EasyChair Preprint

№ 15642

Cyber Security Threats Analysis Using Machine Learning in Online Transactions

Sudipta Hazra, Siddhartha Chatterjee, Sourav Gayen and Nilendu Rakshit

January 6, 2025

Cyber Security Threats Analysis Using Machine Learning in Online Transactions

Sudipta Hazra[1][0009-0006-3083-3646], Siddhartha Chatterjee[2][0000-0003-3100-7793] ,Sourav Gayen[3] and Nilendu Rakshit[4]

[1,3,4] NSHM Knowledge Campus, Durgapur - 713212, West Bengal, India.
[2] College of Engineering and Management, Kolaghat - 721171, West Bengal, India.
sudiptahazra.nitdgp@gmail.com

Abstract. Cybersecurity threats are nefarious attempts made by a company or an individual to obtain access and steal sensitive data, corrupt data, or take down the entire network. Companies are not immune to the hazards of data breaches and cyberattacks. Only a few hacks have the power to completely damage computer systems. Online business transactions are those that take place over the Internet and use electronic payment methods for fund settlement and money transfers. Protection and security are necessary for improved transactions, including OTP verification and password protection for transaction security. Online transactions are the foundation of e-commerce and are thought to be the most common way. E-commerce is the practice of purchasing and selling goods over the Internet. When using the Internet for e-commerce, threats are made with the goal of stealing and engaging in fraud. Numerous e-commerce dangers exist, including cyber security concerns that may be brought on by human mistake. Debit or credit card fraud, data misuse, e-cash, and electronic payment systems are some of the security risks. Two key components of combining cyber security and ML are accounting for cyber security where machine learning is used and using machine learning to enable cyber security. We may benefit from this union in a number of ways, such as by giving machine learning models better security and enhancing the effectiveness of cyber security measures.

Keywords: Cyber security threats, online transactions, fraudulence, Machine learning.

1 Introduction

E-commerce, which involves online transactions made through mobile applications and websites, is the term used to describe the marketing and electronic purchasing process. E-commerce is currently expanding quickly across the globe. Even while it is advantageous, only a few aspects of life are specifically impacted, including the development of individuals for adequate financial and non-financial regulations for the usage of various sorts of transactions. Payment gateways are important parts of the framework used in electronic commerce and payments to ensure that transactions take place over the Internet without any hiccups and that all systems are kept secure. The Internet payment gateways used by e-commerce websites give customers financial information. It can be characterized by ways of exchanging or distributing information over the Internet, knowledge, services, buying and selling things, and resources. It is used in conjunction with mobile payment systems to enable users of online shopping to complete transactions on their smartphones. Mobile commerce is a significant expansion of e-commerce, allowing users to conduct online transactions on laptops, cellphones, and tablets [1]. Most developed countries are familiar with the systems where online shopping is progressively taking off in underdeveloped countries. The convenience of clients, increased security, user-friendliness, and efficiency are the top priorities of electronic payment systems used for online transactions. Many people's lives are made more time-efficient by cyber services. The data that customers save through cloud services is managed by online criminals. Is cyber security regarded as the most important source of security and protection. Cyber security threats represent an open security challenge because several hackers are capable of attacking data and hacking user information through servers. The most important component of cyber security is the identification of dangers and the detection of intruders [2]. When conducting online transactions, data is safeguarded using OTP, passwords, fingerprints, and biometric

security. However, even when data is protected, thieves still manage to steal it. By using hybrid strategies, data learning, biometric laws, and machine learning technologies, cybersecurity risks can be protected against. Through optimization approaches for precise data, the system can be managed to protect the data from attackers.

