International Journal of Engineering & Technology, 7 (3.32) (2018) 123-126

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper

Post Cambridge Analytica Fallout: Observing Facebook Users Awareness Regarding Data Security

Jos Timanta Tarigan*, Elviawaty M. Zamzami

Faculty of Computer Science and Information Technology, Universitas Sumatera Utara
**Corresponding author E-mail: jostarigan@usu.ac.id*

Abstract

In early 2018, Cambridge Analytica, a UK based political consulting group, was in the middle of the spotlight regarding its activity collecting Facebook users' personal data. The data was harvested without users' consent and was used to influence voter opinion to support a presidency campaign. This event triggered numerous movement aimed to inform and motivate the public to be concern of the use of their data in social media. However, whether these actions have raised the awareness of social media users is yet to be investigated. The objective of this paper is to harvest the awareness of social media users in Indonesia. We've performed a survey and collected answers from 312 responders who actively use social media in everyday life. The result shows that most of the responders have a low awareness regarding the event and its impact to social media industry. Over 24% of our responders have heard of the event and only 7% of the responders were able to describe the event correctly. Moreover, most of these users did not aware how their profile information in social media may be used to achieve commercial or political purpose by social media industry or collaborating third party without their consent.

Keywords: *Socio-informatics, Social Media, Social Network Analysis, Big Data*

1. Introduction

Social media has become one of the most significant platforms in recent years. The ability to connect to other people, communicate through text and multimedia content has become a necessity in modern lifestyle. Moreover, users and consumers are increasingly dependent on social media to find proper information in order to make decisions. This fact motivates researchers to find dyadic relations between individual's activity in social media and the decision made related to the activity. The concept Social Network Analysis [1] and Social Set Analysis [2] are two of the most common concept used to solve these problem. While there are many works need to be done to improve the performance, scientists agreed that it is possible to create and support a better conclusion based on these large amounts of social media activity data (usually called Big Social Data). Since then, companies have been gathering social media activity data and trying to create a better understanding regarding their target market. Extracting meaningful facts based on social media activity can give a competitive advantage commercially. Furthermore, companies also actively use social media to interact with the market. With the help of social media industry with "targeted audience" feature, companies may have a more efficient and effective marketing by targeting their ads to relevant prospective customer.

This trend however does not necessarily positives and prune to be misused. By understanding individual's opinion, one can easily use or manipulate subject's opinion by providing specific contents. The use of destructive content and false news (usually called hoax) is common in social media. Moreover, the lack of control from authorities and social media industry combined with the ability to become fully anonymous makes this problem complicated to solve. There are researches and solution that focus on solving this problem [3].

Recent hysteria on Facebook's data breach by Cambridge Analytica [4]–[6], a now defunct [7] UK based political company, gives an opportunity to raise the awareness of social media users regarding the use of their data. However, recent studies show that this event make little to no impact to overall Facebook usage. Especially in Indonesia, a country with one of the most Facebook users, the data breach event did not even affect most of its user. Studies and surveys shows that there are no sign of massive Indonesian users quitting Facebook. This fact motivates us to perform a study on Indonesian Facebook users' awareness regarding the data breach. The study is aimed to answer these question: Is the user aware about the event? Is the user aware about the impact of this event? Was the user affected by this event?

This rest of the paper is organized as follows: Section 2 provides related previous study that may support our works. In section 3, we will describe the structure of our survey form and a set of questions taken from the survey form as samples. The result of the survey will be presented in section 4, followed by conclusion and future works in section 5.

2. Related Works

Understanding user's behaviour in social media is one of the most common topics in socio-informatics research. With massive amount of user's data stored on social media, it is possible to predicts user's opinion, interest, and future behaviour based on their footprint. Asadzadeh et al. presents a work to predict psychological traits based on their recent activity on social media [8]. Based on their study on 92,255 users, Openness and Extraversion are the easiest to be detected by studying their footprints on social media followed by users with Conscientiousness and Neuroticism. Similarly, Halevi et al. examines the correlation between the Big Five personality traits and email phishing response and infor-

Copyright © 2018 Authors. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

mation/news sharing on Facebook [9]. The study aimed to have a better understanding on individuals with a certain trait that prone to performing illegal activity on cyber world.

