International Journal of Engineering & Technology, 7 (4.36) (2018) 463-465

International Journal of Engineering & Technology

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

Research paper

Big Data Social Media Analytics for Purchasing Behaviour

Shahid Shayaa[1]*, Ainin Sulaiman[2], Arsalan Zahid Piprani[2], Mohammed Ali Al-Garadi[2], Muhammad Ashraf[2]

[1]Berkshire Media Sdn Bhd, Malaysia.
**Corresponding author E-mail: shahid@berkshiremedia.com.my*
[2]Faculty of Business and Accountancy, University of Malaya
**Corresponding author E-mail: ainins@um.edu.my*

Abstract

The social media is rich in data and of late its data have been used for various types of analytics. This paper examines the purchasing behavior and sentiments of social media users from Jan - 2015 to Dec – 2016. The purchasing behaviour of the users is categorized into five: buy car, buy house, buy computer, buy hand phone and going for holiday. The paper will also demonstrate the trend of each individual category. The results of the analysis would provide businesses information on the social media users' purchasing behavior, their sentiment thus allowing them to take more appropriate strategies to enhance their competitiveness.

Keywords: *Big data, Social media, Purchasing behavior, Sentiment analysis*

1. Introduction

The Social Media which is built on web 2.0 technologies bring a new way to general public in showing their attitudes towards products and services and may ultimately influence other prospective consumers [1]. Social media now ingrained into human lives and emerged into core constituents of various human activities [6]. We express aspect of our lives through Facebook, Twitter, Blogs and forums. On the other hand, in businesses these platforms are now used to connect to existing customers, market their products and brands and explore new business opportunities.
The massive popularity of the usage of social media channels like Facebook, Twitter, Linkedin, blogs and forums spawned massive amount of data. According to International Data Corporation (IDC), the amount of data generated up till 2010 was 1 ZB of data and data is generated at the exploding rate to 7 ZB by the end of year 2014. Twitter only generates 175 Million tweets on a daily basis and has more than 467 million users [2]. The volume and speed at which data is generated over the years have led to the birth of big data. Moreover, advent of web 2.0 applications like social media platforms augmented the evolution of big data.
Big data can best be defined by its characteristics which make them different from the rest of data. The unstructured nature is one of the biggest trait of big data [5]. Blogs, forums, videos, digital images, sensors data and satellite images are some of the example of unstructured data. These unstructured data in the form of texts, comments, images, videos and sensors data can be useful to both public and private organizations.
With the addition to this, the influx of the social media platform options available to consumers, consumers can easily express their opinion, emotion, feelings and problems about their past purchases. Moreover, the conversations that are posted on different platforms can facilitate the organization to learn shopper's experience and understand consumer's buying behavior resulting improving their marketing and customer services efforts.
The rapid usage of social media platforms drives the application of social media data analytics [9]. The opinion that is mined through social media data analytics can be used to get insights about consumer's sentiments towards any product or services. Firms can capitalize their businesses by analyzing this rich data to obtain valuable insights and hidden knowledge to acquire competitive edge [3]. This social media data analytic technique is referred as sentiment analysis [8]. The intelligence obtained through analyzing sentiments can be the driver for the businesses to gain rich insights for making impact decision
Sentiment Analysis involves natural language processing with machine learning technique and text analytics to evaluate the polarity of sentiments in form positive, neutral or negative about a specific brand, product or service [4]. However, the challenge is to use the right analysis technique to reveal valuable insights and it can take long time to get the desired result. [7]. There are various techniques for sentiment analysis that are document in the literature. Lexicon based, machine learning and hybrid approaches are common techniques that are performed at document, sentence and / or the aspect level.
The objective of this study is to examine purchasing behavior and sentiments through social media data analytics. The study will also demonstrate the trend analysis from Jan - 2015 to Dec – 2016 of each individual category. The purchasing behavior of consumers is influenced by a complex combination of internal [10] and external influences [11] such as consumers' perception/psychology, culture or societal affiliation and online product recommendation. This paper however does not focus on this issue.

