

TOURISTS GEOVISUALIZATION ANALYSIS UTILIZING INSTAGRAM DATA IN CENTRAL JAVA PROVINCE AND SPECIAL REGION OF YOGYAKARTA

Achmad Rofii^{1*}, Totok Wahyu Wibowo, Sudaryatno, Nur Mohammad Farda

Department of Geography Information Science, Universitas Gadjah Mada,
Bulaksumur, Sleman, D.I. Yogyakarta 55281, Indonesia
*achmad.rofii@mail.ugm.ac.id

KEY WORDS: Geovisualization, Hexagonal tessellation, Tourists, Big data, Instagram

ABSTRACT:

The growth of Indonesia's national economy in the sector of tourism has been improving constantly. Based on BPS data year 2015, tourism has contributed to PDB with the amount of 4,25%. Tourists is an important aspect because they play a role in the economic change in tourism. Along with social media development, tourists' distribution can be monitored using big data in social media. Instagram is one of the social media which is often used by tourists to share photos or videos while traveling. Issues that appeared when using big data Instagram to produce spatial data are size, variation, and a big time span. This research is aimed to extract big data Instagram to produce basic tourists' spatial data and analyze tourist attractions using geovisualization hexagonal tessellation in West Java Province and Special Region of Yogyakarta. Generally, the methods used are divided into three which are data mining, graphic visual analysis and geovisualization classification hexagonal tessellation. The Instagram data extraction process uses a web-based software called Netlytic, while pre-processing data to produce tourists' spatial data uses QGIS software. Hexagonal tessellation is tested by graphic visual to ensure the most effective measurement in the scale of 1: 1.250.000. The most effective hexagonal tessellation measurement is used to classify tourist attractions. Based on the results, it shows that the most effective measurement of hexagonal tessellation is 2 Km. The most popular tourist attraction is Malioboro Yogyakarta based on hexagonal tessellation geovisualization analysis.

1. INTRODUCTION

Indonesia is a country with ample natural resources. The advantage of this is sector tourism. Gross Domestic Product (GDP) of tourism is 4,25% on 2015 (BPS, 2018). The highest GDP income is in the province of Java. The high amount of GDP in Java is influenced by the completeness of infrastructure. Currently, the calculation of GDP uses data statistic in the form of tourists' census in the places of tourist attractions. The tourist data in Java is currently uses field survey (Barudin, et al, 2016). This method requires more time, energy and cost if applied in a region which has a wide coverage. Results of survey data is data statistic and does not have any spatial information. The extraction of tourists data uses big data Instagram is one of the most effective and efficient ways to know the spatial pattern of tourists.

The tendency to share photos when on holiday in Indonesia is plenteous. Photo distribution through mobile software allows the data to be collected in the social media server as big data. Big data structure especially social media has a variation of data format such as text, time, location, photo, recording and video. Data variation with huge amount becomes a challenge in data geovisualization to understand spatial pattern of big data (Robinson, et al., 2017). Understanding spatial pattern is necessary to help in making a decision (Perumal, et al., 2015). Big data is a collection of data with a huge amount and centered in one server with various format. People create 2,3 million Terabytes of data each day, more than 90% of data made in two years worldwide (Makhabel, at el., 2017).

Data mining is information mining by using a collection of data with a huge amount or big data. There are 7 steps of data mining which are data cleaning, data integration, data selection, data transformation, data mining, pattern evaluation and knowledge

presentation (Han, et al., 2012). Data cleaning is a process of cleaning out the data that are considered as a disruption and inconsistent. Data integration is to compile all the data to prevent redundancy. Data selection is choosing a collection of data which will be used to analyze. Data transformation is to change the format according to the standard of analysis from those that is considered as unstructured. Data mining is a process of mining the information in data warehouse, according to the filter that has been created previously. Pattern evaluation is to re-evaluate the system that has been made, in case the filter is considered as not good hence the algorithm filter will be changed. Knowledge presentation is the final step of data mining which has been analyzed to produce information and used in making a decision.

