Society 5.0 and the concept of privacy in Data Mining.

An overview of the Privacy-Preserving Data Mining techniques

by

John Rau

Master dissertation submitted in conformity with the requirements
for the MSc in Economics, Finance and Computing
University of Huelva & International University of Andalucia

December 2021

Society 5.0 and the concept of privacy in Data Mining.

An overview of the Privacy-Preserving Data Mining techniques

John Rau

MSc in economics, Finance and Computing

Antonio J. Tallón Ballesteros

University of Huelva and International University of Andalusia

2021

Abstract

In this Master's dissertation, we propose an analysis to counter the threats of undesirable and illegal effects of privacy violation, without obstructing the opportunities for knowledge discovery of data mining technologies. The main idea is to highlight the importance of privacy protection in knowledge discovery technology so that the analysis incorporates the relevant privacy requirements from the outset. Therefore, we describe the privacy-preserving data mining (PPDM) techniques needed in the study of data mining: First, transform the source data into an untraceable version which will guarantee the privacy. Second, ensure that data mining queries can be performed correctly using the transformed data instead of the originals. In addition, this study introduces the concept of Society 5.0, a new vision that arises in the modern era of IoT, Big Data and Automation.

Keywords: Privacy, PPDM, Society 5.0, Privacy Preserving Data Mining.

Resumen

En este trabajo de fin de maaster, proponemos un análisis para contrarrestar las amenazas de efectos indeseables e ilegales de la violación de la privacidad, sin obstruir las oportunidades de descubrimiento de conocimiento (DC) de las tecnologías de minería de datos. La idea principal es resaltar la importancia de la privacidad en la tecnología de DC para que el análisis incorpore los requisitos de privacidad relevantes desde el principio. Por tanto, describimos las técnicas de minería de datos para preservar la privacidad (PPDM) necesarios en el estudio de la minería de datos: Primero, transformar los datos de origen en una versión diferente con garantía de privacidad. Segundo, garantizar que las consultas de minería de datos se puedan realizar correctamente utilizando los datos transformados en lugar de los originales. Esta TFM estudia además el concepto de Sociedad 5.0, una nueva visión que surge en la era moderna de Internet of Things, Big Data y Automatización.

Palabras Clave: Privacidad, PPDM, Sociedad 5.0, Preservación de la privacidad de la minería de datos.

Acknowledgments

I thank Professor Antonio J. Tallón Ballesteros for his patience in supervising my dissertation, his suggestions and professionalism allowed me to improve this study. I am also blessed to have received help and motivation from my family and friends.

Table of contents

List of Tables

List of Figures

1.Introduction

Technology diffuses an increasing number of human activities and information, providing huge amounts of data from different sources. Many actions made on the web can be stored and analyzed, playing an active part in the knowledge on the Internet; for instance, search engines queries can be used to improve the performance of algorithms for information recovery and website ranking the so-called SEO (Search Engine Optimization) (Haider & Sundin, 2019). While Social Media are a useful solution for identifying the topics that are getting the most attention based on their user's engagement and opinions. Furthermore, similar content can be stored and proposed to third parties' corporations that are willing to use metadata, hashtags, images, videos, geolocation information provided in the status of the main social media profiles for their purposes (Foulonneau, Martin & Turki 2014).

The growth of digital data is strictly due to the increase in users and their activities on the web (Digital Global Overview Report, 2021). Figure 1 and Table 1&2 provide information about the web activity per minute and per second; both graphics show data during 2020, however, the numbers are constantly growing. Nowadays, it becomes increasingly necessary not only to provide but also to store large amounts of data (Siddiqa, Karim, & Gani, 2017).

Figure 1: Data every 60 seconds. (Gary Thomas, 2020)

Platforms and information	**Per second**
Linkedin post	2,1333333
Instagram pictures	1.099,53333
Youtube hours	8,33333333
Twitter tweets	5.833,333333
Snapchat stories	4.629,61667
TikTok views	277,766667
Pinterest pins	162,033333
Facebook likes	52.083,3333

Table 1: Average of data shared each second

Description of statistics	**Value**
Internet Users in the world	5,140,428,321
Total number of Websites	1,912,521,924
Emails sent in 1 second	3,095,753
Google searches in 1 second	97,846
Skype calls in 1 second	6,325
Instagram photos uploaded in 1 second	1,122
Tweets sent in 1 second	9,780
YouTube videos viewed in 1 second	93,049
GB of Internet traffic in 1 second	135,959
Electricity used today for the Internet	622,605MWh

Table 2: internet live statistics Internet live stats, 2020 accessed on 15.11.21)

The objective of this study is, to offer a valid support for the understanding of the big data and the possibility to preserve the right to privacy with technical approaches used in data mining and finding the the right support tools for the collection, analysis and management of an increasing amount of data with a protected process.

