HOLISTICA Vol 11, Issue 2, 2020, pp. 161-170 DOI: 10.2478/hjbpa-2020-0026

IS BIG DATA SECURITY ESSENTIAL FOR STUDENTS TO UNDERSTAND?

Rochman Hadi MUSTOFA[1]*

Received: December 2019| Accepted: December 2019 | Published: August 2020
Please cite this paper as: Mustofa, R. H. (2020). Is big data security essential for students to understand?, *Holistica Journal of Business and Public Administration*, vol. 11, iss. 2, pp. 161-170

Abstract

Big Data has become a significant concern of the world, along with the era of digital transformation. However, there are still many young people, especially in developing countries, who are not yet aware of the security of their big data, especially personal data. Misuse of information from big data often results in violations of privacy, security, and cybercrime. This study aims to determine how aware of the younger generation of security and privacy of their big data. Data were collected qualitatively by interviews and focus group discussions (FGD) from. Respondents were undergraduate students who used social media and financial technology applications such as online shopping, digital payments, digital wallet and hotel/transportation booking applications. The results showed that students were not aware enough and understood the security or privacy of their digital data, and some respondents even gave personal data to potentially scam sites. Most students are not careful in providing big data information because they are not aware of the risks behind it, socialization is needed in the future as a step to prevent potential data theft.

Keywords: Big Data, Security and Privacy, Digital Data, Financial Technology

JEL Classification: C83

1. Introduction

Technological development brings an influence on people's behavior and lifestyle. People's daily activities such as paying bills, buying tickets, or ordering food can be done using technology. These activities will create digital recordings known as big data. Big data is the concept of a set of very large amount of information data as a result of the use of information technology that develops modernly (Hussain & Cambria, 2018). Big data is digital and because of its complexity, it is necessary to use applications and meta-analyses to translate it (Strang & Sun, 2016). The activity of extracting information from big data is known as data mining. Big data is widely used for marketing (Moro, Rita, & Vala, 2016) (Amado, Cortez, Rita, & Moro, 2018), academic (Logica & Magdalena, 2015), education (Li

[1] Universitas Muhammadiyah Surakarta, Indonesia, Rochman.hm@ums.ac.id
* Corresponding author

& Zhai, 2018), medical (Knight et al., 2019), government (Anshari & Lim, 2017), and industrial purposes (Nunes, 2018).

Even though the information stored in big data can be utilized in various fields, challenges arising make big data need serious attention. One of them is related to security and privacy (Bao, Chen, & Obaidat, 2018). The famous case of big data misuse is, for example, the Cambridge Analytica Scandal, where it was estimated that 50 million Facebook users' data had been taken without the permission of the owners and used for political purposes (Cadwalladr, Carole; Graham-Harrison, 2018). Thus, to prevent illegal data mining, protection and encrypted storage is needed (Kantarcioglu & Ferrari, 2019). After the data has been stored, it should be noted that only authorized users are permitted to access the data. Data access must consider in what situations and for what purpose the data is used (Qiu, 2015). Risks that arise when the data is accessed by unauthorized parties are data misuse such as privacy violations (Mills, 2018), cybercrime to data mining on personal information (Garcia-Rivadulla, 2016). Snowden's revealing that the National Security Agency had used data mining for surveillance activities caused criticism because it was considered spying on the public, violating privacy ethics (Lyon, 2014). Advanced countries are now very concerned about the privacy and security of their digital data.

For developing countries with high digital penetration rates such as Indonesia, digital data privacy and security are still in the stage of formation. The average Indonesian people spend 8 hours 36 minutes on the Internet per day (Wong, 2019). YouTube, Instagram, and online shopping platforms are favorite destinations while the residents of the 17-25-year age category dominate by 85.4% (Statista, 2019). The government, in this case, is still adapting to the rapid level of usage and making regulations gradually (Idzalika et al., 2014).

