JOURNAL OF AERONAUTICAL MATERIALS
ISSN: 1005-5053

Vol. 43, Issue-01, 2023
pp. 543-560

AN ANALYTICAL RESEARCH BASED ON MITIGATING REVIEW SPAM ON E-COMMERCE SITES

Ganesh S. Wayal
Research Scholar, Department of Computer Science & Engineering, Madhyanchal Professional University, Bhopal, India

Dr. Vijay Bhandari
Associate Professor, Department of Computer Science & Engineering, Bhopal, India

Abstract - Online reviews have become an important aspect of e-commerce, providing valuable information to potential customers about products and services. However, the prevalence of review spam, which are fake or fraudulent reviews, has become a growing concern for both consumers and online retailers. Review spam can manipulate the perceived quality of products and services, influencing customer purchasing decisions and damaging the reputation of online retailers. This paper analyzes the different types of review spam, their impact on e-commerce, and the methods used to mitigate review spam. Through a review of the literature and case studies, this paper identifies key strategies and techniques for mitigating review spam on e-commerce sites, including the use of machine learning algorithms, human review, and the implementation of review policies and guidelines.

1 INTRODUCTION

Online reviews have become an important aspect of e-commerce, providing valuable information to potential customers about products and services. According to a survey by Bright Local, 91% of consumers read online reviews, and 84% trust them as much as personal recommendations. However, the prevalence of review spam, which are fake or fraudulent reviews, has become a growing concern for both consumers and online retailers. Review spam can manipulate the perceived quality of products and services, influencing customer purchasing decisions and damaging the reputation of online retailers. In this paper, we analyze the different types of review spam, their impact on e-commerce, and the methods used to mitigate review spam.

Online reviews play a critical role in e-commerce, influencing customer purchasing decisions and providing valuable feedback to retailers. However, the rise of review spam has become a growing concern for e-commerce sites, with the potential to manipulate the perceived quality of products and services, influencing customer purchasing decisions, and damaging the reputation of online retailers. This paper aims to provide an overview of review spam and its impact on e-commerce, as well as strategies and techniques for mitigating review spam.

1.1 Objective

In general observation a genuine reviewer will not try to comment on the quality of the service or product more than once, until he wants to reply to another customer's comment or he deliberately wants to mislead others by posting hyper or defaming reviews. The features of a spam review and a non-review as discussed in the following section in detail are practically implemented by constructing a dataset of our own.

A web application has been developed that takes the inputs (reviews) from the users, stores the details in the database and then detects the spammers based on the proposed method. Identity of a customer is traced using his email id with which he has logged into the website, his location and his device's IP address.

2 LITERATURE REVIEW

Mamoona et. al. [1] presented a comprehensive survey on the evolution, prevention, and mitigation of ransomware in the context of the Internet of Things (IoT). The paper provided deeper insights into the evolution of ransomware in IoT and discussed various aspects of ransomware attacks, including types of ransomware, current research, prevention and mitigation techniques, ways to deal with affected machines, and the decision about paying the ransom or not. The authors also highlighted future emerging trends of ransomware propagation in IoT and provided a summary of current research to show various directions of research. This survey is expected to be useful for researchers and practitioners who are involved in developing solutions for IoT security.

Sunghoon et. al. [2] proposed an unsupervised model that reduces product rating biases that result from varying degrees of customers' optimism/pessimism. The proposed customer rating analysis model adjusts product ratings based on rating histories and tendencies, instead of human-labeled training data, to provide customers with more objective and accurate feedback. The authors used a case study involving real-world customer rating data from an e-commerce company to validate the method. This model is significant for identifying unbiased customer ratings and true product quality, considering reviewers' rating histories and tendencies because each reviewer has different criteria for buying and rating products.

Sumit et. al. [3] conducted a literature survey to analyze the security in e-commerce systems. The paper illustrated the year-wise publication of various attacks on e-commerce sites over the last 10 years, along with various security measures and challenges. This survey is beneficial for researchers working in the domain of e-commerce system security.

Michael et. al. [4] surveyed prominent machine learning techniques proposed to solve the problem of review spam detection. The authors extracted meaningful features from text using natural language processing and conducted review spam detection using various machine learning techniques. They also discussed how reviewer information, apart from the text itself, can be used to aid in this process. The majority of current research has focused on supervised learning methods, which require labeled data, a scarcity when it comes to online review spam. Research on methods for big data are of interest since there are millions of online reviews, with many more being generated daily. The paper also highlighted the need for developing techniques for detecting review spam since not all online reviews are truthful and trustworthy.

Charlie et al. [5] aimed to showcase a Bayesian reinterpretation of Platt and Burges' NIPS reviewer calibration technique, as well as apply extensions to the model on a novel dataset. The results indicated the existence of groups of reviewers who consistently gave higher or lower ratings than their peers, and that a straightforward Bayesian model could detect this latent bias.

Tanmoy et al. [6] highlighted the growing problem of hate speech on social media platforms, which can cause distress and hostility for individuals and communities. The contextual nature of hate speech and its impact on marginalized groups have prompted interest from the machine learning and data mining communities. The authors discussed methodological challenges in developing automated hate mitigation systems and presented a series of proposed solutions to limit the spread of hate speech on social media.

Hamzah et al. [7] focused on the problem of fake/spam reviews that mislead customers in online shopping. The authors proposed a supervised classification approach to distinguish between spam and non-spam reviews, with a bagging-based method to address the data imbalance issue in training classifiers for spam detection. Their experiments and comparisons showed that their proposed method, iSRD, outperformed baseline methods for review spam detection.

Haitao et. al. [8] found that in online markets, a store's success is closely linked to its reputation. To quickly achieve a high reputation, some sellers turn to an underground marketplace known as a seller-reputation-escalation (SRE) market. This market allows online sellers to use human laborers to conduct fake transactions for improving their store's reputation. In this study, the authors investigate the impact of the SRE service on reputation escalation. They infiltrated five SRE markets and studied their operations using daily data collection over a two-month period. They identified more than 11,000 online sellers posting at least 219,165 fake-purchase tasks on the five SRE markets. The study demonstrated that online sellers using SRE service can increase their stores' reputations at least ten times faster than legitimate ones while only 2.2% of them were detected and penalized. The authors also offer some insights into potential mitigation strategies based on their analysis of the operational characteristics of the underground economy.