Other studies focus on the use of social media as commercial tools. Hu et al. studies how marketing campaign on Facebook fan pages affects ticket-purchase behaviour [10]. The result of this work shows that the behaviour of ticket sales was correlated to marketing activity on Facebook page. Boldt et al. presented a work on predicting sales based on Facebook data [11]. Boldt et al. argues that it is possible to correlate the social set analysis dataset taken from Facebook page (likes, comments, and shares) and product sales. There are also studies that aimed to present methods to optimize the use of social media for commercial purpose. Goncalves et al. evaluated narrowcasting, or target segmentation, in achieving optimal online sharing level [12]. Khosla et al. evaluated the properties of popular images using 2.3 Million images on flicker in order to be able to understand how to make a popular image [13].

Other studies evaluate social media footprint to find it relation to user's opinion regarding a specific subject. Most of these studies are focus on political subject and involving the use of semantic analysis to gather sentiment of the post. Chie et al. perform a study to find the relation of user's political tendency based on Facebook posts [14]. Flesch et al. observed political party social media activity during 2017 German Federal Election [15].

3. The Cambridge Analytica Scandal

There are numerous articles explaining the scandal; some may have a different story from others. Hence, before we go through the survey, it is important to agree on a single story. In this case, we use an article by Alvin Chang [16] as our main source.

The simplest version of the scandal is as follows: 87 Million Facebook users' data were exposed by a UK based political consultant named Cambridge Analytica. These data were later used to support 2016 Trump's Presidency campaign as a foundation to build psychographic profiles on a national scale. Usually, building this profile requires a huge effort in both time and money to perform nation-wide survey. However, Cambridge Analytica found a solution by using a method to map personality traits based on users' activity on Facebook developed by Cambridge University's Psychometrics Centre [17]. The company then approached Dr. Alexandr Kogan, a psychology professor who built a quiz app on Facebook platform that able to gather users' data.

Aside from Kogan's failure to comply with Facebook policy not to sell to third party and delete the data, there is nothing illegal in this activity. While collecting user's data without their consent is unethical, it is not illegal per se. It is important to know that the so-called breach was not a hack [18]. As a free service, Facebook expected its users' data to be harvested. What more important to discuss vast majority of users are unaware on how their personal data can be used as a valuable data.

Upon the blowout of this event, journalists and policy makers have performed thorough investigation concerning the activity related to harvesting social media activity data and how to guard the data and protect users privacy [19].

4. User Awareness Survey

Our objective in this survey is to understand user's understanding regarding the impact of their behaviour in social media. The survey is structured into three parts as follows:

1. User Behaviour on Social Media
   In this part we collected the users' behaviour on social media such as frequency and the content uploaded to their Facebook account.
2. The Cambridge Analytica Scandal
   The objective of this part is to measure users' knowledge regarding the case. Our focus here is to measure user's understanding regarding the case such as the legality aspect of the case, the impact of the case, how their reaction after the case.
3. Information privacy awareness
   This part focuses on their understanding regarding the information privacy in social media. The objective is to understand what social media company can and can not do regarding their data.

The target audience for our samples are university students. The age is varied from 14 to 21. We split the audience into two groups: one with computer science and information technology background and the other without any background in computer science. Total samples taken during the survey is 312.

4.1. Social Media Behaviour

Despite the increase of various social media application available in current market, recent survey shows Facebook is still the mostly used social media [20]. The first part of the survey is to figure out the behaviour of social media activity of our responders. Table 1 shows the result of our survey.

