Twitter Mining to Explore Perceptions of Tourism Objects in Indonesia: A case study of Borobudur and Prambanan Temple

Valentinus Roby Hananto
Dept. of Information System, Faculty of Technology and Informatics
Institut Bisnis dan Informatika Stikom Surabaya
Raya Kedung Baruk 98, Surabaya 60298, Indonesia
valentinus@stikom.edu

***Abstract*— The tremendous growth of social media and User-Generated Content (UGC) can be an opportunity for tourism industry. The purpose of this paper is to use Twitter analysis to explore visitor perceptions of two tourism objects in Indonesia (Borobudur and Prambanan Temple). Using 620 tweets referring to the two biggest temples in Indonesia, this research used text mining and sentiment analysis to find meaningful patterns, popular words, and emotional states in opinions. Twitter users held positive perceptions of Borobudur and Prambanan temple. The number of tweets for Borobudur temple was significantly higher than Prambanan temple. The frequency of tweet from both objects had an upward trend toward the end of the year. This research provides a methodology that future researchers can use in applying social media analytics to explore major issues and extract sentiment information from text messages. Limited research has been conducted applying social media analysis for Indonesia tourism. Therefore, this study fills a gap by using social media analytics with Twitter data to examine the Twitter users' thoughts and emotions for Indonesia tourism objects (Borobudur and Prambanan temple).**

Keywords*— *Social media analysis, Twitter mining, Indonesia tourism

I. INTRODUCTION

Data Mining is a process of analyzing huge data sets to search and find out earlier unidentified patterns and relationships to produce information to support decision making process. Tourism is one of the areas that can gain benefits from Data Mining process. With the help of Data Mining tools and techniques, market tourism product and services can be managed effectively. User-Generated Content (UGC) is the driver for collecting the data to support web mining/analytics. With big data generated from UGC, we can use it to analyze all tourism review from visitors. Therefore, plan and strategy to develop tourism can be discussed. Tools of Big Data analysis can be very useful in analyzing the patterns and trends in the data

The number of internet users has been grown rapidly nowadays. People can easily upload any of their ideas or concerns on the internet. Thus, social media and user-generated content (UGC) also rise and it produces a huge quantity of information on the internet [1]. The social media interaction generates big data that can be stored and computed on the available cloud services. Big data analysis is a way to find trends and patterns in the data. With the emergence of Business Intelligence and Smart cities, big data analysis have been implemented in various sectors such as tourism, healthcare, and other sectors, and there already have been many benefits from this technology [2]. The implementation of information technology has assisted the tourism industry and also the tourists. The conventional business models has been transformed to adapt the usage of the internet.

Smart tourism is a concept that applied to describe the increasing reliance of tourism destinations, their industries and their tourists on emerging forms of ICT that allow for massive amounts of data to be transformed into value propositions [3]. Given the fact that the number of social media users have been grown massively, there have been numerous method to analyze social media content. One of the popular social media is Twitter, a microblogging service that was launched in 2006 [4].

Analyzing customer perception from tweet messages can be useful for researchers to estimate customers' attitudes toward an object. User-generated content in Twitter, or a tweet, includes diverse attributes like message text, screen name of sender, posting time, language type and so on. Given that customers provide honest opinions on products and service, social media analysis using Twitter is important to the tourism industry.

Each tourism object has their own images that can be different with those of the destinations at which the objects are located. In addition, even in the same circumstances or surroundings, every individual is likely to perceive a certain landmark differently. In some cases, the image of landmark naturally overlaps with that of the destination.

This research is designed to implement social media analysis using Twitter, to analyze the perceptions and sentiments toward two biggest temple in Indonesia, Borobudur Temple and Prambanan Temple. Both temple are located in Central Java. The Borobudur Temple is one of the greatest Buddhist monuments in the world, and was built in the 8th and 9th centuries AD during the reign of the Syailendra Dynasty

[5]. Prambanan Temple Compound consist of Prambanan Temple (also called Loro Jonggrang), Sewu Temple, Bubrah Temple and Lumbung Temple. Prambanan Temple itself is a complex consisting of 240 temples. All the mentioned temples form the Prambanan Archaeological Park and were built during the heyday of Sailendra's powerful dynasty in Java in the 8th century AD [6].

