Data Mining Essay:

the use of personal information for purposes other than those it was collected

A.S.Wijewardena

Abstract

Studies have been previously conducted relating to the misuse of personal information in data science. This essay discusses main ethical considerations relating to it, real life examples and solutions introduced combining with data mining algorithms and techniques. Throughout, selected practical research methods and theory based research have been reviewed respective to each topic.Finalizing in the conclusion, the current state of the point in question and future improvement.

1 Introduction

In the field of data science, personal information from users' are being collected for various reasons depending on the purpose it is collected. The purpose could be different from each project/business, data could be collected for marketing strategies or to be utilized to improve user experience. Even though personal information is collected for a said purpose, in some instances the information could be used for other intentions without the acknowledgment of the owner of the information, which would be considered as unethical and it compromises the user's trust towards personal data protection. In most conditions the intention of the usage of obtained information holds a vague nature. Brehmer et al. (2020)

2 Background

The essay focuses on the topic of the use personal information for other than those it was collected, along with main ethical considerations, real world examples of such analysis, solutions being proposed and conclusions.

2.1 Main Ethical Considerations in Privacy, Control and Analysis on personal information

Considering main ethical issues in data science relating to how personal information is used for purposes other than for the reason it was collected, the protection level of acquired personal information is concerning. As per recent studies discuss the advancement of "economic globalization" in the big data era Chen (2021). It addresses one of the main concerns as the utilization of the web. As global communication increases daily, the amount of people who uses web platforms increases. In most cases people are unaware about the personal information that could be leaked unintentionally. A hacker would be able to access this leaked information, which is a violation of privacy and also displays a lack of protection of personal information.

Furthermore, Martin (2015) discusses the ethical concerns that occur with "reselling consumer data" to various markets related to big data. By reselling consumer data without the knowledge of the user, certain companies are able to target specific groups of people to promote their businesses and products. Which could be noted as helpful, but the method of data acquisition is unethical, which is a breach in a user's trust relating to sensitive information.

In Onik et al. (2019), focuses on the ethical challenges that occur in the healthcare sector. Electronic health records are not only shared between the patient and the doctor, but at times records are shared for other purposes like healthcare data analysis, marketing analysis, medication investigations, medical research. Due to the health care division containing extensive amounts of data, it has been a challenge in protecting personal information of

patients. Certain information shared could be used for other purposes, without the patient's knowledge, as addressed above.

Yeong Kwon et al. (2019) highlights the concept of "Data Driven" industries being strongly dependent on the usage of acquired personal information. Commonly most sectors collect personal data in order to manipulate the data for the economic gain. More often than so it conflicts with regards to the protection of such sensitive data. Most sectors obtain data displaying a certain purpose but later on the data is used for other purposes which is a breach in the user's trust regarding the safety of their shared personal information. The paper discusses the debate between protecting the privacy of the information and using such data for economic benefit.

2.2 Real world analysis Examples

There are many real world analysis examples that display multiple purposes of personal information used unethically in the data science field.

Focusing on personal health information gathered for research purposes, "wearable technology and mobile computing" is a central method of gathering data. In Zheng et al. (2018) states that personal data obtained using such devices consist of sensitive information which is owned by the user respectively. Although in concept it is true, in practice currently these data are managed by various service providers, while the data can be considered as a centralized storage, which increases the risk of data being shared among different sources and potential security issues.

Another example would be the use of smart personal assistant applications. Mane et al. (2017) reviews the functionality of such applications and identifies their main goal as "Assisting users in performing tasks". Its addressed that these apps have the capacity to identify the user's preferences, habits and focuses through personal data collected. Its even stated that the information could be utilized to a point where a smart personal assistant would be able to predict their user's actions early on. Gathering information of an individual's life to that extent requires obtaining extremely personal information. Currently these applications store data in databases and access them when required but still a certain amount of private information are considered to be shared for other purposes like marketing and advertisements. While these could be considered as personalized advertisements, the method of data collection in order to produce such adverts is unethical.

Fabricius (2021) Discusses other purposes where personal information gathered through menstrual self tracking apps have been used, while invading the user's privacy and protection towards their own data. The paper discusses about the apps "Flo" and "Ovia", in which the first application shared private user information with various corporations and the second app distributed personal data among the users' management. Essentially these apps were supposed to help cycle tracking and data acquired was supposed to be used for that purpose only, but it was used purely for other purposes. This violation resulted in major issues among the work force, discrimination towards certain genders, pay reductions and limiting access to work resources. This is a precise real world analysis example which indicates the implications of unethical use of personal information in the data science sector.

2.3 Solutions being proposed

The study Wang and Zhang (2022) explores the legal protection aspect of personal information, basing of selected data-mining algorithms and classification methods. While the paper focuses on the lawful security of delicate information, it also discusses privacy protection algorithms used in "frequent pattern mining" and noise addition techniques in standard data mining methods. Mentioned algorithms and models could be enforced in to protecting users' personal information and limiting the usage of unethical practices.