2. The era of a Social Data Science

This section includes an overview of Social Data Science in the context of digital culture introducing two topics covered by this topic.

1. The application of Data in everyday life is an ever-present reality that touches all aspects of society
2. Provide a useful overview of Society 5.0 and its innovative vision

Big Data refers to data sets so extensive that it is impossible to process them and extract knowledge with traditional analysis techniques.

The exceptional size is the defining feature of these datasets, but with Big Data we also refer to data of great complexity and internal variability, produced at a continuous rate (Vassakis, Petrakis, & Kopanakis, 2018).

During the era of Social Data Science, every human activity leaves digital traces (Menchen, 2013) and can be studied quantitatively. In this scenario, the real breaking point is that this data is also available to non-specialized users.

During the last decades, for instance, data on population, natural phenomena and social trends were only accessible by specific institutions which selected only in a limited number. Nowadays, dedicated APIs (Application Program Interfaces) and private data companies allow any user to find, consult and use data on different topics.

Numerous are the advantages offered by the analysis of "social data"; for instance, a music festival in London becomes an opportunity for a study of mobility, both under the form of individual and collective movements. The possible applications are numerous: traffic control, integration with similar events in near neighborhoods, planning the best location for services and refreshment points, etc.

However, the potential of the investigation is not limited to data only for social purposes: the analysis of Big Data allows to predict effects of economic crises, epidemics like Covid19 and the spread of fake news. Nowadays, data can be found in the most various types: GPS signals, health records and data on money transactions. Any phenomenon that can be described by data becomes analyzable at any level. Every aspect of human activity can be studied and can be analyzed (Blumenthal, 2017).

2.1 A Step Towards Society 5.0

Introduced by the Japanese government during the 5th Science and Technology Basic Plan in 2016, the concept of Society 5.0 outlines an ideal vision for the society of the future. It is a real further step in human evolution oriented to the well-being of people and guided by scientific and technological transformations. Society 5.0 wants to create a "Super Smart Society" (Deguchi, 2020), a concept developed by Keidanren (Holroyd, 2020) capable of using the potential of frontier technologies towards solving the needs of society and individuals, helping to make economic development shared and positive progress. This vision would help overcome the barriers of nationality, origin, age and gender (Gladden, 2019), the innovative solution developed by this new vision are aimed at guaranteeing a dignified life for all, ensuring an adequate level of all those goods and services considered essential.

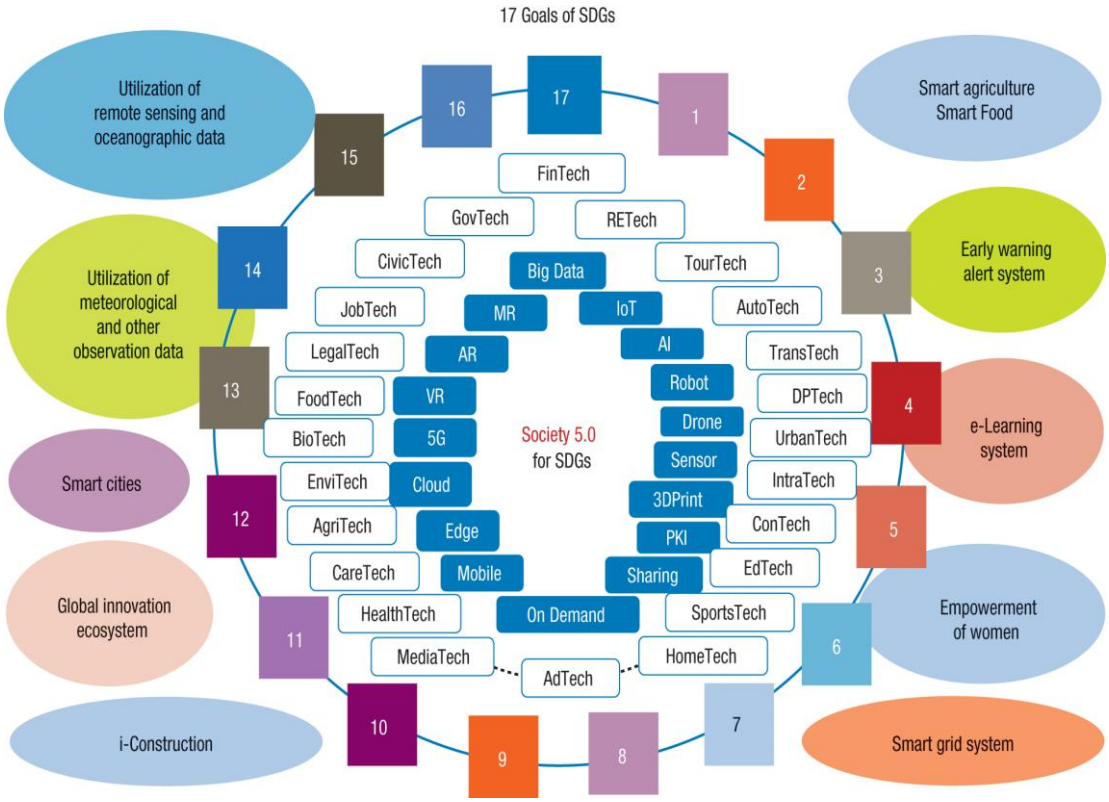

Figure 2: Society 5.0 goals SDGs. (unu-merit, 2018)

