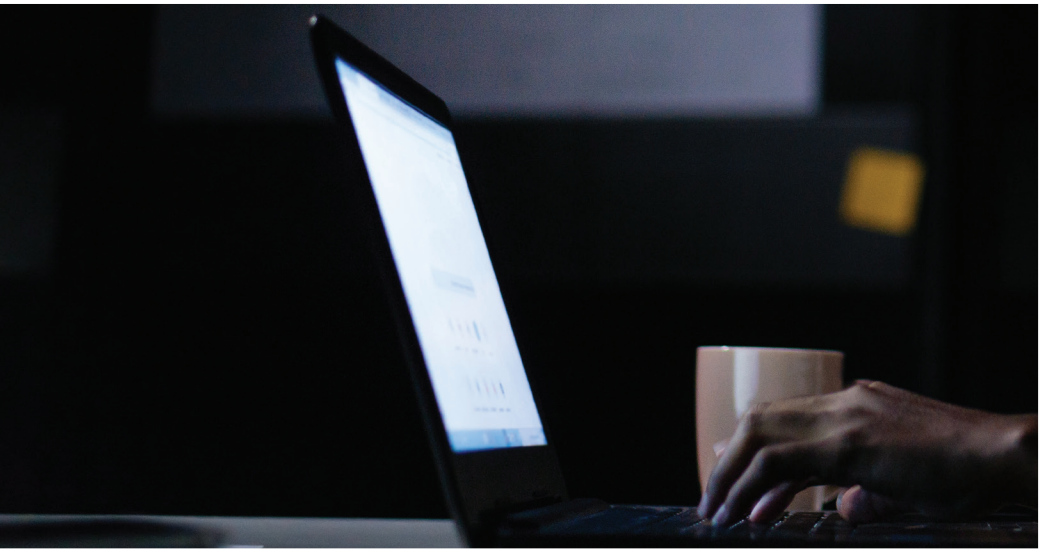
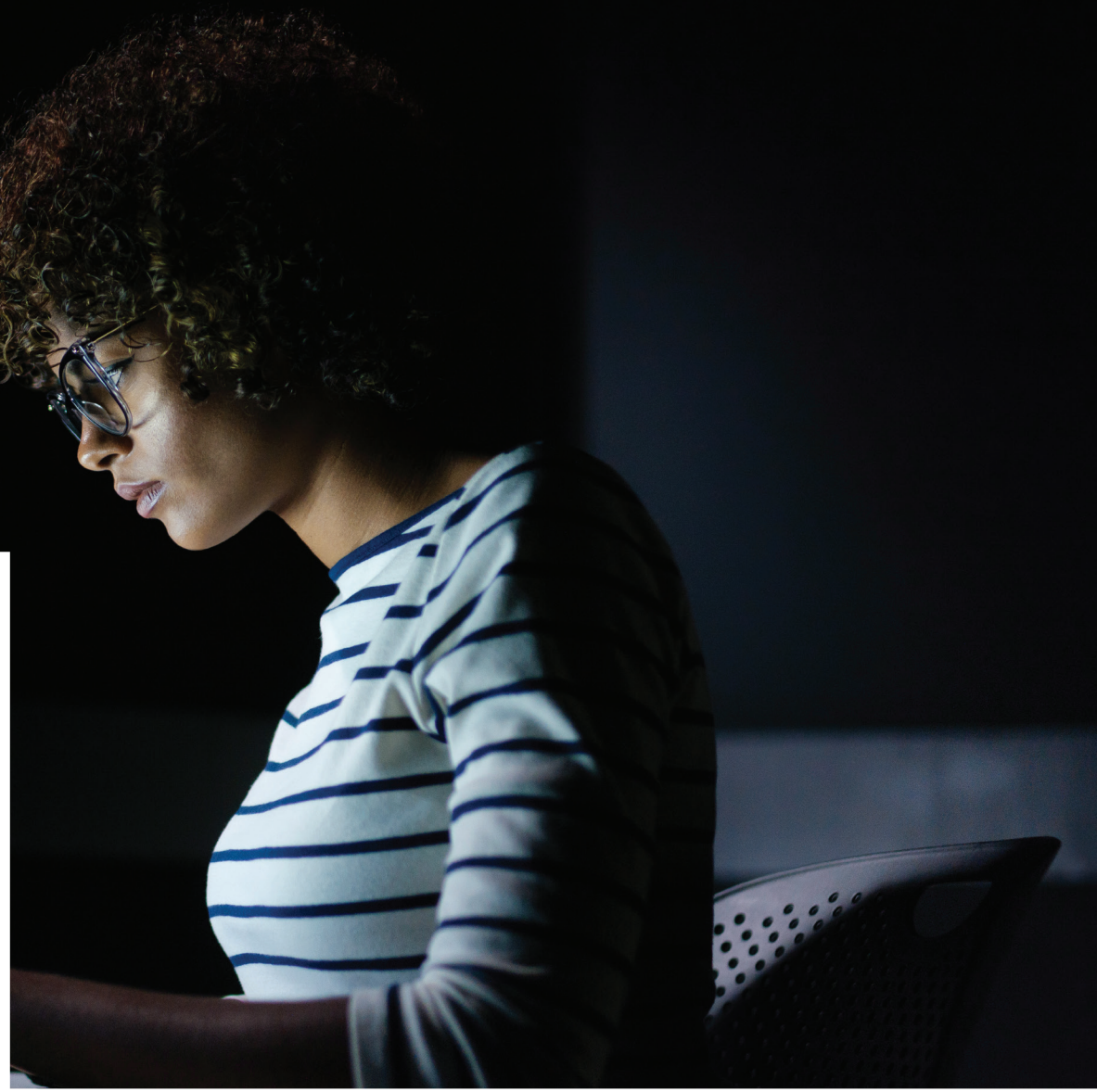
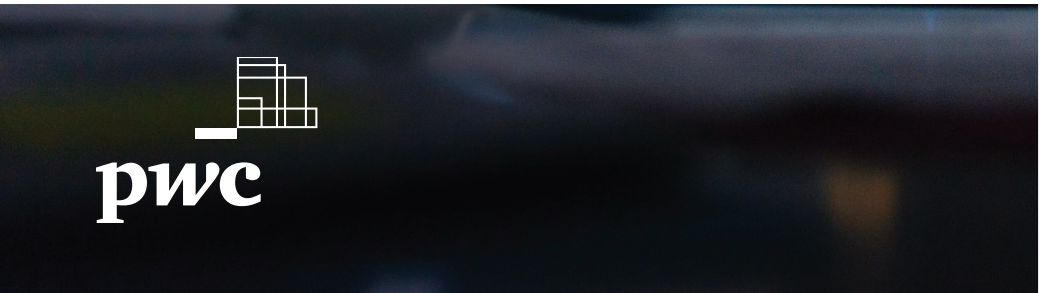
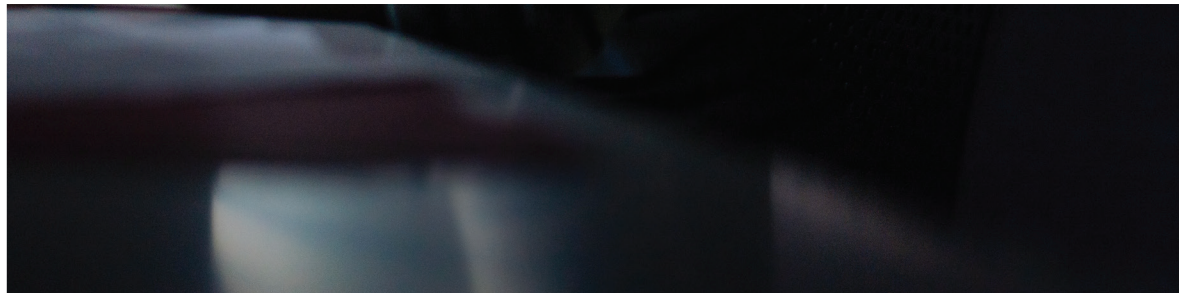
Decoding the dos and don’ts of data profiling