The public views digital data can be protected with the antivirus installed in gadgets or personal computers/notebooks. In fact, handling computer viruses and data mining that leads to cybercrime are completely different. In handling cybercrime issues, a security system is needed to prevent the entry of malware or illegal data mining activities (Cheng, Liu, & Yao, 2017). The potential for data theft will not cause the original data to be corrupted, but the impact is of concern. A hacker might insert a link to record digital activities such as a mobile banking pin via email or a website that has been modified in such a way as to deceive the user (Ronquillo et al., 2018). Email or personal address can be obtained through data mining. In addition, data mining can also be done through daily digital activities without the user being aware of it such as using online shopping platforms, digital payments, and online hotel/accommodation booking services. The use of a Virtual Private Network or VPN is considered to be able to protect privacy because it disguises the location and user through encrypted data. Some argue that VPN can sell user data. Digital activities such as searching for the latest shoes in an online shopping application or airline ticket prices will leave a history or digital footprint stored in a large volume of data. The data will usually be stored on a cloud system. Normally, the application will use the data to facilitate the user by giving a suggestion or price notification based on the user's digital footprint. However, in the case of data mining

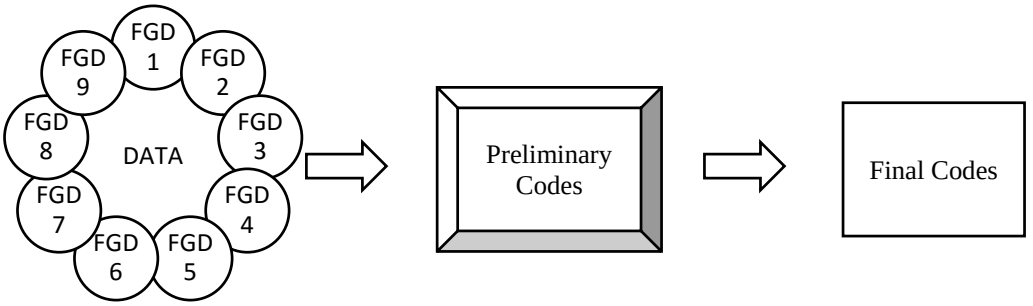
misuse, digital footprints are sold and used by marketers to get users bombarded with advertisements (Saqr, 2017). This is of course unethical business-wise because the user's personal data must remain safe and may not be traded without his/her permission. Both the company, digital platform providers, cloud service providers and data center managers must be more responsible for the security of the user's data.

There are several things that make users vulnerable to data leakage or even theft. The first is because of the user's *faux pas* and the second the misuse by other parties. Some *faux pas* of users includes using Wi-Fi networks in public places to log in web pages (Hu, Myers, Colizza, & Vespignani, 2009). A user may experience attacks by data miners, in this case hackers, deliberately disguising malware to certain sites or applications (Hammouchi et al., 2019) which will be active when the user enters the site. That is why, the ability to distinguish between safe and risky sites is needed (Liu & Zhong, 2017). The younger generation are often tempted to enter their personal data into a scam website that offers photo editing or lottery. Parents are often tempted with click baits. Another faux pas is not installing antimalware on computers/notebooks. Currently, there are many anti-malwares that function to prevent phishing of sensitive information (Joglekar & Pise, 2016).

2. Method

The method used in this research is a case study using a qualitative approach with focus group discussion (FGD). In a qualitative research, the researcher becomes the main research instrument. FGD is conducted to obtain data depth through discussion and sharing of opinions (Nyumba, Wilson, Derrick, & Mukherjee, 2018). The respondents were 74 students. First, the respondents were divided into 8 groups, each of which consisted to 8-9 people. Then, they were given opening questions like "Does anyone know what big data is?" or "Have you ever heard the term Big Data?". After that, the researcher asked them to discuss in groups about what they knew about big data. The researcher asked each group to discuss and share their knowledge about big data. Then, he instructed the respondents to write down the activities carried out on the Internet/digital applications related to entering personal data such as email, full name, surname, date of birth, and mobile phone number. The researcher set the discussion time of approximately 20 minutes. The respondents were asked to write their opinions on a piece of paper and then collected to the researcher for further analysis by coding techniques. Coding in the qualitative research can serve to separate the core of the interview with raw data (Saldana, 2009). Because there are many respondents, the researcher made a coding analysis based on the discussion group. In each discussion group, the researcher coded based on 3 stages, namely raw data, preliminary codes, and final codes.