T. Ravindra et. al. [9] focused on fraudulent activities in E-Commerce companies and presented machine learning models that can detect them for effective mitigation. One particular fraud involves orders placed with randomly typed alphanumeric characters as customer addresses, referred to as monkey-typed addresses. The authors proposed a machine learning-based approach that can classify a given address as normal or monkey-typed with high accuracy. The approach integrates address preprocessing, novel feature generation, and classification. Accurately identifying such addresses can help reduce operational costs.

Andrew et al. [10] discovered that while electronic commerce is popular, it often lacks the traditional information that builds trust between exchange partners. However, digital technologies have created new forms of “electronic word-of-mouth,” which have the potential to gather credible information that influences consumer behavior. Through a nationally representative survey and a focused experiment, the study evaluated how individuals perceive the credibility of online commercial information in comparison to information from more traditional channels. The survey results showed that consumers rely heavily on web-based information and that ratings information is critical in evaluating the credibility of online commercial information. The experimental results showed that ratings are positively associated with perceptions of product quality and purchase intention. However, people may use ratings information suboptimally by potentially privileging small numbers of ratings that could be idiosyncratic, and product quality was shown to mediate the relationship between user ratings and purchase intention. The study provides practical and theoretical implications for ecommerce scholars, consumers, and vendors.

In the study by Raymond et al. [11], it was revealed that in the era of Web 2.0, enormous volumes of consumer reviews are posted to the Internet daily. However, manual approaches to detecting and analyzing fake reviews are impractical due to the problem of information overload. Therefore, the authors designed and developed automated methods to detect fake reviews. The study applied design science research methodology and proposed novel computational models, including a text mining model integrated into a semantic language model, for the detection of untruthful reviews. The proposed models were evaluated using a real-world dataset collected from amazon.com, and the results of the experiments confirmed that the proposed models outperformed other well-known baseline models in detecting fake reviews. This work represents the first successful attempt to apply text mining methods and semantic language models to the detection of fake consumer reviews, and it has significant managerial implications for firms that can use the proposed models to monitor online consumer reviews and develop effective marketing or product design strategies based on genuine consumer feedback.

Li He et al. [12] noted that online reviews play a critical role in determining consumers' purchase choices in e-commerce, yet many online reviews are intentionally created to confuse or mislead potential consumers. Moreover, driven by product reputations and merchants' profits, more and more spam reviews are inserted into online platforms. These reviews may be positive, negative, or neutral but have common features, such as misleading consumers or damaging reputations. The article introduced the task of detecting spam online reviews, provided a common definition of spam reviews, and comprehensively summarized the existing methods and available datasets. The authors also summarized the existing network-based approaches in dealing with this task and proposed some future research directions.

In the Web2.0 era, user-generated content (UGC) has become a valuable source of data for understanding consumers and driving intelligent business. Shugang et. al. [13] propose that text mining techniques, such as semantic and sentiment analysis, can extract meaningful information embedded in UGC for e-commerce business applications. However, this interdisciplinary field lacks a comprehensive summary of research frameworks and application directions. To fill this gap, the authors derive a general framework based on e-commerce practices and discuss mainstream directions of text mining for business applications, including high-quality UGC detection, consumer profiling, product enhancement, and marketing. They also emphasize research gaps and provide suggestions for future work.

In a study by Yehoshua et. al. [14], perceived risk was found to be an important ingredient in the consumer decision-making process related to Internet usage and e-marketing. The authors aim to investigate the perceived barriers to Internet usage and e-marketing by both users and non-users in order to develop more efficient marketing strategies. Using a qualitative research paradigm, they develop a detailed perceived risks map and suggest a model with the factors affecting the Internet's perceived risk elements, including demographic traits and usage behavior characteristics. The model is tested against a sample of 465 employed adults.

Wosah et. al. [15] discuss the increasing use of email as a communication channel for exchanging information and the rise of phishing cybercrime attacks in which attackers send malicious emails to deceive users into disclosing their personal information. The authors examine different approaches that have been developed to combat phishing attacks, highlighting the limitations of existing solutions in identifying phishing emails from legitimate ones. They provide a literature analysis of existing phishing mitigation approaches and suggest ways to improve phishing detection.

Gourab et. al. [16] discovered that online product reviews and ratings have a significant impact on the purchasing decisions of potential customers. However, false reviews can skew the opinions of customers either positively or negatively, making it difficult to trust online reviews and damaging the credibility of e-commerce platforms. The detection of review spam is challenging, as it can be crafted to appear indistinguishable from genuine reviews. This problem is unlike email spam detection because spam emails are typically commercial in nature and contain obvious features.

Haiqin et. al. [17] highlighted the economic benefits of e-commerce, which have attracted malicious merchants to insert fraudulent purchases, fake review scores, and feedback to promote items. Detecting this fraud is challenging due to the difficulty of accessing internal e-commerce data, the variability of e-commerce services used by malicious merchants, and the reluctance of service providers to cooperate. To address this problem, they developed an efficient, platform-independent, and robust e-commerce fraud detection system called CATS. The prototype was implemented and evaluated on popular e-commerce platform Taobao, achieving a high detection accuracy of 91%. They also applied CATS on another large-scale e-commerce platform and achieved an accuracy of 96%, demonstrating its effectiveness in real-world scenarios. Their analysis of reported frauds revealed several abnormal yet interesting behaviors that can help defend against fraud for various e-commerce platforms.

Thomas et. al. [18] suggested that botnets are one of the most prevalent threats to the Internet today, mainly due to the economic incentives behind their operation. Botnet toolkits are readily available for purchase, allowing anyone to generate a customized botnet and become a botmaster. Botnet services sold by botmasters enable criminals to steal identities and credit card information, which is then sold to end-users for unauthorized transactions. However, targeting the authors or botmasters of these toolkits is challenging because they are technology-savvy and elusive. In this paper, the authors propose a bottom-up approach, where end-users of stolen credentials are prosecuted or discouraged, thereby defaming botnet toolkits. They present a case study of applying this approach to Zeus, a popular botnet toolkit, using reverse engineering and behavioral analysis methodologies.