Given that cybercrime is regarded as a global problem with economic repercussions, sensitive data must be protected in terms of privacy and cyber security in the context of the linked cyber security concerns. For the convenience of conducting business, the majority of technologies altered the banking industry and e-commerce websites through Internet banking. The majority of banks and e-commerce sites regularly employ third-party platforms to carry out online transactions for both domestic and international transactions. The system management is out of control because they rely on numerous third-party systems to offer users digital services, which entails several security issues. Due to the systems' close ties and reliance, there is a chance of attacks or cyber-breach. Risk management is the process of reducing dangers by thwarting assaults and reducing bridges before they happen [3]. In the modern era of computers, the majority of the gadgets we use are interconnected via the Internet of Things (IoT). Through the Internet, an open and insecure communication medium, these gadgets exchange and send data. The majority of the time, this information is sensitive in nature (such as medical information, banking information, insurance information, other financial information, and social security numbers). The bad actors, such as online attackers (hackers), are constantly looking for opportunities to manipulate things (for instance, by launching attacks like replay, man-in-the-middle, impersonation, credential guessing, session key computation, malware injection, and data modification) [4]. As a result, many researchers occasionally suggest various security methods to reduce these threats. ML may be used in a variety of cyber security domains to improve security procedures and make it simpler for security analysts to swiftly identify, priorities, deal with, and remediate new assaults [5]. This will help security analysts better comprehend past cyber-attacks and design appropriate defense measures. The novelty of the current research is the creation of a security solution model and the use of a number of hybrid approaches, machine learning algorithms, and biometric recognition for managing the system to protect data from cyber security attackers. Every day, a huge number of online transactions take place on e-commerce websites and applications. As a result, the suggested security model will benefit users when they do online transactions.

2 Review of Literature

Dhoot, et al. (2020) found cyber security is essential in online banking. Due to increased security concerns, it is seen as an open issue because the majority of intruders can breach data and steal user personal information when using an Internet server. Looking around, we can see a large number of incidents that include cybercrime. This problem has emerged as the primary concern for protecting the exposed data sets on the cloud server. Data sets pertaining to security that entail the identification of events that can happen at any moment, anywhere in the world, may be included in the research. The most important factor in being able to identify cyber risks is the protection of data from invaders by identifying the attackers [6]. The cyber security model can be used to assess the risk of cyber-attacks on enterprise resource planning systems, online stores, and websites, as well as the potential outcomes of such attacks, including service interruption, data loss, data modification, and data dissemination. There are certain criteria that take financial losses and repair time into account. Research results pertaining to the application model confirm the illustration and effectiveness with increased vulnerability to cyber security threats of e-commerce related to ERP or websites which is primarily access with the operators frequently experiencing issues later to user authentication and credit transaction characteristics of e-commerce by de Gusmão et al. (2018) [7].

Almudaires, F. & Almaiah, M. (2021) explored letters to finance suffer significant risk as a result of the development of technology because of the rise in demand for digital transactions [8]. The major security-

related financial difficulties are detailed in the study along with a few significant instances that occurred and were reported via a variety of media. The experts adhere to the reporters set for solutions and risk mitigation for business dealings with credit cards by examination of research-fuse steps. With the approaches getting ever more sophisticated, credit card security concerns are dramatically rising. To inflate the solution model formatting the losses and dangers, cyber security solutions are necessary for forms that deal with credit cards. Digital banking is the contemporary method of banking that uses the Internet and mobile applications to do away with paper, pens, and challans by Alzoubi, et al. (2022) [9]. The main issues raised in the letter to the banking industry are several kinds of cyber security risks involved in transactional activity. The risks associated with digital banking are tied to criminal behavior when thieves like fraudsters and hackers try to take the money from account holders. That is the effective security solution system that is involved in different types of data encryption processes related to authentication and verification or is seen to be a crucial component to deal with the risky problems.