Taking into consideration Figure 2, for a Society 5.0 technologies such as Internet of Things, Artificial Intelligence (AI), Robotics and Big Data become key tools for the improvement of the quality of life and are technological areas in which Japan industries are considered excellence on a global level (Holroyd, 2020).

3. The other side of Big Data: the risks of privacy

In this following section, we will study various definitions of privacy and we will highlight the risk to privacy that arises by carrying out activities on the web according to EU legislation and Spanish.

Nowadays, the term privacy has undergone several changes over the last few years and its definition has been affected as can be seen in these numerous

academic studies that try to define the term Privacy without finding a common and fixed definition:

"*Privacy is the right to be free from unwarranted intrusion and to keep certain matters from public view*" (Law, 2015).

"*Privacy is an important element in the autonomy of the individual. Much of what makes us human comes from our interactions with others within a private sphere where we assume no one is observing. Privacy thus relates to what we say, what we do, and perhaps even what we feel*" (MacMenemy 2016).

"*Privacy protects our subjectivity from the pervasive efforts of commercial and government actors to render individual and communities fixed, transparent and predictable. Privacy is an indispensable feature of a democracy where an individual maintains his identity while contributing to their civic duty*" (Cohen 2016).

Many experts define the right to knowledge as a "common good", that is affecting anyone, directly or indirectly. Protecting this common good and ensuring that it is used with respect for all the individuals involved is therefore essential (Berendt, 2019). Therefore, the accessibility of all sorts of information opens countless possible scenarios; for example, predicting the spread of an epidemic disease like coronavirus well in advance has allowed countries to prioritize preventive health interventions in the neediest areas or age groups. Despite having indisputable benefits, however, there is the possibility, to trace the sources of this knowledge and redirect them to any individual. "Personal data" is any information that identifies or makes identifiable an existing person. Generally, a distinction is made between identifying data (personal data, photographs) and sensitive data, which can reveal racial and ethnic origin, religious orientation, gender, political opinions, membership of parties and health information (Irwin, 2021).

It is very common for a dataset to include, in its original form, information that is traceable to specific individuals - a name, a date of birth, and IP address. This becomes potentially dangerous when you can associate data about the same individual from different sources – health information, opinions expressed on social media, data related to online purchases, and payment information. Often, how such information is found is not completely legal and the average user is not always aware of how much and what sensitive data are shared during his daily operations. We are all users and at the same time producers of data, so the dissemination of knowledge about the protection of the private sphere is, at the same time, a prerequisite and objective of Social

Data Science, which has brought humanity back to the center of scientific investigation.

An overview of the different laws in place in the world for the protection of personal data is represented below in Figure 3; warmer colors such as red and orange indicate low level of restriction about data protection and privacy while color such as light blue and dark blue show more restricted regulations on the subjects.