July 2020









pwc

01 Data profiling process

Data profiling is the process of reviewing and analysing raw data collected from multiple channels and sources to understand the content, data structure and interrelationships in data sets. Profiling helps in identifying and characterising the information embedded within the data. It includes different statistical and analytical algorithms that can be leveraged to provide insights into the content of the data and identify qualitative characteristics of underlying data sets.

Data, particularly when aggregated, can reveal a lot of information. If someone enquires about insurance, visits a website of an insurance company and calls an insurance company's customer care, it can be assumed that the person is looking to buy a new insurance plan or renew an existing plan. Profiling helps in automation of such outcomes and results by providing large and diversified data sets sourced from varied structured and unstructured data sources. These data sets include personal data like behavioural, social, location and contacts. Automated data profiling based on machine learning (ML) also provides more comprehensive insights for better decision making. Results of customer age and product usage profiling can be aggregated and used for customer segmentation, customised service offering and digital marketing.

Data profiling has long been considered as a critical activity to understand the structure of large data sets collected using various methods. The following data collection methods are practised by companies.

- Most companies ask for data directly from their customers, usually at the beginning of a business relationship. For example, when someone subscribes to a service or tries to download an application, he/she is usually required to fill out a consent form for the service provider to access information. Additionally, organisations can conduct customer surveys, ask questions and record the responses for customer profiling.
- Personal data can be sourced indirectly or collected from multiple other channels like web portals, company websites and social media platforms. These online sources mostly use embedded cookies and web beacons that help to track browsing histories of any individual. Companies use the collected data to create behavioural or preference patterns
- Customer data can also be acquired from data companies. These companies collect, analyse and sell customer and business data for targeted advertising campaigns.
- Once companies source all types of data, including sensitive and personal data, they often start the profiling process without notifying or obtaining due consent from individuals from whom the data has been collected. Though profiling can lead to better and quicker business decision making, there could be significant health, financial and reputational risks for individuals if organisations use data irresponsibly without establishing proper data usage controls and processes.

02 Data profiling in the era of data privacy

In the context of data privacy, profiling involves the processing of personal data. Therefore, it is important to act in accordance with the privacy laws such as the General Data Protection Regulation (GDPR), California Consumer Privacy Act (CCPA) and Personal Data Protection (PDP) Bill, 2019.

As virtual activities and digital presence become more traceable, data profiling raises pressing policy questions on how people's privacy can be protected when sensitive information can be predicted from aggregating disconnected data. Two pertinent concerns related to data profiling and privacy are mentioned below.

- How can organisations ensure that profiling (and the decisions it informs) is legal, fair and non-discriminatory?
- How can data principals exercise their rights if the processing itself is not transparent?

Many data privacy and protection regulations have designed provisions to address these risks and provide guidance to organisations dealing with data.

Challenges and risks in profiling of personal data

As rules and restrictions related to personal data processing are becoming stringent worldwide, companies are expected to face many issues as they align themselves with regulatory requirements to sustain their core business. Some of the key challenges and risks that organisations need to address are summarised below.

Challenges

- Understanding obligations as a data fiduciary or data processor of personal data and fulfil responsibilities by transforming data profiling processes, as per data privacy guidelines.
- Identifying the significance and impact on concerned persons due to decisions taken on the basis of data generated after profiling personal data.
- Balancing an organisation's interests with the data principal's interests, rights and freedom.
- Drawing a precise boundary for the types of decisions that would qualify as significant and fulfilling the legal basis for processing of personal data.
- Identifying new ways of working with personal data to achieve business objectives like increased customer retention and increased return on investment RoI.
- Assessing how to manage the rights of data principals or data subjects whose personal data has been profiled.
- Ensuring that current legal processes and frameworks are adhered to for data protection while also considering the legitimate interest of profilers.

- Putting additional checks in place for profiling/ automated decision-making systems to protect any vulnerable groups.
- Avoiding irresponsible handling of profiled customer data as it can result in unethical usage.
- Establishing safeguards and measures to be used continuously as a part of data profiling processes.
- Minimising the possibility of non-compliance that can result in hefty monetary penalties and compromise customer trust.