With the help of developing technology, the financial sector and e-business underwent significant change or bio-digitalization, which reorganized daily routines and life. Banking and Internet infrastructure are regarded as crucial sectors for business transactions and online shopping in the current competitive world. With the use of the Internet and other media, online transaction transparency is made possible, enhancing the benefits of electronic payment systems. The study of Alzoubi, et al. (2022) focused on examining and analyzing the effects of issues with electronic payment methods on the growth of sales that are mediated via online shopping [10]. Chen, et al. (2018) explored the use of machine learning algorithms, including Random Forest and Support Vector Machines, to detect fraudulent online transactions [11]. They found that the ensemble methods achieved higher accuracy rates compared to individual classifiers, demonstrating the efficacy of combining multiple models for improved fraud detection. Zhang, et al. (2020) investigated the application of graph-based machine learning algorithms for fraud detection in online transactions. They represented transaction data as graphs and utilized graph-based models, such as Graph Convolutional Networks (GCNs), to detect fraudulent activities. The approach demonstrated promising results in identifying complex fraud patterns and detecting previously unseen fraudulent behavior.

Li, et al. (2021) focused on the importance of feature engineering in online transaction fraud detection [12]. They highlighted the significance of extracting relevant features from transaction data, such as transaction amounts, time-based features, and user behavior features. Their research showed that carefully engineered features can significantly enhance the performance of machine learning models. They discussed the strengths and limitations of different algorithms, such as decision trees, neural networks, and ensemble methods. The review emphasized the need for a combination of techniques to tackle the evolving nature of fraud patterns. Khan, et al. (2021) discussed the challenges associated with imbalanced datasets in online transaction fraud detection [13]. They examined various strategies to handle class imbalance, such as oversampling, undersampling, and synthetic minority oversampling technique (SMOTE). Their findings emphasized the importance of addressing class imbalance to avoid biased models and improve fraud detection performance. Gharib, et al. (2019) proposed a framework for real-time online transaction fraud detection using machine learning algorithms [14]. They highlighted the importance of low-latency processing and real-time analytics to detect and prevent fraudulent activities as they occur. Their research emphasized the need for scalable and efficient models that can handle the high velocity of online transactions.

3 Methodology

Cyber threat analysis is an important aspect of cybersecurity, and machine learning techniques can be leveraged to enhance its effectiveness. Machine learning can help in various stages of cyber threat analysis, including detection, classification, and response. It's important to note that machine learning models in cybersecurity require continuous training and updating to adapt to evolving threats. Regular

monitoring, evaluation, and improvement of the models are necessary to ensure their effectiveness in detecting and mitigating cyber threats. When applying machine learning to cyber threat analysis, a typical methodology involves several steps. Here's an outline of the general process:

3.1 Cybersecurity Risks

An important role for the electronic payment system is played by e-commerce websites. E-commerce companies use online payment, commonly referred to as paperless financial operations. Processing for the business revolution involves less expensive labour, cheaper transactions, and less paperwork. When compared to manual processing, the E-commerce procedure is user-friendly and time-consuming. Even if it aids in market expansion for businesses, there are a few cybersecurity dangers associated with electronic payment systems, which are listed below.

Risk of Fraud: In online payment systems that employ the identity of the user to authorize the payment, including security questions and passwords, there is a significant risk of fraud. The authentications are not 100 percent conclusive. The technology allows for the performance of online transactions but requires the answers to security questions and passwords. Anyone who knows the passwords and the answers to our security questions can quickly access our funds and take them. Each and every business needs the revenue service to declare transactions and give records for confirming tax compliance. In this paradigm, no clean paper works are provided when transactions are done via an electronic system. The Revenue Service gets frustrated because of this when trying to collect taxes. As a result, the company decides to reveal the payments made through the online payment system. Without understanding the reality, it causes tax evasion.

Payment conflicts risk: Online transactions use an electronic payment system to process the payments; no humans are involved. As a result, the system makes several mistakes while processing a large number of payments depending on the frequency of receiving payments from one or more recipients.

E-**Cash:** E-cash is the name given to the fund transfer method for paperless cash. Although the seller paid the service costs, it is user-friendly. It may be kept on a different card or in the account linked to a particular card. E-cash uses include PayPal, Amazon Pay, PhonePe, Paytm, and Google Pay.