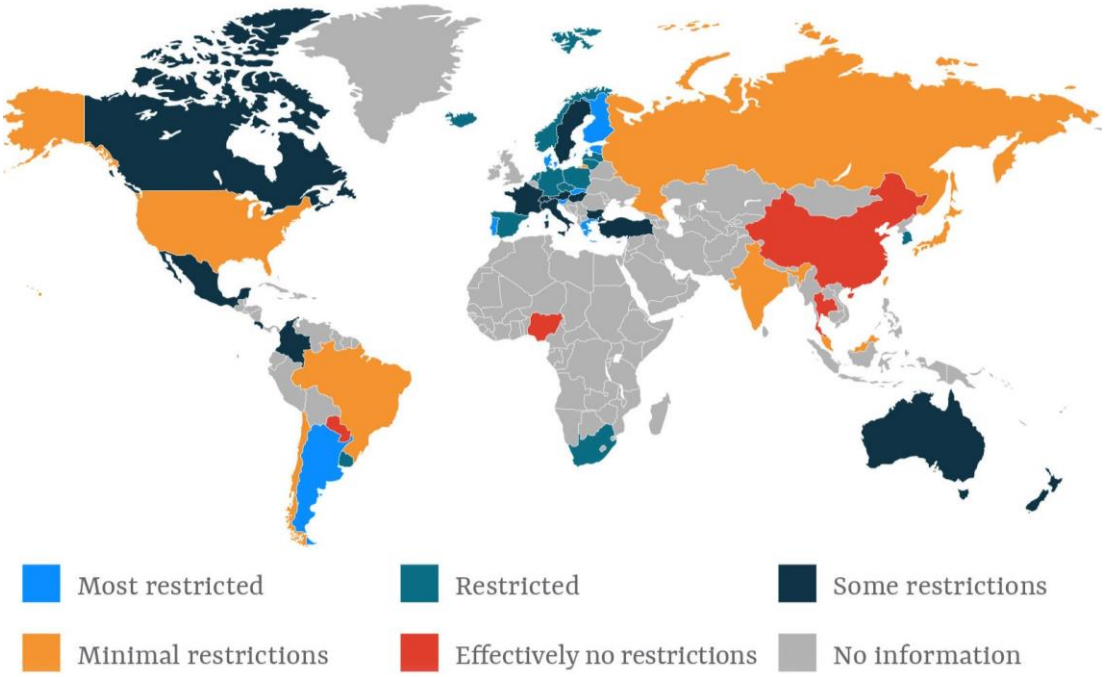

Figure 3: Privacy and data protection by country (Forrester, 2018)

3.1 The European legislation

In this section, the purpose is to clarify the European legislation on privacy, with specific attention to how the problem of privacy is perceived and treated in Spain.

As can be observed in Figure 3, the European framework in terms of interest and commitment to the protection of sensitive data ' is well defined.

The European Union considers that personal data may be collected legally under specific conditions and for legitimate reasons. It also establishes that organizations that collect personal data must commit to protecting it from inappropriate use. In particular, the EU is concerned with ensuring maximum protection of sensitive data in all member countries, in an equal way, discouraging national laws that lower the standards of protection established at the European level.

In May 2016, the European Commission proposed an extensive reform of the legislation on the protection of sensitive data contained in the European Parliament law of 24 October 1995, the so-called data protection package (Otto, 2018). This reformulation became necessary for two reasons:

- Firstly, the previous legislation was inadequate to regulate the right to privacy on the Web and for online activities

- Secondly, the national government of each European country was proceeding in different directions, making an EU reformulation even more urgent.

The basis of this reorganization remains an art. 8 of the Charter of Fundamental Rights of the European Union. The approval of the data protection package requires the intervention of both the European Parliament and the EU Council.

The data protection package uses two different tools:

1. a proposal for a Regulation, which discusses "the protection of each individual about the processing of personal data and the sharing of such data" and will replace the reform 95/46.
2. a proposal concerning the "regulation of the areas of prevention, contrast, and repression of crimes, as well as the implementation of criminal sanctions”

The new EU Regulation on the protection of personal data states:

"*Everyone has the right to protection of data concerning him or her and that the established process must be fair, for specified purposes and based on the consent of the person concerned. Another important part is that it gives people the right to access data concerning themselves and have incorrect information rectified.* "(Article 8, Protection of personal data).

In addition, in 2016 the European Commission promoted various statistical surveys in European countries to test public opinion on web privacy.

The questions were asked such as general use of the network, feelings for common web marketing strategies, personalized ads on certain sites based on online search queries, the results showed that the Spanish population was significantly less hostile (43%) than the European average (54%). The study questioned the concept of providing personal information to obtain free services (56%) while the European average was 29%. The survey revealed a general tendency in Spain to have greater confidence, compared to other countries of Europe, that the processing of their data was lawful and that there were adequate means to control their data on the web. For a part of the sample, the unusual data could also be explained by only a partial understanding of the potential threat to privacy (ePrivacy). Recently a similar survey has been carried out on a sample of 27,000 citizens, known to the public as the "Eurobarometer" (Press corner, 2019). The results were disclosed in June 2019; the main figure, consistent with past surveys, reports that European citizens complain about the lack of control over their data (67%) and do not trust online sellers (62%). As many as 70% of respondents fear that their data is used for purposes other than those mentioned, and even 89% find necessary an equal and common regulation in all countries of the EU. At the same time, respect for citizens' privacy will produce a virtuous circle between the protection of a fundamental right, consumer confidence, and economic growth. (Christou & Rashid, 2021).

3.2 The regulations on privacy in Spain

Surely, there are differences between the various European legislations, but this can be largely a consequence of the different ways in which the authorities interpret certain legal concepts and the very concept of privacy.

In Spain, the **LOPD** 3/2018 has come into force, replacing the previous LOPD 15/1999. This new legislation introduces some relevant innovations:

- It applies to both natural and legal persons, both private and public, who have personal data. The privacy policy must allow citizens to know clearly and easily the most important aspects of the process:

- It specifically recognizes the right of access and, where appropriate, rectification or deletion by those who have links with deceased persons – for family or factual reasons and their heirs – unless the deceased has expressly excluded it.

- Concerning minors, consent can be granted independently at the age of 14. It also expressly regulates the right to request the deletion of data provided to social networks or other information society services by the same minor or by third parties during his or her minor age. It regulates the right to be forgotten on social networks and equivalent information society services.

