International Journal of Computer Applications Technology and Research
Volume 12–Issue 08, 126 - 145, 2023, ISSN:-2319–8656
DOI:10.7753/IJCATR1208.1016

Investigating Potential Applications of Machine Learning Techniques for Personal Data Protection in E-commerce

George N. Wainaina
School of Science Engineering and Technology
Kabarak University
Box: 20157
Nakuru, Kenya

Ruth Oginga
School of Science Engineering and Technology
Kabarak University
Box: 20157
Nakuru, Kenya

Denis K. Kiyeng
School of Science Engineering and Technology
Kabarak University
Box: 20157
Nakuru, Kenya

Abstract:

The increasing demand of e-commerce and online retail services in the past years has required development of strong security protocols able of protecting the privacy and safety of consumers. Machine learning has proven an essential tool in the analysis and reaction to cybersecurity threats, thus ensuring the security and privacy of users of online retail stores all over the globe.

In this sense, machine learning protocols have proven essential in the prevention of cyberattacks like cross site scripting (XSS), in the identification of potential threats affecting the financial results and overall performance of a company, or the surveyance of the behavior of the individuals within an online community to identify excessively authoritative and potentially harmful members within different communities. The present paper analyzes the state-of-the-art advances in this field, providing an overview of the different methods that have proven their validity to develop machine learning-based models to ensure the security and privacy of users of online retail stores. The proposed machine learning methods use validation protocols such as NLP, SVM-based algorithms, neural networks, and text mining. Based on such analysis, the paper suggests potential approaches to a better implementation of machine learning in cybersecurity, enabling a faster and more effective response against the threat.

Keywords:*Machine Learning, Artificial Intelligence, Personal Data Protection, E-commerce, Privacy Preservation, Data Security.*

1. Introduction

E-commerce has become crucial for conducting trade and releasing private information in today's globally linked society. In light of the growing privacy and security concerns, it is essential to look at the use of machine learning algorithms for protecting sensitive information during online transactions. Machine learning and AI methods may secure users' private information on e-commerce platforms. Some of the most important uses of these technologies are as follows: Machine learning algorithms are used in anomaly detection or identifying unusual patterns or behaviors that may indicate fraud or infiltration (Wang, et al., 2021). User Authentication; machine learning might improve user authentication procedures, ensuring that only authorized people can access critical data. Businesses may employ encryption and privacy preservation measures to safeguard sensitive data while it is kept or sent.

According to Yathiraju (2022), user data may be aggregated and

anonymized using methods such as differential privacy, protecting users' privacy while allowing relevant insights to be derived. Fraud Detection; machine learning algorithms that have been taught to detect suspicious financial or online behavior might be used to protect personal information. Privacy Policy Compliance; machine learning methods might be used to monitor and validate privacy laws and policies. NLP algorithms may scan privacy and service agreement terms to identify inconsistencies and security vulnerabilities that might disclose sensitive data. Customers of e-commerce websites may be certain that their personal information is being handled lawfully (Shaba et al., 2022). Threat Intelligence and Predictive Analysis: Large datasets, including known threats and security weaknesses, may be used to train machine learning models for threat intelligence and prediction analysis.

Machine learning algorithms may be applied in several ways to secure the personal information of online purchasers. Some major machine-learning applications are as follows: E-commerce platforms may be required to share client data with third-party partners or service providers for logistics, payment processing, and tailored suggestions, among other things. Federated learning is an example of how users may train models locally on their devices rather than uploading raw data to a centralized server (Shaba et al., 2022).

Customers may obtain fine-grained privacy settings matched to their needs and interests using machine learning. Machine learning models may advise users on proper privacy practices by assessing their actions, preferences, and prior permission. Electronic commerce systems may use machine learning methods to provide intelligent risk assessment. Advanced Fraud Detection: Machine learning methods can greatly improve e-commerce fraud detection systems (Habbat et al., 2022). These models may evolve and change over time, detecting new fraud trends and guarding against new risks as they arise. Machine learning algorithms may avert data breaches and compromised accounts by evaluating user behavior for unusual tendencies. Constant risk assessment Machine learning algorithms might constantly monitor e-commerce platforms for security flaws and data breaches. This research investigated how machine learning and AI may be used to safeguard user data on e-commerce platforms.

1.1 Research Problem

Privacy and data security concerns have grown in response to the exponential growth of online commerce, which has expanded the collection, storage, and processing of personally identifiable information. Traditional data protection processes may need to be more in light of the emerging hazards in the online retail business according to Wang, et al., 2021). As a result, it is critical to investigate how e-commerce platforms may utilize machine learning to protect their consumers' data better.

1.2 Purpose of Research

This study aimed to discover how machine learning methods may be utilized to safeguard client data in the

online retail business. This paper analyzed how machine learning and AI currently secure users' private data. Furthermore, people want to learn more about and get feedback on how machine-learning approaches might be utilized to increase the security of users' data when doing online transactions.

1.3 Study Objectives

I. Examine the current scenario regarding personal data security on e-commerce platforms using machine learning and artificial intelligence.
II. Investigate if machine learning methods can better safeguard sensitive customer data during online transactions.
III. To suggest how e-commerce businesses might use machine learning methods to secure customer data better.