Yen-Hsien et. al. [19] noted that while existing supervised learning techniques are effective in supporting product recommendations in business-to-consumer e-commerce scenarios, they become ineffective in single-class learning scenarios where the training sample only consists of examples pertaining to one outcome class (positive or negative). To address this challenge, the authors propose a COst-sensitive Learning-based Positive Example Learning (COLPEL) technique that constructs an automated classifier from a training sample comprised of positive examples and a large number of unlabeled examples. The proposed technique uses cost-proportionate rejection sampling to derive a subset from the unlabeled examples that likely feature negative examples in the training sample. The technique follows a committee machine approach and constructs a set of classifiers that make joint product recommendations while mitigating potential biases. The authors evaluate the proposed method using book ratings collected from Amazon.com and compare it with two prevalent techniques for benchmark purposes. According to their results, the proposed COLPEL technique outperforms both benchmarks in terms of accuracy and positive and negative F1 scores.

Attila et. al. [20] highlighted that as technology advances and the internet becomes more widely adopted, traditional forms of communication such as parcels have become unpopular, and the internet has become the primary mode of communication. The increased use of the internet has led to modern threats attributed to cybercrime. Nigeria, with an internet penetration of 33.6% and an estimated population of over 205 million people, is experiencing growth in population and internet connectivity. Nigeria is now considered the cradle of cybercrime activities in sub-Saharan Africa, estimated to cost the economy $649 million annually. Cybercrime activities have other impacts, such as impersonation and plagiarism. Given the scale of these impacts, the authors argue that it is imperative to develop adequate capacity to handle emerging issues in cybercrime. They investigated cybercrime preparedness in Nigeria and found that the country has insufficient legal and educational preparedness to mitigate the rising incidences of cybercrime. Nigeria has only one cybercrime act passed in 2015 and five universities that offer cybercrime-specific courses.

In their paper, Colin et al. [21] argue that cybercrime is a complex phenomenon that involves both technical and human aspects, and that the information security and environmental criminology communities have been studying the problem from separate angles. Despite the large amount of work produced by these communities, they have largely remained disjoint, with little cross-fertilization of ideas. The authors suggest that it would be beneficial for the information security community to incorporate the theories and systematic frameworks developed in environmental criminology to develop better mitigations against cybercrime. They provide an overview of the research from environmental criminology and its application to cybercrime, survey research proposed in the information security domain, and draw explicit parallels between the proposed mitigations and environmental criminology theories. Finally, they propose a framework to define cyberplaces as a potential research direction for interdisciplinary efforts in cybercrime research.

Shewangu et al. [22] propose a conceptual study of cyber-banking fraud to mitigate risk. They examine key role players and elements in electronic and online fraud risk management, including e-fraud victims, fraudsters, guardians (banks), environmental factors, and fraud types. The paper concludes by proposing a model to aid financial institutions in mitigating cyber fraud risk by assimilating all the pertinent elements into a proposed model.

Shafiya et al. [23] reveal how email databases are continually updated with the inclusion of active email addresses collected by hackers and spammers for their illicit purpose. They investigate and expose a bot-based technique for email address harvesting from email messages, including chain email messages and emails sent to multiple recipients. The paper includes experimentation results demonstrating the bot's effectiveness in misusing technologies to collect email addresses, and suggests mechanisms that can be put in place to prevent this type of email address harvesting significantly. The paper also proposes a mitigation method to detect and mitigate the designed Bots of this nature.

Abid et al. [24] note the exponential increase in the usage of the internet, including social media platforms like Facebook. They present a novel model to detect and prevent Social Engineering Based Phishing Attacks (SEBPA) on Facebook, which is the first attempt to address SEBPA. They validate the proposed model using four realistic scenarios and suggest that the proposed model provides comprehensive coverage of nearly all occurrences of SEBPA. The proposed model also prefigures the threatening situation to users with different colors during the validation process.

According to Guo-Hua et. al. [25], the rapid growth of e-commerce in recent years has led to a significant number of spammers disrupting the fair order of e-commerce platforms. The ratings provided by these spammers do not match the quality of items, causing confusion regarding good and bad items and posing a serious threat to the interests of merchants and normal users. Although several effective spamming detection algorithms have been proposed, they are not effective in determining the trustworthiness of users with insufficient rating data. To address this issue, the authors propose a method inspired by traditional recommender systems, which completes missing ratings of low-degree users using user similarity to enhance the efficiency of spamming detection algorithms when approaching those users. They propose a new reputation ranking method, IOR_LU, and test their approach against DR, IGR, and IOR on three typical datasets. The experimental results demonstrate that their method combined with IOR has improved by at least 9.68%, 3.29%, and 0.21% in dealing with malicious spammers, respectively. Additionally, their method improves by at least 5.06%, 21.12%, and 4.46%, respectively, in detecting random spammers.

Wahyudi et. al. [26] state that content management systems (CMS) are widely used to create and administer web applications, including e-commerce. However, there is limited knowledge on what framework can guide the selection of appropriate CMS for building a quality e-commerce. To address this gap, the authors selected five CMS e-commerce (WP e-commerce, Woocommerce, VirtueMart, Prestashop, and OpenCart), deployed and customized them, and hosted them on an online server. They then tested the systems against quality criteria and metrics, providing empirical evidence on the applicability of the proposed model for gaining insights into the quality of e-commerce CMS.

J. Efrim et. al. [27] discuss the inherent trade-off between the necessity of providing personal information to consummate an online transaction and the risk of negative consequences from providing such information in e-commerce. They state that the requirement and increased sophistication of companies' personal information gathering have made e-commerce privacy a critical issue, leading to a broad research literature that is reviewed in their paper. They organize key research issues and findings using a framework defined by four key stakeholder groups—companies, customers, privacy solution providers (PSPs), and governments—as well as the interactions among them. The review highlights that published research on e-commerce privacy peaked in the early 2000s and has not addressed many of the technological advances and other relevant developments of the past decade. The authors suggest potential research opportunities for researchers in Management Information Systems (MIS) and Accounting Information Systems (AIS) related to company privacy strategies, operations, disclosures, and compliance practices, customer privacy concerns, privacy-enhancing technologies, controls, and assurance practices developed by PSPs, and privacy regulations relating to various industries, countries, and cultures. They encourage more use of experimental and archival research in this area.

