Date and Time: Wednesday, 1 July, 2020 7:29:00 PM MYT

Job Number: 120264652

Document (1)

1. *The Impact of the Personal Data Protection Act 2010 on Data Analytics in the Retail Industry [2020] 3 MLJ lxii*

Client/Matter: -None-

Search Terms: tay pek san

Search Type: Natural Language

Narrowed by:

Content Type	Narrowed by
MY Secondary Materials	-None-

The Impact of the Personal Data Protection Act 2010 on Data Analytics in the Retail Industry [2020] 3 MLJ lxii

Malayan Law Journal Articles

THE IMPACT OF THE PERSONAL DATA PROTECTION ACT 2010 ON DATA ANALYTICS IN THE RETAIL INDUSTRY

Dr Tay Pek San

Associate Professor

Faculty of Law

University of Malaya

INTRODUCTION

Data analytics is the processing of voluminous data using complex algorithms to analyse datasets for hidden patterns, find correlations in the data or draw inferences about specific matters such as individuals' preferences or behaviour. The objective of data analytics is to draw new knowledge, usually in the form of new trends or patterns based on criteria determined beforehand. In the retail industry, data analytics of customers' preferences and market trends is increasingly used by retailers in recent times to enable them to understand their customers better and make more informed business decisions. Retailers regularly collect data about their customers from various sources, such as purchase transactions, loyalty programmes, credit card transactions, customers' surveys, reviews, emails that ask about customers' preferences and websites or social media pages that track customers' activities. These data are extremely useful to retailers who, by using data analytics, can obtain new insights and predict trends about their customers and the market. Retailers are thus able to equip themselves with a clearer picture of their customers' expectations and predict their likely behaviour. With this, retailers are placed in a more competitive position because they can develop strategies to engage with their customers through targeted advertisements that recommend items specifically for them.

In most cases, retailers who collect personal data of their customers, which are usually in the form of past transaction history, will outsource the data analytics operations to third-party service providers because the latter has a dedicated team of data professionals who are trained to process and analyse the data for meaningful insights and conclusions. The data analysts are accountable for all stages relating to the data processing and are tasked with presenting reports to the retailers on new insights and knowledge obtained from the data analytics. Under the Personal Data Protection Act 2010 ('the PDPA'),[1] the data analysts fall within the category of 'data processors'. While the PDPA imposes duties and obligations on data users, there are no corresponding obligations imposed on data processors. Notwithstanding this, data processors are subjected to the same extent of compliance with the requirements of the Act as data users.[2]

Retail data analytics raises issues of personal data protection under the PDPA. This is because, fundamentally, data analytics is a form of data processing which thrives on collecting large amounts of customers' personal data and analysing them to make inferences and predictions about their purchasing preferences. The inferences and predictions on customers' purchasing preferences and interests derived from the analytics provide retailers with useful insights. These are used as the basis for targeted advertisements which are then sent to the customers. The purpose of this paper is three-fold. First, it examines the impact of the PDPA on data analytics in the retail industry. Secondly, it analyses the legal status of inferences and predictions about customers' personal preference, behaviour or interests that are derived from the analytics. Specifically, it analyses whether the inferences and predictions constitute personal data within the ambit of the PDPA. This latter point is important because it determines the legal obligations which the service providers are required to observe vis a vis the inferences and predictions. In the event these are not personal data, it is permissible for the service providers to share the inferences and predictions with third parties who engage their services, subject to any restriction in the agreement between the service provider and the earlier retailer. On the contrary, if the inferences and predictions are personal data, the service provider is bound by the PDPA and is not allowed to share it with third parties. Thirdly, the paper highlights the inadequacies of the PDPA in dealing with data analytics and suggests improvement to some aspects of the Act so as to provide certainty to stakeholders who are involved in data analytics.

The paper is divided into five parts. The second part provides an overview of the process of data analytics in the retail industry. The third part examines and analyses the provisions in the PDPA 2010 that are relevant to data analytics. The fourth part explores, by way of a comparative study, the legal position of data analytics under the European Union data protection law as embodied in the General Data Protection Regulation 2016/679 ('GDPR').[3] The GDPR came into force across the European Union on 25 May 2018. The fifth part analyses the legal status of customers' inferences and predictions that are derived from the data analytics. The sixth part makes recommendations on putting in place a suitable legal environment to ease the development of data analytics. The final part is the conclusion.

DATA ANALYTICS PROCESS FOR TARGETED ADVERTISING IN THE RETAIL INDUSTRY

Targeted advertising is a complex business activity that typically involves different technologies, a team of individuals within an organisation and numerous consumers. Organisations conducting targeted advertising may also forge close data-sharing partnerships with other organisations. The analytics process involved in converting data into useful knowledge and insights for the purpose of targeted advertisement entails a number of steps.[4]

While different process models of data analytics exist, they essentially build on a generic framework which comprises the following operations. First, the retailer identifies the business problem that needs to be solved. Secondly, suitable data samples need to be prepared. In doing so, data sources from which relevant data may be collected need to be identified. This is an important step because data is the fundamental raw material upon which any analytics exercise is carried out. The data that is collected will have a deterministic impact on the interpretation and, consequently, the output of the analysis. Thirdly, the data from various sources is consolidated in the staging area. The staging area is essentially a temporary storage area for the data that is collected, and it is the place where data that is relevant for the purpose of the analysis is extracted and transformed before it is finally loaded in the data warehouse. The step of extraction involves the algorithm crawling through the unstructured dataset to retrieve relevant information for the purpose of further processing. Once the relevant data is extracted, it goes through the transformation step where the data is cleansed to remove corrupted or inaccurate data in order to improve data consistency. The cleansed data is then converted from one format into another format to make it ready for processing at the data warehouse. Finally, in the analytics step, a suitable analytical model utilising a computer program is developed to process the transformed data in order to yield important insights based on the original business problem. The output from the processing, which has to be represented in a user-friendly way so that the patterns that the analytical model captures may be understood, is then interpreted and evaluated by data analysts and business experts.