- It requires companies to protect their customers' data and at no time share or filter it without their consent, using strong encryption for their protection. Update the guarantees of the right to privacy against the use of video surveillance devices and sound recording in the workplace. It also strengthens the guarantees of the right to privacy with the use of digital devices made available to employees, integrating the regulation of the right to privacy also on the use of geolocation systems in the workplace, of which they must be informed.

3.3 The legislation on cookies

Cookies are information recorded by the browser while browsing specific websites, which report some data of the current session, such as ad example the name and address of the server, identifiers, time of access, e so on.

The importance of cookies can vary and can be directly established by the user. Typical use is to speed up authentication to sites web that requests it and collect information about users and their activity on the web service, as well as on their browsing preferences (yes think for example of the choice of language and currency shown on-site e-commerce).

Beyond this use, cookies can also be used for user profiling, or user profiling intended as monitoring their activities, purchasing habits and/or desires, to assign, for example, a certain user to a certain user profile they have common interests. This happens because the terminal with which the Internet is accessed is also identified, and therefore can be used, together with the rest of the information contained in the cookies, on websites other than those usually consulted. Once the recognition has taken place, the content page, pop-ups and advertisements are adjusted according to the profile building.

The very nature of the so-called “profiling cookies” makes them a potential danger for the privacy of users. European legislation provides that the user must always be aware of and in favor of the collection of information about cookies. In May 2014, the Guarantor decided that every website that uses cookies must declare at the beginning of the session:

a) The use of profiling cookies for purposes advertising.

b) Whether the cookies can be shared to third parties.

4. The treatment of privacy in Data Mining

In this following topic, we will study the various set of approaches in data mining for extracting knowledge from large amounts of data according to the privacy preserving process.

The expression data mining considers a phase of the Process of Knowledge Discovery in Databases (Figure 4), which consists in discovering interesting patterns in large datasets, through the implementation of techniques and tools from the world of Artificial Intelligence, machine learning, and statistics. Data mining is an essential process in every scientific field, which is used both to search for latent knowledge in data, and in semi-automated or automated approaches to try to discover patterns and phenomena of interest in large amounts of data (Alasadi & Bhaya,2017).

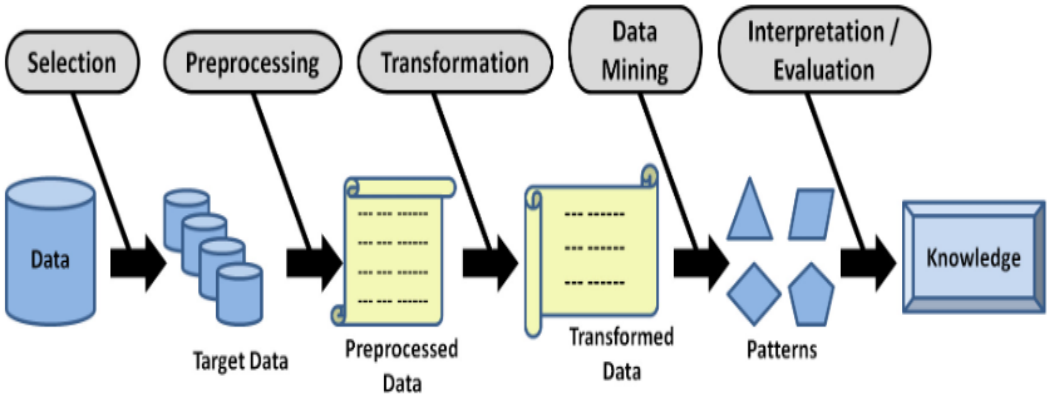

Figure 4: The Knowledge Discovery in Databases (KDD) process (Gullo, 2015)

As discussed previously, through the analysis of massive amounts of data you can discover rules, patterns, recurring sequences in different datasets and coming from different sources. Every aspect of human activity, from transport systems to energy saving, can benefit from the knowledge derived from the analysis of this data through data mining methods.

Data mining techniques are applied to the most varied areas: from business (analysis of customer transactions and search for recurring patterns to plan

marketing actions) to biology and genetics (search for anomalies), to the analysis of social networks (identification of the most important nodes of a network).

The advantages of these techniques are considerable, but many datasets contain sensitive information raising privacy issues (Singh, 2003); for this reason, it is part of the data mining itself the identification, processing (anonymization, elimination, transformation) of sensitive information.

Therefore, it is necessary to introduce the role of privacy preserving data mining, which is the task of producing valid models and patterns starting from data, without disclosing private information. However, the definition of "private information", of course, is regulated from country to country (Agrawal & Srikant, 2000).