According to Greg et. al. [28], E-commerce has faced numerous challenges since its inception, one of which is the lack of user trust created by the risk of phishing. This paper explores the vulnerability of e-commerce to phishing attacks and the methods and techniques used in phishing, such as phishing emails, websites and addresses, distributed attacks, redirected attacks, and the data that phishers seek to obtain. The paper also explores ways to reduce the risk of phishing and increase trust between users and websites, including the importance of Trust and the Uncertainty Reduction Theory and the balance between trust and control. The paper concludes by presenting Critical Success Factors for phishing prevention and control, such as User Authentication, Website Authentication, Email Authentication, Data Cryptography, Communication, and Active Risk Mitigation.

D. Harrison et. al. [29] present evidence that consumers often hesitate to transact with Web-based vendors due to uncertainty about vendor behavior or the perceived risk of having personal information stolen by hackers. Trust plays a central role in helping consumers overcome perceptions of risk and insecurity in e-commerce. This paper proposes and validates measures for a multidisciplinary, multidimensional model of trust in e-commerce, including four high-level constructs and 16 measurable subconstructs. The paper demonstrates the psychometric properties of the measures through use of a hypothetical, legal advice Web site and proposes relationships among the trust constructs as well as relationships between the trust constructs and three other e-commerce constructs.

Bharathipriya et. al. [30] propose the use of a recommender system to filter, prioritize, and personalize appropriate information to increase e-commerce demand. The proposed collaborative filtering approach generates optimal product selection by dynamically solving voluminous data. The recommender system focuses on obtaining a similar group of customers using a novel method, and personalized customer product recommendation is obtained using classification and clustering algorithms. The quality of the product evaluation is measured using metrics like root mean square error (RMSE) and mean square error (MSE). The paper concludes that the recommender system enhances the quality of the decision-making procedure and has a great impact on people’s decision-making.

Shubham et. al. [31] highlight that most businesses and services in the current internet world rely on cloud services for computing and data storage, raising concerns about the safety, reliability, and security of cloud environments. The paper analyzes various Distributed Denial of Service (DDOS) attacks in cloud environments and their mitigation strategies, and proposes a methodology/algorithm that is completely cloud-based, removes some limitations from earlier technologies, and provides access to legitimate users 24/7 while blocking unauthorized users. The paper discusses all types of DDOS attacks and their mitigation strategies.

In a study by Misael et. al. [32], the authors explore the increase in online goods and services worldwide and examine the case of Tango Discos, a small entertainment company in Colombia that receives customer messages through various channels. The dataset of 29,970 messages collected from 2019 to 2021 is categorized as a sale, request or complaint, and different supervised classification models are evaluated to automate the message classification task. As the data set is unbalanced, the different models are evaluated with various data balancing approaches. The best model is the Linear Support Vector Machine using the Random Over Sampler balancing technique, which minimizes false positives in the sales category. This model is deployed in the cloud and exposed through a RESTful interface.

Emmanouil et. al. [33] discuss the growth of e-commerce solutions and the increasing access to web applications. The authors review a variety of fusion-based solutions used in e-commerce applications to address challenges such as data source inaccuracy and unreliability.

A 4-fold categorization approach is introduced, including product-related, economy-related, business-related, and consumer-related solutions. Relevant subcategorizations are based on the challenges faced by e-commerce. The paper includes 65 fusion-related solutions, with a great variety of different fusion applications and machine learning techniques focused on the same e-commerce-related challenge.

Rajesh et. al. [34] focus on exploring the applications of machine learning in e-commerce, as the industry has experienced a rapid growth in online purchases. The authors review the usage of machine learning techniques in various e-commerce applications, such as Product Recommendations, Dynamic Price Adjustment, Supply and Demand Prediction, Fraud Detection and Segmentation, Personalization, and Targeting.

Linda et. al. [35] describe how technology can be used for illegal and unethical activities in e-commerce, such as spying, theft, and spreading false information. The authors discuss the use of bots as one of the leading-edge tools of e-commerce and online information exchange, identifying available security counter-measures to defend against attacks.

Alexy et. al. [36] present a comprehensive review of the most effective content-based e-mail spam filtering techniques, focusing primarily on machine learning-based spam filters and their variants. The paper examines the basics of e-mail spam filtering, the evolving nature of spam, spammers playing cat-and-mouse with e-mail service providers (ESPs), and the machine learning front in fighting spam. The authors conclude by measuring the impact of machine learning-based filters and explore the promising offshoots of latest developments.

Emine et. al. [37] demonstrated the efficacy of deep learning in solving various machine learning tasks, particularly in image classification using Convolutional Neural Networks (CNNs), which have emerged as a dominant architecture in this field. Applications of image classification include medicine, education, and security, where accurate classification is of utmost importance. Although several image classification algorithms have been developed, CNNs provide superior performance due to their ability to extract hidden features, parallel processing, and real-time operation. In this study, the authors employed a LeNet network model and the caffe library to classify cat and dog images from the kaggle dataset.

Lin et. al. [38] proposed a regional credibility evaluation model based on big data to enhance the authenticity and reliability of evaluation data for reflecting the regional credit level in the business platform. The model integrates behavior characteristics of regional economic development and the judgment of the credibility of comments. The authors analyzed regional economic behavior and proposed regional activity, reliability, and contribution as features, while the credibility of comments was based on the quality of regional text. The model features two sets of features: regional behavior and regional text quality, which comprehensively evaluate regional social relations, the quality of regional comments, and regional behavior credibility, solving the problems of incomplete features and data cold start in regional credibility evaluation models.

Mohammed et. al. [39] discussed the rise of spoofed content on the internet, including fake news and fake reviews, and the challenges of distinguishing between genuine and fake reviews. The authors compared two tools for detecting opinion spamming using 100 reviews from travelers who evaluated Istanbul's top five best-value hotels.

Vianka et. al. [40] highlighted the growth of e-commerce sales, which reached $25.6 trillion globally in 2018, up 8% from 2017, and the increasing popularity of e-commerce during the COVID-19 pandemic. However, e-commerce startups face various financial risks, such as fraud, cybersecurity, payment transactions, taxation, and insurance issues, which can impact their success. The authors explored different e-commerce business models and associated financial risks to educate students about the perils of operating in the cyber business space.

