Comparison of Liveable City of Three Cities in Indonesia through Index of Happiness Data from Social Media and Urban Structure

Munawir[1,2], B J Dewancker[1] and M D Koerniawan[3]

[1] The University of Kitakyushu, Japan
[2] The University of Cokroaminoto Palopo, Indonesia
[3] Bandung Institute of Technology, Indonesia

Abstract. Social media has contributed to providing information from the users. Users can expressing and sharing their thoughts and opinions on all kinds of topics and events freely. Twitter offers organizations fast and effective way to monitor the users' feelings towards their mood in the cities. In twitter there are the big information data that be able to download freely, this study collected data from twitter and make the classification based on users' tweets to categorize the index of happiness of the city. The urban structure data are collected from the urban guideline development from each city. The index of happiness becomes the prosperity level paradigm of a city, it becomes a tool to assess and compare how the city is liveable. This study will analyse from the users' tweets on a regular basis in three cities of Indonesia, to the parameters of happiness that will be compared with the urban structure development. Based on data of the index of happiness from social media can offer better information for city planners and developers that can be used to improve planning and quality of life in the city and future of urban development.

1. Introduction

The Urban and regional analysis involves the use of a wide range of approaches to understand and manage complex sectors, such as transportation, environment, health, housing, the built environment, and the economy. The goals of urban research are many, and include the theoretical understanding of infrastructural, physical and socioeconomic systems; developing approaches to improve urban operations and management; long-range plan making, and impact assessments of urban policy [1].

Indonesians were early adopters of Twitter and are among the most prolific Twitter users. In 2010, the comScore report ranked Indonesia as the country with the highest Twitter penetration, with 20.8% of the Internet-using population visiting Twitter in the month of June. In early 2012 a market research firm put the Indonesian Twitter-using population at 29.4 million, the fifth largest in the world in raw numbers [2].

In this paper, we try to introduce the feelings of public on sentiment analysis into the field of computational sustainability. We collected data about their feelings from the twitter and we use tweets classification based on the feelings of the user that is positive, negative and neutral than we can know it is happy or unhappy. Here we use Twitter to study the fine-grained geography and dynamics of sentiment in the greatest of three cities in Indonesia Bandung, Makassar, and Bandar Lampung Bandar Lampung areas, identifying areas and times of positive and negative sentiment.

This paper we currently use data from twitter, due to its wealth and volume. To understand the demographic information of the city area, we can improve the process of making it cheaper and easier to infer existing demographic information and investigating whether Twitter volume or the volume of

Content from this work may be used under the terms of the Creative Commons Attribution 3.0 licence. Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.
Published under licence by IOP Publishing Ltd

a particular discussion topic, correlates with the level of happiness or quality of life. If so, then we can use social media as a proxy for concrete action, and thus collect that information at a low price and in real time.

In this study, we try to find the condition of the three cities based on data from tweets. We want to analyze that their share social in media and for the future data can be a reference to determine the index happiness city and provide that the city that livable or not

The aim of the research is how to use data from social media about their feelings and expected to be a tool to determine the characteristics of the city. Investigate how geographic place correlates with influences societal levels of happiness. The index of happiness becomes the prosperity level paradigm of a city, it becomes a tool to assess and compare how the city is livable, and so it can be used as a reference for future urban development and sustainable urban planning.

2. Methods

2.1. Data Collection

We used data from Twitter. We selected three cities in Indonesia are Bandung, Makassar, and Bandar Lampung. The three cities are 10 largest cities in Indonesia. Table 1 shows the condition of the three regions of each city.

Table 1. The population of three cities, Bandung, Makassar, and Bandar Lampung City

No.	City	Total Area	Population Census 2010	Coordinate
1	Bandung	167.67 km2 (64.74 sq. mi)	2.394.873	6°54′43″S 107°36′35″E
2	Makassar	199.3 km2 (77.0 sq. mi)	1.338.663	5°8′S 119°25′E
3	Bandar Lampung	167.67 km2 (64.74 sq. mi)	881.801	5°27′0″S 105°16′0″E

2.2. Sentiment Analysis

We use twitter to study of the cities based on geography and dynamics of sentiment in the greater area of the city. The relevant tweets can be filtrated basic on feelings or topic and followed by an assessment of whether the tweets are objective or subjective before identification tweets is positive, negative or neutral. The number of words in each text tweets with positive meaning is then compared to the number of words with a negative meaning to give an overall sentiment score [3, 4]. The framework is shown in figure 1, a typical sentiment analysis model that grants access to historical content and to various tools for searching, analysing and reporting based on data.