The growth of internet usage in Indonesia has been increasing every year. In the year 2017, internet usage reached a total of 143,26 million people (APJII, 2017). Based on the total of internet usage, 58,08% user are located in Java (APJII, 2016). Device that are mostly used to access the internet are mobile with 44,16% (APJII, 2016). Whereas the service used to access the internet are social media with a total of 87,13%. In a survey carried out in 2016, Instagram has a total of 15% or 19,9 million user, next is Facebook with 54% 71,6 million user (KataData, 2018). Instagram is one of big data in social media which is designed to share photos and videos. During the beginning, Instagram is designed for photographers with effects and filters included (Boy & Uitermark, 2017).

Instagram provides API (Application Programming Interface) for the developers of software therefore they are able to obtain the data which contains content, profile and spatial information. Extraction of spatial information using *big data* social media can be carried out with various ways, one of them is extraction of coordinate system information. Coordinate system which is

generally used in *big data* social media is geography system coordinate *World Geodetic System* (WGS 1984) (Bivand, et al., 2008). References of the Earth's shape is projection system which is often used worldwide because it is flexible in parts of the world.

Obtaining spatial information using big data social media especially Instagram can be carried out by using the approach of data mining. Instagram provides API access which allows the developers in obtaining the data. Data structure in Instagram contains spatial information in photos which are shared. Those data can be extracted due its crucial part in spatial analysis especially tourists. Instagram data extraction for tourists is not an easy case because of the various data structure and huge amount of data. Monitoring of tourists distribution is important because it concerns with the management of tourist attractions especially with huge area. (Vu, et al., 2013). The search of big data for tourists can be done with the approach of data mining by using the key word “#wisata”. Hashtag tour (wisata) at 08.22 WIB on 6th October 2018 has a total of 1,828,663 posts (Instagram, 2018). The amount kept increasing with the amount of Instagram account in Indonesia also kept increasing. Therefore, analysis spatial pattern of tourists can be explored using the approach of big data and searching for data with hashtag *wisata*,

In the last few years, the role of social media in the development of tourist attraction in Central Java Province and Special Region of Yogyakarta is massive. Many tourist attractions which has been promoted through social media and grow rapidly because there are many tourists that visit those attractions. Instagram is one of the social media which is often used by tourists in sharing object photos. Role of social media Instagram in tourists' trip is pretty significant (Harun & Syam, 2018). Extraction of big data Instagram can be made use to know which tourists object is the most popular according to the amount of tourists that visit and share their photos through social media Instagram. As a result, that method is possible in exploring the potential of tourist attractions through social media especially Instagram.

The result of extraction of spatial information through social media Instagram is in the form of coordinate point data. Data point visualization in a huge amount makes the understanding of the spatial pattern more difficult. Data collection with huge amount in geovisualization process has to undergo generalization which is map with different scale has to have different information (Kraak & Ormeling, 2007). Hexagonal tessellation is one of the forms of geometric iteration to present the data points that are overlapping with generalization according to the scale. With hexagonal tessellation, territory that has many near-by points will appear more clearly. However, hexagonal tessellation is a discreet visual tessellation. Thus, various tessellation measurements will directly affect the geovisualization of tourists. Hexagonal tessellation has not yet been explored despite it can be a better method in geovisualization (Birch, et al., 2007). Geovisualization hexagonal tessellation is best for visualizing earthquake data in Indonesia by interpolating earthquake data specifically the smaller scale (Dharmawan, et al., 2017). Whereas hexagonal tessellation for tourists geovisualization in Indonesia are rarely carried out.

Geovisualization is a combination of two words which are “geo” and “visualization”. The word “geo” is taken from geography which is a study of Earth from the core to atmosphere as well as the phenomenon included which concerns

with physic and non-physic though the perspective of agihan, integration and correlation. The word “visualization” means a visual representation which is supported by computer devices as interactively to strengthen cognition and aimed to ease the understanding and analysing of abstract data (Kerren, et al., 2007).

