# Data Security and Privacy of Individuals in Data Mining: A Critical Analysis of Data Mining in India

**Nilanjan Chakraborty[1], Yogesh Mishra[2], Prattyay Chakraborty[3], Debalina Sengupta[4], Aakash Bag[5], Ankita Mishra[6]**

*[1]Assistant Professor, SOA University, Bhubaneswar, [2]Assistant Professor, KIIT University Bhubaneswar, [3]Assistant Professor, Calcutta University, Kolkata, [4]Assistant Professor, Calcutta University, Kolkata,[5]Assistant Professor, MATS University, Raipur, [6]Research Assistant, KIIT University, Bhubaneswar*

## Abstract

We live in world where every day vast amount of data flows in through various medium. Data mining could be considered because of the natural growth of information technology. The concept of data mining is widely used all over as an instrument to mine hidden information from the database of companies that can be used by the respective companies to deal with their complicated business affairs and improve their services accordingly. Despite that the information revealed by data mining applications, people have shown their increasing concern about the other side of the coin, specifically the privacy threats posed by data mining. An individual's privacy may be hampered or violated due to unauthorized access to their sensitive personal data, unwanted discovery of once embarrassing information etc. In this research paper the author desire to highlight the probable privacy issues that an individual must face due to data mining.

***Keywords:*** *Data mining, Information Technology, Individual's Privacy, Sensitive Personal Data, Unauthorized Access.*

## Introduction

Since ancient times, the search for useful and important information has been carried on manually. With an expeditious growth in the volume of data, it is not possible to mine useful information manually. So, the miners have come up with more efficient and effective technologies or data mining techniques to handle such a search of useful patterns or knowledge. Data mining can be contemplated as an outcome of natural evolution of information technology. Data mining could be web data mining, social data mining, image data mining, healthcare data mining, financial data mining, e-book data mining, SQL data mining, mining data for fraud detection, stock market data mining, text or multimedia data mining, data mining for consumer segmentation, capturing, analysis and interpretation of data, new stories extraction, tracking and analyzing competitor's growth, meta- data extrication from various websites and all. Since data mining deals with processing of "*sensitive personal information*", data privacy and data security concerns are at its heights. With the advent of digital era and development of various technologies the data security, data protection and privacy preservation concern has just increased to another level.

**Privacy, Security and Social Impact of Data Mining:** Dr. B. Thuraisingham explained the main essence of Data mining by applying his own point of view and opinion. He stated, "Most of the time, data mining is part of our lives and we are often unaware of its presence. Data mining is present in many aspects of our lives and can influence our well-being of which we are unaware of such as ubiquitous and invisible data

**Corresponding Author:**
**Nilanjan Chakraborty**
Assistant Professor, SOA National Institute of Legal Studies, Shiksha 'O' Anusandhan University, Bhubaneswar, Odisha–751003, India
e-mail: nilanjanchakraborty@soa.ac.in
Contact No.: +91 8013552943

mining. Invisible data mining can be in the form of smart software, such as search engines, customer adaptive web services, 'intelligent' database system, email managers, ticket masters and so on, incorporate data mining into its functional components".[1]

However, it is important to know that many data mining application don't even touch personal data, such as applications involving natural resources, the predictions of floods and drought, meteorology, astronomy, geography, geology, biology and other scientific and engineering data. Furthermore, most studies in data mining focus on the development of calculable algorithms and do not involve personal data.[2]

Dr. G. Piatetsky rightly stated "The focus of data mining technology is on the discovery of general or statistically significant patterns, not on specific information regarding individuals. For the data mining application that do involve personal data, in many cases, simple method such as removing sensitive IDs from data may protect the privacy of most individuals. Nevertheless, privacy concerns exist wherever personally identifiable information is collected and stored in digital form and data mining programs can access such data, even during data preparation".[3]

Dr. Mark Todd stated, "Improper and nonexistent disclosure can be the root cause of privacy issues. To handle such concerns, numerous data security-enhancing techniques have been developed. There has been a great deal of recent effort on developing privacy preserving data mining method. Many data security-enhancing techniques have been developed to protect data. Database can employ a multilevel security model to classify and restrict data according to various security level, with users permitted access to only their authorized level", wherehe explained the whole concept of Big data and how it is related to modern data science. Encryption is another technique in which individual data items may be encoded. This may include blind signatures (which build up on public key encryption), biometric encryption (where the image of the person's iris or fingerprint is used to encode his/her information) and anonymous database (which permits the consolidation of various database but limit access to personal information to only those who need to know; personal information is encrypted and stored in different location.[4]Intrusion detection is another active area of research that helps protect the privacy of personal data.

