Awareness and Privacy Perceptions among Social Media User towards Social Data Analytics

Yogeswari Suppiah
Faculty of Computer Science and Information Technology,
Universiti Selangor, Malaysia

Raja Mohd Tariqi Raja Lope Ahmad
Faculty of Computer Science and Information Technology,
Universiti Selangor, Malaysia

Mohd Fahmi Mohamad Amran
Computer Science Department,
Faculty of Science and Defense Technology,
National Defence University of Malaysia,
KualaLumpur, Malaysia

Wan Azlan Wan Hassan @ Wan Harun
Faculty of Computer Science and Information Technology,
Universiti Selangor, Malaysia

Suziyanti Marjudi
Faculty of Computer Science and Information Technology,
Universiti Selangor, Malaysia

Azhar Hamid
Faculty of Computer Science and Information Technology,
Universiti Selangor, Malaysia

ABSTRACT

Social Data Analytics represent a new era in the ability of organizations and businesses to tap the potential of the information economy. But, this new opportunity comes with hazards and risks as well. Privacy is the major concern among social media users. Protecting their personal information privacy has become a controversial issue among online social network providers and users. This study aims to determine the awareness of Social Data Analytics among the social media users .The result of this study can be used by future researchers, social media application developers, and providers' to understand the level of awareness of Social Data Analytics among social media users. The result also can be used to develop a new framework or privacy system according to the social media user's perception and requirement.

Keywords

Big Data Analytics, Social Data Analytics, Social Media Users, Social Data Privacy

1. INTRODUCTION

We are living in Digital Era where everything is digital such as an e-library, e-mail, e-shopping, e-ticket, e-payment, e-governance and much more. People use social media sites for entertainment such as Facebook, Twitter, and You Tube for video, photos, twits, and data downloads as well as uploads on the Internet. The Internet has stored a massive amount of data or information that is in the zeta or in Exabyte's and it is nothing but the Big Data [1]. Dynamic nature of social media data is a significant challenge for continuously and speedily evolving social media sites. There are many additional interesting questions related to human behavior can be studied using social media data. Social media can also help advertisers to find the influential people to maximize the reach of their products within an advertising budget. Social media can help sociologists to uncover the human behavior such as in-group and out-group behaviors of users [2].

The business world is undergoing a revolution driven by the use of data and analytics to guide decision-making. While many forces are at work, a major reason for the business analytics revolution is the rapid proliferation of the amount of data available to be analysed. Leading organizations increasingly recognize the importance of leveraging their data as a strategic asset. Some organizations undertake analytics initiatives to improve the quality of customer experience by measuring and acting on sentiments expressed by customers or by linking various divisions and operating units that customers tend to connect with. Others analyse data to predict a customer's propensity to buy new products or services in order to proactively make recommendations for future purchases or offer discounts to encourage a longer-term relationship [3].This study aims to understand big data and social data analytics. It also investigates the privacy concern among social media users on their personal information on these sites.

2. LITERATURE REVIEW

Social media and networks provide powerful systems for businesses to learn and use productively. Based on the pervasive use and prevailing impact social media have on society, our position is that it is essential for leaders to integrate these technologies and seek the best way to use social media and networks to the advantage of the business and for him/herself, personally. They also explain some of the reason using social media which are viral growth, connectivity, ease of use, anonymity, global reach community, relevance, and a "smart" trend. The viral growth of usage requires everyone who wants to compete in the business environment of today to understand and utilize social technologies [4].

Most people have already shared parts of their personal information with several organizations, and have accepted the loss of partial privacy. For example, one may have accepted terms to share personal health information with healthcare networks and insurance agencies, and purchasing power and trends with sales and marketing agencies. However, such data

may be shared between the data holders, either due to mergers or data sharing agreements between the organizations. The combination of such data may result in a complete view of the individual, which may be unacceptable to the concerned individuals, especially if they perceive that it may be used against their interests. Extending this from a practical perspective, it would be interesting to explore economic models that allow incentive or compensation mechanisms for individuals to trade off their privacy [10].

Social media analytics refer to the analysis of structured and unstructured data from social media channels. Social media is a broad term encompassing a variety of online platforms that allow users to create and exchange content. The key characteristic of the modern social media analytics is its data-centric nature. The research on social media analytics spans across several disciplines, including psychology, sociology, anthropology, computer science, mathematics, physics, and economics. Marketing has been the primary application of social media analytics in recent years[7]. social data analytics involves the analysis of social media in order to understand and surface insights which are embedded within the data. Online platform providers process and analyse [5]. Organizations concerned about brand, reputation, and the privacy interests of their customers must bring to bear broad expertise to evaluate their processes and decisions and to foster trusted, ethical outcomes[6].