Table 1: Social Media Activity Behaviour

Accessibility to Facebook Account		
Yes: 265 (85%)	No: 47 (15%)	
Frequency of accessing Facebook		
Daily: 168 (63%)	Weekly: 65 (25%)	Monthly or Less: 32 (12%)
Type of activity spent on Social Media		
Fun/Leisure: 299 (96%)		
Serious Activity: 198 (63%)		
Commercial: 39 (12%)		
Frequency of interacting in Facebook (status, comments, likes, photos, links, new profile)		
Daily: 141 (53%)	Weekly: 79 (30%)	Monthly: 45 (17%)
Media to Access Facebook		
Mobile: 102 (38%)	Computer: 53 (20%)	Both: 110 (42%)
Other Social Media		
Instagram: 228 (73%)	Snapchat: 102 (32%)	
Path: 31 (10%)	Twitter: 189 (60%)	
Others: 135 (43%)	No Social Media: 32 (10%)	

Our samples show the same tendency; 85% of our samples still use Facebook as their social media platform with more than half of the user access Facebook daily. However, only 53% of the 265 responders perform interaction in Facebook such as updating status/profile, giving comments, uploading videos, or sharing links; the rest of them considered themselves as passive user. Most of our responders also admit that they put their real profiles, status, and opinions on social media. They are confident that social media company have good policy and security system in preventing their data to be misused.

Based on the type of activity, most of our responders uses social media for leisure activity. However, 63% of our responders admit that sometimes they use social media for serious purpose such as criticizing, stating political opinion, and reporting cases. Only 12% of our responders uses social media for commercial purpose. This is expected since various e-commerce applications available makes Facebook irrelevant for commercial purpose.

4.2. Awareness on Cambridge Analytica Scandal

On the second part of the survey, our focus is to observe the knowledge of our responders regarding the Cambridge Analytica scandal and its impact. Table 2 below shows the result of the second part of the survey.

Table 2: Awareness on The Data Breach Event

Aware of The Cambridge Analytica Event		
Yes: 31 (10%)	Vaguely: 44 (14%)	No: 237 (76%)
Capability to Describe the Event Correctly (only responders who answered 'Yes' and 'Vaguely' from previous question)		
Correct: 5 (7%)	Average: 23 (31%)	Incorrect: 47 (63%)
Aware of the stolen Facebook Indonesian user's data		
Yes: 84 (27%)	Vaguely: 102 (33%)	No: 126 (40%)
Impact on users upon knowing the event		
Significant: 67 (21%)	Moderate: 171 (55%)	No effect: 74 (24%)

Despite the massive local media blowout regarding the event in early 2018, most of our responders (76%) were unaware of the event. Furthermore, amongst those who aware of the event, only 7% were able to give a correct description regarding the event. Most of the users who acknowledge the event describe the case as a Facebook hack. However, over 50% of our responders are aware that there were data stolen from Facebook. They were aware that a great number of Indonesian users affected to this event. Upon knowing the event, 76% of our responders admitted they were affected and may decide to change their digital behaviour and review the data privacy policy on social media application. The other 24% that not affected by this event agues that their there are no significant information can be harvested in their social media account.

Another interesting fact we found during this part of the survey is all respondents who described the event correctly have a background in computer related study such as Computer Science and Information Technology. Moreover, most of our respondent with no computer related study background had trouble understanding the technical aspect of the event, specifically terms related to computer science such as "Facebook API", "Third Party Developer", or even "Data Security and Privacy".

4.3. Data Sharing Policy

The next section of our survey is to measure users' knowledge regarding how their data is used by social media company and how this action may affect their behaviour. In this survey, we present some common cases of data usage in social media industry practice. The 4 options available are:

1. Aware and Affected if the user has already known this fact and have change their behaviour in social media.
2. Aware and not Affected if the user has already known the face but decided not to change their behaviour.
3. Not Aware and Affected if this is the first time the user has been informed regarding the case and decided to change their behaviour
4. Not Aware and Not Affected if this is the first time the user has been informed regarding the case and decided not to change their behaviour.

Table 3 below shows the result of the third part of the survey.

