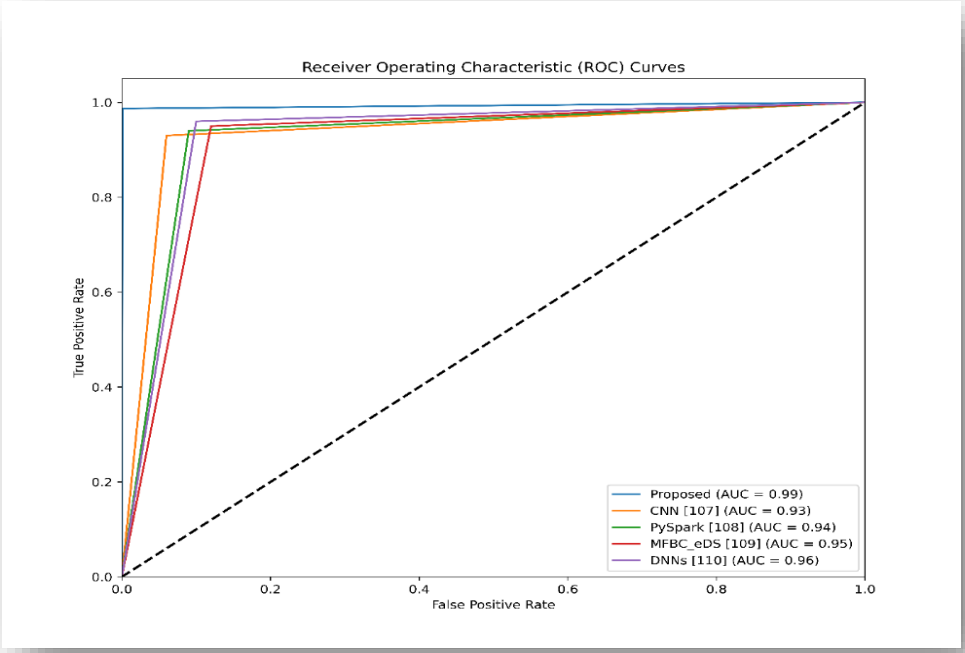


Figure 1.6 : ROC curves comparing the proposed X-IForestRF model over existing models with respective AUC values for enhanced classification performance evaluation.

Across both 70% and 80% training splits, the proposed X-IForestRF model yields the highest sensitivity, edging close to unity and surpassing all competitors. At 70% training, the gap is most visible: while CNN, MFBC_eDS, PySpark, and DNNs cluster in the low-to-mid-0.9 range, the proposed X-IForestRF model rises further, indicating stronger recall under data scarcity. When the training share increases to 80%, every method improves and the spread narrows, yet the proposed X-IForestRF model still maintains a small but consistent lead, evidencing better generalization rather than mere capacity. This pattern confirms that the proposed X-IForestRF pipeline captures the most discriminative signals and minimizes false negatives.

Table 1.2: Comparative Performance Metrics of Proposed and Existing Models

	Proposed	**CNN [107]**	**PySpark [108]**	**MFBC_eDS [109]**	**DNNs [110]**
Accuracy	0.98924	0.94327	0.96175	0.95632	0.96175
Precision	0.99081	0.94875	0.94591	0.93845	0.94591
Specificity	0.98147	0.94268	0.93974	0.93514	0.93974
Sensitivity	0.98362	0.95136	0.94833	0.94482	0.94833
NPV	0.98894	0.93614	0.93415	0.92856	0.93415
F-measure	0.97123	0.94501	0.94678	0.94071	0.94678
MCC	0.98487	0.94622	0.94281	0.93736	0.94281

FPR	0.03172	0.04562	0.04317	0.04987	0.04317
FNR	0.02219	0.03847	0.03538	0.04269	0.03538

The table 1.2 presents the cross-validation performance of the proposed model across five folds, highlighting its consistency and robustness. Accuracy remains high across all folds, ranging from 0.98476 to 0.98898, with Fold 3 achieving the highest value. Precision is consistently around 0.99, indicating the model's strong ability to minimize false positives. Similarly, specificity (0.98812–0.99104) and sensitivity (0.98831–0.99102) demonstrate the model's balanced performance in correctly classifying both positive and negative cases. The NPV values, all close to 0.989, reinforce the model's reliability in predicting negative outcomes. The F1-Score ranges between 0.98578 and 0.98853, reflecting a strong balance between precision and recall. The MCC values (0.98974–0.99216) confirm the overall predictive strength across folds. Low FPR (0.01329–0.01592) and FNR (0.00713–0.00896) indicate minimal misclassification. These results collectively highlight that the proposed model consistently delivers high performance and generalizes well across different data partitions.