The focus is on the concerns that individuals have about who have access to their private information and the information that is being used. Certainly, with the increased use of the Internet, social networking, and other forms of information sharing, concerns over privacy continue to be a source of much research and discussion. Some of the factors that affect privacy concern include information usage, awareness, information sensitivity, familiarity with the firm/organization, and compensation [9]. Protecting the privacy of personal information is one of the biggest challenges facing website developers, especially social network providers. Several researchers have discussed the issue of privacy [8].

3. OBJECTIVES
This study aims to determine the awareness of users of social media about Social Data Analytics. .It also aims to measure

the usage of social media sites offered by social network providers. Lastly, the study aims to measure the importance of privacy for social media users and to define which pieces of personal information they consider more important in terms of privacy protection. This information can be used to develop a new framework or privacy system according to the social media user's perception, concern, and requirement.

4. METHODOLOGY
4.1 Respondents of the study
The study sample was selected from the graduate students in Selangor. The criteria for selecting respondents were that respondents should have at least one social media account. Data were collected from the respondents through Google Forms which were emailed directly with the attached link to the respondents. Data also collected through WhatsApp application by sharing the linked Google forms to the respondents. Data collected was being analysed using SPSS 23.0.

4.2 Survey
In this research, the sampling units consist of 201 sets of questionnaires. They were distributed to the students at Private Colleges in Selangor, Malaysia. Questionnaires used in this present study consist of a cover page and its contents are divided into four section which includes Section A, Section B, Section C, and Section D with Likert-Scale format questions mainly.

5. SURVEY RESULTS
5.1 Summary of descriptive statistics
Based on this study, there is a total of 201 respondents. Among these respondents, there are 144 (71.6 %) of them are in between 17 - 20 years old. While the rest are between the ages of 21 - 40 is 57 or 28.4%, in other words, the minority group is the total of respondents. In term of status, 171 instead of 201 of them are single. Based on this study shows that most of the respondents are single . A total of 89 (44.3%) respondents were male and female respondent which is 122 or 55.7%. Most of the respondents, which are 143 (71.1%) respondents, are the student. Only 30 respondents (14.9%) are working and 27 respondents (13.4%) are having both as their occupation status.

T able 1: Respondent Demographic Profile

Respondent Demographic Profile

		Frequency	Percentage (%)
Gender	Male	89	44.3
	Female	112	55.7
	Total	**201**	**100.0**
Age	17-20 years	144	71.6
	21-25 years	20	10.0
	26-30 years	14	7.0
	31-40 years	23	11.4
	Total	**201**	**100.0**
Status	Single	171	85.1
	Married	28	13.9
	Divorced	1	0.5
	Widowed	1	0.5
	Total	**201**	**100.0**
Occupation	Student	143	71.1
	Working	30	14.9
	Both	27	13.4
	None	1	0.5
	Total	**201**	**100.0**

5.2 Usage of social media sites

Table 2 : Usage of social media sites(Purpose)

Why do you use social media sites ?	Mean	Std. Deviation
1.To make new friends	3.13	1.201
2.To play games	3.11	1.316
3.To make professional and business contacts	2.81	1.126
4.To get opinions	2.68	1.024
5.To share your experiences	2.67	1.185
6.To share videos / pictures / music	2.65	1.139
7.To keep in touch with family and friends	1.86	0.788
8.To find information	1.85	0.867

Based on table 2, it shows that the respondents are using social media sites for making new friends as illustrated. The highest mean which is 3.13 and the standard deviation is 1.201. Then goes on to play games, to make professional   and business contacts , to get opinions and to share experiences fills the list. The lowest mean is 1.85 from the respondent feedback which is respondent using social media sites for finding information.

5.3 Online social media accounts and frequency of usage

To understand the behavior of online social network users, participants were asked about the number of social media account usage and the amount of time spent browsing their account. Figure 1 shows that majority of the respondents, 79.1%   had a total number of 1-10 social media accounts. 14.4% of the users had an average of 11-20 accounts. Figure 2 shows that 43.8 % percent of the respondents were the moderate user of social media accounts. Followed by 28.4 % of the low frequency usage and 23.9% of the high frequency of social media account usage. Based on Figure 8, can be concluded that the respondents are an active user of social media with the total frequency for moderate and high reach at 67.7 %.