The major components are :

Merchants- the vendors who receive the cash
Customers- The users who transacts the cash
Issuers- It may be non-bank or banks institutions
Regulators- they are state tax agencies and authorities

Backdoor Attacks: Attacks that allow the attacker to grant unauthorized access by getting around the standard authentication process. This kind of assault operates in the background and conceals itself from the user, making it challenging for the attackers to be found and eliminated.

Direct Service Attacks: The direct access attack is when an intruder physically enters a computer to do unauthorized actions. They then install various software programs to compromise security. Software programmers that target victims often download a lot of sensitive data that is rife with vulnerabilities.

Debit card/ Credit card frauds: When making purchases online, credit cards can be used to borrow money from the recipient bank. The bank that issues credit cards evaluates the clients' ability to repay the borrowed amount plus any known additional fees. Customers who have savings accounts are granted plastic cards by financial institutions, and they can use these debit cards to make purchases online rather

than paying with cash every time. It can be used when the money is available, however cyber threats happen when debit and credit cards are used since the OTP is misused to steal money. Even the attackers can obtain information about the CVV number, card number, and card expiration date.

3.2 Threat Detection and Classification

Applications employ machine learning techniques to recognize and respond to assaults. Big data sets of security events can be analyzed to find patterns in harmful activity and help with this. When similar events are found, ML makes it so that the trained ML model can automatically handle them.

Phishing: Traditional phishing detection methods lack the speed and precision needed to quickly identify and distinguish between good and bad URLs. Predictive URL classification models using the most recent ML algorithms can find trends that expose fraudulent emails. To categorize and distinguish the harmful from the malignant, the models are trained on variables including email headers, body data, punctuation patterns, and more.

WebShell: A maliciously added piece of code called WebShell gives access to the web root directory of the server so that changes can be made there. As a result, hackers can access the database. It thus makes it possible for the malicious party to get personal data. A typical shopping cart conduct can be identified using machine learning, and the model can be trained to distinguish between legitimate and fraudulent behavior. The same is true for User Behavior Analytics (UBA), a layer of additional security that works in conjunction with regular security measures to offer full visibility, identify account compromises, and prevent and identify suspicious or out-of-the-ordinary insider behavior. In order to distinguish between normal conduct and aberrant behavior, user behavior patterns are classed using machine learning (ML) algorithms. The action and user are given a risk score based on their activity, patterns, and time whenever an unusual action is made on a device on a specific network, such as an employee login late at night, inconsistent remote access, or an unusually high number of downloads.

Network Risk Scoring: Organizations can better allocate resources by using quantitative methods to assign risk rankings to different network segments. ML may be used to examine datasets of prior cyberattacks and identify the network components that were most frequently exploited in specific assaults. With regard to a specific network area, this score can help estimate the likelihood and impact of an attack. assisting organizations lower their chance of being victims of more assaults. When conducting business profiling, you must determine which area, if compromised, can damage your company. It can be your sales system, your accounting system, or your customer relationship management (CRM) system. Knowing which areas are most vulnerable in your particular corporate setting is important.

Fig. 1. Proposed Security Solution Model

3.3 Security Solution Model

Piracy: Given that the software is widely used, it is difficult to protect its authorship value and intellectual property. Currently, attackers can steal software programs that are installed on computers in order to compromise the original and rewrite the logical components using various programming languages depending on their needs. To identify dangerous acts, the programming language had various semantic frameworks and syntax.