1.4 Research Gap

A rising corpus of literature examines the connections between e-commerce, machine learning, and the security of personal data. However, few studies concentrated on finding particular machine learning methods that might be used to strengthen personal data protection in the e-commerce industry. Numerous previous studies either focused on individual data security issues without considering the broader e-commerce data security landscape or describe generic techniques without providing specific tactics (Vinoth et al., 2022). This paper aimed to close this gap by offering a thorough analysis of prospective machine learning applications in e-commerce for the security of personal data. For instance, future studies should concentrate on performing thorough empirical research in developing nations like Kenya, considering the distinct socio-cultural, technological, and regulatory issues. This would help fill in these knowledge gaps. The research should offer helpful tips and recommendations for organizations, especially SMEs, on embracing and using ML and AI technology.

2. Literature Review

The explosive growth of e-commerce in recent years has fundamentally changed how we transact business and share personal data. However, this remarkable growth has also led to severe worries about data security and privacy in the online retail industry. Therefore, it is imperative to investigate cutting-edge methods to safeguard sensitive data during online transactions. In order to overcome these issues and improve the protection of personal data in e-commerce, machine learning techniques have emerged as a possible option (Policarpo et al., 2021). The literature review section critically assesses the current work on machine learning techniques for personal data protection in e-commerce. The review is divided into three parts: Machine Learning and Its Role in Data Protection, E-commerce, and Personal Data Security, and Machine Learning Applications in E-commerce for Personal Data Protection.

2.1 E-commerce and Personal Data Protection

How business is conducted and connected with internet platforms has

International Journal of Computer Applications Technology and Research
Volume 12–Issue 08, 126 - 145, 2023, ISSN:-2319–8656
DOI:10.7753/IJCATR1208.1016

been entirely transformed by e-commerce. The security of personal data has become a significant concern due to the rising reliance on digital transactions and the sharing of personal information. This section focuses on how e-commerce and personal data security intersect, exploring the difficulties and methods of safeguarding sensitive data. The gathering, storing, and processing of personal data has grown exponentially due to the rapid expansion of e-commerce. Regarding user information security and privacy, this has caused severe worries. Misuse of personal data can result in fraud, identity theft, and other nasty things. Personal data security is essential to preserve user confidence and trust in e-commerce platforms (Cheng et al., 2023). According to the study by Sean Michael Kerner. (2023), the crime complaint center reported the top 10 data breaches in history, as indicated in Figure 1 below.

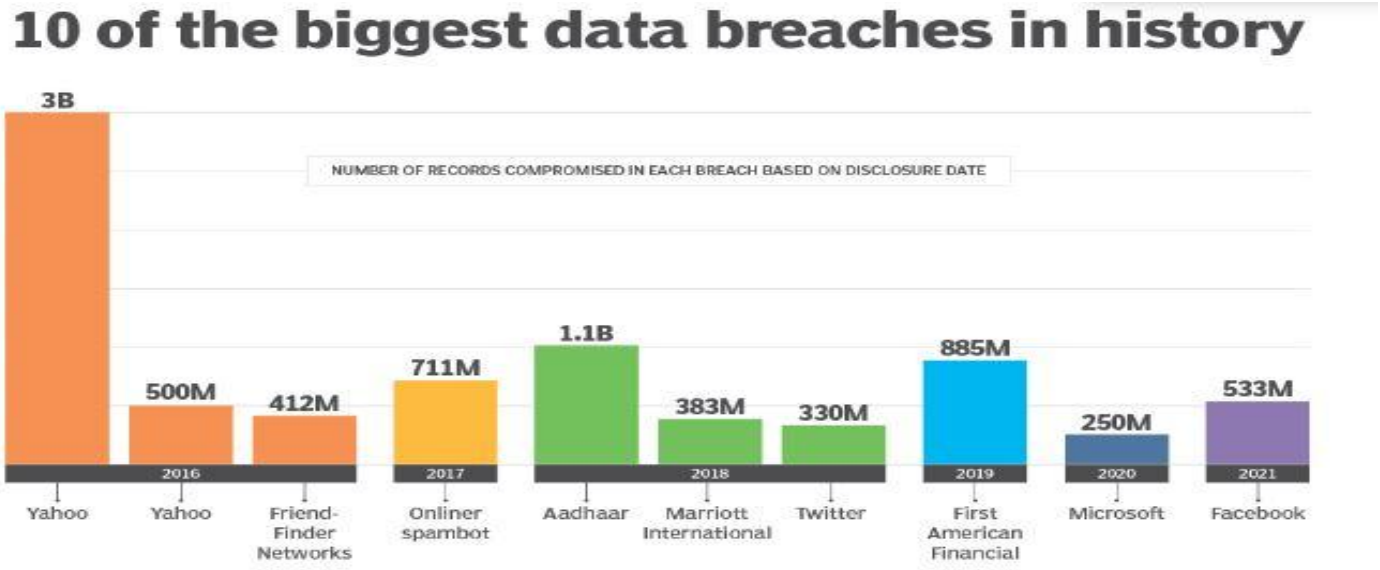

Figure 1; The biggest data breaches in history. Source: (Harford, 2022)

However, e-commerce platforms have introduced several security measures to address these worries. Utilizing encryption methods to safeguard data during transmission and storage is standard practice. Sensitive information is protected by encryption so that only those with the proper permissions can access it (Song, et al., 2022). Secure Socket Layer (SSL) protocols, which encrypt data transferred during online transactions, are frequently used to provide secure connections between users and websites. Implementing authentication systems is another strategy for e-commerce personal data security. User authentication helps to prevent unwanted access to sensitive data by confirming the identity of people using online services. This can involve one-time passwords, biometric authentication, or multifactor authentication (Yin et al., 2021). By guaranteeing that only authorized users can access sensitive information and

conduct transactions, these steps improve the security of personal data. Nevertheless, despite these security precautions, e-commerce companies still face many faculties in protecting user data; the prevalence of data breaches is one significant difficulty. Hackers continuously look for system weaknesses to obtain illegal access to databases and steal personal information. To reduce the danger of data breaches, e-commerce platforms must make significant investments in cybersecurity protections and keep their security systems updated often