Table 3 : Usage of social media sites(Number of sites)

1. How many social media sites communities or groups you have been a member of ?

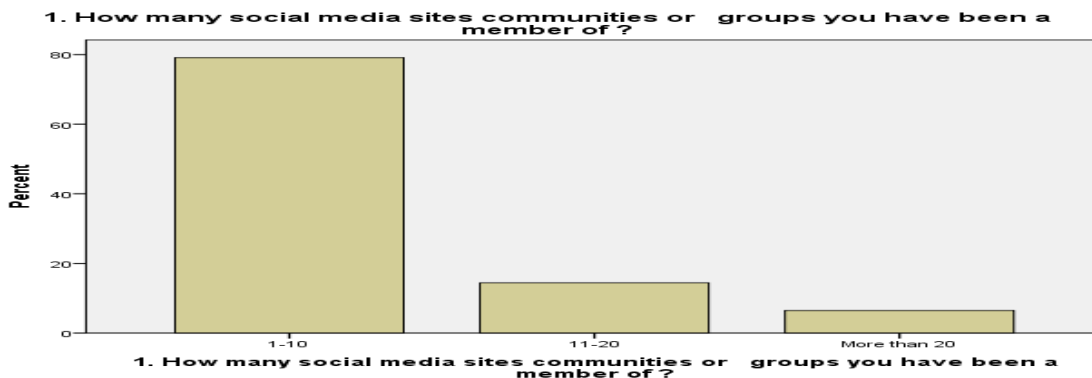|		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1-10	159	79.1	79.1	79.1
	11-20	29	14.4	14.4	93.5
	More than 20	13	6.5	6.5	100.0
	Total	201	100.0	100.0	

Figure 1 : Distribution of number of social media accounts owned by the respondents

Table 4 :Usage of social media sites(frequency)

2. How frequent you use social media sites ?

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Very low	2	1.0	1.0	1.0
	Low	6	3.0	3.0	4.0

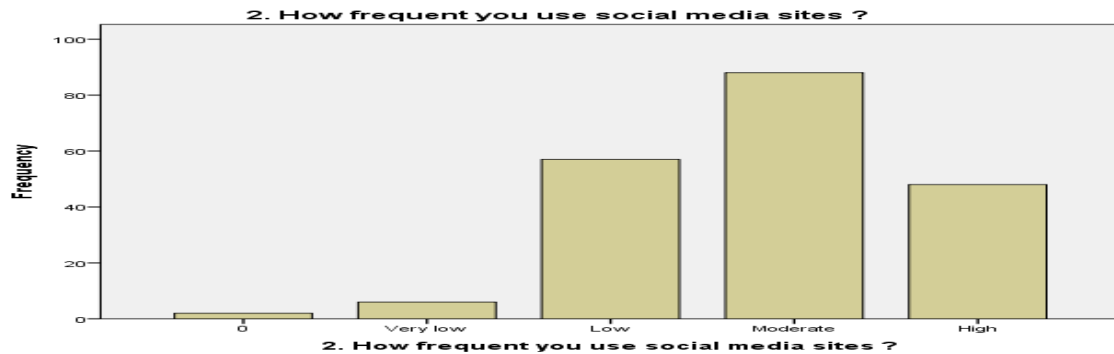|				
Moderate	57	28.4	28.4	32.3
High	88	43.8	43.8	76.1
Very High	48	23.9	23.9	100.0
Total	201	100.0	100.0	

Figure 2 : Distribution of number of the frequency of social media usage by the respondents

5.4 Respondent Awareness of Big Data Analytics

Table 5 : Awareness of Big Data Analytics

What do you know about Big Data Analytics ?	Mean	Std. Deviation
1. I know about Big Data.	3.33	1.124
2. I know about Big Data Analytics.	3.48	1.077
3. I know that social media data are being used for analysis purposes.	2.84	1.033
4. I realized that nowadays businesses are depending on this technology to analyse their customers buying behavior.	2.46	1.091
5. Big Data Analytics is a trending practice that many companies are adopting.	2.87	1.193

Based on table 5, the highest mean is 3.48, this means that most of the respondents are aware of Big DataAnalytics . Followed by the mean of 3.33 which proves that majority of the respondents are aware of Big Data. Most of the respondents have moderate awareness about business using their data on social media for analysis purposes.

5.5 Respondent Privacy concerns

Table 6 : Personal information details ranked according to importance of privacy concerns

1.What is your concern about the following personal information on social media account ?	Mean	Std. Deviation
1. Location	2.77	1.448
2. Friends List	2.68	1.030
3. Address	2.65	1.348
4. Pictures	2.64	1.277
5. Contact Number	2.62	1.416
6. Comments and Posts	2.62	1.066
7. Interest and Activity	2.56	1.033
8. Education and Work	2.54	1.063
9. E-Mail	2.39	1.024
10.Name	2.19	1.108

As shown earlier, the majority of the respondents are concerned about the privacy of personal information on their social network accounts. To determine which piece of personal information they are most concerned about, respondents were asked to rank each item of personal information according to the importance of protecting the

privacy of each item. The average rank (1 means low and 5 means high) of each item is presented in Table 6. Based on table 6, it shows that the respondent is concern about their location in the social media sites as illustrated with the highest mean which is 2.77 and the standard deviation is 1.448. The lowest mean is 2.19 from the respondent feedback whereby

the respondent's name is their concern in social media sites. Users consider contact information, such as current address, physical address, phone number and email address, need more privacy protection than other personal information. Items such as Interest and Activity, Education and Work are not as prominent in terms of the need of privacy protection.