Privacy Issues and Data Mining when Violates Privacy of anIndividual

As the definition goes 'Data Mining' is the mining of personal data stored in database for identifying potential knowledge, formerly unidentified, implicit information. [5] In an information era, preserving privacy refers to concealing personal information and having control over the use of such sensitive personal information. Before understanding how data mining has given rise to privacy issues, let us look at this example and get a hang of it. Example- There is a user who is capable enough to apply data mining software which is easily available online. This user can pose different queries and infer sensitive hypotheses. An inference problemoccurs when a user who is engaged in data mining using various data mining techniques tries to withdraw inferences from the processed data. There are many data mining tools available online.[6]

The inference problem can be solved by building an inference controller which can detect the motives of the miner and prevent or stop the inference problem from occurring. Also, we can have different levels of privacy. For instance- Name, age, gender is comparatively less private than mobile phone numbers, salaries which are more private. Similarly, names and medical records collectively could be more private.[7]

With the origin of World Wide Web, there is now ample of information available about individuals that can be obtained within seconds. Such information could be obtained by way of mining or just from information retrieval. This makes it necessary to enforce controls on data mining technologies as well as databases. Enforcing such a control is practically difficult task with respect to data mining.[8]This calls for developing new techniques to stop users from extracting information from the data whether available on server or web.

Stepping on the social aspect, there prevail cultures where privacy of individual is preached. Also, there exist certain cultures where it is difficult to ensure privacy.[9] That could be with respect to technological or political issues or the fact that people believe that privacy is not essential. Many technologies or technological solutions have been proposed for data security in general and confidentiality could be used for privacy. The big challenge is to prevent violation of privacy without prejudicing the benefits of datamining.[10]

Data mining is often used to detect terrorist

activities by mining data of individuals and finding unusual patterns and fraudulent behavior. Acting in faith of national security, data mining is used for counter-terrorism applications.[11] When at one hand, all these applications can benefit humans and save their lives, at the other hand it possesses threat to the individual's right to privacy. There arises the conflict between two is difficult. This scenario has led to the developing of such privacy preserving measures that would help prevent privacy violation and at the same time will allow us to have the fruits of data miningtechnologies. [12]

Data mining techniques sometimes also discloses business related critical information which compromises free competition and therefore discloser of personal or confidential information should be prevented. [13] Hence, it was important to address the issue of privacy preserving in data mining. Different sanitizing measures were proposed to conceal personal data and sensitive patterns. [14]

Certain ways in which privacy concerns have been raised by data mining techniques are as follows:[15]

The implicit pattern covering information about the data subjects that can be derived from result in the data mining process vs. the explicit nature of personal information extracted by way of traditional database retrieval techniques.

The probable use of only single database to extract information about data subjects vs. use of multiple databases to extract information regarding the data subjects. The use of "open ended" queries to find out information on associations about an individual or groups vs. using specific questions to derive data about relationships that are already known.

The non-predictive aspect of unknown information obtained about an individual from data mining vs. general predictive aspect of data derived from traditional database technique.

Information of public nature is mostly extracted about an individual through data mining process vs. the intimate nature of information retrieved using traditional techniques about individual.

We often talk about privacy getting compromised in the process of data mining. How can we define privacy? We have heard people saying that "*keep the information about me from being available to others*".[16] And the Webster's Dictionary defines privacy as "*freedom from unauthorized intrusion*". In data mining case, the use of personal data in a way has some negative effect on someone's life which creates concern. So here we are trying to identify that line beyond which data mining will lead to violation of right to privacy or threat data security. So basically, if the data is not misused, for most of the people it does not amount to any kind of privacy violation. The problem arises once the data mining results or information is released and is open to the probability of being misused. [17]

It becomes difficult to impose universally acceptable rules to differentiate between right and wrong and demarcate the line beyond which data mining would pose threat to privacy.[18] When the database is handled by an authorized miner for a noble cause to achieve benefits legally, then it is very much within the limits of data protection and data privacy. But the moment data mining technologies are used by unauthorized person or user or local service providers, there arises a probability of individual's right to privacy getting violated. Using this distinction and ensuring that a data mining project will not enable misuse of personal sensitive information and opens opportunities that "*complete privacy*" would present.[19] The same concern arises while collecting data. While collecting data we might come across and learn about individual's data items. An individual might not care about someone knowing some common information about them like- name, date of birth, gender etc., but getting to know all these information leads to identify theft.

