TweetGST: A Web Based Systems for Sentiment Analysis and Opinion Mining for GST in Malaysia

Nurul Misyani Mohd Rafie¹, Kasturi Dewi Varathan¹ and Mohammad Shafenoor Amin²

¹Department of Information Systems, Faculty of Computer Science and Information Technology, University of Malaya, 50603 Kuala Lumpur, Malaysia.
²Department of Computer Science and Engineering, BRAC University, 1212 Dhaka, Bangladesh

nurulmisyani_mr@gmail.co.id*

Abstract. Twitter is one of the most used platform for communication by people around the world. Sentiment analysis and opinion mining on twitter has become an important field of research that sees beyond the number of likes, shares and comments that users have for some topics. Nowadays, business organizations are not aware of emotional aspect of their customers and may take longer time in analyzing the market sentiment in making the necessary business decision. TweetGST is a web based system that enable business user to extract GST based tweets from Twitter as well as it helps in analyzing the tweets accordingly. TweetGST enables for trend analysis and sentiment analysis to be made. These analysis will help the business organization in order to understand customer sentiment and provide for their needs. The analyzed tweets will be used in helping the business organization to make the right decision which are recommendation for sales and stocks. Thus, TweetGST is expected to increase the revenue of the businesses.

1. Introduction
Sentiment analysis refers to a classification problem where the main focus is to predict the polarity of words and then classify into positive, negative or neutral feelings with the aim of identifying attitude and opinions that are expressed in any form or language. (Farhan Hassan, Saba Bashir, Usman Qamar, 2013). Moreover, opinion mining is the field of study that analyzes people's opinions, sentiments, evaluations, attitudes, and emotions from written language. (Bing Liu, 2012). The emergence of social media like Facebook, Twitter and blogs as medium that users use to express feelings, opinions, thoughts about different topics such as current issues, entertainment industries and etc.

This research is conducted in order to know the sentiments of the people towards the implementation of Goods and Service Tax(GST) as well as to help business organizations to be aware of the emotional aspects of their customers (Asfaneh Nahavandi, Robert B. Denhardt, Janet V. Denhardt, Maria P. Aristigueta, 2014). Thus, the purpose of developing TweetGST is to analyze the sentiment and opinion of public users in Twitter pertaining the products that were sold which are related to GST in Malaysia. This analysis will help the business organizations to understand customers’ need based on their sentiments. In addition this system will also be useful in making
decisions for business organizations for example identifying promotional items and managing stocks among business organizations.

2. Related Works
Sentiment analysis and opinion mining is a field of study that analyzes peoples’ opinions, sentiments or thoughts, evaluations, attitudes and emotions from written language. This is one of most active research area in natural language processing and is also widely studied in data mining, web mining and text mining. It is the field of study that analyzes people's opinions, evaluations, attitudes, and emotions from written language. (Bing Liu, 2012). Maria Ogneva(2010) has indicated that sentiment is believed to play a vital role in the decision making process.

There are some existing systems that analyzes sentiments. Sentiment140(Alec Go.,2010) is also known as Twitter Sentiment allows users to discover the sentiment of brands, topics or products in Twitter. This system caters for English and Spanish languages. However, this system does not cater for sentiment of products or brands in Malaysia. The other systems such as SnetimentViz, SocialMention, StreamCrab(Timor Abdulaive., 2011), Twitrat, UberVu as well as Twends provides tweets extraction, sentiment analysis, search and product analysis. All the systems mentioned above do not provide features on trend analysis, recommendation on products and stocks. Apart from all those features one of the important features of TweetGST is that it provides analysis based on user generated content from Malaysia.

| Table 1. Feature comparison between sentiment analysis systems |
|---------------------------------------------------------------|
| **Systems** | **Tweet Extraction** | **Sentiment Analysis** | **Search** | **Product Analysis** | **Trend Analysis** | **Malaysians Tweets** | **Recommendation (products)** | **Recommendation (stocks)** |
| Sentiment140  | V | N | N | N | V | - | - | - |
| Sentiment Viz | V | N | N | N | N | V | - | - |
| SocialMention | V | N | N | N | N | N | V | - |
| StreamCrab | V | N | N | N | N | N | N | N |
| Twitrat | V | N | N | N | N | N | N | N |
| UberVu | V | N | N | N | N | N | N | N |
| **TweetGST** | V | N | N | N | N | N | N | N |

3. Methodology
Tweets are extracted and analyzed in this research. GST based tweets are collected from 31 January 2015 until 20 May 2015. All the tweets are collected from a set of keywords that is related to GST such as GST, Good&ServicesTax, Good&Services Tax, #GST, myGST, GST2015, gst and #myGST. Total data collected was 125,465 tweets. The conceptual model of TweetGST is shown in Figure 1.
By utilizing the Twitter APIs, the GST based tweets will be retrieved by the connection of APIs and the database filtered based on the GST and goods and services tax keywords identified. After that, all the tweets are analyzed and reports are generated for end users. Apart from sentiment analysis module, there will be recommendation modules which recommends sales and stocks of the following months by using the results of sentiment and sales of products. Users also can search products and view sentiment analysis of public users or customers based on tweets on the products. All the modules will generate report and display these reports to the users. Moreover, admin has three main modules which are Sales Management, Stocks Management and Stock Sentiment Preference. In these three modules, admin can insert, update and delete data regarding sales and stocks.