In their study, Bou-Harb et al. [41] suggested that the 4G mobile network would be a user-centric system that incorporates various transmission technologies, which could expose the network to serious IP-based attacks. They proposed a distributed architecture for the LTE network that could efficiently mitigate such attacks by solving the over dimensioning problem and using low-cost hardware in the distributed nodes. Through simulations and analysis, they demonstrated the feasibility and effectiveness of their approach in itigating the impact of SMTP SPAM flooding attacks on the LTE network.

Rong-Ruey et al. [42] highlighted the importance of establishing control in e-commerce to gain acceptance and trust of participants, which requires expanding the traditional view of internal control to encompass the activities of customers, suppliers, and other “outside” users. They presented a framework for analyzing control in online auctions and suggested ways of controlling risks such as privacy, authentication, and denial-of-service attacks, using eBay's control practices as an illustrative example. They also analyzed assurance services available in 2002 for privacy, integrity, and security of online transactions and discussed challenges and opportunities facing existing services such as WebTrust.

Haitao et al. [43] addressed the issue of web bot traffic, which consumes considerable resources at web servers, resulting in high workloads and longer response time, while not bringing in any profit. They proposed an efficient approach to detect web bot traffic in a large e-commerce marketplace and characterized the behavioral patterns of web bots different from normal users. Their detection approach consisted of an Expectation Maximization (EM)-based feature selection method, a gradient-based decision tree, and a threshold estimation mechanism. They applied their approach on Taobao/Tmall platforms and demonstrated its effectiveness in identifying a considerable amount of web bot traffic. Their findings provided new insights for public websites to improve web bot traffic detection and protect valuable web contents.

According to Parag et al. [44], e-commerce is a prominent aspect of the IT industry and Artificial Intelligence has greatly contributed to its success by efficiently fulfilling customer demands. AI plays a crucial role in managing large amounts of product data and helps to personalize and retarget potential customers, thus creating a favorable environment for customers to make purchases. This paper explores the application of AI in the e-commerce sector, its role in the industry, and some fundamental e-commerce models.

In a study by Tulsi et al. [45], e-commerce applications and email communications are popular in modern society. However, these applications are vulnerable to various types of attacks if proper security protocols are not in place. For instance, the MITM attack and spoofing attacks are major threats to e-commerce and email applications. The paper reviews the SSL and TLS protocols and commonly used security attacks for these applications. The authors implemented an HTTP and HTTPS web server for the e-commerce application and secured it using the SSL/TLS protocol. They also implemented a Javamail API for the email application and secured it using the TLS protocol to prevent MITM attacks. The results showed that with

SSL/TLS protocol enabled, data transmission was encrypted, and MITM attacks were successfully blocked, and spoofing attacks were also mitigated.

Sutirtha et al. [46] highlight the popularity of e-commerce due to open competition and minimal barriers to entry. However, recent surveys indicate a lack of consumer confidence in transacting online. This paper investigates the core inefficiencies in e-commerce that prompt such consumer uncertainty, including seller anonymity, lack of product transparency, and lack of process transparency. The authors argue that consumer uncertainty is not an inherent buyer characteristic but is contingent on the information specificity of products that consumers transact in B2C and C2C e-commerce. By integrating behavioral economics, the paper proposes a novel perspective on electronic market inefficiencies and their effects on consumer uncertainty. In addition to proposing a model of consumer uncertainty in e-commerce, the study provides empirical validation of the proposed model, which suggests that anonymity and lack of product and process transparencies cause consumer uncertainty. The findings further reveal that buyer uncertainty increases when purchasing products with high information specificity, especially when product transparency is lacking.

In their research paper, Mohammad et al. [47] discuss how confidentiality and availability are the primary concerns for website stakeholders, and improving security is necessary despite service-security trade-offs. While the commonly used CAPTCHA security framework is considered secure by many, attackers can now break it in milliseconds using innovative technologies such as deep learning. To overcome this issue, the authors propose a robust and simple image-based CAPTCHA using an Arabic scheme and noise to make it difficult to break even for deep learning approaches. The proposed model is evaluated for security and usability using an empirical experiment with 16 participants and breaking the model using Convolutional Neural Network (CNN). The results demonstrate the superiority of the proposed model.

Glorin et al. [48] argue that organizations struggle to find the right balance between flexibility and security for remote work. With a significant portion of the workforce working from home due to the COVID-19 pandemic, employers may not have prepared adequately for this scale of load on their IT infrastructure and cybersecurity. Cybercriminals have exploited this unpreparedness, and the authors propose a WFH cyber-attack mitigation framework with eight simple yet effective steps to mitigate and prevent cyber-attacks.

Yusuf et al. [49] highlight that developing countries are transitioning from cash to electronic-based economies, leading cybercriminals to exploit possible loopholes in the electronic payment system to perpetuate fraud. To prevent unauthorized online banking withdrawal and transfer and detect phishing attempts, the authors propose an enhanced approach that uses Semantics Content Analysis, Earth Mover Distance, and Biometric Authentication with fingerprint to construct a model. The model's efficacy is demonstrated through experiments conducted, achieving good and considerable results.

In their paper, Samuel et al. [50] argue that e-commerce can be a curse for brand names due to information asymmetries and the existence of indifferent consumers, leading to counterfeiting of branded products and infringement of brand names in e-markets. Legal measures may not be efficient in dealing with brand name problems in e-markets. The authors propose market mechanisms such as information syndication, pricing of e-markets services, and vendor malpractice as effective measures in deterring counterfeiting and brand name infringement, which can reduce sales and profits of brand name holders. They hope their positional contribution will inspire others to find innovative mechanisms to safeguard online transactions.

2.1 Types of Review Spam

Review spam can take many forms, including fake reviews, biased reviews, incentivized reviews, and reviews from competitors. Fake reviews are written by individuals who have never used the product or service, and are often created by businesses or individuals looking to manipulate the perceived quality of the product or service. Biased reviews are written by individuals with a vested interest in the product or service, such as employees, friends or family members of the business owner, or paid reviewers. Incentivized reviews are written by individuals who have received incentives, such as discounts or free products, in exchange for writing a positive review. Finally, reviews from competitors are written by individuals affiliated with a competing business, with the intention of damaging the reputation of their competitor.