**Individual Rights in Relation to Processing of their Personal Data:** Individuals do possess certain rights regarding their personal data or information they share with service providers.[20]An individual need to be aware of those rights to protect their sensitive personal information from getting misused. Some of such rights to be kept in mind are:

**Access to data:** Rule 5, subsection 6 of the IT Rules directs that any person or body corporate on its behalf must allow providers of information or data subjects to examine the information they may have given. The above sectionprovides:

*"Body corporate or any person on its behalf permit the providers of information, as and when requested by them, to review the information they had provided and ensure that any personal information or sensitive*

*personal data or information found to be inaccurate or deficient shall be corrected or amended as feasible: Provided that a body corporate shall not be responsible for the authenticity of the personal information or sensitive personal data or information supplied by the provider of information to such boy corporate or any other person acting on behalf of such body corporate".*[21]

## Correction and Deletion

Rule 5, subsection 6 of the IT Rules[22]says that "the data subjects must be permitted to access to the data provided by them to check that any information found to be inexact or deficient shall be rectified or amended as feasible. Although the rules don't directly talk about the deletion of data, they state in Rule 5, subsection 1, which corporate institution or person representing them must secure written consent from data subjects regarding the use of the sensitive information they provide". The above subsection provides:

"Body corporate or any person on its behalf shall obtain consent in writing through letter or Fax or email from the provider of the sensitive personal data or information regarding purpose of usage before collection of such information."

**Objection to processing:** Rule 5 of the IT Rules[23] says that the data subject or provider of information shall have the option to later withdraw consent which may have the choice to later withdraw the consent which may have been issued to the corporate institution previously; such withdrawal of consent should be stated in writing to the corporate body. On withdrawal of consent, the corporate body is forbidden from processing the personal information or data in question.

**Disclosure of data:** Data subjects have rights with respect to disclosure of the information they provide. The disclosure of sensitive personal information needs the provider's prior permission unless either:

i. Disclosure has already been agreed to be in the contract between the data subject and the data controller or[24]

ii. Disclosure is required for compliance with a legal obligation.

There are exceptions to this rule, if an order under law has been made, or if a disclosure must be made to Government agencies mandated under the law to obtain information for the purpose of: -

a. Verification of identity.

b. Prosecution or punishment of offences

c. Detection, prevention and investigation of crime

Recipients of this sensitive personal information are barred from further disclosing said information.

## Suggestions and Discussions

As the questions have arisen regarding the definition and limitations of privacy in data mining, let us give a bird eye view on the newly invented technology of **Secure Multiparty Computation (SMC)**.[25] The basic idea behind SMC is that the parties those who are involved acquire nothing but the results.

The technology of SMC makes sure that the involved party learns or acquires nothing but only the results. SMC involves the involvement of the third party. There may be substantial communication between the involved parties, but SMC makes sure that the parties do not learn anything from the concerned communication.

From much possible approach one of the reliable approaches is **Constraint-based data mining**. "*This part of research is concerned with improving the efficiency of the algorithms and the understandability of results through providing up-front constraints on what results would be of interest.*" [26]

## Conclusion

Our main task for now is to determine that "line" beyond which data mining would result in the violation of an individual's right to privacy. When a database or data set is accessed from a system by an authorized person or an authorized miner with clear intention, it is pretty much within the limits of data security. But the moment, the information stored in the database or data sets are accessed by unauthorized persons for their personal use, there lies a probability of the information getting misused or misplaced. Nowadays data mining software are available online, so it is easier for those unauthorized dealers and miners to access the information and exploit them. This has increased the concern for data protection and privacy preservation among the individuals.