The sharing of personal data with outside service providers also raises privacy issues. Information on customers must be shared since e-commerce platforms frequently work with logistics partners, payment processors, and marketing firms. Protecting personal data requires ensuring these third-party organizations follow stringent data protection laws and uphold strong security measures (Luo et al., 2022). Expanding e-commerce has created severe difficulties in protecting personal information. E-commerce systems use encryption, SSL protocols, and authentication processes as critical instruments to safeguard sensitive data. Data breaches and the sharing of data with outside suppliers, however, continue to be issued (Cao et al., 2022). To keep users' faith and confidence in the safety of their personal data, e-commerce platforms must continuously evaluate and improve their security procedures.

2.2 Machine Learning: An Overview

2.2.1 Different Types of Machine Learning Techniques: (Supervised Learning, Unsupervised Learning, Reinforcement Learning, and Deep Learning)

Machine learning is a branch of artificial intelligence (AI) that focuses on creating models and algorithms that let computers learn and anticipate the future or make decisions on their own. It entails studying data to spot patterns, draw forth essential insights, and base judgments on the patterns found. Data and algorithms are the foundation of machine learning. The algorithms are trained using labeled data, where the input data is associated with corresponding desired outputs, in a process known as supervised learning (Sajja et al., 2021). Based on the patterns found in the labeled training data, the algorithm learns to predict or categorize new, unobserved data. Typical applications of this method include classification and regression. The algorithms are trained on unlabeled data without the expected results in unsupervised learning. Without explicit instructions, the algorithms seek to identify any inherent structure or patterns in the data. Unsupervised learning is frequently used for clustering and dimension reduction. Reinforcement learning: An agent learns to make decisions and behave in a way that will maximize a reward signal using this learning paradigm (Pallathadka et al., 2023). The agent learns the best tactics over time by receiving feedback as rewards or penalties based on behavior. Games and autonomous systems are only two examples of applications where

reinforcement learning has been successful.

2.2.2 Application of Machine Learning in Various Industries with a Focus on E-Commerce.

Due to its capacity to analyze massive amounts of data, draw conclusions, and generate precise forecasts, machine learning has been applied in numerous industries, including e-commerce. The following objectives are the main ones for which the e-commerce business has applied machine learning techniques: systems for providing advice E-commerce companies frequently employ machine learning algorithms to offer users personalized product recommendations based on their browsing history, purchasing patterns, and preferences (Chatterjee et al., 2021). These algorithms examine user data and behavior patterns to make pertinent product recommendations that improve the shopping experience and boost client happiness. Fraud detection: Machine learning algorithms are used to identify fraudulent activity and guard against financial losses in e-commerce (Kumar et al., 2021). These algorithms can discover patterns suggestive of fraudulent activities, such as odd buying habits or abnormal transactions, and flag them for further inquiry by evaluating transaction data and user behavior.

2.3 Machine Learning for Personal Data Protection.

2.3.1 Use of Machine Learning in Personal Data Protection: Existing Research and Findings.

In many fields, including e-commerce, machine learning approaches have shown promise in strengthening personal data protection. The application of machine learning algorithms for anomaly detection, user authentication, privacy preservation, fraud detection, and policy compliance in data protection has been investigated in previous research (Pallathadka et al., 2023). Finding anomalous patterns or behaviors that could be signs of fraud or security breaches is known as anomaly detection. In order to identify anomalies and initiate the necessary security procedures, machine learning algorithms, such as clustering and classification models, can examine user behavior, transaction data, and network traffic. Another area where machine learning can help to protect data is user authentication. By examining past user data, machine learning models may create user profiles and estimate the likelihood of a person's identification based on their behavioral patterns, device information, and other contextual elements (Boissay et al., 2021). By doing so, authentication processes may be strengthened, and the chance of unwanted access to personal data can be decreased. A crucial part of protecting personal data is maintaining privacy. Differential privacy is one example of a machine-learning technique that can aggregate and anonymize user data while still enabling the extraction of valuable insights. This protects users' privacy

while maintaining the usefulness of the data for analysis and decision-making. Machine learning algorithms have shown their efficiency in detecting suspect financial or online actions (Policarpo et al., 2021). Fraud detection is a common worry in e-commerce. Machine learning may detect new fraud cases by training models on previous data containing well-known fraud tendencies. This allows for the proactive protection of personal data.

2.3.2 Effectiveness of Machine Learning Techniques in Protecting Personal Data: Comparative Analysis with Traditional Methods.

Machine learning techniques offer significant benefits over conventional approaches to protecting personal data. Machine learning algorithms may process large data sets, spot intricate patterns and adjust to changing security risks (Liu et al., 2021). They are particularly suited for dynamic and quickly changing e-commerce settings because they can learn from historical data and continuously improve their performance. The constraints of conventional approaches, such as rule-based systems that rely on predetermined rules and signatures, may also be solved by machine learning techniques. Machine learning offers more precise and effective identification of security incidents by spotting abnormalities and trends that may be challenging to identify using established rules. Machine learning approaches can also enhance data protection by personalizing and adapting it (Sarker et al., 2021). While machine learning algorithms can customize protection, mechanisms based on individual user profiles, behavior, and contextual information, traditional methods frequently apply standard security measures for all users. The effectiveness of data protection can be increased, and the user experience as a whole can be enhanced with this customized approach. The quality and representativeness of the training data, the choice of algorithms, and the deployment tactics are just a few examples of the many variables that affect how well machine learning approaches protect personal data. Additionally, adversarial attacks, in which bad actors try to influence the behavior of the models or exploit flaws, may be possible against machine learning models. It is essential to comprehend these issues to ensure the stability and dependability of machine learning-based data protection systems.