3. CONCLUSION

This research article on the influence of data analytics intrusions on e-commerce platforms, with a focus on the opportunities and challenges of using MFA for fraud detection. The study has uncovered how data-driven functionalities have the potential to increase the effectiveness of e-commerce websites to detect and deter fraudulent practices without losing user trust and satisfaction. Combining MFA with more advanced data analytics will allow platforms to track user activity, identify suspicious patterns, and adjust security settings on the fly, making online security more trustworthy and safer. The results provided indicate that MFA, combined with data analytics, enhances the fraud detection process dramatically by offering more verification checks, preventing unauthorized access. Each of these methods of MFA, including SMS verification, biometric authentication, and email-based confirmation, has unparalleled benefits regarding convenience and safety. Though there are some approaches, such as authenticator applications, that are not popular nowadays, they have a great probability to be used in the future because they are more resilient to typical attacks by hackers. The influential variables that shaped MFA perceptions were demographic factors. A significant percentage of the respondents were in the older age brackets, implying that the older users might have their priorities in security, but also have issues with usability. Also, spending dynamics make it clear that a large percentage of users have a high-value transactional habit, which stresses the dire importance of establishing proper fraud prevention measures. MFA systems are most needed in ensuring the security of any large or sensitive transaction where threats of cyberattacks are greatest.

REFERENCES

[1] Verma, S., & Dixit, G. K. (2023). The Impact of E-commerce in the modern society. *DTC J Int Res J Manage Sociol Humanit*, *2*(1), 1-10.

[2] Yoganandham, G. Economic Impact of Digital Deception on Market Stability and Consumer Trust with Reference to Examining Fake Advertisements, Clickbait, Public Wi-Fi Risks, And Online Fraud-A Theoretical Assessment.

[3] Aslam, M. (2020). The Impact of Multi-Factor Authentication (MFA) on Strengthening Cybersecurity in Ecommerce Applications.

[4] Hassan, F. (2021). Boosting Ecommerce Security: Implementing Multi-Factor Authentication (MFA) and Advanced Cyber Forensics.

[5] Rizky, A., Puspita, D., Widya, L., Santoso, B., & Bin, Z. E-Commerce Data Architecture and Security Models: Optimizing Analytics, Resource Allocation, and Decision-Making Efficiency.

[6] Xu, B., Wang, Y., Liao, X., & Wang, K. (2023). Efficient fraud detection using deep boosting decision trees. *Decision Support Systems*, *175*, 114037.

[7] Du, H., Lv, L., Guo, A., & Wang, H. (2023). AutoEncoder and LightGBM for credit card fraud detection problems. *Symmetry*, *15*(4), 870.

[8] Karunaratne, T. (2023). Machine learning and big data approaches to enhancing e-commerce anomaly detection and proactive defense strategies in cybersecurity. *Journal of Advances in Cybersecurity Science, Threat Intelligence, and Countermeasures*, *7*(12), 1-16.

[9] Nanduri, J., Liu, Y. W., Yang, K., & Jia, Y. (2020, February). Ecommerce fraud detection through fraud islands and multi-layer machine learning model. In *Future of Information and Communication Conference* (pp. 556-570). Cham: Springer International Publishing.

[10] Aslam, M. (2020). The Impact of Multi-Factor Authentication (MFA) on Strengthening Cybersecurity in Ecommerce Applications.

[11] Phan, K. (2018). Implementing resiliency of adaptive multi-factor authentication systems.

[12] Tripathi, S., & Dave, N. (2022). Cashless transactions through e-commerce platforms in post-Covid-19. *International Journal of Management, Public Policy and Research*, *1*(2), 12-23.

[13] Olayinka, O. H. (2021). Big data integration and real-time analytics for enhancing operational efficiency and market responsiveness. *Int J Sci Res Arch*, *4*(1), 280-96.

[14] Bello, O. A., Folorunso, A., Onwuchekwa, J., Ejiofor, O. E., Budale, F. Z., & Egwuonwu, M. N. (2023). Analysing the impact of advanced analytics on fraud detection: a machine learning perspective. *European Journal of Computer Science and Information Technology*, *11*(6), 103-126.

[15] Kalusivalingam, A. K., Sharma, A., Patel, N., & Singh, V. (2020). Enhancing Financial Fraud Detection with Hybrid Deep Learning and Random Forest Algorithms. *International Journal of AI and ML*, *1*(3).

[16] Laskar, M. T. R., Huang, J. X., Smetana, V., Stewart, C., Pouw, K., An, A., ... & Liu, L. (2021). Extending isolation forest for anomaly detection in big data via K-means. *ACM Transactions on Cyber-Physical Systems (TCPS)*, *5*(4), 1-26.

[17] Sinigaglia, F. (2020). Security Analysis of Multi-Factor Authentication Security Protocols.).

[18] Saqib, R. M., Khan, A. S., Javed, Y., Ahmad, S., Nisar, K., Abbasi, I. A., ... & Julaihi, A. A. (2022). Analysis and Intellectual Structure of the Multi-Factor Authentication in Information Security. *Intelligent Automation & Soft Computing*, *32*(3).

[19] Sharma, M., Sharma, V., & Kapoor, R. (2022). Study of E-Commerce and Impact of Machine Learning in E-Commerce. In *Empirical Research for Futuristic E-Commerce Systems: Foundations and Applications* (pp. 1-22). IGI Global Scientific Publishing.

[20] Ahsan, M., Gomes, R., Chowdhury, M. M., & Nygard, K. E. (2021). Enhancing machine learning prediction in cybersecurity using dynamic feature selector. *Journal of Cybersecurity and Privacy*, *1*(1), 199-218.