As for personal medical information, Abdulshaheed et al. (2022) proposes a strategy of separating the utilization and storage of patients' sensitive data from common information and implementing data mining algorithms to encode personal data in a specific order, so that

obtaining private information unethically will be difficult due to the hierarchical algorithms in place. Zheng et al. (2018) discusses about a "block-chain based personal health data sharing system", which provides the user an understanding of how their personal information would be used for different purposes like medical research, marketing analysis.

As one of the recent solutions being proposed the concept of "cybersecurity data science" can be considered. The study Sarker et al. (2020) addresses the process of gathering personal data, analysing them and developing purpose focused data-driven models, which minimize the security threats towards users' personal information. This approach is considered to be inventive and holds a potential in the future, if implemented accurately, obtained user information would only be used for the purpose it was initially collected for.

Considering the digital economy division, personal information obtained has a much higher value according to Yao (2021), which acknowledges that the field works with highly sensitive data. Therefore, they are required to have much more efficient security protection towards gathered user information. The paper discusses about the "Chinese Civil Code" , how it aids in protecting private data legally and implementing such measures to banks as well. It could be utilized to limiting unapproved use of personal information.

3 Conclusion

In conclusion, data science is a vast field of study, which consists of various topics. On an ethical aspect, considering the issue of personal information being used for different purposes other than for the intention it was collected for, has been and currently is one of the main principal issues. There have been countless instances where sensitive information was unethically accessed and used for diverse objectives. This contributes negatively towards any data central businesses and projects. Recent studies have been done to implement data mining algorithms and models along with legal protection laws of personal information , in order to manage this point in question. Some solutions are currently in practice successfully and a considerable amount of them are still being researched. In general solutions being proposed are promising and have potential for the future.
To finalize, the use of personal information is needed in data science. The grey area is the intention behind the utilization of collected data. While there are solutions being practised and researched, it seems to be a constant ethical issue in the field of data science, which could be managed to a certain extent.

Word Count (including references) - 1,968 words

References

Abdulshaheed, H. R., Al-Juboori, S. A. M., Al Sayed, I. A., Al Barazanchi, I., Gheni, H. M., and Jaaz, Z. A. (2022). Research on optimization strategy of medical data information security and privacy. In *2022 9th International Conference on Electrical Engineering, Computer Science and Informatics (EECSI)*, pages 132–136. IEEE.

Brehmer, M., Lee, B., Isenberg, P., and Choe, E. K. (2020). A comparative evaluation of animation and small multiples for trend visualization on mobile phones. *IEEE Transactions on Visualization and Computer Graphics*, 26(1):364–374.

Chen, X. (2021). A study on personal information protection mechanism of web platform in big data age. In *2021 5th Annual International Conference on Data Science and Business Analytics (ICDSBA)*, pages 120–124.

Fabricius, A. (2021). Privacy is a feminist issue: Reconsidering data sharing in menstrual self-tracking apps. In *2021 IEEE International Symposium on Ethics in Engineering, Science and Technology (ETHICS)*, pages 1–1.

Mane, P., Sonone, S., Gaikwad, N., and Ramteke, J. (2017). Smart personal assistant using machine learning. In *2017 International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS)*, pages 368–371.

Martin, K. (2015). Ethical issues in the big data industry. *MIS Quarterly Executive*, 14:2.

Onik, M. M. H., Aich, S., Yang, J., Kim, C.-S., and Kim, H.-C. (2019). Chapter 8 - blockchain in healthcare: Challenges and solutions. In Dey, N., Das, H., Naik, B., and Behera, H. S., editors, *Big Data Analytics for Intelligent Healthcare Management*, pages 197–226. Academic Press.

Sarker, I. H., Kayes, A., Badsha, S., Alqahtani, H., Watters, P., and Ng, A. (2020). Cybersecurity data science: an overview from machine learning perspective. *Journal of Big data*, 7:1–29.

Wang, L. and Zhang, T. (2022). The application of data mining algorithm in the legal protection of personal data. In *2022 IEEE 2nd International Conference on Data Science and Computer Application (ICDSCA)*, pages 1339–1342.

Yao, T. (2021). Mathematical statistics and analysis on the path mechanism of protecting personal information relying on information digitization and big data. In *2021 IEEE International Conference on Emergency Science and Information Technology (ICESIT)*, pages 729–733.

Yeong Kwon, H., Young Min, K., and Ae Chun, S. (2019). Data industry and legislations for personal information protection. In *Proceedings of the 20th Annual International Conference on Digital Government Research*, pages 529–531.

Zheng, X., Mukkamala, R. R., Vatrapu, R., and Ordieres-Mere, J. (2018). Blockchain-based personal health data sharing system using cloud storage. In *2018 IEEE 20th International Conference on e-Health Networking, Applications and Services (Healthcom)*, pages 1–6.