Risks

- Profiling and automated decision making may pose risks to data principals. Individuals exposed to such risks may be discriminated against or subjected to abuse and stereotyping.
- Inappropriate results of personal data profiling can cause much harm to data principals and lower customer trust. Low customer trust can cause reputational damage to an organisation.
- Automated data profiling results can be considered unethical and even illegal in certain scenarios. This can occur when a profiling exercise is focused on gathering data on ethnicity, gender, religion or sexual orientation of data principals.
- Data profiling results and their significance largely depend on the authenticity of data source and underlying quality of data. Usage of low-quality data can lead to inaccurate results and impact data principals personally and organisations financially as they have to face monetary and legal consequences.

03 Personal Data Protection Bill, 2019 – a view on data profiling

The PDP Bill, 2019, defines profiling as any form of processing of personal data that analyses or predicts aspects concerning the behaviour, attributes or interests of a data principal.[1] The bill provides guidance to protect an individual's privacy related to data collection processes, data usage and sharing of personal data. The bill has defined the rights of data principals and empowers them to control and avoid misuse of their data by any third party, individual or organisation.

Profiling dos and don'ts in accordance with the PDP Bill

Most organisations currently rely on email marketing and online campaigns to target/reach out to customers, using data profiling up to an extent. Profiling is important to address customer needs in a personalised and relevant manner. More personalised communication along a customer's journey results in higher RoI. The prospect of higher RoI has resulted in organisations often not establishing controls, standards and measures before processing personal and sensitive customer data. In the recent past, regulators, along with individuals, have realised the potential negative impact of uncontrolled access to data and its processing by organisations. This resulted in introduction of multiple data privacy laws worldwide. These laws define stringent rules and restrictions on collecting, storing and processing of personal data.

The introduction of the PDP Bill in India is expected to compel organisations to revamp their business and underlying data management processes. In the following sections, we have discussed what is permissible under the proposed new legislation and what is no longer permitted.

Dos of data profiling

The activities detailed below are mandatory for personal data processing and profiling by data fiduciaries.

- Notify the data principal about profiling and its purpose at the time of data collection.
- Bring in transparency in the processing of personal data by requiring organisations and entities who act as data fiduciaries to inform the data principal and make information available. The data principal should also be able to give, withdraw, review and manage his/her consent through an accessible platform.
- Establish suitable safeguards for profiling, including anonymisation or pseudonymisation as components of profiling-based activities.
- Implement technical and organisational security controls to correct inaccuracies and avoid errors, thus minimising the risk of inaccurate profiling.
- Ensure quality of stored personal data. It should be accurate, complete and consistent with the information provided by the data principal.

1 http://164.100.47.4/BillsTexts/LSBillTexts/Asintroduced/373_2019_LS_Eng.pdf

- Establish frequent checkpoints to validate collected data and update it if required before profiling begins.
- Carry out profiling based on legitimate purposes.
- Use appropriate ML algorithm based on data categories like Sensitive personal data and critical personal data for automated data profiling. The algorithm should internally use processing logic which produces appropriate results and does not pose any risk to data principals.
- Inform data principals about the existence and type of ML algorithms used for processing. There is a wide range of different ML algorithms available, e.g. k-means clustering, k-nearest neighbor, support-vector machine, logistic regression and market basket analysis.
- Data fiduciaries should implement appropriate measures to mitigate the risk of algorithms working on incomplete or unrepresentative data that may generate spurious correlations, resulting in unjustifiable decisions.
- Test the algorithm/logic utilised for automated profiling and decision making to ensure the system performs as intended and does not produce any discriminatory, erroneous or unjustified results.
- Utilise third-party auditing of such systems (where decision making based on profiling has a high impact on individuals) to receive an independent view on the processing and potential consequences of automated profiling and decision-making activities on data subjects.
- Implement technical and administrative measures for data minimisation to incorporate clear retention periods for profiles and for any personal data used when creating or applying the profiles.
- Liaise with ethical review boards and external counsel to assess potential harms and benefits to data principals on using particular applications for profiling.
- Train employees/contractors to make them aware on data privacy aspects such as data profiling, data retention, data subject access requests and data security.
- Implement retention policies in order to ensure that a cookie is read and is automatically deleted after a certain period (necessary for the purpose of processing).
- The following factors should be taken into consideration while processing personal data and taking decisions based on profiling results:
  - the level and comprehensiveness of the profiling results
  - the impact of the profiling results and decision making on the data subject
  - the measures required to avoid bias, non-discrimination and inaccuracy in profiling.