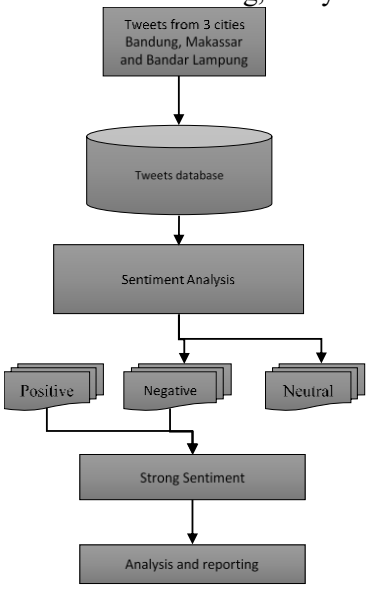

Figure 1. The Frameworks of sentiment analysis model

2.3. Text Mining

In this research, we use Naïve Bayes classification for text classification. Naive Bayes classifier is a simple model for classification. It is simple and works well on text classification. It is a probabilistic classifier based on applying Bayes' theorem with strong independence assumptions.

This is the simplest form of Bayesian Network, in which all attributes are independent given the value of the class variable. This is called conditional independence. It assumes each feature is conditional independent to other features given the class. A Naive Bayes classifier is a technique that applies to a certain class of problems, namely those that phrased as associating an object with a discrete category. We make two classes positive and negative tweets [5]

A conditional probability is a probability that event X will occur, given the evidence. So, our initial formula looks like this:

$$P(sentiment|sentence) = \frac{P(sentiment)P(sentence|sentiment)}{P(sentence)}$$

Process stemming is each tweet was stemmed into the group Indonesian words, a match of each word was searched in the lexicon database, scoring Positive and negative matches were summed to define a score of each tweet and polarity: (P-N)/ (P+N), where P=total sum of positive sentiment words; N=total sum of negative sentiment words.

3. Analysis and Discussion

We collected twitter data from three cities in Indonesia are Bandung, Makassar, and Bandar Lampung during September and October 2015. Table 2 shows the total of tweets from the three cities.

Table 2. The Total of tweets data from three cities during September and October 2015

No.	Location	Total Tweets	Total Users
1	Bandung	544,791	100,417
2	Makassar	222,353	27,802
3	Bandar Lampung	50,456	9,076

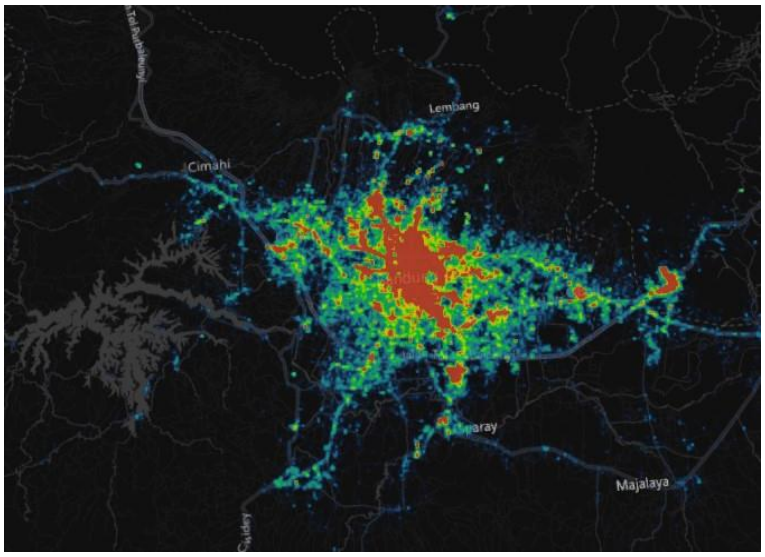

Figure 2. Twitter activity of Bandung city. Each point by geolocated of users. Maps were created using