Figure 1 Focus group discussions stage and codes



Source: (Nyumba et al., 2018; Saldana, 2009)

As shown in Figure 1, the researcher adopted the focus group discussion technique to obtain the data. This method is considered more effective in obtaining data depth with large respondents (Nyumba et al., 2018). The data in each group was then coded. The data obtained from the FGD was still raw data. The researcher examined each group's note and included it in the preliminary code category or stage. The next step was to make the core data segment from preliminary codes.

Table 1 Coding results from each FGD groups

FGD Groups	Preliminary Codes	Final Codes
FGD 1	"Actively using social media"	Generates big data;
	"Using online shopping applications"	Low level of awareness;
	"Making digital transactions"	High risk.
	"Possibly ever accessing third-party applications"	
	"Frequently use of WiFi in public areas"	
	"Heard big data from news on media"	
FGD 2	" Actively using social media "	Generates big data;
	"Making digital transactions"	Low level of awareness;
	"Frequently use of WiFi in public areas "	High risk.
	"Most have never heard of big data"	
FGD 3	"Actively using social media"	Generates big data;
	"Making digital transactions"	Low level of awareness;
		High risk

	"Frequently using online shopping applications"	
	"Frequently use of WiFi in public areas"	
	"Possibly ever accessing third-party applications"	
FGD 4	Actively using social media"	Generates big data;
	"Making digital transactions"	Low level of awareness;
	"Frequently using online shopping applications"	High risk
	"Frequently use of WiFi in public areas"	
	"Possibly ever accessing third-party applications"	
FGD 5	"Actively using social media"	Generates big data;
	"Making digital transactions"	Low level of awareness;
	"Frequently using online shopping applications"	High risk
	"Frequently using WiFi in public areas"	
	"Possibly ever accessing third-party applications"	
FGD 6	"Actively using social media"	Generates big data;
	"Making digital transactions"	Low level of awareness;
	"Frequently using online shopping applications"	High risk
	"Frequently using WiFi in public areas"	
	"Possibly ever accessing third-party applications"	
FGD 7	"Actively using social media"	Generates big data;
	"Making digital transactions"	Low level of awareness;
	"Frequently using online shopping applications"	High risk
	"Frequently using WiFi in public areas"	
	"Possibly ever accessing third-party applications"	

FGD 8	"Actively using social media"	Generates big data;
	"Making digital transactions"	Low level of awareness;
	"Frequently using online shopping applications"	High risk
	"Frequently using WiFi in public areas"	
	"Possibly ever accessing third-party applications"	
FGD 9	"Actively using social media"	Generates big data;
	"Making digital transactions"	Low level of awareness;
	"Frequently using online shopping applications"	High risk
	"Frequently using WiFi in public areas"	

Source: Mustofa, 2019

In Table 1, the researcher did not include raw data because the researcher has many FGD notes. Based on Table 1, it is known that the results of the analysis have similar patterns such as using social media, using at least one online shopping application for shopping, the habit of using WiFi in public places, and making digital transactions. Most of the FGD groups have accessed third-party applications, especially through those on Facebook. In addition, the respondents often used online photo-editing applications that have the potential to be scams (Webb, 2019). After the researcher observed the results of the FGD, then he conducted an interview only to several respondents to explore deeper understanding (Dilshad & Latif, 2013).

3. Results and Discussion

The data obtained from the respondents shows a lack of understanding of big data. At least 9 respondents answered that they had never heard the term big data before. When conducting further interviews related to the possibility of accessing news or television sites in the past 3 months, the respondents could not give the exact answer of when they accessed the media the last time.

All the respondents who have never heard the term big data are the students who live in boarding houses or rented houses that do not provide television access. In this case, the researcher asked additional questions as an assumption that the term big data might be reported through the media. Note that all 9 respondents have smartphones with at least one type of financial technology application installed, actively using in social media, have an internet connection, and have done digital transactions at least once in the last 3 months. Thus, the researcher concluded that the 9 respondents have done digital activities that produce or generate big data, but they are not aware of it while other 65

respondents claimed to have at least heard the term big data in the media. From the perspective of the respondents, big data is a complex data that contains financial records such as bank documents and complex documents that are only owned by certain people.