2. Methodology

The data used for this paper was extracted in two phases. In the first phase, Berkshire Media Sdn Bhd, a social analytics firm extracted posts from Twitter, forums, mainstream media, blogs, Facebook, online comments and Youtube. These posts are easily accessible and frequently accessed by the public and extracted using selected keywords that reflects purchasing behaviour such as purchase, buy, *beli* (in Bahasa Malaysia). Only posts made public are extracted. Each post is given a sentiment score based on how positive or negative the posts are. Positive posts are given a positive sentiment score; neutral posts are given a zero sentiment score, while negative posts are given negative sentiment scores. A larger positive sentiment score will be given for more positive posts, and a more negative sentiment score will be given for more negative posts. Data from January 2015 to December 2016 was obtained from the various social media platform. The second phase of the data extraction involve categorizing the purchasing behavior posts into five categorizations 1) buy car 2) buy house 3) buy computer 4) phone and 5) going for holiday. The extraction was carried out using automatic categorization based on the keywords in English and Bahasa Malaysia: 1- Car (*KERETA*, K*ereta, kereta, kreta Kerete, kerete, kete, Kete, keta*, car). 2- Home (house, *rumah, rmh*). 3- Computer (computer, *komputer*, *kmputer*, printer, Laptop, laptop, lptop). 4- Phone (phone, iphone, Iphone, handphone, ipad, Ipad). 5- Holiday (travel, trvel, holiday, Holiday, *cuti, Cuti*).

3. Discussion and Conclusion

3.1. Platform Distributions

A total of 44909 posts was analyzed in this study, 34813 were from Twitter, 7759 from Forum, 105 from Facebook and 2232 from others. The Twitter application program interface (API) enables the researcher to extract public posts easily [13, 14]. Twitter has been used as source of data for many researchers compared to other social media websites [15, 16] . However, even though a platform such as Facebook is a popular social media site nevertheless due to restriction of data extraction and privacy issue (most of the users keep their profile in private mode leading to restrictions of the data extraction) consequently the data extracted for the Facebook is less compared to other platforms. Fig 1 shows the histogram of overall distributions of 44909 posts according to the five categories of purchase.

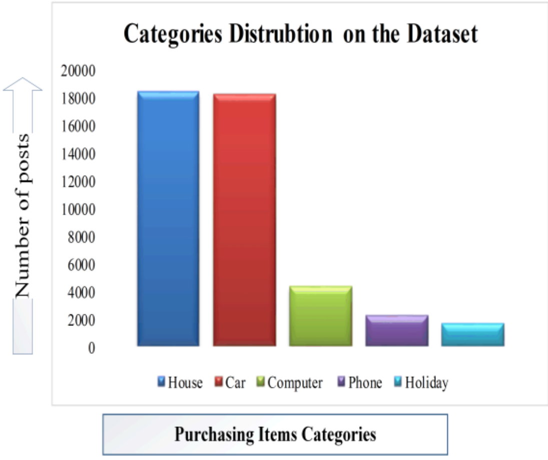

Figure 1. Distribution of purchasing items categories

3.2. Sentiment analysis by each category

The sentiment analysis for each product was analysed and classified into positive, negative, or neutral. Figures 2 to 6 shows the sentiment analysis of purchasing behavior towards home, car, computer, phone and holiday respectively.