5.6 Pearson Correlation Analysis

Table 7: Correlation between Usage ,Awareness and Privacy Concerns

Correlations

		All_Usage	All_Awareness	All_Privacy
All_Usage	Pearson Correlation	1	.380**	.292**
	Sig. (2-tailed)		.000	.000
	N	201	201	183
All_Awareness	Pearson Correlation	.380**	1	.236**
	Sig. (2-tailed)	.000		.001
	N	201	201	183
All_Privacy	Pearson Correlation	.292**	.236**	1
	Sig. (2-tailed)	.000	.001	
	N	183	183	183

**. Correlation is significant at the 0.01 level (2-tailed).

Based on table 7, the correlation between the variable Usage and Privacy Concerns is r = 0.292. The variable shows that the positive correlation but the strength of the correlation is under weak positive correlation. This result was supported the

H1a: There is a relationship between Usage and Privacy Concerns among social media users at Kolej MASA in Selangor.

5.7 Results of research hypotheses

Table 8 : Results of research hypotheses

Hypothesis	Results
H0a : There is no significant relationship between usage of social media sites and privacy concern among social media user	Rejected
H1a : There is a significant relationship between usage of social media sites and privacy concern among social media user	Accepted
H0b: There is no significant relationship between awareness of big data analytics and privacy concern among social media user.	Rejected
H1b: There is a significant relationship between awareness of big data analytics and privacy concern among social media user.	Accepted

6 LIMITATIONS OF THE STUDY

This study had a few limitations. Firstly, the population sample is skewed towards the students in Selangor only, so the result that the researchers get may not be extremely accurate. It is because the perspective of these respondents does not mean that is same with the remaining population. Secondly, the difference in backgrounds of respondents in different research will show a different result. For example, in this study, the majority of the respondents are students, students will have different perceptions from those who are workings, thus, resulted that perception has a significant relationship with occupation and age group which are inconsistent with results of previous researchers.

7 RECOMMENDATIONS FOR FUTURE RESEARCH

The result also can be used to develop a new framework or privacy system according to the social media user's perceptions, concerns and requirements. There is a lack of research about Social Data Analytics. Hence, this research could contribute to the research papers in Malaysia and future

researchers who are interested to study on the Social Data Analytics could use this study as a reference.

8 CONCLUSION

In a nutshell, the main objective of this study is to determine the level of usage, awareness and privacy concerns among social media user. Social media users have an average knowledge and understanding and awareness about Social Data Analytics. This study has a total of 201 respondents and the data collected was being analysed using SPSS 23. The level of awareness is just at a moderate level among social media users. As for privacy concern, it is varied among the respondents. Thus, this study has met its objective by knowing the level of awareness of Big Data Analytics, usage and privacy concern among social media users. The result of the study can be used to develop a new framework or privacy system according to the social media user's perception and requirement.

9   REFERENCES

[1] Gondaliya, T. P. (2015). New Big Things in Era of Digital Data : " Big Data " & Big Data Challenges with its, 496–507.

[2] Gundecha, P., & Liu, H. (2012). Mining Social Media : A Brief Introduction, (Dmml), 1–17.

[3] Gopalkrishnan, V., & Steier, D. (2012). Big data, big business: bridging the gap. *BigMine '12:/ Proceedings of the 1st International Workshop on Big Data, Streams and Heterogeneous Source Mining*, 7–11. http://doi.org/10.1145/2351316.2351318

[4] Billington, M. G., & Billington, P. J. (2012). Social media tools for leaders and managers. *Journal of Leadership, Accountability & Ethics*, *9*(6), 11–19.

[5] De Prato, G., & Simon, J. P. (2015). The next wave: "big data"? (*). *Communications & Strategies*, (97), 15–39,153,155. Retrieved from http://search.proquest.com.idpproxy.reading.ac.uk/docview/1678885914?accountid=13460

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

[7] Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management*, *35*(2), 137–144. http://doi.org/10.1016/j.ijinfomgt.2014.10.007

[8] Aldhafferi, N., Watson, C., & Sajeev, A. S. M. (2013). PERSONAL INFORMATION PRIVACY SETTINGS OF ONLINE SOCIAL NETWORKS AND THEIR, *2*(2), 1–17.

[9] Tan, X., Qin, L., Kim, Y., & Hsu, J. (2012). Impact of privacy concern in social networking websites. *Internet Research*, *22*(2), 211–233. http://doi.org/10.1108/10662241211214575

[10] Cuzzocrea, A. (2014). Privacy and Security of Big Data : Current Challenges and Future Research Perspectives, 45–47.