4.1 Approaches of Privacy-Preserving Data Mining (PPDM)

The purpose of this subsection is to discuss the problem of the processing and securing of sensitive information in data mining and describe the most well-known PPDM techniques

Firstly, it is necessary to introduce four fundamental approaches to protecting privacy in Data Mining, defined as follows:

1. Data Perturbation and Obfuscation: The purpose of this family of approaches is to protect individual privacy. The input data are, also in this case, disclosed after ad hoc changes, to make it impossible to recover the original records, which would fail in protecting the privacy of the citizens. The biggest challenge of this approach is to maintain a distribution of data (also limited to some attributes of interest) such that it is still possible to extract significant patterns, rules and patterns.

2. Distributed Privacy Preserving Data Mining: This term encompasses all the methodologies in which the original dataset is partitioned, and partitions are distributed to different sources. The data mining task takes place locally for each dataset deployed, and finally, the results are combined, making sure that no party has access to data contained in other partitions.

3. Privacy-aware Knowledge Sharing: It is about sharing the knowledge resulting from the analysis, taking care instead of hiding the original data. The crucial ethical question in this group of algorithms is whether the very fact of subjecting data to a certain analysis task constitutes a violation of privacy.

4. Knowledge Hiding: It is a matter of disseminating data in suitably modified inputs since it is known that certain analyzes can bring to light rules, patterns, or models (therefore information) that should instead remain secret. The modification operation, in this case, is known as data sanitization.

As stated previously, the main problems of privacy protection are its lack of homogeneity and the difficulty to know which information should be protected from disclosure without considering the type of input dataset and the type of analysis required.

Figure 5 provides an overview of various additional techniques to protect privacy in Data Mining, seven of them are going to be discussed more deeply in the following sections:

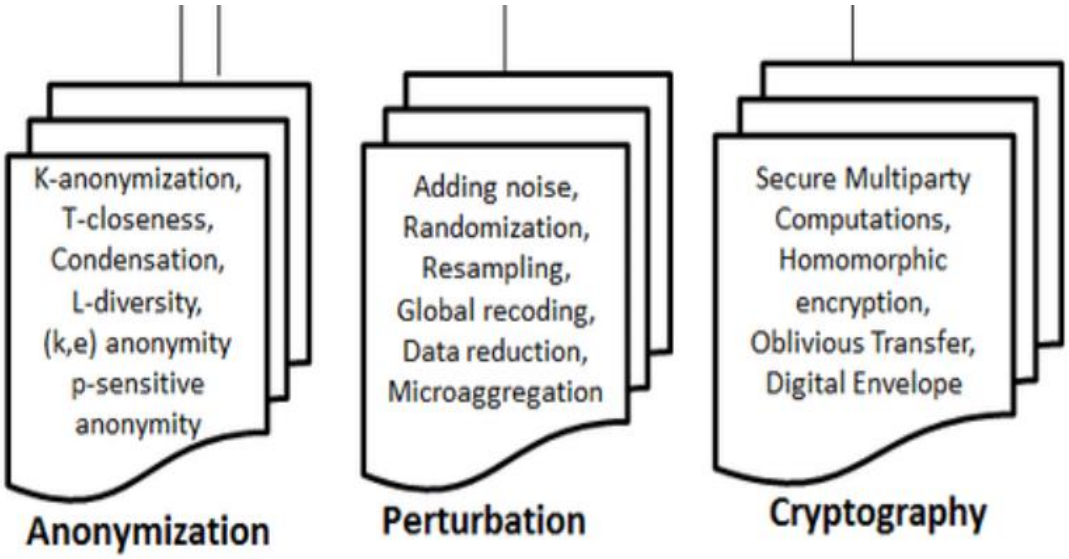

Figure 5: techniques in PPDM (Alpa Shah, 2016)

4.1.1 Randomization

The main idea of randomization' is to manipulate the data to be disclosed, distorting its nature just enough to eliminate any too explicit link between the sensitive data and the individual (crucial especially in the case of EHRs, Electronic Health Records) (Shukla & Sadashivappa, 2014).

The strength of this method is that it can be performed in the data collection phase, as it does not require previous knowledge on the distribution of the same. There is also another side of the coin: the local density of records is not considered, so records that describe outlier phenomena are treated like the others. However, some experts have shown that there are cases in which the original dataset can be reconstructed.

4.1.2 Adding noise

A particular type of randomization is to add “noise” to the data so that the original data can no longer be reconstructed. How heavily such manipulation is established depends on the level of accuracy of the knowledge to be released and the impact on the privacy of the subjects. Noise insertion is often accompanied by other anonymizing techniques, such as the removal of attributes and quasi-identifiers (Machanavajjhala & Reiter, 2012)

4.1.3 Differential privacy

This approach responds to cases where the dataset is not directly released or modified but can be queried by third parties. Given a user's query, the results received are based on the original dataset but are altered to an extent based on the differential privacy model while remaining useful for research. The advantage of this approach is that the dataset is never fully released or made available in modes other than desired; however, continuous monitoring of submitted queries is needed to prevent episodes of inference or linking to real data (Shrivastva & Singh, 2014).