Table 3: Awareness on Data Usage in Social Media

Social Media Company uses your data to perform a more effective advertising (narrowcasting)		
	Affected	**Not Affected**
Aware	64 (20%)	132 (42%)
Not Aware	56 (18%)	60 (19%)
Facebook Application such as quiz app allows its developer to access your data		
	Affected	**Not Affected**
Aware	45 (14%)	61 (20%)
Not Aware	86 (26%)	120 (38%)
Your public social media data can be harvested without permission from the Social Media Company		
	Affected	**Not Affected**
Aware	42 (13%)	27 (9%)
Not Aware	156 (50%)	87 (28%)
There are studies/researches/methods to find relation between social media activity and user's psychographic profile.		
	Affected	**Not Affected**
Aware	8 (3%)	3 (1%)
Not Aware	130 (42%)	171 (55%)
You can modify the accessibility of your social media data on most social media application		
	Affected	**Not Affected**
Aware	90 (29%)	87 (28%)
Not Aware	65 (21%)	79 (22%)
Companies/Organization may use your data for commercial purpose		
	Affected	**Not Affected**
Aware	56 (18%)	86 (26%)
Not Aware	87 (28%)	83 (27%)
Companies/Organization may use your data for psychographic which can be used for political purpose		
	Affected	**Not Affected**
Aware	55 (17%)	67 (21%)
Not Aware	99 (31%)	91 (29%)

The result in this survey shows that most of our users (62%) aware of their activity may be used to perform a targeted advertising. However, most of these users does not affected by this fact since advertising in social media does not bother their activity. Most of the responders however, reacted positively to advertising since it is important to keep the service free. The result also shows that most of our responders were unaware that social media application gives access to third party developer to access their data by using a Facebook application. Moreover, most of the respondent also did not aware the value of their social media activity data and the use of those data for commercial and political purpose. Nearly 70% of our responder were unaware that their social media activity can be used to map their psychological profile, which is important to predict their opinion and action regarding a specific matter.

Unsurprisingly, 97% of our responders did not know that there are studies, researches, and methods dedicated to performing psychographic profile mapping based on their social media activity. While we expected that most of our responders were unaware regarding the study, the amount is too far from our expected value. We assume that this fact may be the cause of the unawareness of our responders regarding the value of their social media activity data.

5. Conclusion and Future Works

The Cambridge Analytica scandal opened an opportunity for computer educator and government to educate social media users concerning their digital activity. The massive media blowout was a rare and good opportunity to highlight the impact of digital activity in the real world. In Indonesia, this event also caught local media attention and caused a have managed to trigger the government to pay attention to Indonesian Facebook users.

However, based on our survey, the effect of the event to Indonesian active Facebook users is minimal. We have performed a survey and collected 312 samples of active social media users. The results show that most of the responders in our survey have a very low awareness regarding the event and its impact in social media activity. Moreover, they were unaware the value of their data and how it can be harvested and used unethically for commercial or political purpose.

Based on this research, we suggest that the society need to be educated concerning their digital lifestyle and the impact of it in the long run. The senate hearing against Facebook CEO Mark Zuckerberg [21] and the introduction of General Data Protection Regulation (GDPR) by the European Union [22] are two of the most recent breakthrough in raising the awareness of users digital data. However, it is safe to conclude that even after the media blowout of these events, local government should consider taking a significant action to raise users concern about protecting their data.

Acknowledgement

This research is funded and supported by author's institution, Universitas Sumatera Utara. We would like to thank Universitas

Sumatera Utara who facilitate all the work in this research. Furthermore, authors would like to thank the staff of Faculty of Computer Science and Information Technology for the support during the research.