2.2 Impact of Review Spam on E-commerce

The prevalence of review spam has a significant impact on e-commerce, affecting both consumers and online retailers. For consumers, review spam can lead to inaccurate information about the quality of products and services, influencing purchasing decisions and potentially leading to negative experiences. For online retailers, review spam can damage their reputation, reducing customer trust and ultimately leading to a loss of business.

2.3 Mitigating Review Spam on E-commerce Sites

Mitigating review spam on e-commerce sites is a complex process that requires a combination of strategies and techniques. One effective approach is the use of machine learning algorithms, which can analyze large amounts of data to identify patterns and anomalies in reviews. Machine learning algorithms can be trained to identify specific characteristics of review spam, such as excessive use of keywords or similarities in language between reviews, allowing for automated detection and removal of spam reviews. However, the use of machine learning algorithms requires continuous monitoring and tuning to ensure their effectiveness.

Another approach is human review, which involves manual review of reviews to identify and remove spam reviews. Human review can be time-consuming and labor-intensive, but it allows for a more nuanced analysis of reviews, and can help identify patterns or trends that machine learning algorithms may not detect.

Finally, the implementation of review policies and guidelines can help prevent review spam by providing clear rules and expectations for reviewers. Policies and guidelines can include requirements for verified purchases, restrictions on incentivized reviews, and prohibitions on biased or fake reviews. By setting clear expectations for reviewers, online retailers can reduce the likelihood of review spam, and promote more accurate and helpful reviews.

2.4 Review Spam: Definition and Impact

Review spam refers to fake or fraudulent reviews that are designed to manipulate the perceived quality of products or services. Review spam can take many forms, including fake positive reviews, negative reviews posted by competitors, and reviews posted by individuals who have not actually used the product or service. Review spam can have a significant impact on e-commerce, with the potential to:

- Manipulate the perceived quality of products and services: Review spam can create a false impression of the quality of products and services, leading to customers making purchasing decisions based on inaccurate information.
- Influence customer purchasing decisions: Review spam can influence customer purchasing decisions by either promoting or dissuading customers from purchasing a product or service.
- Damage the reputation of online retailers: Review spam can damage the reputation of online retailers by creating a perception that their review systems are not trustworthy, leading to a loss of customer trust.

2.5 Strategies and Techniques for Mitigating Review Spam

Mitigating review spam requires a combination of strategies and techniques, including the use of machine learning algorithms, human review, and the implementation of review policies and guidelines.

2.6 Case Studies

There have been several case studies of successful review spam mitigation on e-commerce sites. For example, Amazon has implemented several strategies to detect and remove review spam, including machine learning algorithms and human review. Amazon also requires reviewers to have a verified purchase before submitting a review, and prohibits incentivized reviews. As a result of these strategies, Amazon has been able to maintain a high level of trust in their review system, with 90% of customers stating that they trust Amazon reviews.

Similarly, Trip Advisor has implemented a range of strategies to mitigate review spam, including machine learning algorithms, human review, and review policies and guidelines. Trip Advisor uses machine learning algorithms to identify suspicious reviews, and human review to verify these suspicions and remove spam reviews. Trip Advisor also prohibits incentivized reviews and requires reviewers to have a verified stay before submitting a review. As a result of these strategies, Trip Advisor has been able to maintain a high level of trust in their review system, with 90% of customers stating that they trust Trip Advisor reviews.

2.7 Machine Learning Algorithms

Machine learning algorithms can be used to detect patterns of review spam and identify suspicious reviews. These algorithms can analyze various features of reviews, such as the language used, the length of the review, and the reviewer's behavior, to identify potential instances of review spam. Machine learning algorithms can also be used to identify clusters of reviews that have similar patterns, which may indicate a coordinated effort to post review spam. However, machine learning algorithms are not perfect, and there is always the risk of false positives and false negatives.

2.8 Human Review

Human review involves having trained individuals review suspicious reviews to determine if they are legitimate or spam. Human review can be more effective than machine learning algorithms in detecting review spam, as humans are better able to identify context and nuance in reviews. However, human review can be time-consuming and expensive, and may not be scalable for larger e-commerce sites.

3 CONCLUSION

In conclusion, review spam is a growing concern for e-commerce sites, with the potential to manipulate the perceived quality of products and services, influencing customer purchasing decisions, and damaging the reputation of online retailers. Mitigating review spam requires a combination of strategies and techniques, including the use of machine learning algorithms, human review, and the implementation of review policies and guidelines. Through a review of the literature and case studies, this paper has identified key strategies and techniques for mitigating review spam, which can help maintain customer trust in e-commerce review systems and promote accurate and helpful reviews.

However, there is still much research to be done in this area, particularly in developing more sophisticated algorithms for identifying review spam, as well as exploring the effectiveness of different strategies for promoting authentic reviews.

Future research could also explore the impact of review spam on customer behavior, such as the extent to which customers are influenced by fake reviews and how this influences their purchasing decisions. Additionally, research could explore the role of social media in spreading review spam and how e-commerce sites can leverage social media to detect and mitigate review spam.

Overall, mitigating review spam is a critical issue for e-commerce sites, and implementing effective strategies and techniques is essential for maintaining customer trust in review systems. By staying up to date with the latest research and continuously refining and improving review spam mitigation strategies, e-commerce sites can promote authentic and helpful reviews that accurately reflect the quality of products and services, benefitting both customers and retailers alike.

REFERENCES

1. Mamoona Humayun, NZ Jhanjhi, Ahmed Alsayat, Vasaki Ponnusamy, “Internet of things and ransomware: Evolution, mitigation and prevention”, Egyptian Informatics Journal 22 (2021) 105–117.
2. Sunghoon Lim, Conrad S. Tucker, “Mitigating Online Product Rating Biases through the Discovery of Optimistic, Pessimistic, and Realistic Reviewers”.
3. Sumit Badotra, Amit Sundas, “A systematic review on security of E-commerce systems”, International Journal of Applied Science and Engineering, Vol. 18(2) 2020323, https://doi.org/10.6703/IJASE.202106_18(2).010.
4. Michael Crawford, Taghi M. Khoshgoftaar, Joseph D. Prusa, Aaron N. Richter and Hamzah Al Najada, “Survey of review spam detection using machine learning techniques”, Crawford et al. Journal of Big Data (2015) 2:23, DOI 10.1186/s40537-015-0029-9.