2.4 Gaps in the Literature and Future Research Directions.

2.4.1 Identification of Gaps in the Existing Literature.

Despite the growing interest in applying machine learning techniques for data protection in e-commerce, there are still several gaps in the existing literature. These gaps can be identified as follows:

Few studies explore the practical difficulties and factors in deploying machine learning-based data protection systems in real-world e-commerce contexts, instead focusing primarily on theoretical elements and proof-of-concept implementations. Lack of

complete evaluation metrics: Standardized evaluation metrics are required to measure machine learning algorithms' efficacy, efficiency, and usefulness to protect sensitive data (Cheng et al., 2022). The evaluation criteria used in existing studies often vary, making comparing and generalizing the results challenging. User perspectives are not sufficiently considered: While machine learning techniques can improve data safety, it is crucial to take users' perceptions, attitudes, and concerns into account while utilizing these methods (Ramesh et al., 2022). User acceptance, privacy attitudes, and trust in machine learning-based data protection systems in e-commerce are not fully understood by the literature.

2.4.2 Importance of Filling These Gaps for Better Understanding and Application of Machine Learning in E-Commerce Data Protection.

For several reasons, filling these gaps in the literature is imperative. A more accurate evaluation of machine learning-based data protection systems' viability, efficacy, and scalability will be possible after addressing the practical issues and concerns surrounding their use in e-commerce. This will make implementing and incorporating machine learning methods into existing e-commerce platforms easier (Sundararajan et al., 2021). Second, adopting uniform evaluation metrics would encourage a more organized and consistent examination of the efficacy of machine learning algorithms for data security. This will make it easier to compare various methods, make benchmarking easier, and encourage the creation of best practices for assessing machine learning-based data protection systems in e-commerce (Elen et al., 2021). Third, creating user-centric data protection systems requires a thorough grasp of user views and concerns. It will be feasible to design machine learning solutions that meet user needs, improve user trust, and promote user adoption of e-commerce platforms by considering user attitudes, preferences, and expectations. Generally, the currently available literature offers convincing proof of the promise of ML and AI in e-commerce data protection. However, it also shows that there is still a lot to learn about the best use of these technologies. More study in many contexts is required, emphasizing weighing the advantages of these technologies against any privacy issues they may pose.

3. Methods:

Mokbal, F. M. M., Wang, D., & Wang, X. (2022). Detect cross-site scripting attacks using average word embedding and support vector machine. *International Journal of Network Security*, *24*(1), 20-28.

While machine learning has gained a prominent place in cybersecurity, it is still unable of accurately detecting cross-site scripting attacks with often affect online retail services. In this sense, the results obtained thus far cause many attacks to be undetected and show an outstandingly high rate of both false positive and false negative results. In such a scenario, Mokbal et al. developed and validated a method

based on the use of support vector machine algorithms using a min max approach to provide a more accurate identification of XSS attacks. Once properly trained by using a training dataset of over 20 thousand payload secondary data samples containing both malicious and benign payload texts. Once trained, the researchers validated the developed SVM algorithm by analyzing the precision rate, detection rate, false positive and false negative rates, the F-score, and the misclassification rate of benign and malicious data.

Cao, Y., Shao, Y. and Zhang, H., 2022. Study on early warning of e-commerce enterprise financial risk based on deep learning algorithm. *Electronic Commerce Research*, *22*(1), 21-36.

The authors used deep learning algorithms based on long and short-term memory networks (LSTM) and convolutional neural networks (CNN) to analyze and identify potential dangers to the financial secondary data obtained from e-commerce sites such as their profitability, liquidity, solvency, and operational efficiency. The first method enabled Cao et al. (2022) to include specialized memory cells into the machine learning algorithm that could store the information over long sequences of data. Thus, the algorithm developed enabled the authors to evaluate how the financial data collected from the e-commerce sites evolved over time, enabling a simpler identification of potential threats whenever any of these data showed a change in the general trend stored in the LSTM neural network. The second method, on the other hand, provided the authors with a strong hierarchical classification of the financial data, enabling them to detect any potential interactions between the collected information. In this sense, CNN allowed for the identification and reaction of potential threats that caused a new interaction between the variables to appear that resulted in a change in the hierarchical classification of the secondary data analyzed.

Chatterjee, S., Goyal, D., Prakash, A., & Sharma, J. (2021). Exploring healthcare/health-product ecommerce satisfaction: A text mining and machine learning application. *Journal of Business Research*, *131*, 815-825.

The authors evaluated a combination of text mining and machine learning techniques to explore the factors that influenced consumer's happiness in healthcare and health e-commerce. For this purpose, they used a dataset of secondary data consisting on 186,057 reviews on 619 e-commerce firms operating in 29 different subcategories of the healthcare and health-product industry. The reviews used in the analysis were posted by the consumers on a review-website between 2008 and 2018. The data analysis involved the use of text-mining protocols to identify the emotions and sentiments of the reviewers. The data used in this study are secondary in the sense that they were collected from a review website.

Liu, F., Li, Z., Wang, B., Wu, J., Yang, J., Huang, J., Zhang, Y., Wang, W., Xue, S., Nepal, S. and Sheng, Q.Z., 2022. Eriskcom: An e-commerce risky community detection platform. *The VLDB Journal*, *31*(5), 1085-1101.