The process model outlined above is depicted in Figure 1 below. It should be noted that the process model is iterative in nature in that it may be necessary at any stage to go back to a previous step during the exercise, for example, to identify additional information, which may then require additional cleansing and transformation.[5] The iterative process is inherent in the analytics procedure so that patterns, correlation or insights may be extracted to help solve the business problem.

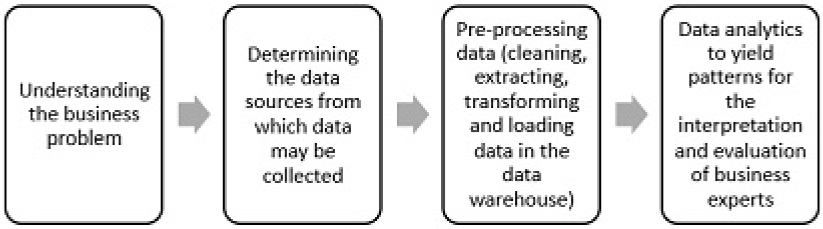
Understanding the business problem
Determining the data sources from which data may be collected
Pre-processing data (cleaning, extracting, transforming and loading data in the data warehouse)
Data analytics to yield patterns for the interpretation and evaluation of business experts

Figure 1

In targeted advertising, the marketing strategy is aimed at addressing specific customer groups or individual customers who possess certain traits that are determined by their purchasing behaviour and history. The data analytics that is carried out for the purpose of targeted advertising will invariably reveal new insights into the customers' interest trends and behavioural patterns. This can contribute significantly to a more precise advertising direction and improve marketing accuracy because business experts are placed in a more informed position to match the interests to individual customers or specific target groups of customers with shared traits. Personalised advertisements, content or product recommendations can be directed to the target groups with a greater degree of precision to meet the customers' preferences. The advertisements may be addressed anonymously to the customers within the target group or, in the case of one-to-one marketing where the customer is known by his name, personalised advertisements may be addressed to him individually. In this way, the advertisements that are delivered have more relevance to the customers concerned since those who are likely to have a strong preference for a product will receive the advertisements instead of those whose preferences do not match the product's attributes.

THE IMPACT OF THE PDPA ON DATA ANALYTICS

The second, third and final steps in a data analytics process outlined above involve the processing of data which, in the context of the retail industry, would comprise personal data unless the data that is processed has been anonymised.[6] Where the data comprises personal data, the data analytics process is subject to the PDPA. 'Personal data' is defined in the PDPA as any information in respect of commercial transactions that relates to a data subject who is identified or identifiable from that information or from that and other information in the possession of a data user.[7] The definition also covers expressions of opinion about data subjects. Examples of personal data includes name, age, date of birth, physical and email addresses, telephone number, national registration identity card number, employees' provident fund number, passport number, SOCSO number, driving licence number, bank card number and internet protocol address.

The Act further defines 'processing' as the act of collecting, recording, holding, storing or carrying out any operation on personal data. There are four types of operation on personal data that are explicitly spelt out in the Act which constitute the processing of personal data, but these are clearly non-exhaustive instances of data processing. The first type of operation is the organisation, adaptation or alteration of personal data. The second is the retrieval, consultation or use of personal data. The third is the disclosure of personal data by transmission, transfer, dissemination or otherwise making the personal data available. The fourth is the alignment, combination, correction, erasure or destruction of personal data. In data analytics, the second step in the process outlined above is the integrated collection of personal data phase and this is a prerequisite initial step in the strategy for delivering targeted advertisements. The third and final steps are the profiling phase where the collected data is organised and structured according to their applicability. This is then analysed for interest and behavioural patterns in order to generate segmented or individual user profiles which are then used for targeted advertisements. These steps undoubtedly constitute data processing and fall within the ambit of the PDPA.

Before discussing the impact of the PDPA on data analytics in the retail sector, it should be noted that the Act has a specific provision which empowers a data subject to give a notice in writing to a data user requiring him to cease processing his personal data for purposes of direct marketing.[8] Direct marketing is defined to mean the communication by whatever means of any advertising or marketing material which is directed to particular individuals.[9] Targeted advertising is direct marketing as the activity falls within the above definition. An individual whose personal data is processed for the purpose of targeted advertising may notify the data user to cease processing his personal data. A data subject who is dissatisfied with the failure of the data user to comply with his notice to cease processing his personal data may submit an application to the Personal Data Protection Commissioner to require the data user to comply with the notice.[10] If the Commissioner is satisfied that the application is justified, he may require the data user to take steps to comply with the notice.[11] A data user who fails to comply commits an offence which carries a fine not exceeding RM200,000 or imprisonment not exceeding two years or both.[12]

The ensuing discussion in this paper presupposes that the data subject has given his consent to the processing of his personal data for marketing purpose and has not withdrawn his consent.

