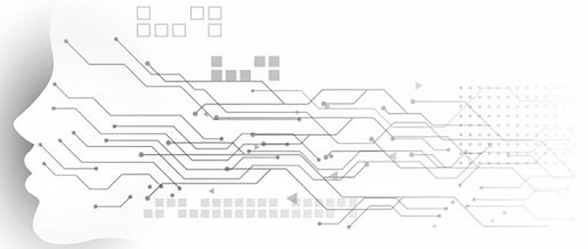
Bölüm
1



BÜYÜK SAĞLIK VERİLERİNDE MAHREMİYET

Yavuz CANBAY[1]

ÖZET

Gelişen sağlık sistemi ve teknolojileri ile beraber üretilen verinin miktarı ve hacmi her geçen gün artmaktadır. Böylesi verilerin analiz edilmesinde geleneksel teknoloji ve alt yapılar yetersiz kaldığı için bu sorunu aşmak amacıyla büyük veri kavramı ortaya çıkmış ve bu alandaki ihtiyaçların karşılanmasında güçlü bir teknoloji olmuştur. Büyük sağlık verilerinden değer üretmek her ne kadar kıymetli olsa da, bu verilerin sahiplerinin de mahremiyetini korumak aynı derecede önemlidir. Günümüzde kişisel verilerin sıklıkla üretildiği bir alan olan sağlıkta, verilerin toplanmasından depolanmasına, analiz edilmesinden yayınlanmasına kadar geçen süreçte veri mahremiyetine en üst seviyede riayet edilmesi gerekmektedir. Bu çalışmada, büyük sağlık verilerinin yayınlanmasında veri mahremiyeti problemi ele alınmış, büyük sağlık verilerine yönelik mahremiyet saldırıları ve koruma modelleri açıklanmış ve çeşitli değerlendirmeler sunulmuştur.

GİRİŞ

Bilgi toplumunun gelişmesiyle beraber, çeşitli kaynaklar tarafından üretilen verinin boyutu, hacmi, çeşitliliği ve hızı artmaktadır. Bu tür verileri işleyebilmek için yeni platformlara, donanımlara, model, sistem ve paradigmalara ihtiyaç duyulduğu aşikârdır. Bu ihtiyaçları karşılamak için büyük veri kavramı ortaya çıkmış ve yeni platformları, yazılım ve donanımları da beraberinde getirmiştir. Geleneksel yollarla işlenemeyen büyük veriyi, yapısı gereği yüksek işleme kapasitesi ve hızına sahip yeni donanımlarla ve platformlarla işlemek mümkündür.

Pek çok alanda olduğu gibi sağlık alanında da üretilen ve saklanan veri, büyük veri olarak değerlendirilmektedir. Örneğin; yayınlanan istatistiklere göre 2015

[1] Kahramanmaraş Sütçü İmam Üniversitesi, Bilgisayar Mühendisliği Bölümü

- Ülkemiz için önemli bir yapı olan sağlık hizmetlerinin, sistemlerinin, teknoloji ve girişimlerinin güçlendirilmesi ve iyileştirilerek yeniden organize edilmesi amacıyla veri mahremiyetini sürecini sağlık birimlerinde her alanda en iyi şekilde uygulanması gerektiği,
- KVKK ve GDPR gibi yasal mevzuatların, sağlık verilerinin korunması konusunda ülkemiz için önemli bir unsur olduğu ve
- Ülkemizdeki sağlık verilerinin mahremiyetinin korunarak merkezi bir yapı tarafından yayınlanması halinde katma değeri yüksek pek çok çıktının elde edileceği değerlendirilmektedir.

KAYNAKLAR

1. Ş. Birinci. (08.08.2017). *Sağlıkta Büyük Veri*. Available: http://www.acibadem.edu.tr/doc/SuayipBirinci_SagliktaBuyukVeri.pdf
2. (11.03.2020). *Kişisel Verilerin Korunması Kanunu*. Available: http://www.resmigazete.gov.tr/eskiler/2016/04/20160407-8.pdf
3. (12.03.2020). *General Data Protection Regulation*. Available: https://eur-lex.europa.eu/eli/reg/2016/679/oj
4. H. K. Patil and R. Seshadri, "Big Data Security and Privacy Issues in Healthcare," in *IEEE International Congress on Big Data* Anchorage, AK, USA, 2014, pp. 762-765.
5. C. Lin, Z. Song, H. Song, Y. Zhou, Y. Wang, and G. Wu, "Differential Privacy Preserving in Big Data Analytics for Connected Health," *Journal of medical systems,* vol. 40, pp. 1-9, 2016.
6. P. Jain, M. Gyanchandani, and N. Khare, "Big Data Privacy: A Technological Perspective and Review," *Journal of Big Data,* vol. 3, p. 25, 2016.
7. I. Olaronke and O. Oluwaseun, "Big Data in Healthcare: Prospects, Challenges and Resolutions," in *Future Technologies Conference*, San Francisco, USA, 2016, pp. 1152-1157.
8. M. A. Beyer and D. Laney, "The Importance of Big Data: A Definition," Gartner2012.
9. K. Nandini Prasaad and T. Pratheek, "Providing Anonymity Using Top Down Specialization on Big Data Using Hadoop Framework," in *IEEE Annual India Conference*, New Delhi, India, 2015, pp. 1-6.
10. N. Victor, D. Lopez, and J. H. Abawajy, "Privacy Models for Big Data: A Survey," *International Journal of Big Data Intelligence,* vol. 3, pp. 61-75, 2016.
11. X. Zhang, C. Yang, S. Nepal, C. Liu, W. Dou, and J. Chen, "A Mapreduce Based Approach of Scalable Multidimensional Anonymization for Big Data Privacy Preservation on Cloud," in *International Conference on Cloud and Green Computing*, Karlsruhe, Germany, 2013, pp. 105-112.
12. W. Li and H. Li, "LRDM: Local Record-Driving Mechanism for Big Data Privacy Preservation in Social Networks," in *IEEE International Conference on Data Science in Cyberspace*, Changsha, China, 2016, pp. 556-560.
13. M. Tanwar, R. Duggal, and S. K. Khatri, "Unravelling Unstructured Data: A Wealth of Information in Big Data," in *International Conference on Reliability, Infocom Technologies and Optimization*, Noida, India, 2015, pp. 1-6.
14. W. Vorhies. (2014). *How Many V's in Big Data? The Characteristics that Define Big Data*. Available: https://www.datasciencecentral.com/profiles/blogs/how-many-v-s-in-big-data-the-characteristics-that-define-big-data