The authors used an algorithm based on SALSA to evaluate the threats affecting e-commerce communities. This algorithm focuses on the use of the theory of Markov chains to evaluate the nodes of the community by a random analysis of the attributes and attitudes of the different members of the community. The SALSA method proposed by the authors used Markov chains to analyze the existence of excessively authoritarian individuals within the community that could potentially represent a threat to the safety and wellbeing of the rest of the community members, thus representing an effective tool to recognize and monitor potentially hazardous e-commerce community settings.

4. Results

Title of Article Reviewed	What authors did.	General Findings from the article.	What they did not do.	How we intended to solve in our research.
Mokbal, F. M. M., Wang, D., & Wang, X. (2022). Detect cross-site scripting attacks using average word embedding and support vector machine. *International Journal of Network Security*, *24*(1), 20-28.	The authors developed and validated a method based on support vector machine algorithm to identify potential payload XSS attacks.	A comparison of the accuracy of the SVM algorithm in the detection of malicious XSS attacks suggests that this machine learning method is considerably better than alternative machine learning protocols based on logistic regression, linear discriminant analysis, K-neighbors clustering, Cart decision trees, Gaussian NB, AdaBoost, Gradient Boost, and Random Forest models.	Though the authors validated the SVM algorithm to demonstrate its higher accuracy rates, they did not seem to apply it to primary data, but rather on a secondary dataset.	Based on the positive results obtained, it would be interesting to evaluate the applicability of SVM machine learning algorithms on the evaluation of cyberattacks on primary data.
Cao, Y., Shao, Y. and Zhang, H., 2022. Study on early warning of e-commerce enterprise financial risk based on deep learning algorithm. Electronic Commerce Research, 22(1), 21-36.	The authors used deep learning algorithms to create an early warning system for financial risk in e-commerce firms. They gathered financial information from e-commerce enterprises, including characteristics such as sales, costs, and profitability. The authors then used deep learning techniques, notably deep neural networks, to evaluate and simulate the financial data. They created deep learning models to detect patterns	The developed model based on deep learning using LSTM and CNN neural networks enables the identification of potential threats that affect both the temporal sequence of financial data and the interaction between the	The authors did not go into the implementation specifics of the deep learning algorithm utilized, nor do they provide an in-depth examination of the findings produced from using the	In this research, we conducted a thorough examination of the outcomes produced from applying the framework to real-world e-commerce data. To do so, we investigated the implementation specifics of the deep learning algorithm employed, including its architecture, parameters, and training technique. We obtained insights into the algorithm's strengths, limitations, and potential areas for development by thoroughly

	and signals suggestive of financial danger in e-commerce businesses. The study's findings help to improve risk management methods and decision-making processes for e-commerce businesses, allowing them to address potential financial concerns more effectively.	different variables affecting such data. From this point of view, the combined deep learning method provides an invaluable tool to detect signals of financial danger in e-commerce buildings, thus enabling the implementation of faster and more effective corrective actions.	framework to real-world e-commerce facts.	studying its performance. In addition, weundertook a thorough examination of the findings collected, focusing on critical parameters such as accuracy and precision.
Chatterjee, S., Goyal, D., Prakash, A., & Sharma, J. (2021). Exploring healthcare/health-product ecommerce satisfaction: A text mining and machine learning application. *Journal of Business Research, 131*, 815-825.	The authors used text mining and machine learning approaches to investigate customer satisfaction in the healthcare/health-product e-commerce market. They compiled a collection of customer reviews and ratings from numerous healthcare/health-product e-commerce platforms. The authors then used text mining approaches to preprocess the data, such as text cleaning and stemming, in order to extract relevant information from the reviews. They used machine learning algorithms, notably the Support Vector Machine (SVM) classifier, to assess the data and forecast customer happiness based on the linguistic content of the reviews. The authors	The paper demonstrates the existence of a relationship between the level of customer satisfaction with the emotions expressed by reviewers using text-mining. Thus, it is possible to use logistic regression methods coupled to text-mining approaches to predict the customer satisfaction.	The authors did not include an in-depth description of the text mining and machine learning techniques used, the broader implications of the results on the healthcare sector and consumer satisfaction, a comprehensive examination of the study's limitations, and a review of different explanations or factors affecting healthcare/health-product ecommerce	Weprovided a full explanation of the text mining and machine learning approaches utilized in this research, including specific methodology, algorithms, and feature selection processes. We highlighted the larger implications of our findings for the e-commerce business, emphasizing how the findings could help influence decision-making, boost e-commerce services, and improve overall consumer experience.