This study makes important theoretical contributions to literature on social media mining in two ways. First, this study is one of the first to use Twitter analysis for exploring the perceptions and sentiments of tourism objects in Indonesia. This study extends the literature on Borobudur and Prambanan Temple by analyzing visitors' perceptions and emotions on Twitter. In particular, analyzing emotions in tweets provides insight into how people perceive Indonesia tourism. There were still few researches that have analyzed and compared customer perceptions of or attitudes toward different tourism objects in Indonesia.

The structure of paper is as follows. This paper first reviews the current stream of social media analytics in tourism research and previous discussion of customer perceptions of Borobudur and Prambanan Temple. Secondly, this study demonstrates social media analysis (e.g. text mining techniques, frequent word analysis and trend analysis) to assess customer perceptions of and attitudes toward Borobudur and Prambanan Temple. Lastly, managerial implications driven by research outcomes are offered to researchers and experts in Indonesian tourism management and marketing.

II. LITERATURE REVIEW

Twitter is a real-time social media with microblogging service that allows users to post a short status update in 140 characters. An understanding of tweets structure is important in order to use Twitter's API effectively. A tweet is not only about the 140 characters, it also contains text content associated with a user's status update and also more metadata. In addition to the textual content of a tweet itself, there are also two essential metadata within a tweet: entities and places. Tweet entities are consisted of the user mentions, hashtags, URLs, and media that may be associated with a tweet, whereas places are geo-location information that attached in a tweet. However, a place is not always the actual location in which a tweet was authored, but it might also refers to the place described in a tweet [7].

There were some steps in mining data from Twitter, initially begin with collecting data. In order to have access to Twitter data programmatically, an app is required to interact with the Twitter API (http://apps.twitter.com). Twitter provides REST APIs to interact with their service. There is various Python-based clients available that can be used with Twitter REST API. In particular, Tweepy is one of the most interesting and straightforward Python-based clients to use [8].

After the tweet data have been collected, the tweet structure should be analyzed. These are the key attributes of a tweet:

- text: the text content of the tweet
- created_at: the date of creation
- favorited, retweeted: boolean stating whether the authenticated user (you) have favorited or retweeted this tweet
- favorite_count, retweet_count: the number of favorites and retweets
- lang: acronym for the language
- id: the tweet identifier
- place, coordinates, geo: geo-location information if available
- user: the author's profile
- entities: list of entities like URLs, mentions, symbols and hashtags
- in_reply_to_user_id: user identifier if the tweet is a reply to a specific user
- in_reply_to_status_id: status identifier if the tweet is a reply to a specific status

The data analysis can be stated by breaking the text down into words. Tokenization is one of the most important steps in text analysis. The goal of tokenization is to split a stream of text into smaller units called tokens, usually words or phrases. While this is a well understood problem with several out-of-the-box solutions from popular libraries, Twitter data pose some challenges because of the nature of the language.

This research is based on the methodology by [9], they implemented the methodology of analyzing Twitter to explore perceptions of Asian restaurants. This research adapted the methodology to implement the analysis for tourism objects.

III. METHODOLOGY

The objective of this study is to propose a social mining method to estimate the visitors' percetions toward a tourism object. To this end, tweets of two biggest temples in Indonesia (Borobudur and Prambanan) were collected and then analyzed by counting the specific words that frequently mentioned. This research will go through two steps: data collection and data analysis.

Data collection

Two tourism objects, the Borobudur Temple and the Prambanan Temple, were selected for data collection because of their fame as Indonesian landmark. Data were collected from December 10th, 2017 to January 1st, 2018. The total number of tweets was 620. Table I presents the data collection periods of each dataset and shows the grand total of tweet messages used for this research.