Looking at the concern for the privacy among the individuals various Privacy Preserving Data Mining technologies (PPDM) have been developed and introduced. Secure Multiparty Computation, Data Obscuring, anonymization, perturbation and

augmentation have also been implemented. But it should surely be appreciated how the competent authorities are taking necessary measure and ways to prevent any such data mining which is infringing individuals right to privacy and I must say that the there is a long way to go and to develop the erring rod to maintain the check and balance in the abstract world by cyberspace.

**Ethical Clearance:** Not required as the researcher has just referred to some published works. The research is doctrinally undertaken, completely by the researcher himself**.**

**Source of Funding:** Self

**Conflict of Interest:** Nil

## References

1. Thuraisingham B. Data Mining, National Security, Privacy and Civil Liberties. SIGKDD. 2015;4(2):1-5.
2. O'Leary D. Some Privacy Issies in Knowledge Discovery: OECD Personal Privacy Guidlines. IEEE Expert. 2016;10(2):48-52.
3. Clifton C, Marks D. Security and Privacy Implications of Data Mining.Data Mining and Knowledge Discovery, Montreal, Canada, 1996.
4. Singh DK, Swaroop V. Data Security and Privacy in Data Mining: Research Issue & Preparation. International Journal of Computer Trends and Technology. 2013;4(4):194-200.
5. Yan W, Royakkers L. Ethical Issues in Web Data Maining. Ethics and Information Technology. 2017;6(2):129-40.
6. Castro VE, Brankoyic L. Data Swapping: Balancing privacy against precision in mining for general patterns. Department of Computer Science. 2011; New York.
7. Aggarwal R, Srikant R. Privacy Preserving Data Mining. SIGMOND. 2000;10:120-21.
8. Kaufmann M. Data Mining concepts and Techniques. Cyber Security and Technology. 2012;6(2).
9. Clifton C, Marks D. Security and Privacy implications of Data Mining. Data mining and Knowledge Discovery. 1996; Montrel, Canada.
10. Chawla S. Towards Privacy in public Databases. TCC. 2005; California.
11. Bettini C. Protecting privacy against location based personal Identification. Secure Data Management. IJCRTE. 2005;7(4).
12. Fung B. Top Down Specialization for Information and Privacy Preservation. ICDE Conference. 2005;Tokyo.
13. Sengupta S. DATA PROTECTION [Internet]. ICLG. 2018 [cited 17 March 2017]. Available from: https://iclg.com/practice-areas/data-protection/data-protection-2017/india#chaptercontent14.
14. The Information Technology (Reasonable security practice and procedure and sensitive personal data or information) Rule, 2011.
15. The Information Technology (Reasonable security practice and procedure and sensitive personal data or information) Rule, 2011.
16. Dunham M. Dtata Mining - Introductory and Advanced Topics. 2009. Chicago.
17. Kargupta H, Joshi A, Sivakumar K, Yesha Y. Introduction to Data Mining. IJRTE. 2011;14(5):45-9.
18. Okar MC. Ethical and Legal Aspects of Data Mining. Journal of Yasar University. 2008;9(3):234-8.
19. Ashwinikumar UM, Anandakumar KR. Towrads Implementation of Ethical Principles: Privacy and Confidentiality in Medical Data Mining - Indian Scenario. International Journal of Pure and Apllied Mathematics. 2009;10(4):46-50.
20. Klosgen W. KDD: Public and Private Concerns. IEEE Expert. 1995;10(2):55-7.
21. Van WL, Royakkers L. Ethical Issues in Web Data Mining. Ethics and Information Technology. 2001;6(4):129-40.
22. The Information Technology (Reasonable security practice and procedure and sensitive personal data or information) Rule, 2011.
23. The Information Technology (Reasonable security practice and procedure and sensitive personal data or information) Rule, 2011.
24. Shah A, Zacharias N. Right to Privacy and Data Protection [Internet]. Security and Privacy Concern. 2000. [cited 24 March 2018] Available from: http://www.nishithdesai.com/fileadmin/user_upload/pdfs/Right_to_Privacy_-_data_protection.pdf.
25. Yao A. How to Generate and Exchange Secrets. IEEE. 1986;27:162-7.
26. Sweeney L. Computer Disclosure Control: A primer on Data Privacy Protection. MIT. 2001.