5. Charlie Walker, Scott Buttinger, Sam Wharton, “Towards Mitigating Bias in Online Reviews: An Application to Amazon.com”.
6. Tanmoy Chakraborty and Sarah Masud, “Nipping in the Bud: Detection, Diffusion and Mitigation of Hate Speech on Social Media”, arXiv:2201.00961v1 [cs.SI] 4 Jan 2022.
7. Hamzah Al Najada and Xingquan Zhu, “iSRD: Spam Review Detection with Imbalanced Data Distributions”, IEEE IRI 2014, August 13-15, 2014, San Francisco, California, USA.
8. Haitao Xu, Daiping Liu, Haining Wang, Angelos Stavrou, “E-commerce Reputation Manipulation: The Emergence of Reputation-Escalation-as-a-Service”, WWW 2015, May 18–22, 2015, Florence, Italy. ACM 978-1-4503-3469-3/15/05. http://dx.doi.org/10.1145/2736277.2741650.
9. T. Ravindra Babu, Vishal Kakkar, “Address Fraud: Monkey Typed Address Classification for e-Commerce Applications”, Proceedings of the SIGIR 2017 eCom workshop, August 2017, Tokyo, Japan, published at http://ceur-ws.org.
10. Andrew J. Flanagin, Miriam J. Metzger, Rebekah Pure, Alex Markov, Ethan Hartsell, “Mitigating risk in ecommerce transactions: perceptions of information credibility and the role of user-generated ratings in product quality and purchase intention”, Electron Commer Res (2014) 14:1–23, DOI 10.1007/s10660-014-9139-2.
11. Raymond Y. K. Lau, S. Y. Liao and Ron Chi-Wai Kwok, Kaiquan Xu, Yunqing Xia, Yuefeng Li, “Text Mining and Probabilistic Language Modeling for Online Review Spam Detection”, 2011 ACM 2158-656X/2011/12-ART25 $10.00, DOI 10.1145/2070710.2070716 http://doi.acm.org/10.1145/2070710.2070716.
12. Li He, Xianzhi Wang, Hongxu Chen, Guandong Xu, “Online Spam Review Detection: A Survey of Literature”, Human-Centric Intelligent Systems (2022) 2:14–30, https://doi.org/10.1007/s44230-022-00001-3.
13. Shugang Li, Fang Liu, Yuqi Zhang, Boyi Zhu, He Zhu and Zhaoxu Yu, “Text Mining of User-Generated Content (UGC) for Business Applications in E-Commerce: A Systematic Review”, Mathematics 2022, 10, 3554. https://doi.org/10.3390/math10193554.
14. Yehoshua Liebermann and Shmuel Stashevsky, “Perceived risks as barriers to Internet and e-commerce usage”, Qualitative Market Research: An International Journal, Volume 5, Number 4, 2002 . 291-300.
15. Wosah Peace Nmachi and Thomas Win, “Mitigating Phishing Attack in Organisations: A Literature Review”, pp. 75-83, 2021. CS & IT - CSCP 2021 DOI: 10.5121/csit.2021.110105.
16. Gourab Nath, Anagha Karanam, Rohit Akkenapalli, Sandilya Machiraju, “Network Based Approach to Detect Spam Reviews: A Critical Analysis”.
17. Haiqin Weng, Shouling Ji, Fuzheng Duan, Zhao Li, Jianhai Chen, Qinming He, Ting Wang, “CATS: Cross-Platform E-commerce Fraud Detection”, 2019 IEEE 35th International Conference on Data Engineering (ICDE)”, 2375-026X/19/$31.00 ©2019 IEEE, DOI 10.1109/ICDE.2019.00203.
18. Thomas Ormerod, Lingyu Wang, Mourad Debbabi, “Defaming Botnet Toolkits: A Bottom-Up Approach to Mitigating the Threat”.
19. Yen-Hsien Lee, Paul Jen-Hwa Hu, Tsang-Hsiang Cheng, Ya-Fang Hsieh, “A cost-sensitive technique for positive-example learning supporting content-based product

recommendations in B-to-C e-commerce", Decision Support Systems 53 (2012) 245–256.

20. Attila Mátė Kovács, "Here there be Dragons: Evolution, Potentials and Mitigation Opportunities of Cybercrime in Nigeria A Review, Analysis, and Evaluation", JCEEAS – Journal of Central and Eastern European African Studies – ISSN 2786-1902.
21. Colin C. Ife, Toby Davies, And Steven J. Murdoch, Gianluca Stringhini, "Bridging Information Security and Environmental Criminology Research to Better Mitigate Cybercrime", arXiv:1910.06380v2 [cs.CR] 14 Jul 2022.
22. Shewangu Dzomira, "Cyber-banking fraud risk mitigation- conceptual model", Banks and Bank Systems, Volume 10, Issue 2, 2015.
23. Shafiya A. Sheikh, M. Tariq Banday, "Mitigating BOT-based Methods for Email Address Harvesting and Spamming", August 19th, 2022, DOI: https://doi.org/10.21203/rs.3.rs-1954011/v1.
24. Abid Jamil, Kashif Asif, Zikra Ghulam, Muhammad Kashif Nazir, Syed Mudassar Alam, Rehan Ashraf, "MPMPA: A Mitigation and Prevention Model for Social Engineering Based Phishing attacks on Facebook", 2018 IEEE International Conference on Big Data (Big Data).
25. Guo-Hua Li, Jun Wu and Hong-Liang Sun, "Identifying Spammers by Completing the Ratings of Low-degree Users", September 2022.
26. Wahyudi Agustiono, "An Open Source Software Quality Model and Its Applicability for Assessing E-commerce Content Management Systems", Atlantis Highlights in Engineering (AHE), volume 1, 2018.
27. J. Efrim Boritz, Won Gyun No, "E-Commerce and Privacy: Exploring What We Know and Opportunities for Future Discovery", Journal of Information Systems American Accounting Association Vol. 25, No.2 2011 pp. 11–45.
28. Greg Megaw, Stephen V. Flowerday, "Phishing within E-Commerce: A Trust and Confidence Game", 978-1-4244-5494-5/10/$26.00 ©2010 IEEE.
29. D. Harrison McKnight, Vivek Choudhury, Charles Kacmar, "Developing and Validating Trust Measures for e-Commerce: An Integrative Typology", Information Systems Research, 2002 INFORMS, Vol. 13, No. 3, September 2002, pp. 334–359.
30. C. Bharathipriya, B. Swathi, X. Francis Jency, "Product Recommendation Framework Based on Customer Review Using Collaborative Filtering Techniques", J. Mech. Cont.& Math. Sci., Special Issue, No.- 7, February (2020) pp 58-71.
31. Shubham Kumar Sahu, Dr. R. K. Khare, "DDOS Attacks & Mitigation Techniques in Cloud Computing Environments", Gedrag & Organisatie Review - ISSN:0921-5077, VOLUME 33 : ISSUE 02 – 2020.
32. Misael Andrey Alba˜nil S´anchez, and Ixent Galpin, "Classifying Incoming Customer Messages for an E-Commerce Site using Supervised Learning".
33. Emmanouil Daskalakis, Konstantina Remoundou, Nikolaos Peppes, Theodoros Alexakis , Konstantinos Demestichas, Evgenia Adamopoulou and Efstathios Sykas, "Applications of Fusion Techniques in E-Commerce Environments: A Literature Review", Sensors 2022, 22, 3998. https://doi.org/10.3390/s22113998.