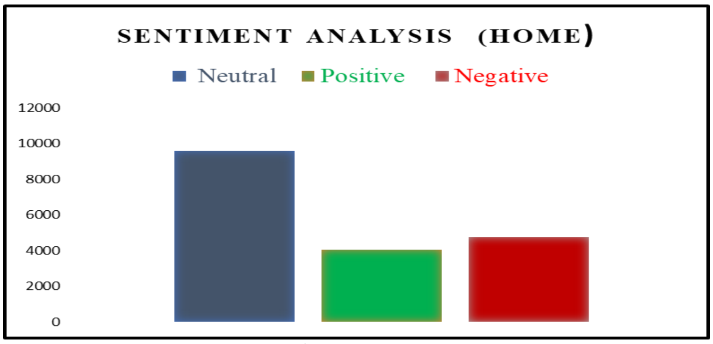

Figure 2. Sentiment analysis of purchasing behaviour toward home

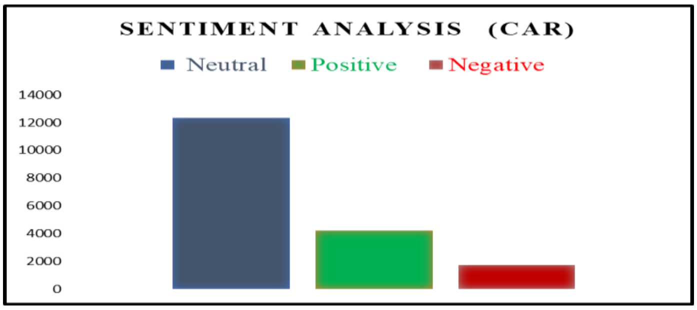

Figure 3. Sentiment analysis of purchasing behavior toward car

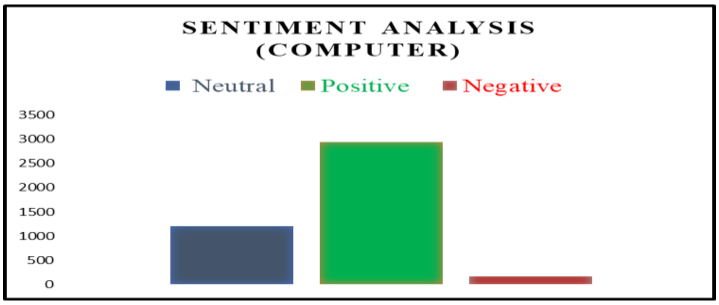

Figure 4. Sentiment analysis of purchasing behavior toward computer

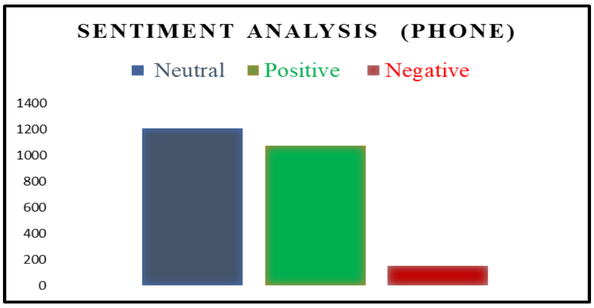

Figure 5. Sentiment analysis of purchasing behavior toward phone

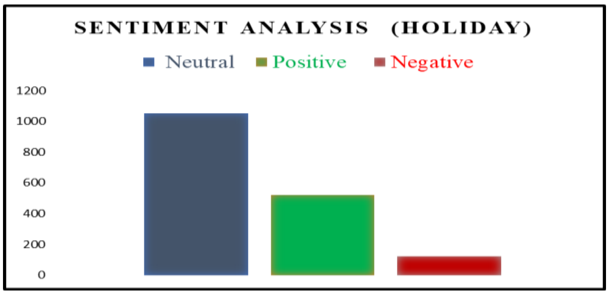

Figure 6. Sentiment analysis of purchasing behavior toward holiday

From the Figures 2 to 6, there is more neutral sentiment for all categories except for computer whereby the positive sentiment is higher. This may be due to the fact that computer is seen to be a necessity while the other categories are not. The Figures also illustrated that the negative sentiments are the lowest for all categories except home, however the difference between positive and negative sentiment is small.

3.2. Trend analysis over time

The sentiment scoring of the analyzed posts are correlated with the months of 2015 and 2016. Figures 7 and 8 show the sentiment scoring of purchasing behavior of all categories over 2015 and 2016 respectively. From the Figures, it can be seen that the sentiment scores in 2016 are relatively higher than 2015. It is also shown that the sentiment scores are highest at the end of September and beginning of October.