	were able to acquire insights into the elements impacting client satisfaction in the healthcare product e-commerce area and find patterns and trends within the dataset by employing these computational tools.		satisfaction.	
Liu, F., Li, Z., Wang, B., Wu, J., Yang, J., Huang, J., Zhang, Y., Wang, W., Xue, S., Nepal, S. and Sheng, Q.Z., 2022. Eriskcom: An e-commerce risky community detection platform. The VLDB Journal, 31(5), 1085-1101.	The authors designed the Eriskcom technology to discover dangerous networks in e-commerce. To accomplish this, they first gathered data from e-commerce platforms, such as user profiles, product information, and transaction records. They then used several data processing techniques to preprocess the data, such as data cleansing and feature extraction. Following that, they used community detection algorithms to identify communities within the e-commerce network based on user-product interactions and transaction patterns. They implemented a risk assessment mechanism to measure the danger level of each neighborhood. The authors deployed their proposed platform and conducted experiments to assess its performance in spotting dangerous populations.	Compared to alternative algorithms as HITS or TKC, the SALSA algorithm enables the researchers a more accurate evaluation of the e-commerce communities, as it allows the identification of a higher number of situations in which there is an excessively authoritarian role within the community.	The authors fail to provide detailed details about the specific steps taken or approaches used to reach the outcomes indicated. The authors do not disclose the platform's algorithms, strategies, or approaches, making it difficult to grasp the underlying mechanisms that lead to the detection of unsafe communities in e-commerce.	Weshared extensive details regarding the processes taken and approaches used to attain the results. Wecomprehensively detailed the platform's algorithms, strategies, and approaches, shedding light on the fundamental mechanisms that allowed the detection of unsafe communities. Readers will obtain a clear knowledge of how the platform works and the elements that contributed to its efficacy by providing a detailed overview. Furthermore, we went into the methodology' intricacies, outlining the rationale for their selection as well as prospective modifications or enhancements.

(Sources: Mokbal et al., 2022; Cao et al., 2022; Chatterjee et al., 2021; Liu et al., 2022)

International Journal of Computer Applications Technology and Research
Volume 12–Issue 08, 126 - 145, 2023, ISSN:-2319–8656
DOI:10.7753/IJCATR1208.1016

5. Way Forward in the E-Commerce Sector in Kenya.

Figure 2: Penetration internet rate in Kenya. (Junior, 2021)

As shown in the figure above, the study by USA Business Team. (2021) noted that the e-commerce market in Kenya had expanded significantly in recent years due to technological improvements, rising internet usage, and shifting consumer preferences. In order to take advantage of this expansion and further improve the e-commerce ecosystem, several crucial areas need attention and concentration.

The following are future directions for Kenya's e-commerce industry going forward: Enhanced infrastructure: For the e-commerce industry to expand, a robust digital infrastructure must be developed. This entails enhancing broadband coverage, increasing internet connectivity, and supplying dependable logistics and delivery services. Infrastructure improvements will make online transactions more frictionless and help e-commerce companies more effectively reach clients in outlying areas (Makokha et al., 2021).

Enhanced payment systems: The success of e-commerce depends on a reliable and secure payment system. Developing and promoting user-friendly, user-accessible digital payment systems should be a joint effort between the government and financial institutions. Consumer confidence in online purchases will increase by promoting the use of mobile money platforms, increasing the interoperability of payment systems, and assuring safe online payment choices.

More effective cybersecurity measures are required as the e-commerce industry expands due to an increased risk of cyberattacks and data breaches. Prioritizing cybersecurity measures is essential to safeguarding consumer data and fostering trust in online transactions. Fostering a secure e-commerce environment would require

strong data protection policies, raising awareness of cyber dangers, and establishing methods to resolve cybersecurity events (Achiando, 2019).

The government should create a favorable regulatory environment to encourage the e-commerce industry. Regulations should find a balance between promoting company expansion and protecting consumers. A direct taxation, licensing, international trade, and dispute resolution framework will create a favorable climate for e-commerce companies to operate and grow. Promoting digital skills development and entrepreneurial support is crucial for fostering innovation and growth in the e-commerce industry (Rani et al., 2021).

Digital literacy programs, e-commerce entrepreneur training courses, and SMEs' access to financing are all initiatives that will enable people and companies to take advantage of the opportunities provided by e-commerce. In addition, Implementing Privacy by Design: Privacy considerations must be incorporated into every level of system development to fully utilize ML and AI for data security. To guarantee that their systems are developed with privacy in mind, businesses should closely collaborate with ML and AI solution providers, reducing the risk of data leaks or breaches.
Another direction is collaboration and partnerships: It is essential to develop the e-commerce sector for stakeholders, including governmental organizations, e-commerce platforms, financial institutions, logistical providers, and industry groups, to work together (Nurcahyo et al., 2021). Collaboration can help enhance the growth and sustainability of e-commerce enterprises in Kenya by fostering information sharing, resource pooling, and creating cooperative initiatives to address shared difficulties. Building consumer trust is essential for the long-term viability of the e-commerce industry. Consumer safety is also important. Consumer confidence will be increased through establishing mechanisms for protecting consumers, such as efficient dispute resolution processes, transparent pricing, and clear product information (Missault et al., 2021). Programs that educate consumers about their rights and obligations in e-commerce transactions can also increase that awareness. Generally, Kenya is poised to benefit from using ML and AI in e-commerce data protection. All interested parties must work together, including the government, industry, academia, and consumers.

6. Conclusion

To conclude, the enormous potential of machine learning (ML) and artificial intelligence (AI) to improve data security in e-commerce has been highlighted by this research, along with the obstacles that must be overcome for its deployment. As e-commerce expands and takes a more significant role in world trade, safeguarding users' personal information's more critical than ever. Traditional security measures frequently fall short in the face of complex and developing cyber threats in today's quickly digitizing environment. Therefore, using cutting-edge technology like ML and AI to protect sensitive client data during online transactions is more important

than ever. With their capacity to learn from and adapt to massive amounts of data, ML and AI have the potential to enhance the current security procedures greatly. Anomaly detection, user authentication, encryption, compliance with privacy policies, and predictive analysis are just a few of the security-enhancing applications for which these technologies are employed, according to our examination of the literature.