The PDPA lays down seven personal data protection principles and these form the core values of data protection law in Malaysia.[13] Of these seven principles, this paper discusses four of them which are significant to retail analytics. Nevertheless, the other principles are equally relevant to data analytics as with other processing of personal data activities. For instance, the data integrity principle requires a data user to take reasonable steps to ensure that the personal data is accurate, complete, not misleading and kept up-to-date by having regard to the purposes for which the personal data was collected.[14] The data user in a retail analytics case needs to comply with this principle at all stages, such as when collecting data, analysing data, building a profile for an individual or applying a profile to make a decision about the individual.

General principle

Pursuant to s 6(1) of the PDPA, a data user shall not process any personal data about a data subject unless the data subject has given his consent to the processing of the personal data.[15] Consent serves as a legal basis for processing under the PDPA. Consent is a mechanism through which the Act empowers individuals to exercise autonomy over their privacy generally and personal data specifically. However, the Act is silent on what constitutes consent. Notwithstanding this, 'consent' implicitly conjures up the notion of permission or agreement to do something. A more definitive meaning of 'consent' is provided in the European Union General Data Protection Regulation 2016/679. Article 4 of the Regulation defines 'consent' to mean 'any freely given, specific, informed and unambiguous indication of the data subject's wishes by which he or she, by a statement or by a clear affirmative action, signifies agreement to the processing of personal data relating to him or her'. Clearly, for a data subject to consent to the processing of his personal data, he must be informed at least of the processing and its purpose so that he knows what he consents to. A broadly-phrased consent agreement, such as one that provides a blanket consent to processing without specifying its purpose, is arguably so vague as to be devoid of any legal value and, accordingly, is vulnerable to legal challenge. Unless the personal data is sensitive personal data, the PDPA does not require the consent for processing to be explicitly given.[16] The consent may be orally given or implied. A data subject may by notice in writing withdraw his consent to the processing of his personal data.[17]

In data analytics where data users obtain digital consent from their customers, questions may arise whether the consent, in reality, is an informed one. This is mainly because of the manner in which the consent is obtained. For instance, it is common to find online notices which state that the submission of data to an organisation's website or the continued use of a website without leaving it is tantamount to giving consent that the personal data may be processed in the manner described in the organisation's privacy policy. The privacy policy is presented on a page which users can find by clicking the 'privacy policy' link. The privacy policy invariably contains provisions that are not readily understood by most data subjects, the vast majority of whom are not trained in the law. In many cases, the privacy policy is long, written in legalistic terms or are intended primarily to protect the data user rather than to inform the data subject. In such a case, obtaining digital consent from him may be an empty formality and a futile exercise.

Apart from the above, online retailers often require users to opt-in or opt-out of certain data collection and processing activities. In opt-in consent, customers grant express permission to the processing of their personal data, such as by ticking a box on a webpage. In opt-out consent, the responsibility is upon the customer to withdraw his permission, rather than on the data user to obtain such permission from the customers to process their personal data.[18] An instance of opt-out consent may be found in AEON Co (M) Bhd's Personal Data Privacy Notice which states that when customers provide their personal data, they are taken to have consented to the processing of their personal data in accordance with the purposes identified in the notice.[19] The privacy notice further states that if the customer does not want to have his personal data processed, he has to give a written notice to withdraw his consent. Although the PDPA allows both forms of consent, opt-in consent grants customers the maximum level of control over their personal data. While this model is a means of providing customers control over the use of their data, in many cases, the customer is compelled to consent to the processing of his personal data as otherwise he would not be able to use the retailers' services. In such situations, the customers have marginal leeway in deciding which data may be collected or processed.

Apart from the consent issue discussed above, the PDPA's General Principle requires that the personal data that is collected must be adequate and not excessive in relation to the purpose of collection.[20] Retail analytics, on the other hand, rely on the collection and use of as much data as possible about the customers and their past purchases so as to improve the accuracy of the derived predictions and inferences. Arguably, this practice runs counter to the PDPA's General Principle.

Notice and choice principle

Section 7(1) of the PDPA requires the data user, inter alia, to give the data subject a written notice informing him that his personal data is being processed and the purpose or purposes for which the personal data is collected and processed. This requirement to notify the data subject of the purpose of processing is circumscribed by s 6(3)(a) and (b). Section 6(3)(a) requires the processing of the personal data to be for a lawful purpose while s 6(3)(b) mandates that the processing of personal data can only be carried out if the processing is necessary for or directly related to that purpose. Notwithstanding this requirement to inform the data subject about the purposes of the processing, it is not uncommon for privacy policy notices to be drafted very broadly to enable data users to process personal data in an all-encompassing manner. For instance, many privacy policy notices provide that the personal data may be used and further processed for, but not limited to, a list of specified purposes. Such a practice does not appear to be in line with s 7(1)(b) of the PDPA which requires data users to be clear with the data subjects on the purposes of the processing.