34. M. V. Rajesh and S. Rao Chintalapudi, “A Review on Applications of Machine Learning In E-Commerce”, Advances and Applications in Mathematical Sciences, Volume 20, Issue 11, September 2021, Pages 2831-2841.
35. Linda A. Bressler, Martin S. Bressler, “Beware the evil bots: e-commerce thieves and spreaders of “fake news”, Journal of Technology Research Volume 8, January 2019.
36. Alexy Bhowmick, Shyamanta M. Hazarika, “Machine Learning for E-mail Spam Filtering: Review, Techniques and Trends”, arXiv:1606.01042v1 [cs.LG] 3 Jun 2016.
37. Evangelos Moustakas, C. Ranganathan, Penny Duquenoy, “E-mail marketing at the crossroads A stakeholder analysis of unsolicited commercial e-mail (spam)”, Internet Research, Vol. 16 Iss 1 pp. 38 – 52 Permanent link to this document: http://dx.doi.org/10.1108/10662240610642532.
38. Lin Li, FangQin, Can Wang, Jianyan Sun, Wei Jia Zeng, Lin Lin Yu, “Regional Development of E-Commerce Based on Big Data Evaluation Model”, Journal of Physics: Conference Series 1883 (2021) 012113, IOP Publishing, doi:10.1088/1742-6596/1883/1/012113.
39. Mohammed Awad, Khouloud Salameh, Assamahou Malika Ngoungoure, Maryam Abdullah, “Opinion Spamming: Analyzing the Accuracy of Online Detection Tools”, CEEeGov, September 22, 23, 2022.
40. Vianka Esteves Miranda, Elizabeth Abington Prejean, Carmella Parker, Weiwen Liao, “Exploring Financial Risk Management for E-Commerce Startups”, American International Journal of Business Management (AIJBM), ISSN- 2379-106X, www.aijbm.com Volume 5, Issue 12 (December-2022), PP 112-123.
41. Elias Bou-Harb, Makan Pourzandi, Mourad Debbabi and Chadi Assi, “A secure, efficient, and cost-effective distributed architecture for spam mitigation on LTE 4G mobile networks”, Security and Communication Networks, Security Comm. Networks 2013; 6:1478–1489.
42. Rong-Ruey Duh, Shyam Sunder, “Control and Assurance In e-Commerce: Privacy, Integrity, and Security at eBay”, September 13, 2002.
43. Haitao Xu, Zhao Li, Chen Chu, Yuanmi Chen, Yifan Yang, Haifeng Lu, Haining Wang, and Angelos Stavrou, “Detecting and Characterizing Web Bot Traffic in a Large E-commerce Marketplace”.
44. Parag G Gidh, “A Multi-Dimensional Research Study in E-Commerce to Capture Consumer Expectations”, International Journal for Research in Applied Science & Engineering Technology (IJRASET), Volume 8 Issue XI Nov 2020.
45. Tulsi Pawan Fowdur, Muhammad Shafeeq Aumeeruddy, Yogesh Beeharry, “Implementation of SSL/TLS-based security mechanisms in e-commerce and e-mail applications using Java”, Journal of Electrical Engineering, Electronics, Control and Computer Science –, JEEECCS, Volume 4, Issue 11, pages 13-26, 2018.
46. Sutirtha Chatterjee, Pratim Datta, “Examining Inefficiencies and Consumer Uncertainty in E-Commerce”, Communications of the Association for Information Systems: Vol. 22, 2008 Article 29.
47. Mohammad Al-Fawa'reh, Malik Qasaimeh, Ibrahim Abu Arja, Mustafa Al-Fayoumi, “Mitigating Deep learning Attacks Against Text Image CAPTCHA Using Arabic

Scheme", International Journal on Communications Antenna and Propagation (I.Re.C.A.P.), Vol. 11, N. 4, ISSN 2039 – 5086 August 2021.

48. Glorin Sebastian, "Descriptive Study on Cybersecurity Challenges of Working from Home during COVID-19 Pandemic and a Proposed 8 step WFH Cyber-attack Mitigation Plan", Communications of the IBIMA, https://ibimapublishing.com/articles/CIBIMA/2021/589235/, Vol. 2021 (2021), Article ID 589235, 7 pages, ISSN: 1943-7765, DOI: 10.5171/2021.589235.
49. Yusuf Simon Enoch, Adebayo Kolawole John, Adetula Emmanuel Olumuyiwa, "Mitigating Cyber Identity Fraud using Advanced Multi Anti-Phishing Technique", (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 4, No. 3, 2013.
50. Samuel Otim & Varun Grover, "E-commerce: a brand name's curse", Electron Markets (2010) 20:147–160, DOI 10.1007/s12525-010-0039-6.