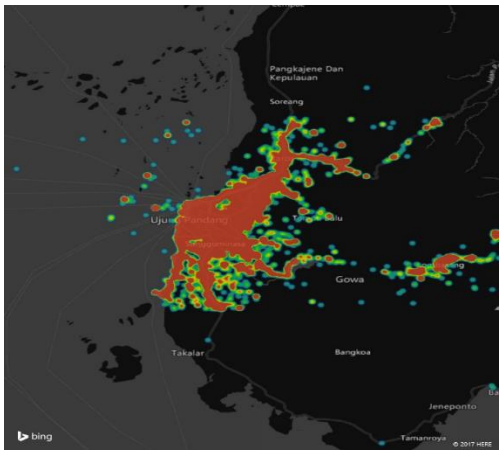

Figure 3. Twitter activity of Makassar city, each point by geolocated of users. Maps were created using Power Map.

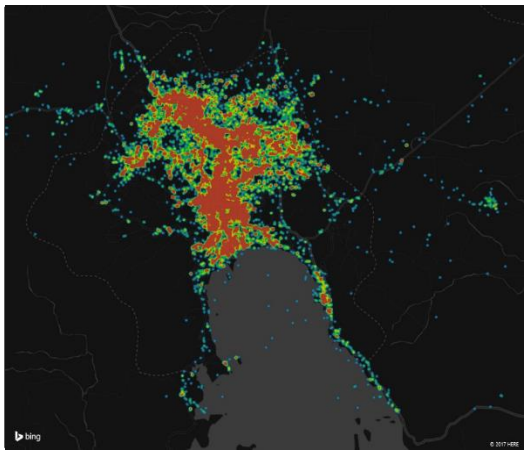

Figure 4. Twitter activity of Bandar Lampung city, each point by geolocated of users. Maps were created using Power Map.

We collected tweets by name of user, text tweets, date, time, longitude, and latitude. From this data we classification which tweets positive or negative based on geolocation. From the figure 2, 3, and 4, we can see the tweet activity reflects population density each cities. Most of the all city very active in social media twitter.

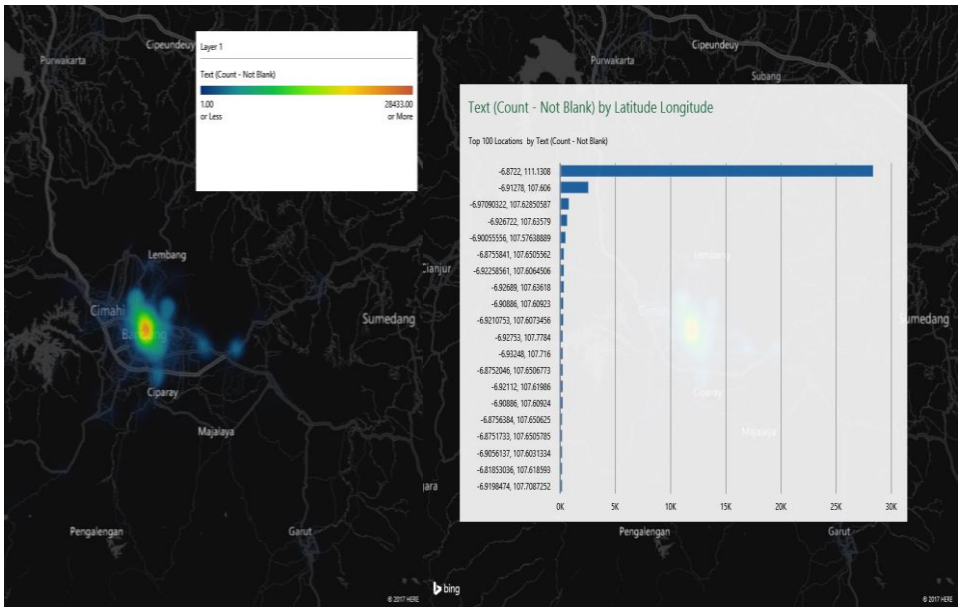

Figure 5. The frequency of tweets by geolocated, Bandung City

We calculated user's position by the number of tweets. In figure 5, 6, and 7 shows there are positions where users often upload their tweets at coordinate points. It is that points we can analysis about their feelings are more concerned their city. We can select any location by looking at the number of frequencies tweets and more detail what is their feelings.