Security principle

Section 9(1) of the PDPA imposes on a data user the responsibility to take practical steps to protect the personal data from any loss, misuse, modification, unauthorised or accidental access or disclosure, alteration or destruction. Where processing of personal data is carried out by a data processor on behalf of the data user, s 9(2)(a) requires the data user to ensure that the data processor provides sufficient guarantees in respect of the technical and organisational security measures governing the processing. This allows for protection of personal data from accidental loss, unauthorised disclosure, amendment or destruction The data user is also required by s 9(2)(b) to take reasonable steps to ensure that the data processor complies with those measures. This is important in retail analytics where retailers almost always outsource the analytics to service providers. The outsourcing increases the risk of data breaches and data leakages. In this respect, s 9(1)(e) requires the data user to take measures to ensure the secure transfer of the personal data.

Retention principle

Pursuant to s 10(1) of the PDPA, the personal data processed for any purpose shall not be kept longer than is necessary for the fulfilment of that purpose. Section 10(2) imposes on the data user the duty to take all reasonable steps to ensure that all personal data is destroyed or permanently deleted if it is no longer required for the purpose for which is was to be processed. Nevertheless, as the capacity to store data increases with advancement in technology and the cost of storing data decreases, data users may be motivated to retain the data as long as possible for yet unidentified future purposes.[21] This is especially so since the intrinsic value of each dataset is not limited by one or a few purposes but could potentially be useful for other new purposes as well.

Finally, in so far as the rights of data subjects under the PDPA are concerned, s 30(1) confers on data subjects the right to be informed by the data user that his personal data is being processed by or on behalf of the data user. Section 34(1) provides data subjects the right to correct any inaccuracies in his personal data. A data subject may, by notice in writing, withdraw his consent to the processing of his personal data under s 38(1), in which case the data user shall cease further processing of his personal data, delete the data from the database and terminate all notifications to the data subject. Further, where the targeted marketing is carried out via advertising agencies, data processing and the targeted advertising should cease as well.

THE IMPACT OF THE EU GENERAL DATA PROTECTION REGULATION ON DATA ANALYTICS

The EU General Data Protection Regulation 2016/679 (GDPR) is EU's framework for data protection law and it sets guidelines for the collection and processing of personal data from individuals who are residents or citizens of the EU. It became directly applicable in all EU member states from 25 May 2018. Unlike the PDPA, the GDPR contains specific provisions which impose obligations on organisations that process personal data in a big data environment. While the PDPA uses the term 'data user', the corresponding term in the GDPR is 'controller'. The controller is the person who determines 'the purposes and means' of processing personal data.[22] The term 'data subject' is not defined in the GDPR but essentially means any resident or citizen of the EU whose data is being processed, regardless of where he is physically located at the time of processing. Under article 4(8), a 'processor' is an entity that processes personal data on behalf of the controller. A 'data subject' and 'processor' under the GDPR have the same connotation as in the PDPA.

When a controller engages a processor to carry out processing on its behalf, the controller is required to enter into a written contract with the processor.[23] The contract should contain terms that require the processor to, inter alia, process the personal data only on documented instructions from the controller, observe confidentiality of the data, ensure security of the processing, not engage another processor without the prior authorisation of the controller and delete or return all the personal data to the controller after the end of the processing. As with the PDPA, the GDPR requires the controller to use only processors that provide sufficient guarantees to implement appropriate technical and organisational measures that meet the GDPR requirements.[24]

The GDPR specifically addresses profiling and automated decision-making, both activities being relevant to retail analytics. Profiling is defined in the GDPR as 'any form of automated processing of personal data consisting of the use of personal data to evaluate certain personal aspects relating to a natural person, in particular to analyse or predict aspects concerning that natural person's performance at work, economic situation, health, personal preferences, interest, reliability, behavior, location or movements'.[25] Based on this definition, profiling has to involve some form of automated processing and making predictions or drawing conclusions about a person in order to place them into a certain category or group. The controller is under a duty to inform the data subject that automated decision-making, including profiling is taking place, and provide him with meaningful information about the logic involved, as well as the significance and the envisaged consequences of such processing for the data subject.[26] Individuals have the right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or similarly significantly affects him.[27]

The GDPR outlines six data protection principles which a controller must comply with when processing personal data.[28] The first is lawfulness, fairness and transparency. This requires the personal data to be processed lawfully, fairly and in a transparent manner in relation to the data subject.[29] The second is purpose limitation principle, which mandates that personal data may only be collected for specified, explicit and legitimate purposes and not be further processed in a manner that is incompatible with those purposes.[30] The third is data minimisation principle, which states that the data that is processed must be adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed.[31] The fourth is the accuracy principle, which requires all personal data that are processed to be accurate and kept up to date. Further, every reasonable step must be taken to ensure that personal data that are inaccurate, having regard to the purposes for which they are processed, are erased or rectified without delay.[32] The fifth is the storage limitation principle which mandates that personal data are to be kept no longer than is necessary for the purposes for which the personal data are processed unless for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes.[33] The sixth, and final, principle is the security principle which provides that personal data must be processed in a manner that ensures appropriate security, including protection against unauthorised or unlawful processing and against accidental loss, destruction or damage.[34]

In the context of profiling and automated decision-making by a retailer in order to target its marketing, a number of the data protection principles under the GDPR are particularly relevant. Notwithstanding this, as profiling and automated decision-making tantamount to processing of personal data, all the principles relating to the processing of personal data in article 5 of the GDPR apply. The discussion below highlights key principles that are particularly significant to profiling and automated decision-making.