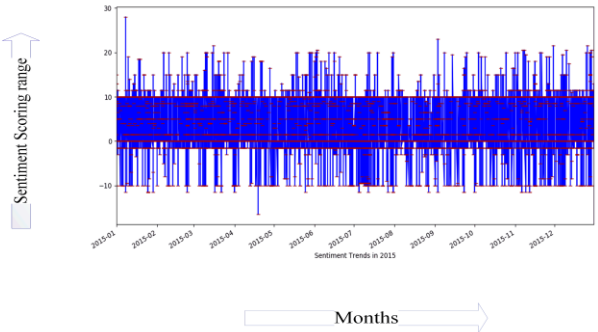

Figure 7. Sentiment scoring over 2015

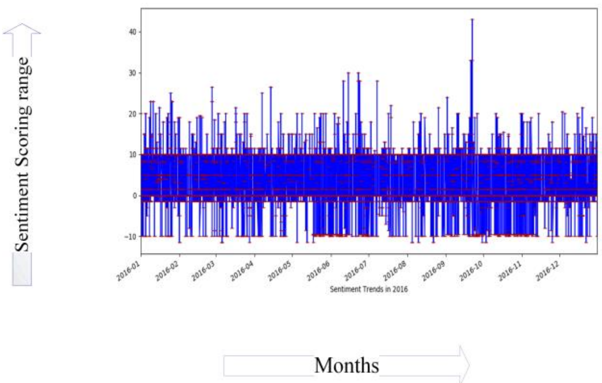

Figure 8. Sentiment scoring over 2016

3.3. Suggestion, Conclusion and Future Studies

This paper focused on data extracted from various social media platforms. Data extraction was carried out in two phases. In phase one, the data was extracted from posts that were related to purchasing behavior. In the second phase, the posts extracted were then classified into five purchasing behavior: car, home, computer, phone and holiday. It was found that a higher percentage of the posts were extracted from Twitter. The sentiment analysis has shown some interesting trends. For example, in terms of purchasing behavior, a higher percentage of the posts were related to car and home purchase. Future research should be carried out to understand why such trends occur. For example, examine whether the sentiment analysis is influenced by consumer confidence index. When it is low, the purchasing behaviour sentiment is low. Future studies should look into this in more detail. It is suggested that future studies to include posts from other countries so as comparisons between countries can be carried out. It would be beneficial to businesses to observe whether purchasing behavior within a period of time is similar for the same categories of products. Future studies should also include a more in-depth analysis of the brands of the products for each category of product as prior studies [10, 11] reported that purchasing behavior changes as product type changes.

Acknowledgement

This research is funded by University of Malaya external grant PV003-2017. We would like to thank Berkshire Media Sdn Bhd for providing the grant

References

[1] D. Arora, K. F. Li and S. W. Neville. Consumers' sentiment analysis of popular phone brands and operating system preference using Twitter data: A feasibility study. Proceedings of the 29th IEEE International Conference of Advanced Information Networking and Applications, (**2015**) March 680-686; Taipei, Taiwan

[2] Deep Analytics. Unstructured data: A big deal in big data. Available at: http://www.digitalreasoning.com/resources/Holistic-Analytics.pdf (**Accessed 22-09-2016**).

[3] W. He, S. Zha and L. Li. Social media competitive analysis and text mining: A case study in the pizza industry. Intl J. of Info Mngt, 33(3), 464-472 (**2013**).

[4] M. Hu and B. Liu. Mining and summarizing customer reviews. Proceedings of the 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, (**2004**) August 168-177; Seattle, USA

[5] Judith Hurwitz, Alan Nugent, Fern Halper, and Marcia Kaufman. Understanding Big Data for Dummies, http://www.dummies.com/how-to/content/ (**Accessed 22-09-2016**).

[6] N.B. Lassen, R. Madsen and R. Vatrapu. Predicting iphone sales from iphone tweets. Proceedings of 18th IEEE International Conference on the Enterprise Distributed Object Computing (EDOC), (**2014**) September 81-90; Ulm, Germany

[7] A. Semenov. Principles of social media monitoring and analysis software. Jyväskylä studies in computing. 168, 1456-5390 (**2013**)

[8] M. Taboada, J. Brooke, M. Tofiloski, K. Voll, and M. Stede. Lexicon-based methods for sentiment analysis. Computational linguistics, 37(2), 267-307 (**2011**).

[9] R. Thackeray, B. L. Neiger, C. L. Hanson, and J. F. McKenzie. Enhancing promotional strategies within social marketing programs: use of Web 2.0 social media. Health promotion practice. 9(4), 338-343 (**2008**).

[10] A. Benlian, R. Titah, and T. Hess. Differential effects of provider recommendations and consumer reviews in E-commerce transactions: An experimental study. J. of Mngt Inf Sys. 29(1), 237-272 (**2012**).

[11] L. Huang, C.-H. Tan, W. Ke, and K.-K. Wei. Comprehension and Assessment of Product Reviews: A Review-Product Congruity Proposition. J. of Mngt Info Sys. 30(3), 311-343 (**2013**).

[12] E. Kouloumpis, T. Wilson, and J. D. Moore. Twitter sentiment analysis: The good the bad and the OMG! Proceedings of the 5th International AAAI Conference on Weblogs and Social Media, (**2011**) July 538-541; Barcelona, Spain.

[13] A. Pak and P. Paroubek. Twitter as a corpus for sentiment analysis and opinion mining. Proceedings of the International Conference on Language Resources and Evaluation, (**2010**) May 17-23; Valletta, MaltaIn.

[14] M. A. Al-garadi, K. D. Varathan, and S. D. Ravana. Cybercrime detection in online communications: The experimental case of cyberbullying detection in the Twitter network. Computers in Human Behavior. 63, 433-443 (**2016**).

[15] M. A. Al-garadi, M. S. Khan, K. D. Varathan, G. Mujtaba, and A. M. Al-Kabsi. Using online social networks to track a pandemic: A systematic review. J. of Biomedical Informatics. 62, 1-11 (**2016**).