Business intelligence for social media interaction in the travel industry in Indonesia

Michael Yulianto\textsuperscript{a}, Abba Suganda Girsang\textsuperscript{a*} and Reinert Yosua Rumagit\textsuperscript{a}

\textsuperscript{a}Computer Science Department, BINUS Graduate Program-Master of Computer Science, Bina Nusantara University, Jakarta, Indonesia; \textsuperscript{*}agirsang@binus.edu

This article is Open Access, in compliance with Strategy 2 of the 2002 Budapest Open Access Initiative, which states:

Scholars need the means to launch a new generation of journals committed to open access, and to help existing journals that elect to make the transition to open access. Because journal articles should be disseminated as widely as possible, these new journals will no longer invoke copyright to restrict access to and use of the material they publish. Instead they will use copyright and other tools to ensure permanent open access to all the articles they publish. Because price is a barrier to access, these new journals will not charge subscription or access fees, and will turn to other methods for covering their expenses. There are many alternative sources of funds for this purpose, including the foundations and governments that fund research, the universities and laboratories that employ researchers, endowments set up by discipline or institution, friends of the cause of open access, profits from the sale of add-ons to the basic texts, funds freed up by the demise or cancellation of journals charging traditional subscription or access fees, or even contributions from the researchers themselves. There is no need to favor one of these solutions over the others for all disciplines or nations, and no need to stop looking for other, creative alternatives.

Please scroll down for article.
Business intelligence for social media interaction in the travel industry in Indonesia

Michael Yulianto\textsuperscript{a}, Abba Suganda Girsang\textsuperscript{a*} and Reinert Yosua Rumagit\textsuperscript{a}

\textsuperscript{a}Computer Science Department, BINUS Graduate Program-Master of Computer Science, Bina Nusantara University, Jakarta, Indonesia 1148

Corresponding author (*): mailto:agirsang@binus.edu

Received 14 May 2018 Accepted 25 July 2018

ABSTRACT Electronic ticket (eticket) provider services are growing fast in Indonesia, making the competition between companies increasingly intense. Moreover, most of them have the same service or feature for serving their customers. To get back the feedback of their customers, many companies use social media (Facebook and Twitter) for marketing activity or communicating directly with their customers. The development of current technology allows the company to take data from social media. Thus, many companies take social media data for analyses. This study proposed developing a data warehouse to analyze data in social media such as likes, comments, and sentiment. Since the sentiment is not provided directly from social media data, this study uses lexicon based classification to categorize the sentiment of users’ comments. This data warehouse provides business intelligence to see the performance of the company based on their social media data. The data warehouse is built using three travel companies in Indonesia. As a result, this data warehouse provides the comparison of the performance based on the social media data.

KEYWORDS Business intelligence, lexicon based classification, sentiment analysis, social media

1. INTRODUCTION

The development of air transportation and airlines in Indonesia is increasing. This is marked by the growing number of airlines that have sprung up by offering both domestic and international travel routes that make the competition more competitive. With competitive competition, many airlines offer promotions that can be an attraction for consumers. This is certainly a great opportunity for business people to use information technology. The development of telecommunication and computer technology led to changes in the pattern of instant purchasing, online reservations, and the ticketing process, which in the aviation world is often called the online system or electronic ticketing (Atmadjati, 2012).

In Indonesia electronic ticketing providers are becoming more common, so competition is increasing. Because business competition requires price matching, companies must compete to attract consumers as much as possible in order to survive. Many companies use media for marketing. This includes social media, like Facebook and Twitter. With social media, customers can easily contact the company (customer service). Businesses start looking at such technologies as effective mechanisms to interact more with their customers (Ali Abdallah Alalwan, et al. 2017).

Social media has become the largest data source of public opinion (Shuyuan Deng, 2017).
Indonesia has the fourth most Facebook users in the world. Therefore, this study focuses on the relationship of social media use, namely Facebook and Twitter, to see the interaction between companies and consumers.

Data that exist in social media can help us to do the analysis to help companies get feedback from consumers. The data that can be retrieved include "like, comment, and share" information. Sentiment analysis can be used to process comments in order to get feedback on the nature of the comment, good or bad (He, Zha, & Li, 2013). Poor comments can be used as advice and input for the company in the future (Saragih & Girsang, 2017).

In this study, using existing data in social media Facebook and Twitter is expected to create business intelligence that can help analyze travel business companies in Indonesia with social media data interaction.

2. CONCEPTUAL BACKGROUND

In this chapter, we examine the concept and characteristics of business intelligence and sentiment analysis using lexicon based classification.

2.1 Business Intelligence

Business information and business analysis in the context of business processes are the key that leads to decision-making and actions that lead to improved business performance. Business intelligence can be defined as "a set of mathematical models and analytical methodologies used to exploit the data available to produce information and knowledge useful for complex decision-making processes" (Vercellis, 2006, Williams, S., and Williams, N, 2006).

