IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 4, May 2023

A Deep Dive into Data Leakage in Machine Learning: Causes, Consequences And Countermeasures

I.V. Dwaraka Srihith[1], L. Rajitha[2], K. Thriveni[2], A. David Donald[3],P. Blessy[3]
[1]Alliance University, Bengaluru, Karnataka
[2,3]Ashoka Women's Engineering College, Dupadu, Andhra Pradesh

Abstract*:* In recent years, machine learning has revolutionized many industries, from healthcare to finance to entertainment. However, as the use of machine learning has grown, so too has the risk of data leakage, which can have serious consequences for individuals and organizations alike. In this article/presentation, we take a deep dive into data leakage in machine learning, exploring its causes, consequences, and potential solutions. We begin by defining data leakage and discussing why it's important to prevent it. Next, we examine the various causes of data leakage, from human error to technical vulnerabilities to malicious attacks. We also discuss the different types of data leakage and their consequences, such as loss of privacy, reduced accuracy, and model poisoning. After analyzing the challenges of data leakage prevention, we explore the latest techniques and best practices for minimizing the risk of data leakage in machine learning, including data masking, encryption, and access control. We also discuss the role of data governance and data management in preventing data leakage, as well as the importance of transparency and accountability in detecting and responding to data leakage incidents. Finally, we look to the future of data leakage prevention, discussing emerging technologies and new regulations that may help mitigate the risks of data leakage in machine learning.

Keywords: Data leakage, Machine learning, Privacy, Security

REFERENCES

[1]. Abadi, M., Chu, A., Goodfellow, I., McMahan, H. B., Mironov, I., Talwar, K., & Zhang, L. (2016). Deep learning with differential privacy. Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security, 308-318.

[2]. Bagdasaryan, E., Veit, A., Hua, Y., Estrin, D., & Shmatikov, V. (2020). Differential privacy has disparate impact on model accuracy. Proceedings of the Conference on Fairness, Accountability, and Transparency, 77-86.

[3]. Goodfellow, I. J., Shlens, J., & Szegedy, C. (2019). Explaining and harnessing adversarial examples. International Conference on Learning Representations.

[4]. Kaur, H., & Singh, A. K. (2020). Machine learning security: A systematic literature review. Journal of Ambient Intelligence and Humanized Computing, 11(3), 987-1006.

[5]. Papernot, N., McDaniel, P., & Goodfellow, I. (2018). Transferability in machine learning: from phenomena to black-box attacks using adversarial samples. arXiv preprint arXiv:1605.07277.

[6]. Srinivas, T. Aditya Sai, M. Monika, N. Aparna, Keshav Kumar, and J. Ramprabhu. "A Methodology to Predict the Lung Cancer and its Adverse Effects on Patients from an Advanced Correlation Analysis Method." In 2023 International Conference on Intelligent Data Communication Technologies and Internet of Things (IDCIoT), pp. 964-970. IEEE, 2023.

[7]. Rana, M. A., Imran, A., & Khan, I. (2020). Secure machine learning for health data analytics: A review. Journal of Ambient Intelligence and Humanized Computing, 11(2), 503-516.

[8]. Vepakomma, P., Gupta, O., Babu, V. S., & Motwani, R. (2018). Federated learning with non-iid data. arXiv preprint arXiv:1806.00582.

[9]. Xu, W., Tao, D., Xu, X., & Zhang, Z. (2020). A survey on adversarial machine learning in healthcare. Journal of Biomedical Informatics, 103476.

[10]. Yeom, S., Fredrikson, M., Jha, S., & Seshadri, A. (2018). Privacy risk in machine learning: Analyzing the connection to overfitting. Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security, 1287-1301.

[11]. Ramasubbareddy, Somula, Evakattu Swetha, Ashish Kumar Luhach, and T. Aditya Sai Srinivas. "A multi-objective genetic algorithm-based resource scheduling in mobile cloud computing." International Journal of Cognitive Informatics and Natural Intelligence (IJCINI) 15, no. 3 (2021): 58-73.

[12]. Zhang, Y., Chen, H., Zou, J., Wang, R., & Tan, Y. (2021). An Overview of Data Leakage Detection Techniques in Machine Learning Systems. IEEE Access, 9, 54169-54186.

[13]. Zhu, Y., Li, J., Wang, K., & Zhang, J. (2021). A survey on data security in machine learning. IEEE Transactions on Knowledge and Data Engineering.

[14]. Bagdasaryan, E., Shmatikov, V., & Veit, A. (2021). The curse of concentration in private decentralized learning. arXiv preprint arXiv:2103.04228.

[15]. Carlini, N., & Wagner, D. (2017). Towards evaluating the robustness of neural networks. Proceedings of the IEEE Symposium on Security and Privacy, 39-57.

[16]. Giorgi, R., & Mauro, N. (2019). Adversarial attacks and defenses in deep learning: A survey. Computers & Security, 87, 101-121.

[17]. Shareefa, P., P. Uma Maheshwari, A. David Donald, T. Aditya Sai Srinivas, and T. Murali Krishna. "Forecasting the Future: Predicting COVID-19 Trends with Machine Learning."

[18]. Kang, B. M., & Park, S. (2021). Deep learning security: a review. Information Security Journal: A Global Perspective, 30(1-2), 1-22.

[19]. Wang, Q., Zhang, L., & Guo, Y. (2020). Privacy-preserving machine learning: Threats and solutions. IEEE Network, 34(6), 180-186.

[20]. Yang, Y., Ye, M., Xu, Y., & Liu, Y. (2019). A survey on security and privacy issues in machine learning. IEEE Access, 7, 146490-146509.

[21]. Bharathi, B., P. Shareefa, P. Uma Maheshwari, B. Lahari, A. David Donald, and T. Aditya Sai Srinivas. "Exploring the Possibilities: Reinforcement Learning and AI Innovation."