finally, we found the total of sentiment classification from three cities, Bandung, Makassar, and Bandar Lampung. Figure 8, 9, and 10 shows the total of sentiment each city. We can see that from tweets

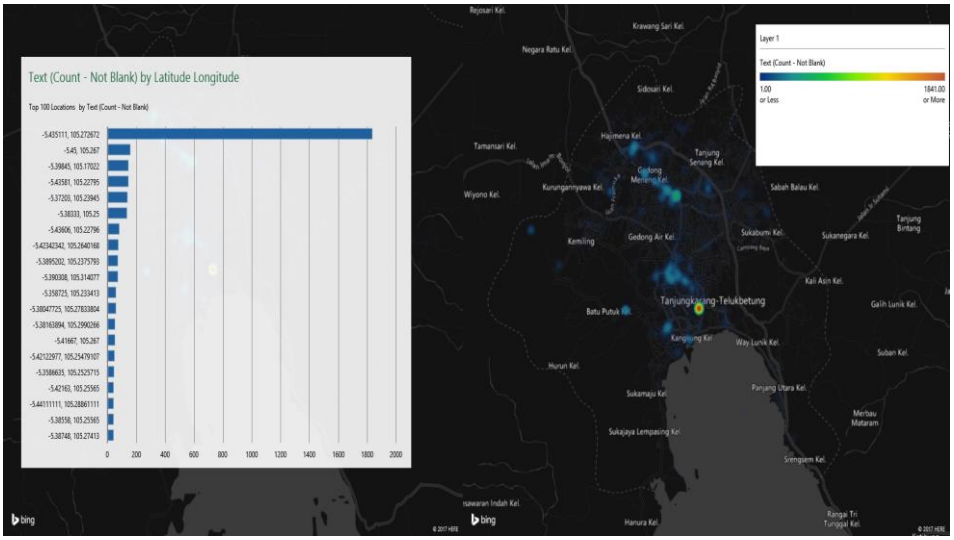

Figure 6. The frequency of tweets by geolocated, Bandar Lampung City

The totally from three cities the percentage of positive tweets more than negative tweets. For example in Makassar city the total of tweets collected 222,353 tweets and the positive tweets is 46391(21%), negative sentiment is 3776(2%) and 77 % is neutral. In this analysis why the neutral sentiment is very high because the system can not calculate half of the tweets the meaning, sometimes the characteristic different write of tweets, users used their local language or symbol.