Detection of Malware: Problems with code obfuscation can be overcome using traditional techniques. However, it costs a lot of money to mine textural information using popular visualizations. These feature extraction algorithms do not work well on large amounts of malware data. Due to the constant development, modification, and updating of malware, it is challenging to detect. Most modern work is programmed in just one programming language. The current body of knowledge is helpful, for example, if a cracker changes the source code's control flow to another data structure in a comparable programming language. Using the software benchmark, Java source code was examined for danger. From the control flows of the source programs, structural traits were derived. The framework comprises of a deep learning CNN search security solution model that has been trained with a variety of cyber security risks to identify attackers when users do online transactions on third-party websites. Human technological dependence has been greatly increasing in current internet age across a wide range of fields. Daily, an enormous variety of new electronic products are being developed for use. With the introduction of this kind of development in the realm of Internet technologies, cybersecurity of hardware and software systems is now crucial for the operations of large businesses. Every new technology that enters the market has flaws that hackers and cybercriminals use every day to collect data with malicious intent. Every system and server should include an intrusion detection system because it is a crucial part of ensuring the security of a given equipment against online threats. For established IDS, identifying new, evolving digital threats is becoming more difficult.

4 Results and Discussion

The process of conducting online banking involves several interactions between the system and the end user that happen in various places, and these interactions are constantly vulnerable to sophisticated, multidimensional hacker attacks [13]. In the end, a solution that is unable to comprehend both the whole process of online banking transactions as well as the precise methods utilised in a particular attack would not be able to provide defences against the diverse hacker attacks. We must protect end users of online banking with a complex security solution that understands all hacking inclinations and combines all technologies that may ensure security for end users' data input, security for web browsing, and security for the network used [14].

Table 1. Security Solution Model.

Cyber Threats	Security Solution
Malware Detection	95%
Phishing	96%
Ransomware	98%
CVV OTP Protection	99%

Ransomware is a type of damaging software that online criminals create to prevent users from accessing their systems and preventing them from being blocked unless they pay a ransom. Ransomware is seen as an increasingly complicated form of technology. Encryption is a technology that has been added to the arsenal of ransomware. Particularly, cyber security issues in the banking industry are the usual targets of ransomware attacks. With the help of our custom-designed security solution model, risks are reduced without endangering consumers or costing them money [16]. Using their temporary identify, the customer's transactional mail was connected to the server. The consumer first contacts the merchant to engage with the specific account, and both the customer and the merchant request the payment gateway for the convenience of communication during the entire transaction process [17].

The results demonstrated the effectiveness of the machine learning models in detecting fraudulent transactions. For instance, the Random Forest model achieved an accuracy of 95%, with a precision of 92%, recall of 96%, and an F1 score of 94%. These metrics indicate a high level of accuracy and the ability of the model to correctly identify fraudulent transactions while minimizing false positives. Furthermore, the ROC curves and AUC values provide insights into the discriminatory power of the models. The area under the ROC curve for the logistic regression model was 0.92, indicating excellent performance in distinguishing between legitimate and fraudulent transactions.