Advantages of business intelligence:

- Increase Profits: Business intelligence can help business clients to evaluate customer value and desire for short-term profits and to use the knowledge used to differentiate between profitable customers and non-profitable customers.
- Reduced costs: Reducing the investment needed to use sales, business intelligence can be used to assist in evaluating the organization's costs.
- Develop Customer Relationship Management (CRM): This is essentially a business intelligence application that applies customer information collection analysis to provide responsible customer service responsibilities that have been developed.
- Reduce the Risk: Applying the business intelligence method to enter data can develop a credit risk analysis, looking at the analysis of consumer activity, producers, and reliability can provide insight into how to shorten the supply chain.

2.2 Sentiment Analysis

Sentiment analysis or opinion mining is a process of understanding, extracting and processing textual data automatically to get sentiment information contained in an opinion sentence. Sentiment analysis is done to view opinions or opinion tendency of a problem or object by someone. Sentiment analysis can be distinguished based on the data source, some of the level that is often used in research sentiment analysis is sentiment analysis at document level and sentiment analysis at sentence level (Bo, P et al. 2002).

The lexicon-based approach depends on the words in the opinion (sentiment), specifically words that usually expresses a positive sentiment or negative sentiment. Words that describe the desired state (e.g. great, good) have positive polarity, whereas the words describing the unwanted state have negative polarity (e.g. bad, horrible). One common approach used in performing sentiment analysis is using a dictionary based approach. Because this research is based on Indonesia, the dictionary will use Indonesian words. Figure 1 is a positive dictionary and Figure 2 is a negative dictionary.

3. METHODOLOGY
Research conducted begins based on the interest of the writer about the data that exist on social media. Therefore, through this research, the author wants to create a data warehouse for social media data in order to perform analyses related to social media interactions. These include an analysis of how actively the company replies or communicates with its customers on social media such as Facebook or Twitter.

### 3.1 Crawling Data

Data retrieval is done from selected social media platforms such as Facebook and Twitter via the social media API available on each platform.

Data retrieval is done periodically by crawlers. The data is taken every Wednesday and Saturday. This is done because the data provided by the Twitter API only retrieves data up to seven days old. For example, data retrieved on October 18, 2017 from Twitter can only go back as early as October 11, 2017. Data before that date cannot be retrieved.

From the data that was regularly taken by the crawler, was stored on in the form of excel files.

The types of data stored on each social media platform are different:

- Facebook: post, comment, reply, like
- Twitter: tweet, retweet, mention

Crawling data in this research uses Rstudio, for crawling Facebook the Library Rfacebook was used and for Twitter, TwitterR was used.

#### 3.1.1 Crawling Facebook

In this research, will use three months of data, from September 2017 to December 2017 from three companies. The pseudocode used to get data using Rfacebook in Rstudio was:

- Load Rfacebook
- Connect to Facebook API using fbOauth
- Get Page from Official Facebook Page using function GetPage
- Get all Post in Page use GetPost
- Get Like and Comment from Post (post$Likes & post$Comments)
- Get Like and Reply from Comment using getCommentReplies
- Export to Csv format

#### 3.1.2 Crawling Twitter

TwitterR uses the Twitter API to get the data. Because of this, there is a seven day limitation from the day we request data. The pseudocode to get the data using Rfacebook in Rstudio was:

- Load TwitterR
- Connect to Twitter API using setup_twitter_oauth
- Search @ from Twitter example from:traveloka
- Search “@” example @traveloka
- Search “to” tweet example to:traveloka
- Export to csv format

### 3.2 Sentiment Analysis

#### 3.2.1 Preprocessing

Preprocessing data data comments from Facebook and Twitter social media is done by preprocessing before sentiment analysis. Figure 4 shows the preprocessing stages.

The first step is case folding. Case folding is the process of converting words into lowercase. The purpose of turning words into lowercase is to eliminate case sensitive errors. The next step is to filter the sentence. Written words are
punctuation, number, and website address. The process of separating sentences into individual words is usually called tokenization. The easiest way to turn a sentence into words is to separate them with spaces. Stemming is the process of converting words into basic words.

### 3.2.2 Lexicon Based Algorithm

The lexicon algorithm converts data via a function that will process every sentence in the data source. Figure 5 is the pseudocode for the sentiment analysis using the lexicon based algorithm (Chopra and Bhatia, 2016).

1. Enter the text as input.
2. Divide this paragraph into tokens and store the words in an array list.
3. Select the first word from array list.
4. Fetch the words of database in second array named as database array.
5. Check whether selected paragraph word matched with each word of database array.
   (i) If match found
       (a) Find the sentiment of word from database whether it is positive/negative or neutral.
       (b) Find the exact position of word in the paragraph.
       (c) Highlight the word according to their sentiment; make it green if it is positive, red if it is negative and blue if it is neutral.
       (d) Calculate the score of sentence.
       (e) Store the results in database.
   (ii) Else match not found
       (a) Select next word from the array
       (b) Go to step 5.
6. Display the result to the user.
7. Plot the graph according to the results.