Lawful, fair and transparent processing

The GDPR requires controllers to process personal data lawfully, fairly and in a transparent manner.[35] Transparency indicates a high degree of disclosure of information by the controller to the data subject on matters relating to the processing of his personal data,

One of the bases for processing personal data lawfully is the consent given by the data subject for one or more specific purposes.[36] The data subject's consent must be an indication that is freely given, specific, informed and unambiguous by which he signifies agreement to the processing of his personal data through a statement or a clear affirmative action.[37] Accordingly, an organisation that relies on consent as the basis for processing in a big data context must inform the data subjects how the organisation will use their personal data and there must be a clear indication that the data subjects consent to the processing. If personal data is collected for one purpose and the controller decides to process it for a completely different purpose, the controller needs to inform the data subjects and obtain their further consent. It is unlawful for organisations to collect, combine and share personal data by drafting wide consents in their privacy policies that are vague. This follows from a controversy between Google and EU privacy regulators in 2012 where Google combined 70 of its existing policies from its other digital platforms into one. The consequence is that Google could collect user data on one service and use it on another of their service, for instance, from Google Search to YouTube. The policy was found by the privacy regulators of a number of EU countries to be so vague that it was no longer clear to data subjects how their data was collected and shared.[38]

Another basis for processing, apart from consent, is that the processing is carried out for the purposes of the legitimate interests of the controller, unless such interests are overridden by the interests or fundamental rights and freedoms of the data subject.[39] To assess whether processing is lawful on the ground of legitimate interests, the controller must carry out a balancing exercise to determine whether its interests are overridden by the data subject's interests or fundamental rights and freedoms.[40] The considerations to be taken into account include the level of detail of the profile (whether at a broad or granular level), the comprehensiveness of the profile (whether the profile only describes a small aspect of the data subject or a more comprehensive one), the impact of the profiling on the data subject and the safeguards aimed at ensuring fairness, non-discrimination and accuracy in the profiling process.[41] Profiling customers in order to target a retailer's marketing has been identified as a legitimate interest.[42] However, in the case of intrusive profiling and tracking practices for marketing or advertising purposes, such as those that involve tracking individuals across multiple websites, locations, devices or services, it would be difficult for controllers to justify using legitimate interests as a lawful basis for processing.[43]

Purpose limitation

When processing personal data, there must be a clearly defined purpose at the time of data collection and the data cannot be re-used for a different purpose which is incompatible with that original purpose.[44] It is not possible for a controller to obtain from the data subject consent to a very broad category of purposes so as to enable them to repurpose personal data on the ground that there is another purpose covered by the consent. This is because the GDPR requires consent to be informed and unambiguous.[45]

Situations may arise where personal data collected for a particular purpose is subsequently used for another purpose that does not necessarily fall within the original purpose or is compatible with it. This is particularly so in the big data context where huge volumes of personal data are collected containing intrinsic value that enables machine-learning algorithms to build other correlations and detailed profiles of individuals. For instance, health applications collect data about device users' physical fitness and activities. The data that is collected may also be used to build a profile on the data subjects for marketing purposes, such as targeted advertisement to promote health equipment and healthy foods. While the GDPR does not prevent the re-use of personal data for another compatible purpose, difficulties arise in establishing which purposes are compatible with the original purpose. Guidelines provided in the GDPR for determining this include the existence of any link between the new and original purpose, the reasonable expectations of data subjects as to their further use, the nature of the personal data, the consequences of the new processing on the data subjects and the existence of appropriate safeguards by the controller to ensure fair processing.[46]

Where the later purpose is incompatible with the original purpose, the data subjects should be informed of this and their consent obtained so that the processing is fair and compliant with the purpose limitation principle.

Data minimisation, accuracy and storage limitation

A feature of big data analytics is the collection of as much data as possible for analysis and this poses the question of whether it is necessary for the purposes of the processing. The UK Information Commissioner's Office recommends controllers to articulate at the outset why they need to collect and process particular datasets.[47] The controllers need to be clear about what they expect to learn or be able to do by processing that data, and thus satisfy themselves that the data is relevant and not excessive, in relation to that purpose.[48]

Data subjects have the right to have inaccurate and out-of-date data about them corrected. This covers all stages of the analytics process, including collecting data, analysing data and building an individual's profile. Clearly, where the data used in the analytics is inaccurate, the profiling will likewise be flawed.

Personal data can only be retained as long as it is necessary for the purpose for which it was collected. Consent by the data subject is not a ground to extend the retention of personal data beyond the period required for normal business purposes. The GDPR provides for the right to be forgotten which gives data subjects the right to have their data erased, for example, where the data is no longer necessary for the purpose for which it was collected or the data subject withdraws his consent.[49]

Security of data

Controllers and processors are required to implement appropriate technical and organisational measures to ensure a level of security appropriate to the risk.[50] This should take into account the state of the art, the costs of implementation and the nature, scope, context and purposes of processing. The technical and organisational measures include the pseudonymisation and encryption of personal data, ensuring ongoing confidentiality and integrity of processing systems, the ability to restore the availability and access to personal data in a timely manner in the event of a physical or technical accident and a process for regularly testing the effectiveness of the technical and organisation measures for ensuring the security of the processing.[51]

The GDPR also has a security breach notification provision which requires controllers to report personal data breaches to the relevant supervisory authority.[52] Specific security threats can arise in big data processing. For instance, where big data analytics are outsourced to a processor, the high level of replication in big data storage increases the risk of breaches, data leakages and degradation.[53] Controllers need to address this as part of their risk assessment in order to put in place appropriate security measures to protect personal data.[54]

As with the PDPA, data subjects are given several rights under the GDPR. They have the right to be informed about the processing. In the case of profiling-based decision making, the controller needs to make clear to the data subject that the processing is for the purposes of both profiling and making a decision based on the profile generated.[55] Data subjects are given the right of access in order to obtain details of any personal data collected about them for profiling,[56] the right to rectify any inaccuracies,[57] the right to erasure or right to be forgotten which gives data subjects the right to have their data erased, for example, where the data is no longer necessary for the purpose for which it was collected or the data subject has withdrawn his consent [58] and the right to object to processing including profiling.[59]

CUSTOMERS' INFERENCES AND PREDICTIONS — ARE THEY PERSONAL DATA?