> *".. As I know from television, big data is a big amount of data, maybe like in the bank, and not everyone can access it"*-5th semester student, female-

> *" I don't think I can learn big data now. It's too hard. Maybe when I work, I'll consider learning about it ...*" -3rd semester student, male-

> *".. Yes, I have used the application. First, we first in the biodata for an account. Then, we can use it to order tickets train ..."* -5th semester students, male-

It can be implied that the respondents have done digital activities that create big data such as signing up social media, login on websites or applications, Internet search history, online payment and shopping, and booking train tickets. However, most of them do not understand that digital data is part of big data. The researcher assumes that the respondents do not care enough about the term big data they hear, so they never look for further definitions and examples.

> *".. Yes, I've heard the news, the case of a bank account burglary. He said it was hacked anyway. That's why, I stay alert when using an ATM "*

> *" Do you think the data on social media can be used for crime?"* - Researcher-

> *" Hmmm, I don't know. It seems like it's just a case of falsification of identity. After all, I only post picture on Instagram."* -5th semester student, female-

The respondents know cybercrime cases on the news, but perceive that data is stolen because of hackers who tap the ATM (Automatic Teller Machine). They do not realize that data on social media can be misused for broader matters. Other respondents also show similar reactions. The researcher believes that there are still many students who are not aware of the huge risk of big data leakage. In fact, social media store a lot of valuable information as big data (Dhawan & Zanini, 2014). The data obtained from social media can be used to predict user's behaviour (Portela et al., 2016) and read his/her pattern (Esfahani, Tavasoli, & Jabbarzadeh, 2019). It is dangerous that the respondents are not cautious in maintaining the confidentiality of their data on social media.

> *"Yes, I used to use the application on Facebook to see who often sees my profile. At first, I was asked to sign up for an account, but it seemed the application didn't exist anymore on Facebook ..."* -7th semester student, female-

From the respondent's opinion, it is known that in fact she is not aware of sending her personal data on a third-party application. It is quite concerning because Facebook has stated that it has blocked third-party applications that violate privacy (Lapowsky, 2019). Third-party applications often appear in the form of quizzes, games, character reading, who I am in the future, even photo editing that tempt users to click on a user license agreement (Kozlowska, 2018). Activities of the younger generation like the most today

are related to self-actualization on social media (Bergagna & Tartaglia, 2018) and they tend to trust product based on visual appearance (Permatasari & Kartikowati, 2018). Based on these observations, the respondents do not understand well the risks when entering their personal data into third-party applications. Data protection and caution in sending personal information can prevent the misuse of big data (Waterman & Bruening, 2014).

4. Results and Discussion

Based on the results of FGD and interviews, the researcher arrives at several conclusions. First, the respondents, in this case the students, do not yet understand what big data is overall. Second, all respondents have conducted activities that can produce big data, but do not understand how big data can be utilized. Third, the respondents do not understand the risk of big data misuse. This makes students more vulnerable to the scamming trap through the latest applications. It is necessary to give understanding to the younger generation as the biggest and potential users of digital data to reduce the risk of data mining and as a preventive measure.

5. Disclosure Statement

The author declares no conflict of interest and responsible of the content and writing of this article.

References

Amado, A., Cortez, P., Rita, P., & Moro, S. (2018). Research trends on Big Data in Marketing: A text mining and topic modeling based literature analysis. *European Research on Management and Business Economics*, *24*(1), 1–7. https://doi.org/10.1016/j.iedeen.2017.06.002

Anshari, M., & Lim, S. A. (2017). E-Government with Big Data Enabled through Smartphone for Public Services: Possibilities and Challenges. *International Journal of Public Administration*, *40*(13), 1143–1158. https://doi.org/10.1080/01900692.2016.1242619

Bao, R., Chen, Z., & Obaidat, M. S. (2018). Challenges and techniques in Big data security and privacy: A review. *Security and Privacy*, *1*(4), e13. https://doi.org/10.1002/spy2.13

Bergagna, E., & Tartaglia, S. (2018). Self-esteem, social comparison, and facebook use. *Europe's Journal of Psychology*, *14*(4), 831–845. https://doi.org/10.5964/ejop.v14i4.1592

Cadwalladr, Carole; Graham-Harrison, E. (2018, March 17). Revealed: 50 million Facebook profiles harvested for Cambridge Analytica in major data breach. *The Guardian*. Retrieved from https://www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election

Cheng, L., Liu, F., & Yao, D. D. (2017). Enterprise data breach: causes, challenges, prevention, and future directions. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *7*(5), 1–14. https://doi.org/10.1002/widm.1211

Dhawan, V., & Zanini, N. (2014). Big Data and Social Media Analytics. *Research Matters: A Cambridge Assessment Publication.*, (18), 36–41.