Additionally, they can significantly advance fraud detection systems, a crucial component of e-commerce data protection.However, there is a glaring void in the body of knowledge about the use of ML and AI in the e-commerce industry of developing nations like Kenya. Different sectors' distinct technological, infrastructure-related, and skill-related issues need in-depth research to create the most efficient ML and AI implementation solutions.
Further investigation is necessary on how cheap and practical these cutting-edge tactics are for small and medium-sized e-commerce businesses. The future of Kenya's expanding e-commerce industry requires a diverse strategy. Critical tactics for maximizing the potential of ML and AI in data protection include developing a supportive regulatory environment, acquiring the necessary technical skills, investing in a robust data management infrastructure, putting privacy by design into practice, raising consumer awareness, and encouraging partnerships and collaboration. Even though ML and AI provide exciting ways to improve data security, it's crucial to recognize and solve any privacy issues they can raise. In this context, privacy-preserving machines and federated learning are significant new trends. It is imperative to apply these technologies fairly and morally, ensuring that their benefits do not compromise personal privacy.

ML and AI provide an immediate fix and a continual learning and adapting mechanism to counter new dangers in the quickly changing e-commerce landscape. We can foresee a future of e-commerce that is safer, highly tailored, effective, and user-centric as firms continue to innovate and adopt this cutting-edge technology. The research provided here is a step toward a more comprehensive understanding of the use of ML and AI in e-commerce data protection. However, rather than being viewed as a conclusion, it should be a starting point for further study. To stay ahead of new risks and guarantee that the e-commerce landscape remains secure for customers and enterprises, ongoing research and investigation are required.

A supportive regulatory environment will provide the required framework for businesses to exist and thrive with open and equitable regulations. For instance, programs for skill development and encouragement of entrepreneurship will enable people and organizations to participate in the e-commerce sector and foster innovation actively. Stakeholders must work together, including governmental organizations, financial institutions, and trade groups, to address shared difficulties and promote the expansion of e-commerce companies. The evolution of e-commerce should prioritize protecting consumer trust and

putting procedures in place to resolve complaints and guarantee openness.

7. References

[1] Achiando, H. A. (2019). E-Commerce Strategy Adoption and Performance of Micro and Small Enterprises: A Case of private Security Firms in Nairobi County, Kenya. IOSR Journal of Business and Management, 21(7), 35-64. https://doi.org/10.13140/RG.2.2.35333.09442

[2] Boissay, F., Ehlers, T., Gambacorta, L., & Shin, H. S. (2021). Big Techs in Finance: On the New Nexus between Data Privacy and Competition. Springer International Publishing.

[3] Cao, Y., Shao, Y. and Zhang, H., 2022. Study on early warning of e-commerce enterprise financial risk based on deep learning algorithm. *Electronic Commerce Research*, 22(1), 21-36. https://doi.org/10.1007/s10660-020-09454-9

[4] Chatterjee, S., Goyal, D., Prakash, A., & Sharma, J. (2021). Exploring healthcare/health-product ecommerce satisfaction: A text mining and machine learning application. *Journal of Business Research*, 131, 815-825. https://doi.org/10.1016/j.jbusres.2020.10.043

[5] Chen, L., Zhang, W., & Wang, Q. (2020). Privacy-Preserving Personalization in E-commerce Using Machine Learning: A Comparative Study. *IEEE Transactions on Knowledge and Data Engineering*, *32*(8), 1507-1520.

[6] Chen, Y., Luo, H., Chen, J., & Guo, Y. (2022). Building data-driven dynamic capabilities to arrest knowledge hiding: A knowledge management perspective. *Journal of Business Research*, 139, 1138–1154. https://doi.org/10.1016/j.jbusres.2021.10.050

[7] Cheng, L., Guo, R., Moraffah, R., Sheth, P., Candan, K. S., & Liu, H. (2022). Evaluation methods and measures for causal learning algorithms. IEEE Transactions on Artificial Intelligence, 3(6), 924-943. https://doi.org/10.1109/TAI.2022.3150264

[8] Cheng, Xusen, Jason Cohen, and Jian Mou. (2023). AI-enabled technology innovation in e-commerce. *Journal of Electronic Commerce Research*, 24(1), 1-6. http://www.jecr.org/node/674

[9] Elen, A., &Avuçlu, E. (2021). Standardized variable distances: A distance-based machine learning method. *Applied Soft Computing*, 98, 106855. https://doi.org/10.1016/j.asoc.2020.106855

[10] Khan, D. S. W. (2019). Cyber security issues and challenges in E-commerce. *In Proceedings of 10th international conference on digital strategies for organizational success.*

[11] Kumar, M. R., Venkatesh, J., & Rahman, A. M. Z. (2021). Data mining and machine learning in retail business: Developing efficiencies for better customer retention. *Journal of Ambient Intelligence and Humanized Computing*, 1-13. https://doi.org/10.1007/s12652-020-02711-7

[12] Junior, N. (2021, February 19). Report: Internet penetration in Kenya at 40pc in January 2021. Techspace Africa. https://tech-

space.africa/report-there-were-22-86-million-internet-users-in-kenya-in-january-2020/

[13] Liu, B., Ding, M., Shaham, S., Rahayu, W., Farokhi, F., & Lin, Z. (2021). When machine learning meets privacy: A survey and outlook. *ACM Computing Surveys (CSUR)*, 54(2), 1-36. https://doi.org/10.1145/3436755

[14] Liu, F., Li, Z., Wang, B., Wu, J., Yang, J., Huang, J., Zhang, Y., Wang, W., Xue, S., Nepal, S. and Sheng, Q.Z., 2022. Eriskcom: An e-commerce risky community detection platform. *The VLDB Journal*, 31(5), 1085-1101. https://doi.org/10.1007/s00778-021-00723-z

[15] Luo, S. and Choi, T.M., 2022. E-commerce supply chains with considerations of cyber-security: Should governments play a role? *Production and Operations Management*, 31(5), 2107-2126. https://doi.org/10.1111/poms.13666

[16] Makokha, T., Asenahabi, B., &Makokha, A. (2021). Electronic Commerce: The Evolution and State in Kenya. *International Journal of Research in Education Humanities and Commerce*, 2(4), 18-22.