**Figure 5** Pseudocode for the sentiment analysis using the lexicon based algorithm.

### 4. RESULTS

Result from the methodology above are shown in Figure 6. There are two table facts and five dimension tables. The two fact tables are: the fact company activity and fact user activity. The five dimension tables are: dim user, dim sentiment, dim company, dim media social, and dim time.

Dashboard admin activity consists of four reports (Figure 7). The first report is the report of admin activity trends during the month, the second report provides an overview of the activities undertaken by the admin, the third report is a report of activity per day while the latter is an hourly activity. Uniquely by using the business intelligence program tableau all existing reports can affect each other, for example when we click on the first report graph on the line Traveloka and September all reports on this page will show Facebook Traveloka data in September.

Dashboard user activity consists of five reports (Figure 8). The first report is the report of user activity trends during the month, the second report is sentiment analysis report, the third report is the most active user in social media, the fourth is user activity by day and the last is an activity report by hour. With this dashboard we can analyze who is active during the month or day or time we choose in the dashboard.

On the dashboard the activity of the companies assessed can be seen. Facebook social media shows that the company Pegi Pegi is the most active compared to other companies. In September it was found that Pegi-Pegi made a social media strategy change, which can be seen in October with a rise of almost 368.81%. The company, Ticket, had the lowest activity. In this company there is even a decline in October and December.

On Twitter, Traveloka has the most activity compared to other companies. Traveloka has more than 1,000 activities per month. Other companies have almost 10 times less activity than Traveloka. Pegi-Pegi and Ticket had an
increase in November and December. In
November there was a decrease in activity.
Figure 9 summarizes the company activity on
social media.

The most frequent Facebook activity by
companies is reply to comments from
customers. This was most frequently done by
Traveloka, followed by Pegi-Pegi and Ticket. At
Pegi-Pegi the most most common activity was
liking comments from its customers. Figure 10
shows activity by hour.

The companies’ Facebook and Twitter
activity peaked at 16:00-16:59. Traveloka’s
activity peaked at 19.00 - 19.59 while Pegi-Pegi
was most active at 16.00 - 16.59 and Ticket was
most active at 12.00 - 12.59 (Figure 11).

Research conducted during four months
of social media data collection on Facebook and
Twitter, obtained 28,445 comments and

Figure 6 Star schema.

Figure 7 Dashboard company activity.
2,379,107 liked statuses by the users (Figure 12). This figure is very high, and reflects how enthusiastic the users with activities performed by the company. On social media Facebook, Traveloka has more enthusiastic users than the other two companies, this is evidenced by the existence of 1,386,318 user activity data points, of which 942,769 activities occurred in October. When viewed in more detail, Pegi-Pegi has more active users than Traveloka in the last two months (November and December).

From 28,445 comments, Traveloka has the most negative sentiment with an average of 14.26% negative, 34.51% positive sentiment and 51.23% neutral sentiment on Twitter. Tickets have the best positive analytical sentiment with a value of 44.05%, compared with negative sentiment which is only 14.10% and a neutral value of 41.85%. Figure 13 shows the results of the lexicon-based sentiment analysis.

The last four months’ data got the names of users who most actively made comments or liked a status or comment. In every form of social media there were users who engaged in more than 100 activities in the last 4 months (Figure 14). On Traveloka, the top ten people engaging had an average activity of 200 interactions, while Pegi-Pegi had an average of 168 activities and Ticket has the lowest average of 84.

5. RECOMMENDATION

From the dashboard analysis various recommendations for companies studied were obtained.

5.1 Traveloka

On Facebook social media needs to be improved again because from November there was a significant decline (23%) compared to the previous month. At 19.00 - 19.59 the activities of the Traveloka are recommended to have more human resources in order to help solve customer problems.
5.2 Ticket

On Facebook, social media needs to be improved. In September there were 94 activities, but this declined considerably to 74 activities in December. On Twitter, engagement should be improved again as compared to Traveloka, as the activity of Ticket is lagging behind. For Twitter we suggest human resources should be available in the early hours, as in December at 00.00 - 07.00 there were only seven activities, compared with user activity on Ticket’s Twitter feed of as much as 85 activities.

5.3 Pegi-Pegi

For Twitter, we suggest increased human resources in early hours. In December at 00.00 - 07.00 there were 55 activities only compared with user activity on Twitter Pegi - Pegi as many as 244 activities.

6. CONCLUSION

Based on the results of the research, there are several conclusions. By using business intelligence conducted in this research, Traveloka has the most interaction in social media, as compared with Pegi-Pegi and Ticket.com.

This research provides some suggestions for the development of business intelligence for social media interaction. The classification accuracy can be further improved by using algorithms and machine learning such as naive base classification and in the future data could also be analyzed to include emoticons for more complete information from Facebook.
REFERENCES