Dilshad, R. M., & Latif, M. I. (2013). Focus Group Interview as a Tool for Qualitative Research: An Analysis. *Pakistan Journal of Social Science*, *33*(1), 191–198. Retrieved from https://www.bzu.edu.pk/PJSS/Vol33No12013/PJSS-Vol33-No1-16.pdf

Esfahani, H. J., Tavasoli, K., & Jabbarzadeh, A. (2019). Big data and social media: A scientometrics analysis. *International Journal of Data and Network Science*, (January), 145–164. https://doi.org/10.5267/j.ijdns.2019.2.007

Garcia-Rivadulla, S. (2016). Personalization vs. privacy: An inevitable trade-off? *IFLA Journal*, *42*(3), 227–238. https://doi.org/10.1177/0340035216662890

Hammouchi, H., Cherqi, O., Mezzour, G., Ghogho, M., & Koutbi, M. El. (2019). Digging Deeper into Data Breaches: An Exploratory Data Analysis of Hacking Breaches Over Time. *Procedia Computer Science*, *151*(2018), 1004–1009. https://doi.org/10.1016/j.procs.2019.04.141

Hu, H., Myers, S., Colizza, V., & Vespignani, A. (2009). WiFi networks and malware epidemiology. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(5), 1318–1323. https://doi.org/10.1073/pnas.0811973106

Hussain, A., & Cambria, E. (2018). Semi-supervised learning for big social data analysis. *Neurocomputing*, *275*, 1662–1673. https://doi.org/10.1016/j.neucom.2017.10.010

Idzalika, R., Pramestri, Z., Amin, I., Riyadi, Y., & Hodge, G. (2014). Big Data for Population and Social Policies. *Pulselabjakarta.Org*. Retrieved from https://pulselabjakarta.org/assets/uploadworks/2019-01-24-08-58-31.pdf

Joglekar, P., & Pise, N. (2016). Solving Cyber Security Challenges using Big Data. *International Journal of Computer Applications*, *154*(4), 9–12. https://doi.org/10.5120/ijca2016912080

Kantarcioglu, M., & Ferrari, E. (2019). Research Challenges at the Intersection of Big Data, Security and Privacy. *Frontiers in Big Data*, *2*(February), 1–6. https://doi.org/10.3389/fdata.2019.00001

Knight, S. R., Ots, R., Maimbo, M., Drake, T. M., Fairfield, C. J., & Harrison, E. M. (2019). Systematic review of the use of big data to improve surgery in low- and middle-income countries. *The British Journal of Surgery*, *106*(2), e62–e72. https://doi.org/10.1002/bjs.11052

Kozlowska, I. (2018). Facebook and Data Privacy in the Age of Cambridge Analytica. Retrieved September 28, 2019, from The Henry M Jackson School of International Studies website: https://jsis.washington.edu/news/facebook-data-privacy-age-cambridge-analytica/

Lapowsky, I. (2019, March). *In Latest Facebook Data Exposure, History Repeats Itself*. Retrieved from https://www.wired.com/story/facebook-apps-540-million-records/

Li, Y., & Zhai, X. (2018). Review and Prospect of Modern Education using Big Data. *Procedia Computer Science*, *129*, 341–347. https://doi.org/10.1016/j.procs.2018.03.085

Liu, W., & Zhong, S. (2017). Web malware spread modelling and optimal control strategies. *Scientific Reports*, *7*(February), 1–19. https://doi.org/10.1038/srep42308

Logica, B., & Magdalena, R. (2015). Using Big Data in the Academic Environment. *Procedia Economics and Finance*, *33*(15), 277–286. https://doi.org/10.1016/s2212-5671(15)01712-8