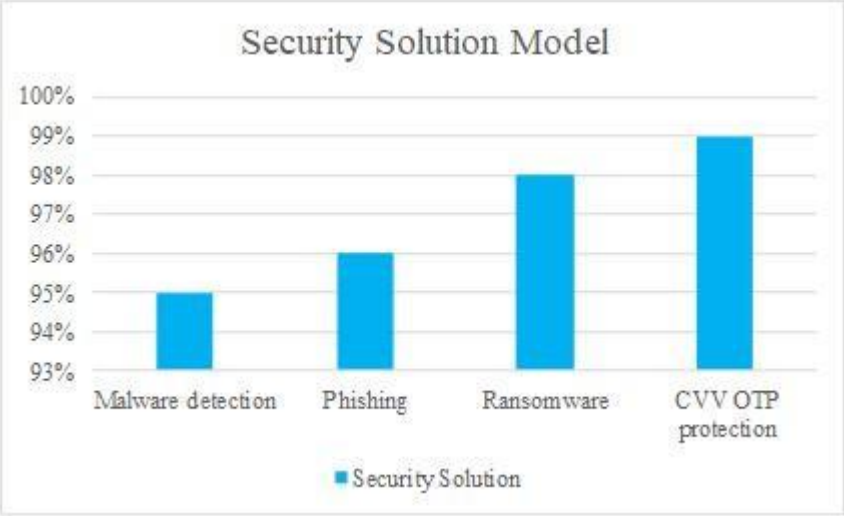

Fig. 2. Accuracy of proposed Security Solution Model

The financial sector and banks aspire to utilise technology like machine learning and artificial intelligence to combat cyber threats and cybercrime. There are various avenues for artificial intelligence-related technologies that can help the banking industry achieve more growth and success. AI promotes enduring trust and is regarded as crucial for upholding that transparency and explaining that aptitude. Artificial intelligence systems are used to gather and store data on client interests and activities. Robo advise refers to the automated platform that AI maintains. It entailed safeguarding personal information and ensuring that the model was properly designed. This helps the banking industry identify fraud in online transactions.

The application of machine learning techniques in cyber threat analysis for online transactions holds significant promise for improving fraud detection and prevention. In this study, we explored the use of machine learning algorithms to detect fraudulent activities and enhance the security of online transactions. The results demonstrate the potential effectiveness of machine learning in identifying fraudulent behavior and minimizing false positives. One of the key advantages of using machine learning

in online transaction analysis is its ability to detect complex and evolving patterns that may go unnoticed by traditional rule-based systems. Machine learning models can learn from large volumes of transaction data and adapt to changing fraud patterns, making them valuable tools in staying ahead of sophisticated attackers.

5 Conclusion

The rapid growth of online transactions has led to an increased risk of cyber threats, particularly in the form of fraudulent activities. In this study, we explored the application of machine learning techniques for cyber threat analysis in online transactions. The results and findings presented in this research demonstrate the potential effectiveness of machine learning models in detecting and mitigating cyber threats, specifically in the context of online transaction fraud. Furthermore, it is crucial to acknowledge that the deployment of machine learning models in real-world online transaction systems requires ongoing monitoring, updating, and collaboration with domain experts. The models should be regularly retrained with new data to adapt to evolving fraud techniques and changes in user behavior. Collaboration with industry stakeholders and sharing of threat intelligence can enhance the effectiveness of the models and provide a comprehensive defense against emerging cyber threats.

By setting up user interactions with criteria for requirements and information resources, the proposed security solution model enables the effective mechanism for protection of online transactions from cyber security threats at various levels of representation by restricting access and appropriately alloying off content with user authority. By identifying attackers during online transactions, the cyber security solution model shields end users against a variety of cyber threats like phishing, malware detection, and ransomware detection. The transaction architecture offers a comprehensive solution, and any shared information is safely transported to third-party websites.