4.1.4 Distributed cryptographic methods

As we described, in distributed approaches the information of interest is accurately partitioned and localized in parts and stored in different data repertoires. It is possible to select different data sources using cryptographic protocols that recompose the requested information without revealing any confidential data (Khari, 2019). These methods are used in all those cases in which different groups see a purpose, and wish to share only aggregated data, without giving up the original data of which they own (Mehmood & Guo, 2016).

4.1.5 Anonymization

Data anonymization is a data protection process focused on protecting privacy. The function of data anonymization is to encrypt or remove personally identifiable information from datasets. In this way, the people from whom the data was collected remain anonymous.

4.1.6 K-anonymity

Some data attributes can be pseudo identifiers if used together and they can uniquely identify records (zip code and date of birth) (Rajendran, 2017). The idea behind a k-anonymity algorithm is to decrease the specificity of the information described by a given record (i.e., the set of attributes-values referring to an instance of the data) so that it can no longer be distinguished from at least other k - 1 record (Sweeney, 2002).

4.1.7 T-closeness

In even more complex scenarios, the party involved in the private information contained in the database could have previous information on the distribution of the values of a target attribute (for example by using external datasets that can be linked to the dataset object of the attack). The scope of t-closeness is a very simple measure: the distribution of the values of the "sensitive" attribute in each group of identifiers must be close to the distribution of the values of that attribute in the entire table; in this way, any previous information available to the hacker is no longer informative (Li & Venkatasubramanian, 2007).

5 Conclusions

The data protection package has a scope that goes far beyond the protection of the privacy of European citizens: ensuring that personal data is protected, anonymized, and not traceable by none avoiding any competitive advantage derived from the knowledge extracted from this data. The public sector is then the ideal stage for mining and KDD applications, as it serves a vast audience and offers applications that bring immediate benefit to all the actors involved.

The key factor of privacy protection lies in planning the processes and tools involved in the knowledge extraction process; all techniques adopted later to randomize the data or information extracted from them must rather be conceived as additional security and robustness measures of the framework, and not as the main protection tool.

Giving out our data is not yet completely safe; it is essential to identify at an early stage any potential risk associated with the release of certain information.

Bibliography

Agrawal, R., & Srikant, R. (2000). Privacy-preserving data mining. In Proceedings of the 2000 ACM SIGMOD international conference on Management of data (pp. 439-450).

Alasadi, S. A., & Bhaya, W. S. (2017). Review of data preprocessing techniques in data mining. Journal of Engineering and Applied Sciences, 12(16), 4102-4107.

Article 8 - Protection of personal data. European Union Agency for Fundamental Rights. Retrieved 9 December 2021, from https://fra.europa.eu/en/eu-charter/article/8-protection-personal-data.

Berendt, B. (2019). AI for the Common Good?! Pitfalls, challenges, and ethics pen-testing. Paladyn, Journal of Behavioral Robotics, 10(1), 44-65.

Blumenthal, M. S. (2017). Data, Data, Everywhere-How shall we live with it?

Christou, G., & Rashid, I. (2021). Interest group lobbying in the European Union: privacy, data protection and the right to be forgotten. Comparative European Politics, 19(3), 380-400.

Cohen, J. E. (2013). What privacy is for (November 5, 2012). Harvard Law Review, 126.

Deguchi, A., Hirai, C., Matsuoka, H., Nakano, T., Oshima, K., & Tai, M. (2020). Society 5.0 A People-centric Super-smart Society. Hitachi-UTokyo Laboratory (H-UTokyo Lab.) The University of Tokyo Bunkyo-ku, Tokyo, Japan. Springer open.

Digital 2021: Global Overview Report — DataReportal – Global Digital Insights. DataReportal – Global Digital Insights. (2021). 13 November 2021, from https://datareportal.com/reports/digital-2021-global-overview-report.

ePrivacy: consultations show confidentiality of communications and the challenge of new technologies are key questions - Digital Single Market - European Commission. Digital Single Market - European Commission. Retrieved 9 November 2021, from https://wayback.archive-it.org/12090/20190630043525/https://ec.europa.eu/digital-single-market/en/news/eprivacy-consultations-show-confidentiality-communications-and-challenge-new-technologies-are.

Foulonneau, M., Martin, S., & Turki, S. (2014, February). How open data are turned into services? In International Conference on Exploring services science (pp. 31-39). Springer, Cham.

Gladden, M. E. (2019). Who will be the members of Society 5.0? Towards an anthropology of technologically posthumanized future societies. Social Sciences, 8(5), 148.

Gullo, F. (2015). From patterns in data to knowledge discovery: What data mining can do. Physics Procedia, 62, 18-22.