Lyon, D. (2014). Surveillance, Snowden, and Big Data: Capacities, consequences, critique. *Big Data & Society*, *1*(2), 205395171454186. https://doi.org/10.1177/2053951714541861

Mills, K. A. (2018). What are the threats and potentials of big data for qualitative research? *Qualitative Research*, *18*(6), 591–603. https://doi.org/10.1177/1468794117743465

Moro, S., Rita, P., & Vala, B. (2016). Predicting social media performance metrics and evaluation of the impact on brand building: A data mining approach. *Journal of Business Research*, *69*(9), 3341–3351. https://doi.org/10.1016/j.jbusres.2016.02.010

Mustofa, R. H. (2019). *Coding Results from Focus Group Discussion (FGD) about Big Data*.

Nunes, M. B. (2018). Understanding Big Data for Industrial Innovation and Design: The Missing Information Systems Perspective. *Journal of Data and Information Science*, *2*(4), 1–6.

https://doi.org/10.1515/jdis-2017-0017

Nyumba, T. O., Wilson, K., Derrick, C. J., & Mukherjee, N. (2018). The use of focus group discussion methodology: Insights from two decades of application in conservation. *Methods in Ecology and Evolution*, *9*(1), 20–32. https://doi.org/10.1111/2041-210X.12860

Permatasari, A., & Kartikowati, M. (2018). The influence of website design on customer online trust and perceived risk towards purchase intention: A case of O2O commerce in Indonesia. *International Journal of Business and Globalisation*, *21*(1), 74–86. https://doi.org/10.1504/IJBG.2018.094097

Portela, J., Villalba, L. G., Trujillo, A. S., Orozco, A. S., & Kim, T.-H. (2016). Estimation of Anonymous Email Network Characteristics through Statistical Disclosure Attacks. *Sensors*, *16*(11), 1832. https://doi.org/10.3390/s16111832

Qiu, J. L. (2015). Reflections on Big Data: 'Just because it is accessible does not make it ethical.' *Media, Culture and Society*, *37*(7), 1089–1094. https://doi.org/10.1177/0163443715594104

Ronquillo, J. G., Winterholler, J. E., Cwikla, K., Szymanski, R., & Levy, C. (2018). Health IT, hacking, and cybersecurity: national trends in data breaches of protected health information. *JAMIA Open*, *1*(1), 15–19. https://doi.org/10.1093/jamiaopen/ooy019

Saldana, J. (2009). *The Coding Manual for Qualitative Researchers* (First Edit). Retrieved from http://www.ghbook.ir/index.php?name=فرهنگ و رسانه های دیجیتال&option=com_dbook&task=readonline&book_id=13650&page=73&chkhashk=ED9C9491B4&Itemid=218&lang=fa&tmpl=component

Saqr, M. (2017). Big data and the emerging ethical challenges. *International Journal of Health Sciences*, *11*(4), 1–2. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/29085259%0Ahttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC5654190

Statista. (2019). Share of internet users in Indonesia in 2019, by age group. Retrieved September 18, 2019, from Statista Research Department website: https://www.statista.com/statistics/997264/share-of-internet-users-by-age-group-indonesia/

Strang, K. D., & Sun, Z. (2016). Meta-analysis of big data security and privacy: Scholarly literature gaps. *Proceedings - 2016 IEEE International Conference on Big Data, Big Data 2016*, (February), 4035–4037. https://doi.org/10.1109/BigData.2016.7841101

Waterman, K., & Bruening, P. J. (2014). Big Data analytics: risks and responsibilities. *International Data Privacy Law*, *4*(2), 89–95. https://doi.org/10.1093/idpl/ipu002

Webb, K. (2019, July 18). *The Russian Photo App That Makes You Look Old Is Probably Keeping Your Data*. Retrieved from https://www.sciencealert.com/viral-russian-app-that-makes-you-look-old-is-probably-keeping-your-data

Wong, E. (2019, March 18). How Indonesians Embrace the Digital World. *The Jakarta Post*. Retrieved from https://www.thejakartapost.com/academia/2019/03/18/how-indonesians-embrace-the-digital-world.html