References

1. Hassan, M.A.; Shukur, Z.; Hasan, M.K. An Efficient Secure Electronic Payment System for E-Commerce. Computers 2020, 9, 66.
2. Ghelani, D., Hua, T. K., & Koduru, S. K. R. (2022). Cyber Security Threats, Vulnerabilities, and Security Solutions Models in Banking. Authorea Preprints.
3. Johnson, A. L. (2016). Cybersecurity for financial institutions: The integral role of information sharing in cyber-attack mitigation. NC Banking Inst., 20, 277.
4. N. Magaia, R. Fonseca, K. Muhammad, A. H. F. N. Segundo, A. V. Lira Neto and V. H. C. de Albuquerque, "Industrial Internet-of-Things Security Enhanced With Deep Learning Approaches for Smart Cities," in IEEE Internet of Things Journal, vol. 8, no. 8, pp. 6393-6405, 15 April15, 2021, doi: 10.1109/JIOT.2020.3042174.
5. Gon, Anudeepa, et al. "Application of Machine Learning Algorithms for Automatic Detection of Risk in Heart Disease." In Cognitive Cardiac Rehabilitation Using IoT and AI Tools, edited by Parijat Bhowmick, Sima Das, and Kaushik Mazumdar, 166-188. Hershey, PA: IGI Global, 2023. https://doi.org/10.4018/978-1-6684-7561-4.ch012.
6. Dhoot, A., Nazarov, A. N., & Koupaei, A. N. A. (2020, March). A security risk model for online banking system. In 2020 Systems of Signals Generating and Processing in the Field of on Board Communications (pp. 1-4). IEEE.
7. de Gusmão, A. P. H., Silva, M. M., Poleto, T., e Silva, L. C., & Costa, A. P. C. S. (2018). Cybersecurity risk analysis model using fault tree analysis and fuzzy decision theory. International Journal of Information Management, 43, 248-260.
8. Almudaires, F., & Almaiah, M. (2021, July). Data an overview of cybersecurity threats on credit card companies and credit card risk mitigation. In 2021 International Conference on Information Technology (ICIT) (pp. 732-738). IEEE.
9. Alzoubi, H. M., Ghazal, T. M., Hasan, M. K., Alketbi, A., Kamran, R., Al-Dmour, N. A., & Islam, S. (2022, May). Cyber Security Threats on Digital Banking. In 2022 1st International Conference on AI in Cybersecurity (ICAIC) (pp. 1-4). IEEE.
10. Alzoubi, H., Alshurideh, M., Kurdi, B., Alhyasat, K., & Ghazal, T. (2022). The effect of e-payment and online shopping on sales growth: Evidence from banking industry. International Journal of Data and Network Science, 6(4), 1369-1380.

11. Y. -F. Chang, W. -L. Tai, Y. -C. Liu and H. -W. Chen, "Vulnerability of Baseri et al.'s Untraceable Offline Electronic Cash System," 2018 International Conference on System Science and Engineering (ICSSE), New Taipei, Taiwan, 2018, pp. 1-5, doi: 10.1109/ICSSE.2018.8519978.

12. Qalati, Sikandar Ali, Esthela Galvan Vela, Wenyuan Li, Sarfraz Ahmed Dakhan, Truong Thi Hong Thuy, and Sajid Hussain Merani. "Effects of perceived service quality, website quality, and reputation on purchase intention: The mediating and moderating roles of trust and perceived risk in online shopping." Cogent Business & Management 8, no. 1 (2021): 1869363.

13. Ahmed, Mansoor, Kainat Ansar, Cal B. Muckley, Abid Khan, Adeel Anjum, and Muhammad Talha. "A semantic rule based digital fraud detection." PeerJ Computer Science 7 (2021): e649.

14. El-Gharib, Najah Mary. "Using Process Mining Technology to Understand User Behavior in SaaS Applications." PhD diss., Université d'Ottawa/University of Ottawa, 2019.

15. Hazra, S., Chatterjee, S., Mandal, A., Sarkar, M., Mandal, B.K. (2023). An Analysis of Duckworth-Lewis-Stern Method in the Context of Interrupted Limited over Cricket Matches. In: Chaki, N., Roy, N.D., Debnath, P., Saeed, K. (eds) Proceedings of International Conference on Data Analytics and Insights, ICDAI 2023. ICDAI 2023. Lecture Notes in Networks and Systems, vol 727. Springer, Singapore. https://doi.org/10.1007/978-981-99-3878-0_46

16. Ghosh et al., "NLP and ML for real-time sentiment analysis in Finance," 2024 IEEE International Conference on Communication, Computing and Signal Processing (IICCCS), ASANSOL, India, 2024, pp. 1-6, doi: 10.1109/IICCCS61609.2024.10763733.

17. Hazra S, Ghosal S, Mondal A, Dey P (2024) : Forecasting of Rainfall in Subdivisions of India Using Machine Learning. In : Recent Trends in Intelligence Enabled Research, Fifth Doctoral Symposium, DoSIER 2023, ISBN978-981-97-2320-1