Geovisualization is graphic software designed to facilitate the understanding of spatial pattern in terms of condition, process, concept and activities happening in human world (Dodge, et al., 2008). Geovisualization develops rapidly along with the development of technology hence visualization media are more various and interactive. A good geovisualization is not just about representing actual and real data. Nevertheless also to be understood by all readers of maps. Dodge at al, (2008) divides geovisualization into three based on the ability of looking, querying and questioning. Looking is a map which is represented graphically, thematic and chart based on the geography system coordinate. This map only able to describe a geographic appearance based on the condition in the field. Querying is a map designed to not just picture the geographic phenomena but also provides interface for the search. Questioning is a map designed to answer the questions of what and why. The map in this step can be used for spatial analysis until spatial modelling.

2. STUDY AREA AND DATASET

Areas that are used for this research are Central Java Province and Special Region of Yogyakarta (Figure 2.1). Special Region of Yogyakarta area has an area of 3.185,8 Km², whereas Central Java Province has an area of 32.544,12 Km². The areas are chosen because of their completeness in infrastructure that supports the building of big data Instagram. Therefore, the data collected in server Instagram are ample and various. Other than that, the amount of tourists in Central Java Province and Special Region of Yogyakarta is considered huge compared to other provinces.

The amount of tourists in Special Region of Yogyakarta in the year 2016 is 20.933.798 people. Meanwhile, in Central Java Province in the year 2016 is 36.899.776 people. Tourist attractions in Central Java and Special Region of Yogyakarta has various types starting from culture, nature and education. Special Region of Yogyakarta is known for its tourism and education while Central Java Province has a wide area compared to Special Region of Yogyakarta. However, tourist attractions in Central Java Province has only a few that are known for tourism activities.



Figure 2.1 Map of research area

Materials used in this research are data from the hashtag wisata in Instagram and tourist attraction statistic in Central Java Province and Special Region of Yogyakarta in the year 2016 as well as administration border data of Central Java Province and Special Region of Yogyakarta. The main material is the data of hashtag wisata obtained from the access of Instagram API. The process of obtaining the data uses a web-based software, Netlytic. This software facilitates obtaining Instagram data with a maximal of 5 dataset. Each dataset has a maximal of 10000 posts. Data format obtained are in the form of table. Administration border data are obtained from geoportal BIG with data format shapefile. Meanwhile, data of tourist attractions obtained from Tourism Departement of Central Java Province and Special Region of Yogyakarta.

3. METHODOLOGY

Big data is a data base which is collected in a huge amount. Knowledge extraction and spatial pattern are useful for the interests of research, especially geovisualization. The tendency of tourists to share their photos to social media Instagram triggers the development of big data for tourism. Big data geovisualization of tourists can be used in the policy of tourism management. The general method concerns with the research process and analysis of data spatial specially tourism. This research focuses on geovisualization of tourists using big data in social media Instagram. Basic geovisualization big data uses hexagonal tessellation with the best measurement based on the data test. The method of testing uses visual graphic analysis to know the comparison of the amount of tourists using big data Instagram extraction along with tourist statistic data. Device used in this research are computer, QGIS Bonna software, a web-based software Netlytic and Google My Map. Further explanation about research method are as follows:

3.1 Data Mining

Data mining is acquisition of knowledge using analysis of big data. In this research, Netlytic software is used to obtain the data in social media. Netlytic software are able to mine the data automatically with the duration of each day. As a result, the data can be obtained each time with the terms where before all the data are obtained, it needed to be uploaded. Time duration of mining starts from 28th August 2018 until 11th December 2018. Table of data results extraction big data Instagram is selected between data that are coordinated and those that do not. This research uses point-based analysis hence the data must obtain geospatial references. Selection process uses ‘select by attribute’ which is available on QGIS attribute table. Tool ‘select by attribute’ in QGIS software uses a simple logic expression. Data that does not have any information on geography coordinate are chosen using ‘select by attribute’ expression and then erased. Next, the data that will be extracted contains geography coordinate information for data cleaning process.