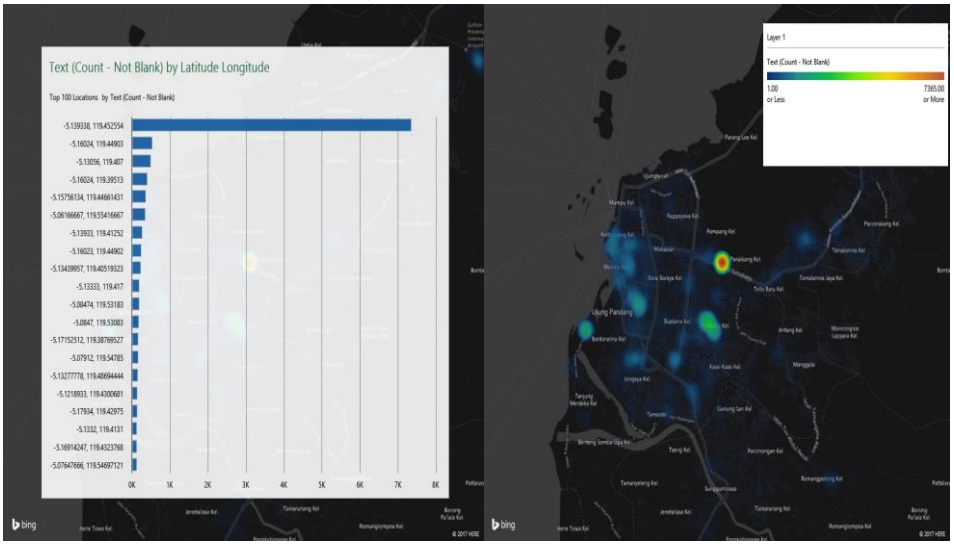

Figure 7. The frequency of tweets by geolocated, Makassar City

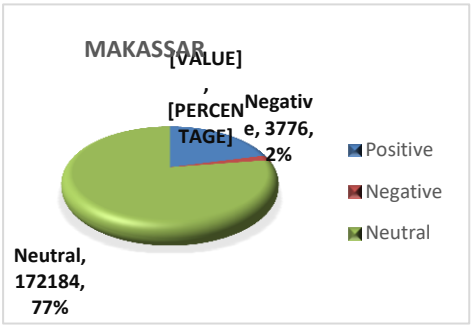

Figure 8. Overall sentiment classification tweets of Makassar.

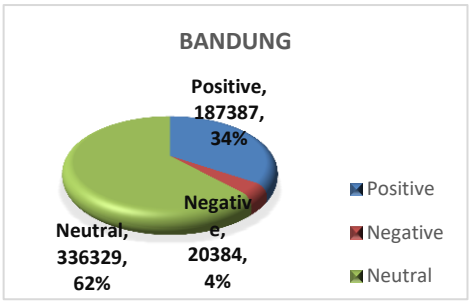

Figure 9. Overall sentiment classification tweets of Bandung

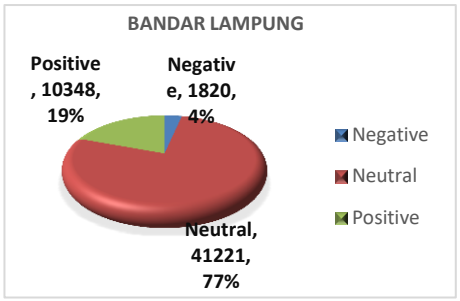

Figure 10. Overall sentiment classification tweets of Bandar Lampung.

4. Conclusions and Future Works

This study is an initial attempt to investigate citizen feelings in the city, determine the index of happiness city based on social media data. Potential of tweets identification their feelings by geolocation on sentiment analysis, as well as offering possible new direction for calculating the well-being of the city and parameter index of happiness city. Using an opinion mining to determine sentiment and impact on the decision-making process can be used and to better information for a city planner.

The report could be effective in generating participation but faced substantial technical, analytical, and communication barriers to influencing decision making. Social media it is very popular in this cities, Bandung, Makassar and Bandar Lampung, that we can see the population of the user using the Twitter is very high. The overall sentiment classification from three cities Bandung, Makassar, and Bandar Lampung shows that positive sentiment more than negative sentiment, it is mean the majority of people happy their life.

For the future work, based on the data we can analyze more detail what the issue make their happy and unhappy related to the city. It is hoped that the data obtained provide better information, understand the people and area of the city are based on the feelings of their mood so that they can accelerate in the city planning process itself. The data can be utilized for the development of sustainable urban and operates efficiently.

Acknowledgments

This study was supported by The University of Cokroaminoto Palopo Indonesia, The University of Kitakyushu, and Directorate General of Higher Education of Indonesia (DIKTI).

References

[1] Thakuriah, P., N. Tilahun and M. Zellner (2015). Big Data and Urban Informatics: Innovations and Challenges to Urban Planning and Knowledge Discovery. In Proc. of NSF Workshop on Big Data and Urban Informatics, pp. 4-3

[2] Kathleen M. Carley, Momin Malik, Mike Kowalchuk Jürgen Pfeffer, Peter Landwehr "Twitter Usage in Indonesia"Center for the Computational Analysis of Social and Organizational Systems.2015

[3] Geetika, V., & Sangharsh, T. (2014). Facebook as a Corpus for Emoticons-Based Sentiment Analysis. International Journal of Emerging Technology and Advanced Engineering, 904-908.

[4] Jasleen, K., & Jatinderkumar, R. S. (2014). On Classifying Sentiments and Mining Opinions. International Journal of Emerging Technologies in Computational and Applied Sciences (IJETCAS), 210-214.

[5] Pablo Gamallo and Marcos Garcia "Citius: A Naive-Bayes Strategy for Sentiment Analysis on English Tweets" Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014), pages 171–175, Dublin, Ireland, August 23-24 2014