[17] Mokbal, F. M. M., Wang, D., & Wang, X. (2022). Detect cross-site scripting attacks using average word embedding and support vector machine. *International Journal of Network Security*, *24*(1), 20-28.

[18] Nurcahyo, R., & Putra, P. A. (2021). Critical factors in Indonesia's e-commerce collaboration. *Journal of Theoretical and Applied Electronic Commerce Research*, 16(6), 2458-2469.

[19] Pallathadka, H., Ramirez-Asis, E. H., Loli-Poma, T. P., Kaliyaperumal, K., Ventayen, R. J. M., &Naved, M. (2023). Applications of artificial intelligence in business management, e-commerce and finance. *Materials Today: Proceedings*, 80, 2610-2613.

[20] Paul, H. and Nikolaev, A., 2021. Fake review detection on online E-commerce platforms: a systematic literature review. *Data Mining and Knowledge Discovery*, 35(5), 1830-1881.

[21] Policarpo, L. M., da Silveira, D. E., da Rosa Righi, R., Stoffel, R. A., da Costa, C. A., Barbosa, J. L. V.,...&Arcot, T. (2021). Machine learning through the lens of e-commerce initiatives: An up-to-date systematic literature review. *Computer Science Review*, 41, 100414.

[22] Ramesh, T. R., Lilhore, U. K., Poongodi, M., Simaiya, S., Kaur, A., & Hamdi, M. (2022). Predictive analysis of heart diseases with machine learning approaches. *Malaysian Journal of Computer Science*, 1, 132-148. https://doi.org/10.22452/mjcs.sp2022no1.10

[23] Rani, S., Kataria, A., Sharma, V., Ghosh, S., Karar, V., Lee, K., & Choi, C. (2021). Threats and corrective measures for IoT security with observance of cybercrime: A survey. *Wireless Communications and Mobile Computing*, 2021, 1-30.

[24] Sajja, G.S., Mustafa, M., Ponnusamy, R. and Abdufattokhov, S., 2021. Machine learning algorithms in intrusion detection and classification. *Annals of the*

Romanian Society for Cell Biology, 25(6), 12211-12219.

[25] Sean Michael Kerner. (2023). 34 cybersecurity statistics to lose sleep over in 2023. WhatIs.com; TechTarget. https://www.techtarget.com/whatis/34-Cybersecurity-Statistics-to-Lose-Sleep-Over-in-2020

[26] Shaba, M., Roland, A., Simon, J., Misra, S., & Ayeni, F. (2022, March). A real-time sentimental analysis on e-commerce sites in Nigeria using machine learning. *In Hybrid Intelligent Systems: 21st International Conference on Hybrid Intelligent Systems (HIS 2021),* December 14–16, 2021 (pp. 452-462). Cham: Springer International Publishing.

[27] Song, Y., Liu, J., Zhang, W. and Li, J., 2022. Blockchain's role in e-commerce sellers' decision-making on information disclosure under competition. *Annals of Operations Research*, 1-40.

[28] Sundararajan, K., Garg, L., Srinivasan, K., Bashir, A. K., Kaliappan, J., Ganapathy, G. P., ... & Meena, T. (2021). A contemporary review on drought modeling using machine learning approaches. *CMES-Computer Modeling in Engineering and Sciences*, 128(2), 447-487.

[29] Vinoth, S., Vemula, H.L., Haralayya, B., Mamgain, P., Hasan, M.F. and Naved, M., 2022. Application of cloud computing in banking and e-commerce and related security threats. *Materials Today: Proceedings*, 51, 2172-2175.

[30] Wang, S., Chen, Z., Xiao, Y., & Lin, C. (2021). Consumer privacy protection with the growth of AI-empowered online shopping based on the evolutionary game model. *Frontiers in public health*, 9, 705777.

[31] Yathiraju, N. (2022). Investigating the use of an Artificial Intelligence Model in an ERP Cloud-Based System. *International Journal of Electrical, Electronics and Computers*, 7(2), 1-26.

[32] Zhang, F. and Yang, Y., 2021. Trust model simulation of cross border e-commerce based on machine learning and Bayesian network. *Journal of Ambient Intelligence and Humanized Computing*, 1-11.

[33] USA Business Team. (2021, January 6). Online selling platforms in Kenya - E-commerce platforms in Kenya. https://www.usabusiness.co.in/online-selling-platforms-marketplaces-kenya/

[34] Yin, X., He, J., Gao, Y. and Li, J., 2021, April. Mass tourism data analysis API based on E-Commerce platform. In Journal of Physics: Conference Series, 1881(2), 022065. IOP Publishing.

[35] Habbat, N., Anoun, H., &Hassouni, L. (2022, January). LSTM-CNN deep learning model for french online product reviews classification. *In Advanced Technologies for Humanity: Proceedings of International Conference on Advanced Technologies for Humanity (ICATH'2021) (pp. 228-240*). Cham: Springer International Publishing.

[36] Harford, I. (2022, July 29). 10 biggest data breaches in history, and how to prevent them: TechTarget. Security. https://www.techtarget.com/searchsecuri

ty/feature/10-biggest-data-breaches-in-history-and-how-to-prevent-them