TABLE I. TWEET DATASET

Searching Words	Collection Period	No. of tweets in English
Borobudur	Dec 10th, 2017 – Jan 1st, 2018	448
Prambanan	Dec 10th, 2017 – Jan 1st, 2018	172
(Total)		620

Data analysis

Subsequently, word counts were performed in terms of how many times the distinct words appeared in the twitter. Sentiment analysis (opinion mining or subjectivity analysis) is a method for extracting information related with the subjectivity of an object (positive, negative, or neutral). Text mining transforms an unstructured dataset (natural language messages) into structured data or measureable values [10]. For text mining and word frequency analysis, Python, a well-known software for data mining and analytics, was used to break tweet messages into single words (tokenization) and then to count frequencies of each word. In particular, the word frequency test ranks the most frequently mentioned words in the tweet data.

IV. RESULTS AND DISCUSSIONS

In the first analysis, this study examined perceptions of Indonesian tourism objects in general. The three sub-datasets were compiled into a single dataset to rank frequent words. The text mining and word frequency analysis showed words appearing most frequently in the data set. To better understand major perceptions, certain unnecessary words were deleted: "temple", "the", "…", and "RT" (Table II).

TABLE II. FREQUENCY ANALYSIS

Rank	Word	Frequency
1	Borobudur	333
2	Prambanan	134
3	photo	56
4	#borobudur	44
5	Buddhist	42
6	Indonesia	40
7	Yogyakarta	38
8	#indonesia	33
9	#prambanan	30
10	#travel	29
11	Compounds	29
12	near	26
13	biggest	24
14	video	22
15	Sunrise	22
16	Java	20
17	#art	20
18	visit	19
19	#Borobudur	19
20	Borobudhur	19

The first finding was that Borobudur is more popular than Prambanan. The number of Borobudur word was 333, in addition to some hashtags i.e. #borobudur (44) and #Borobudur (19). However, some people misspelled it with Borobudhur (19). Prambanan was mentioned 134 times, in addition to the hashtag #prambanan (30).

There were also some themes in the overall data: tourism object related themes like photo (56), #travel (29), video (22), and visit (19), and also location such as Indonesia (40), Yogyakarta (38), #indonesia (33), and Java (20). Buddhist was also frequently mentioned (41), given the fact that Borobudus is one of the biggest Buddhist temples in the world. "Biggest" was also among the top words (24).

This paper also compared the frequency of words between Borobudur and Prambanan using text mining and frequency analysis. For this analysis, the common words from each subset will be omitted, thus the difference can be distinguished.

TABLE III. FREQUENCY ANALYSIS (BOROBUDUR DATASET)

Rank	Word	Frequency
1	Sunrise	22
2	Magelang	18
3	Cultural	13
4	Great	12
5	Wonder	10
6	Largest	10
7	#asia	10
8	Simple	10
9	Buddha	9
10	Wonderful	8

TABLE IV. FREQUENCY ANALYSIS (PRAMBANAN DATASET)

Rank	Word	Frequency
1	Hindu	11
2	Awesome	9
3	Borobudur	7
4	Largest	6
5	First	6
6	Complex	6
7	Love	5
8	Sunset	4
9	#unescoheritage	4
10	Faithful	4

The top words of each subset reflected the characteristic of each object. For instance, sunrise was found in Borobudur dataset (first), whereas sunset was found in Prambanan dataset (eight). It indicates that people prefer to enjoy a sunrise at Borobudur and enjoy a sunset at Prambanan.

From the data in Borobudur dataset, the objects characteristics were frequently mentioned: Magelang (the locations) and Buddha (the religion). Some perceptions toward the Borobudur were great (fourth), wonder (fifth), largest (sixth), and wonderful (tenth).

On the other hand, the data from Prambanan dataset reflected different words and perceptions. The top word was Hindu, in accordance with the fact that Prambanan is a Hindu temple. However, Borobudur was also mentioned frequently here (third), indicated that people visiting Prambanan usually also visited Borobudur. Some perceptions toward Prambanan were awesome (second), complex (sixth), love (seventh), and

faithful (tenth). While there were different perceptions with Borobudur, both objects got positive reviews from the tweet.