Where customers' inferences and predictions drawn from data analytics of past transaction history target a broad section of the community that is classified according to categories such as age group, education level, economic background or race, such inferences or predictions do not constitute personal data because the information does not relate to any specific individual. No individual can be identified from that information, whether directly or indirectly.

The legal position is less clear where the inferences and predictions relate to specific individuals. For instance, it is possible for retail analytics to yield an individual customer profile, derived from his past transaction history, which predicts the types of entertainment that interest him, food that he likes, fashion trends that he prefers, hotels that he is likely to stay in and holiday destinations that excite him. The PDPA does not address whether such inferred data constitutes personal data.

The difficulty of determining whether inferences are personal data may be illustrated by two contradictory decisions of the European Court of Justice ('ECJ'). Both cases were decided under Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data. This directive was the data protection law in Europe immediately prior to the GDPR coming into force.[60] The first case is *YS v Minister voor Immigratie, Integratie en Asiel*,[61] which concerned a third country national who applied for a residence permit in the Netherlands. His application was rejected and he requested the relevant authority for the minutes relating to the decision. This request was refused and he brought an action on the ground that he could not lawfully be refused access to the minutes under data protection law. A question before the ECJ was whether the legal analysis contained in the minutes, which was based on the applicant's personal data, constituted personal data. The ECJ decided that the legal analysis contained in the minutes did not constitute personal data. However, the data relating to the applicant contained in the minutes and in the legal analysis were personal data. Thus, the case took the position that the legal analysis of personal data and the consequent inference connected to an individual are not personal data.

In contrast, in the second case, *Peter Nowak v Data Protection Commissioner*,[62] the ECJ took the view that an assessor's comments, opinions and assessments of an exam script constituted personal data. In that case, Nowak was a candidate in an accountancy examination. He failed the examination and made a request for access to his examination script. The request was rejected on the ground that the examination script did not contain personal data. The question before the ECJ was whether the written answers provided by a candidate at an examination and any comments made by an examiner with respect to those answers constituted information relating to that candidate and, hence, is his personal data. In answering this in the affirmative, the ECJ decided that information is personal data if, by reason of its content, purpose or effect, it is linked to a particular person. The ECJ referred to the definition of personal data which included data 'in the form of opinions and assessments, provided that it relates to the data subject'. The ECJ held that the answers provided by the candidate and the comments made by the assessor on the exam script were linked to a particular person, that is, the candidate, by reason of its content. Accordingly, they were personal data.

Both the above cases demonstrate the uncertainty that shrouds the determination of whether inferences and predictions are personal data. While a cogent case may be put forward that inferences drawn from data analytics in fields, such as genetics, healthcare, crime-prevention and finance are personal data, it is submitted that the position is different in retail analytics. The inferences and predictions in retail analytics are not facts that are necessarily linked to any specific individual, unlike genetic or health data. Instead, they are a set of selected products, which through analytics of the customer's past transactions, are predicted to be the preference of those customers. This information is not capable of identifying any specific individual.

Many a times, the inferences and predictions from retail analytics are 'guesses', the accuracy of which are not verifiable. A customer profile of this nature is usually not privacy-invasive and does not impact on the customers' private lives, identity or reputation. Arguably, a profile which shows that a customer has a preference for a particular type of entertainment, food, holiday destination, hotel or the like is non-specific in nature because such preferences could well be shared by many others. Moreover, the predictions or inferences may not always be accurate about the individual.

The definition of 'personal data' in s 4 of the PDPA extends to include 'expression of opinion about the data subject'. The *Oxford English Dictionary* defines 'opinion' as 'what or how one thinks about something, judgment or belief'. Opinions embody aspects of feelings or thoughts about a matter rather than a systematic analysis of facts or data. By its definition, some human input is required. In retail analytics, the human input in the derived inferences or predictions is missing and it would appear that they do not amount to opinions.

For the above reasons, it is submitted that inferences and predictions from retail analytics do not constitute personal data.

TOWARDS AN IMPROVED LEGAL ENVIRONMENT FOR DATA ANALYTICS

Based on the above discussion, a number of issues pertaining to the intersection between data protection law and data analytics merit legislative clarification. Although the discussion in this paper focuses on retail analytics, many of the issues that arise are shared by data analytics in other industries as well. First, while the PDPA provides for consent as an important basis for the processing of personal data, it does not define what amounts to real and informed consent. As noted above, for consent to be meaningful, the data subject should be informed that data analytics will be carried out in respect of his personal data, what data will be analysed and how it will be used. This necessarily requires adequate disclosure to be made by the data user. Even where the data user discloses these in the organisation's privacy policy, it is not uncommon to find in it accompanying terms that are convoluted and daunting to the layperson. Under such circumstances, it remains questionable whether it would be reasonably practicable to expect the layperson, at that point in time, to review, rationally assess the risks and harms of his submission of personal data to the website when exercising his consent entitlement. Accordingly, it is submitted that data users should be under a legal obligation to disclose to data subjects relevant matters pertaining to the use of their personal data for data analytics in a simple and readily understood form for laypersons.