3. Results and Discussions

For location analysis the graph shows the numbers of GST tweets based on specific locations. Figure 2 and Figure 3 are the examples of results of location analysis. This feature help users to track the most popular location by states.
Figure 3 shows the percentage of tweets based on the states in Malaysia. Pulau Pinang obtained the highest percentile in tweets by 47% followed by Melaka obtaining 19%, while Kuala Lumpur and Sabah receiving 10%. Other states have tweets percentile below than 6%.

Based on the GST tweets obtained by location, Table 2 indicates the percentage of population by top 3 race in Malaysia for top 4 states with GST based tweets. It shows that in Pulau Pinang and Kuala Lumpur the percentage of Malays and Chinese tweets on GST are almost similar.

| States      | Malays(%) | Chinese(%) | Indian(%) |
|-------------|-----------|------------|-----------|
| Pulau Pinang| 43.6      | 45.6       | 10.4      |
| Melaka      | 66.9      | 26.4       | 6.2       |
| Kuala Lumpur| 45.9      | 43.2       | 10.3      |
| Sabah       | 84.8      | 12.8       | 0.3       |

TweetGST system also provides trend analysis of GST based tweets by months. Figure 4 shows the analysis from Month of February to month of May. It shows that the peak of the tweets generated was in March 2015(a month before the implementation of GST in Malaysia).

Figure 5 shows frequency of selected products which are Milk, Book, Topup for phone, Sardine, Burger and Instant Noodle. From the graph, the highest frequency of the graph is Topup for phone. This shows that many people expressed their opinions about the implementation of GST in Top up for phone via Twitter.
The sentiment of users based on products are also analyzed in TweetGST system. Figure 6 and Figure 7 shows the sentiment analysis on Milk products and Books.

From the analysis made, it shows that there is 40% negative sentiments on the implementation of GST on milk products. On the other hand, GST imposed on books hits 57% of negative sentiments. From the results of sentiment analysis of each products, there will be a recommendation from TweetsGST system. The recommendation is based on the number of sales and sentiment analysis of each product. If there is a drop of sales and the product has negative sentiment, then TweetGST will make recommendation for the product to be on sale. However, the percentage of drop of sales and negative sentiment that triggers the promotions of products are determined by the business organization itself via the admin module in TweetGST. This analysis and recommendations will be done on monthly basis on each of the products. In addition, there will be a recommendation in purchasing stocks of products. The purchase of stocks for the upcoming months depends on the increase of sales from the previous months as well as the sentiments that are portrayed on the product itself.

4. Conclusion
The features that are contained in this system enables extractions of tweets and analyse the sentiments that are contained by the tweets based on products. Besides that, the analysis on trends and location provides useful insights to the businesses. The recommendation based on products and stocks gives significant impact on the management of the warehouse. TweetGST can be used by organizations that conducts business like franchise, online business and retailers. TweetGST is an application that help business organizations in making decision based on opinion and sentiment of public tweets. This system will be able to help business organizations in analyzing the customers’ needs based on the
tweets generated by the users and make the necessary decisions resulting in increase of the revenue. Although, TweetGST have its own strengths but it also has its limitations. In TweetGST, sales are recorded per month and this system do not cater an analysis for festive seasons such as Chinese New Year, Deepavali and Hari Raya Aidilfitri.

References
1. Farhan Hassan, Saba Bashir, & Usman Qamar. (2014). TOM: Twitter opinion mining framework using hybrid classification scheme. Decision Support Systems, 57(2014), 245-257. www.elsevier.com/locate/dss
2. Bing Liu. (2012). Human Language Technologies: Sentiment Analysis and Opinion Mining. Chicago: Morgan And Claypol Publisher.
3. Asfaneh Nahavandi, Robert B. Denhardt, Janet V. Denhardt, Maria P. Aristigueta. (2014). Organizational Behavior. SAGE Publications.
4. Alec Go. (2010). Sentiment140 Retrieved by http://www.sentiment140.com/.
5. Timor Abdulaive. (2011). Streamcrab. Twitter Sentiment Analysis. Retrieved by http://www.streamcrab.com/
6. Maria Ogneva. (2010). How companies can use sentiment analysis to improve their business. Retrieved by http://mashable.com/2010/04/19/sentiment-analysis/.