Don’ts of data profiling

As per the PDP Bill, 2019, the following activities related to data processing and profiling have been prohibited.

- The bill prohibits processing of personal data without any specific, clear and lawful purpose. Hence, data cannot be profiled unless explicit consent has been obtained and a specific, clear and lawful purpose for profiling has been established.
- The draft bill lists out certain cases which provide for processing of personal data without consent. Recruitment and termination of employment have also been brought under categories of processing personal data. However, if such data meets the criteria of being sensitive data, then such processing cannot be done without prior consent. Thus, profiling should be limited to personal data and special categories of data should not be profiled, even under scenarios in which processing such data is exempted from consent requirement.
- Profiling of special categories of personal data (e.g. racial, ethnic, sexual orientation and health) is not permitted. Such data can only be profiled if the data principal has provided explicit consent or processing is necessary for public interest.
- Processing of sensitive personal data and critical personal data outside India is prohibited. Sensitive personal data may be transferred outside India, subject to conditions for transfer. However, it shall continue to be stored within India. Critical personal data (the definition of which is to be notified by the Central Government) can only be processed in India.
- Any third party that is processing personal data for profiling should not use said data for personal purposes. Contractual obligations by data fiduciaries with regard to processing and profiling of personal data is to be noted in agreements.

In addition to the above dos and don’ts, the PDP Bill, 2019, also highlights below points:

- The Data Protection Authority (DPA) of India may specify additional safeguards and restrictions for the purpose of repeated, continuous or systematic collection of sensitive personal data for profiling.
- For large-scale profiling of sensitive personal data such as genetic data or biometric data, data fiduciaries must undertake a data protection impact assessment. Data fiduciaries are barred from profiling, tracking or behaviourally monitoring children and undertaking any processing of personal data that can cause significant harm to them.

Way forward

Data profiling is important for organisations to understand customer behaviour in this digital age. And today, due to increased emphasis on ethical and legitimate processing of personal data practices, there is a change of paradigm in the field of profiling. Companies must adopt new ways of managing personal data, work within the limitations of data privacy laws and at the same time, utilise collected data to achieve business objectives to the best extent possible. It is important for companies to ensure that they engage in data processing and profiling only after undertaking the data protection impact assessment in accordance with applicable data privacy laws. To deal with advancements and changes in data privacy, companies do not have to change their entire profiling practices. However, it is recommended that they evaluate their current data management practices against defined requirements in the PDP Bill, assess identified gaps and make necessary changes. They should also define robust monitoring processes to ensure continuous compliance with regulatory obligations.

About PwC

At PwC, our purpose is to build trust in society and solve important problems. We're a network of firms in 157 countries with over 276,000 people who are committed to delivering quality in advisory and tax services. Find out more and tell us what matters to you by visiting us at www.pwc.com.

In India, PwC has offices in these cities: Ahmedabad, Bengaluru, Bhopal, Chennai, Delhi NCR, Hyderabad, Kolkata, Mumbai, Pune and Raipur. For more information about PwC India's service offerings, visit www.pwc.in

PwC refers to the PwC network and/or one or more of its member firms, each of which is a separate legal entity. Please see www.pwc.com/structure for further details.

Contact us

Dhritimaan Shukla
Partner and Privacy Leader
PwC India
Mobile: +91 98990 38326
dhritimaan.shukla@pwc.com

Sonali Saraswat
Associate Director, Privacy
PwC India
Mobile: +91 96207 01515
sonali.saraswat@pwc.com

This article has been authored by **Pratik Bajoria, Ambrish K. Anand and Abhishek Chaurasia.**

The views expressed in this article are personal opinions of the authors and should not be construed as legal interpretation.

KS/July 2020-M&C 6950