Secondly, the PDPA's Notice and Choice Principle mandates that data users should, by written notice, inform data subjects the purposes for which the personal data is processed. It is insufficient for data users to merely state, as one of the purposes of data processing, that data analytics will be performed. This point relates closely to the consent issue raised in the preceding paragraph. Data subjects need to know the purpose or purposes of the data analytics and what personal data is used for the analytics. Very often, privacy policies provide for a long list of purposes which cumulatively cover such a wide range of purposes as to be all encompassing. Such vague and ambiguous notification is clearly unfair to data subjects.

Thirdly, the inherent feature of data analytics is that as much data as possible is collected so as to improve the accuracy of the derived inference. Undoubtedly, the datasets have potential value which enable unexpected correlations between data in the datasets to be discovered and this can form the basis to further use the datasets for analytics for other unrelated purposes. Such arbitrary re-use of personal data should be prohibited unless further consent from the data subjects is obtained for the new purposes.

Fourthly, while the PDPA imposes legal obligations on data users, it is silent on the obligations of data processors. The only provision that is of some relevance is s 9(2) of the Act which imposes obligations on data users to ensure that data processors provide sufficient guarantees in respect of the technical and organisational security measures governing the processing. The section also requires data users to ensure that data processors take reasonable steps to comply with those measures. Although it may be argued that data processors are implicitly bound by the data protection principles laid down under the PDPA despite the statute's silence on this matter, it is submitted that the legal obligations of data processors should be clearly spelt out to avoid any potential disputes that could arise.

Finally, the uncertainty surrounding the legal status of inferences and predictions derived from data analytics needs to be addressed by the Act. More specifically, it is unclear whether inferences and predictions constitute personal data. It is submitted that in determining whether inferences and predictions constitute personal data, the test should be whether they identify or are able to identify a specific individual. This would vary on a case-by-case basis, with reference to the specific circumstances and context of the situation, such as the industry and purpose of the data analytics. Some inferences such as those derived from genetic analytics are more likely to constitute personal data as compared to inferences from retail analytics. Some inferences or predictions provide information at group-level, such as young adults in the age group of 20–30 year olds have a preference for certain types of entertainment or food. Such inferences are clearly not personal data. Legislative guidelines for the determination of this in the context of data analytics are needed. With the increasing use of data analytics in many industries, it behoves

Parliament to provide certainty on this matter so that stakeholders have a clear understanding of their rights and obligations vis a vis inferences and predictions. Otherwise, this may stifle the development of data analytics in this country.

CONCLUSION

Data analytics, as an important means for organisations to harness their data and identify new opportunities, is used in a wide range of industries, including retail, healthcare, life sciences, manufacturing and finance. Data analytics commonly involves the processing of personal data and this sets into motion the operation of the PDPA. The aim of the PDPA is to regulate the processing of personal data in commercial transactions. However, the PDPA does not contain provisions that deal with issues presented by data analytics and this lacuna needs to be addressed by Parliament. The issues highlighted in Part VI of this paper are some of the more evident concerns and it is timely for the PDPA to include a part dealing with the legal obligations of stakeholders in the processing of personal data for analytics purposes. Equally important is the need to clarify the legal status of inferred data.

ACKNOWLEDGEMENT

The author gratefully acknowledges the funding support for this research (UM project PV019-2019) from ZillionQuest Sdn Bhd (Company No 1310834-D) which made the research possible.

[1]
(Act 709).

[2] Munir, Abu Bakar & Mohd Yasin, Siti Hajar, *Personal Data Protection in Malaysia* (Sweet & Maxwell, 2010) at 75. See also, the Personal Data Protection Act 2010 s 9(2) which imposes obligations on data users vis a vis data processors.

[3] In Australia, although the Privacy Act 1988 does not specifically deal with data analytics unlike the GDPR, the Office of the Australian Information Commissioner has released a *Guide to Data Analytics and the Australian Privacy Principles* to help organisations comply with the Privacy Act 1988. The position taken by the Guide is similar to that set out in the GDPR. See Office of the Australian Information Commissioner, *Guide to Data Analytics and the Australian Privacy Principles.* Available at *https://www.oaic.gov.au/privacy/guidance-and-advice/guide-to-data-analytics-and-the-australian-privacy-principles/#part-1-introduction-and-key-concepts* accessed on 3 February 2020.

[4] See Baesens, Bart, *Analytics in a Big Data World: The Essential Guide to Data Science and its Applications* (SAS Institute Inc, 2014) pp 4–6. Available at *https://support.sas.com/content/dam/SAS/support/en/books/analytics-in-a-big-data-world/excerpt.pdf* accessed on 3 February 2020.

[5] Ibid.

[6] Where the data that is collected has been anonymised by removing directly identifying information or information that may allow an individual to be identified, processing of the data is not subject to the PDPA.