Haider, J., & Sundin, O. (2019). Invisible search and online search engines: The ubiquity of search in everyday life. Routledge.

Haoxiang, W., & Smys, S. (2021). Big Data Analysis and Perturbation using Data Mining Algorithm. Journal of Soft Computing Paradigm (JSCP), 3(01), 19-28.

Holroyd, C. (2020). Technological innovation and building a 'super-smart society: Japan's vision of society 5.0. Journal of Asian Public Policy, 1-14.

Holroyd, C. (2020). Technological innovation and building a 'super-smart society: Japan's vision of society 5.0. Journal of Asian Public Policy, 1-14.

Internet Live Stats. Internet Live Stats. Retrieved 15 November 2021, from https://www.internetlivestats.com/watch/electricity-used/.

Irwin, L. (2021). GDPR | Personal Data vs Sensitive Data: What's the Difference?. IT Governance UK Blog. Retrieved 13 November 2021, from https://www.itgovernance.co.uk/blog/the-gdpr-do-you-know-the-difference-between-personal-data-and-sensitive-data.

Khari, M., Garg, A. K., Gandomi, A. H., Gupta, R., Patan, R., & Balusamy, B. (2019). Securing data in the Internet of Things (IoT) using cryptography and steganography techniques. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 50(1), 73-80.

Law. J. (2015) Oxford Dictionary of Law. Oxford: Oxford University Press

Li, N., Li, T., & Venkatasubramanian, S. (2007, April). t-closeness: Privacy beyond k-anonymity and l-diversity. In 2007 IEEE 23rd International Conference on Data Engineering (pp. 106-115). IEEE.

Machanavajjhala, A., & Reiter, J. P. (2012). Big privacy: protecting confidentiality in big data. XRDS: Crossroads, The ACM Magazine for Students, 19(1), 20-23.

MacMenemy, D. (2016). (title) Accessed online on 05/04/2018 https://pure.strath.ac.uk/portal/files/54531639/McMenemy_IFLA_2016_rights_to_priv acy_and_freedom_of_expression_in_public_libraries.pdf

Mehmood, A., Natgunanathan, I., Xiang, Y., Hua, G., & Guo, S. (2016). Protection of big data privacy. IEEE Access, 4, 1821-1834.

Menchen-Trevino, E. (2013). Collecting vertical trace data: Big possibilities and big challenges for multi-method research. Policy & Internet, 5(3), 328-339.

Otto, M. (2018, September). Regulation (EU) 2016/679 on the protection of natural persons about the processing of personal data and the free movement of such data (General Data Protection Regulation–GDPR). In International and European Labour Law (pp. 958-981). Nomos Verlagsgesellschaft mbH & Co. KG.

Press corner. European Commission - European Commission. (2021). Retrieved 13 December 2019, from https://ec.europa.eu/commission/presscorner/detail/en/IP_19_2956.

Rajendran, K., Jayabalan, M., & Rana, M. E. (2017). A study on k-anonymity, l-diversity, and t-closeness techniques. IJCSNS, 17(12), 172.

Shrivastva, K. M. P., Rizvi, M. A., & Singh, S. (2014, November). Big data privacy based on differential privacy is a hope for big data. In 2014 International Conference on Computational Intelligence and Communication Networks (pp. 776-781). IEEE.

Shukla, S., & Sadashivappa, G. (2014, March). A distributed randomization framework for privacy preservation in big data. In 2014 Conference on IT in Business, Industry and Government (CSIBIG) (pp. 1-5). IEEE.

Siddiqa, A., Karim, A., & Gani, A. (2017). Big data storage technologies: a survey. Frontiers of Information Technology & Electronic Engineering, 18(8), 1040-1070.

Singh, D. (2003). Data Security and Privacy in Data Mining: Research Issues & Preparation. Ijcttjournal.org. Retrieved 13 November 2021, from http://ijcttjournal.org/Volume4/issue-2/IJCTT-V4I2P129.pdf.

Sweeney, L. (2002). k-anonymity: A model for protecting privacy. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 10(05), 557-570.

Thomas, G. (2021). What Happens Every 60 Seconds Online?. Gtweb.co.uk. Retrieved 13 December 2021, from https://gtweb.co.uk/what-happens-every-60-seconds-online.

UNU-MERIT » Steering the course of innovation – towards sustainable development. Merit.unu.edu. (2018). Retrieved 13 December 2021, from https://www.merit.unu.edu/steering-the-course-of-innovation-towards-sustainable-development/.

Vassakis, K., Petrakis, E., & Kopanakis, I. (2018). Big data analytics: applications, prospects and challenges. In Mobile big data (pp. 3-20). Springer, Cham.

Web, M. (2021). Maps on the Web. Maps on the Web. Retrieved 13 November 2021, from https://mapsontheweb.zoom-maps.com/post/172521439389/privacy-and-data-protection-laws-by-country-via.