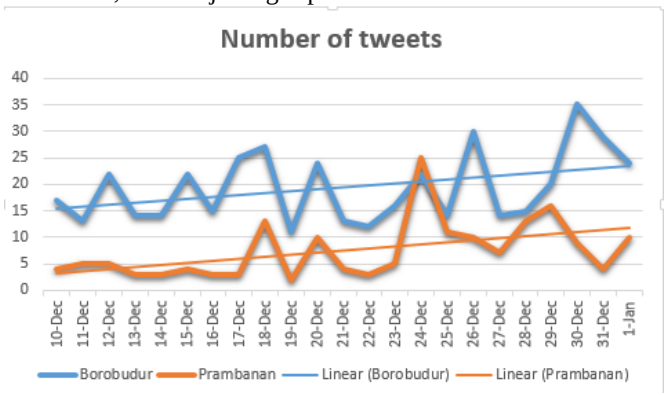

Fig. 1. The number of tweets.

This research also analyzed the frequency of tweets per day. The figure shows the number of tweets from each dataset during the period (Dec 10th – Jan 1st). In general, Borobudur had higher number of tweets except on Dec 24th. Interestingly, the frequency of both datasets had upward trend line and reaching their peak at the end of the year (Fig 1). The holiday season at the end of the year (Christmas and New Year) could be the factor that supports this increment.

V. CONCLUSIONS

Social media bring new opportunities for researching customer opinions. The idea of smart tourism and its implementation using social media has made a huge impact in the industry. This benefits maximize the usage of technology and internet to collect data and analyze it, then it can improve the way of conventional business models work. User-generated content (UGC) is one of the producers of a huge quantity of information on the internet. This can be used as a source of web analytics to gather information or knowledge about tourism that can support decision making process. By examining Twitter messages, this study provides some indications of major perceptions and emotional states for two of tourism objects in Indonesia (Borobudur and Prambanan temple) mentioned in Twitter. The results are noticeable in terms of the improvements in tourist experiences. Positive perceptions were mentioned in both Borobudur and Prambanan dataset, and there were increase in the frequency of tweets at the end of the year.

The theoretical implication from this study is that travelers' perceptions toward tourism objects can be estimated as the basis of studying tourism objects' characteristics. Some practical implications include how decision makers can take action to boost or improve the image of a tourism object. The obstacle of this study was the limited number of the collected tweets. In order to get a bigger dataset, the collection period should be longer, thus the robustness of the study can be more attainable. Future research can also apply some sentiment analysis methods to better interpret the customers' perceptions toward an object.

REFERENCES

[1] E. Marine-Roig and S. A. Clave, "Tourism Analytics with Massive User-Generated Content: A Case Study of Barcelona," *Journal of Destination Marketing & Management*, 2015.

[2] K. Kaur and R. Kaur, "Internet of Things to promote Tourism: An insight into Smart Tourism," *International Journal of Recent Trends in Engineering & Research*, pp. 357-361, 2016.

[3] U. Gretzel, M. Sigala, Z. Xiang and C. Koo, "Smart tourism: foundations and developments," *Electron Markets*, pp. 25: 179-188, 2015.

[4] A. Bruns and J. Burgess, "Researching news discussion on Twitter: new methodologies," *Journalism Studies*, pp. 13:801-804, 2012.

[5] UNESCO, "Borobudur Temple Compounds," 2012. [Online]. Available: http://whc.unesco.org/en/list/592. [Accessed 31 12 2017].

[6] UNESCO, "Prambanan Temple Compounds," 2012. [Online]. Available: http://whc.unesco.org/en/list/642. [Accessed 31 12 2017].

[7] M. A. Russell, Mining the Social Web, 2nd Edition, O'Reilly Media, Inc., 2013.

[8] M. Bonzaini, Mastering Social Media Mining with Python, PACKT, 2016.

[9] S. B. Park, J. Chang and C. M. Ok, "Analyzing Twitter to explore perceptions of Asian restaurants," *Journal of Hospitality and Tousrism Technology*, pp. 7:405-422, 2016.

[10] S. Weiss, N. Indurkhya, T. Zhang and F. Damerau, Text Mining: Predictive Methods for Analyzing Unstructured Information, New York: Springer, 2010.