15. (05.08.2017). *Big Data: Data Wrangling Boot Camp Big Data Vs*. Available: http://www.cs.odu.edu/~ccartled/Teaching/2017-Spring/DataWrangling/Presentations/030-bigDataVs.pdf
16. Y. Canbay, "Aykırı Veri Yönelimli Fayda Temelli Büyük Veri Anonimleştirme Modeli," Doktora Tezi, Fen Bilimleri Enstitüsü, Gazi Üniversitesi, Ankara, 2019.
17. B. Fung, K. Wang, R. Chen, and P. Yu, "Privacy-Preserving Data Publishing: A Survey of Recent Developments," *Computing Surveys,* vol. 42, p. 14, 2010.
18. R. C. Wong, A. W. Fu, K. Wang, and J. Pei, "Minimality Attack In Privacy Preserving Data Publishing," in *International Conference on Very Large Data Bases*, Vienna, Austria, 2007, pp. 543-554.
19. G. Duncan and D. Lambert, "The risk of disclosure for microdata," *Journal of Business & Economic Statistics,* vol. 7, pp. 207-217, 1989.
20. C. Skinner and D. J. Holmes, "Estimating the Re-Identification Risk per Record in Microdata," *Journal of Official Statistics,* vol. 14, pp. 361-372, 1998.
21. F. K. Dankar, K. El Emam, A. Neisa, and T. Roffey, "Estimating the Re-Identification Risk of Clinical Data Sets," *Bmc Medical Informatics and Decision Making,* vol. 12, p. 66, 2012.
22. W. Winkler, "Masking and Re-Identification Methods for Public-Use Microdata: Overview and Research Problems," in *International Workshop on Privacy in Statistical Databases*, Barcelona, Spain, 2004, pp. 231-246.
23. J. Domingo-Ferrer and V. Torra, "A Critique of k-Anonymity and Some of Its Enhancements," in *2008 Third International Conference on Availability, Reliability and Security*, Barcelona, Spain, 2008, pp. 990-993.
24. X. Sun, L. Sun, and H. Wang, "Extended k-Anonymity Models Against Sensitive Attribute Disclosure," *Computer Communications,* vol. 34, pp. 526-535, 2011.
25. M. E. Nergiz, M. Atzori, and C. Clifton, "Hiding the Presence of Individuals from Shared Databases," in *ACM SIGMOD International Conference on Management of Data*, Beijing, China, 2007, pp. 665-676.
26. B. Chen, K. LeFevre, and R. Ramakrishnan, "Privacy Skyline: Privacy with Multidimensional Adversarial Knowledge," in *International Conference on Very Large Data Bases*, Vienna, Austria, 2007, pp. 770-781.
27. L. Sweeney, "Computational Disclosure Control: A Primer on Data Privacy Protection," Ph. D. Thesis, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, USA, 2001.
28. A. Machanavajjhala, J. Gehrke, D. Kifer, and M. Venkitasubramaniam, "l-Diversity: Privacy Beyond k-Anonymity," presented at the International Conference on Data Engineering, Atlanta, USA, 2006.
29. X. Zhang, L. T. Yang, C. Liu, and J. Chen, "A Scalable Two-Phase Top-Down Specialization Approach for Data Anonymization Using Mapreduce on Cloud," *IEEE Transactions on Parallel and Distributed Systems,* vol. 25, pp. 363-373, 2014.
30. B. C. Fung, K. Wang, A. W. Fu, and S. Y. Philip, *Introduction to Privacy-Preserving Data Publishing: Concepts and Techniques*. USA: CRC Press, 2010.
31. L. Sweeney, "k-Anonymity: A Model for Protecting Privacy," *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems,* vol. 10, pp. 557-570, 2002.
32. N. Li, T. Li, and S. Venkatasubramanian, "Closeness: A New Privacy Measure for Data Publishing," *IEEE Transactions on Knowledge and Data Engineering,* vol. 22, pp. 943-956, 2010.