[7] Personal Data Protection Act 2010 s 4. A ‘data subject’ is defined in the section as an individual who is the subject of the personal data. The term ‘data user’ is in turn defined as a person who either alone or jointly or in common with other persons processes any personal data or has control over or authorises the processing of any personal data. A ‘data user’ does not include a data processor. Instead, a ‘data processor’ is defined as any person who is not an employee of the data user but who processes the personal data solely on behalf of the data user. A data processor must not process the personal data for any of his own purposes.

[8] Personal Data Protection Act 2010 s 43(1).

[9] Personal Data Protection Act 2010 s 43(5).

[10] Personal Data Protection Act 2010 s 43(2). On 18 September 2018, the Personal Data Protection Commissioner had issued standard forms for the notices that a data subject could use to issue a notice to cease further processing under the Personal Data Protection Act 2010 ss 43(1) and 43(2).

[11] Personal Data Protection Act 2010 s 43(3).

[12] Personal Data Protection Act 2010 s 43(4). The Malaysian Code of Advertising Practice, Appendix L on Database Marketing, article 2 also provides that unsolicited advertisements are not to be sent unless consent is obtained. Further, advertisements are not to be sent where consumers have asked not to receive them.

[13] The seven personal data protection principles are:

(a) the General Principle;

(b) the Notice and Choice Principle;

(c) the Disclosure Principle;

(d) the Security Principle;

(e) the Retention Principle;

(f) the Data integrity Principle; and

(g) the Access Principle.

[14] Personal Data Protection Act 2010 s 11.

[15] Personal Data Protection Act 2010 s 6(1) further provides a more stringent standard with respect to sensitive personal data. Sensitive personal data about a data subject cannot be processed except in accordance with the restricted circumstances spelt out in s 40(1) of the Act. The term ‘sensitive personal data’ is defined in the Personal Data Protection Act 2010 s 4 as ‘any personal data consisting of information as to the physical or mental health or condition of a data subject, his political opinions, his religious beliefs or other beliefs of a similar nature, the commission or alleged commission by him of any offence or any other personal data as the Minister may determine by order published in the *Gazette*’. Where retail analytics is concerned, the data that is processed will usually not be sensitive personal data unless the retailer is in the healthcare sector. The discussion in this paper does not extend to sensitive personal data.

[16] Personal Data Protection Act 2010 s 40(1)(a).

[17] Personal Data Protection Act 2010 s 38(1).

[18] Lynskey, Orla, *The Foundations of EU Data Protection Law* (Oxford University Press, 2015) p 187.

[19] See AEON Personal Data Privacy Notice. Available at http://www.aeonretail.com.my/ accessed on 3 February 2020.

[20] Personal Data Protection Act 2010 s 6(3)(c).

[21] Lynskey, Orla, *The Foundations of EU Data Protection Law* (Oxford University Press, 2015) p 41.

[22] GDPR article 4(7).

[23] GDPR article 28(3).

[24] GDPR article 28(1).

[25] GDPR article 4(4).

[26] GDPR article 13(2)(f) applies in the case where the personal data was collected from the data subject. In the case where the personal data was not obtained from the data subject, GDPR article 14(2)(g) applies.

[27] GDPR article 22(1).

[28] GDPR article 5.

[29] GDPR article 5(1)(a).

[30] GDPR article 5(1)(b).

[31] GDPR article 5(1)(c).

[32] GDPR article 5(1)(d).

[33] GDPR article 5(1)(e).

[34] GDPR article 5(1)(f).

[35] GDPR article 5(1)(a).

[36] GDPR article 6.

[37] GDPR article 4(11).

[38] BBC, 'Google told to fix privacy policy by EU regulators' (*BBC*, 16 October 2012). Available at *https://www.bbc.com/news/technology-19959306* accessed on 3 February 2020.

[39] GDPR article 6(1)(f).

[40] UK Information Commissioner's Office, *Big data, artificial intelligence, machine learning and data protection*, (2017) p 33 para 68. Available at *https://ico.org.uk/media/for-organisations/documents/2013559/big-data-ai-ml-and-data-protection.pdf* accessed on 3 February 2020.

[41] Article 29 Data Protection Working Party, Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/678, adopted on 3 October 2017 (last revised and adopted on 6 February 2018) p 14.

[42] UK Information Commissioner's Office, *Big data, artificial intelligence, machine learning and data protection*, (2017) p 33 para 67.

[43] Article 29 Data Protection Working Party, Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/678, adopted on 3 October 2017 (last revised and adopted on 6 February 2018) p 15.

[44] GDPR article 5(1)(b).

[45] GDPR Recital 32.

[46] GDPR Recital 50.

[47] UK Information Commissioner's Office, *Big data, artificial intelligence, machine learning and data protection*, (2017) p 41 para 89.

[48] Ibid.

[49] GDPR article 17.

[50] GDPR article 32(1).

[51] Ibid.

[52] GDPR article 33.

[53] UK Information Commissioner's Office, *Big data, artificial intelligence, machine learning and data protection*, (2017) p 49 para 109.

[54] Ibid, p 50 para 110.

[55] GDPR article 13(2)(f).

[56] GDPR article 15.

[57] GDPR article 16.

[58] GDPR article 17.

[59] GDPR article 21(1).

[60] The GDPR, like its predecessor, also does not address the legal status of inferred data.

[61] Case C-141/12 (ECJ).

[62] Case C-434/16 (ECJ).