References

[1] S. P. Borgatti, A. Mehra, D. J. Brass, and G. Labianca, "Network Analysis in the Social Sciences," *Science*, vol. 323, no. 5916, pp. 892–895, Feb. 2009.
[2] R. Vatrapu, R. R. Mukkamala, A. Hussain, and B. Flesch, "Social Set Analysis: A Set Theoretical Approach to Big Data Analytics," *IEEE Access*, vol. 4, pp. 2542–2571, 2016.
[3] P. Pourghomi, F. Safieddine, W. Masri, and M. Dordevic, "How to stop spread of misinformation on social media: Facebook plans vs. right-click authenticate approach," 2017, pp. 1–8.
[4] N. Confessore, "Cambridge Analytica and Facebook: The Scandal and the Fallout So Far," *The New York Times*, 08-Jun-2018.
[5] P. Greenfield, "The Cambridge Analytica files: the story so far," *The Guardian*, 26-Mar-2018.
[6] A. Langone, "Facebook's Cambridge Analytica Controversy Could Be Big Trouble for the Social Network. Here's What to Know," *What to Know About Facebook's Cambridge Analytica Problem | Time*, 04-Apr-2018. [Online]. Available: http://time.com/5205314/facebook-cambridge-analytica-breach/. [Accessed: 10-Jun-2018].
[7] N. Confessore and M. Rosenberg, "Scandal Sinks Firm That Mined Facebook to Target U.S. Voters," *The New York Times*, 03-May-2018.
[8] A. Laleh and R. Shahram, "Analyzing Facebook Activities for Personality Recognition," 2017, pp. 960–964.
[9] T. Halevi, J. Lewis, and N. Memon, "A pilot study of cyber security and privacy related behavior and personality traits," 2013, pp. 737–744.
[10] H.-W. Hu, C.-H. Cheng, Y.-C. Chung, and C.-Y. Lee, "Ticket-purchase behavior under the effects of marketing campaigns on facebook fan pages," 2017, pp. 2746–2751.
[11] L. C. Boldt *et al.*, "Forecasting Nike's sales using Facebook data," 2016, pp. 2447–2456.
[12] J. Goncalves, V. Kostakos, and J. Venkatanathan, "Narrowcasting in social media: effects and perceptions," 2013, pp. 502–509.
[13] A. Khosla, A. Das Sarma, and R. Hamid, "What makes an image popular?," 2014, pp. 867–876.
[14] S.-I. Chiu and K.-W. Hsu, "Predicting Political Tendency of Posts on Facebook," 2018, pp. 110–114.
[15] B. Flesch, R. Vatrapu, and R. R. Mukkamala, "A big social media data study of the 2017 german federal election based on social set analysis of political party Facebook pages with SoSeVi," 2017, pp. 2720–2729.
[16] A. Chang, "The Facebook and Cambridge Analytica scandal, explained with a simple diagram - Vox," *The Facebook and Cambridge Analytica scandal, explained with a simple diagram | Vox*, 02-May-2018. [Online]. Available: https://www.vox.com/policy-and-politics/2018/3/23/17151916/facebook-cambridge-analytica-trump-diagram. [Accessed: 10-Jun-2018].
[17] M. Rosenberg, N. Confessore, and C. Cadwalladr, "How Trump Consultants Exploited the Facebook Data of Millions," *The New York Times*, 02-Apr-2018.
[18] A. Romano, "The Facebook data breach wasn't a hack. It was a wake-up call.," *Vox*, 20-Mar-2018. [Online]. Available: https://www.vox.com/2018/3/20/17138756/facebook-data-breach-cambridge-analytica-explained. [Accessed: 10-Jun-2018].
[19] S. Frenkel, "Scholars Have Data on Millions of Facebook Users. Who's Guarding It?," *The New York Times*, 09-May-2018.
[20] A. Smith, Monica, and erson, "Social Media Use in 2018," *Pew Research Center: Internet, Science & Tech*, 01-Mar-2018. .
[21] Z. Teachout, "Mark Zuckerberg's Facebook hearing was an utter sham | Zephyr Teachout," *The Guardian*, 11-Apr-2018.
[22] J. Powles, "The G.D.P.R., Europe's New Privacy Law, and the Future of the Global Data Economy," *The New Yorker*, 25-May-2018.