Data mining is acquisition of knowledge using analysis of big data. In this research, Netlytic software is used to obtain the data in social media. Netlytic software are able to mine the data automatically with the duration of each day. As a result, the data can be obtained each time with the terms where before all the data are obtained, it needed to be uploaded. Time duration of mining starts from 28th August 2018 until 11th December 2018. Table of data results extraction big data Instagram is selected between data that are coordinated and those that do not. This research uses point-based analysis hence the data must obtain

geospatial references. Selection process uses ‘select by attribute’ which is available on QGIS attribute table. Tool ‘select by attribute’ in QGIS software uses a simple logic expression. Data that does not have any information on geography coordinate are chosen using ‘select by attribute’ expression and then erased. Next, the data that will be extracted contains geography coordinate information for data cleaning process.

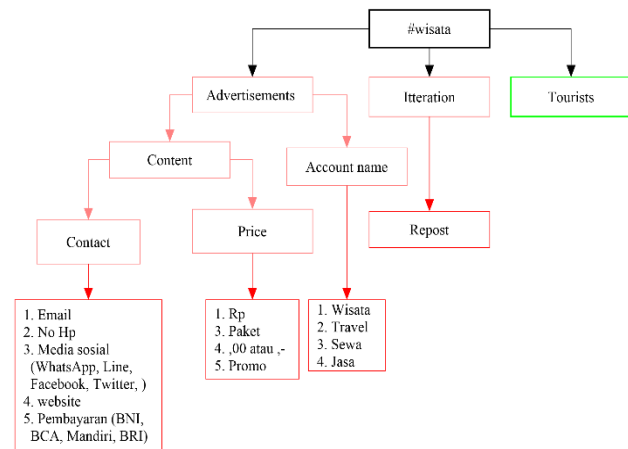


Figure 3.1 Data cleaning pattern using text

Transformation is a process of change for data adjustment to make it a standard data to be further analyzed. In this research, there are two transformations which are result of data mining Instagram which has went through the process of data cleaning in data format of non-spatial hence the format is changed to produce spatial data. This is carried out by defining the coordinates of latitude and longitude. Instagram data uses global projection system which is WGS 1984. Tourist attractions data does not have any geographic coordinate system and in the form of statistic book. Moreover, there is a need to convert tourist attractions using web-based software Google My Map. This software able to search geographic coordinate system by algorithm with the approach of keywords. Dataset is downloaded in the format kmz or kml to be processed in visual graphic using QGIS software.

3.2 Hexagonal Tessellation Geovisualization

Making of tessellation cell using QGIS software based on administration data of Central Java and Special Region of Yogyakarta. Tessellation cell made is hexagonal using ‘create grid’ tool. In this analysis, tessellation cells are made in three sizes with the scale of 1:1.250.000 which are 1 km, 1,25 km and 2 km. Tessellation cells size of 1,25 km is obtained from the rule of Waldo Tobler (1987) who divided the scale denominator to 1000 and the result obtained with one meter (Baja, 2012). This research uses 1:1.250.000 scale, so 1.250.000 is divided by 1000 is equals to the size of the cell which is 1.25 km. Size of tessellation cell 1 km is assumed to be below the average area of tourist attractions which are less than 1 km. Meanwhile, size of tessellation cell 2 km refers to the rule of Shelton et al (2014) which states that the size of hexagonal tessellation cell in national scale is 65 km whereas for urban areas is 2 km (Shelton, et al , 2014).

The research area based on the theory of Taylor Shelton (2014) is more close to urban areas than national. Henceforth, the size of hexagonal tessellation cell uses 2 km. Distribution of data points resulted from the process of data Instagram

transformation as an input for estimation calculation of each hexagonal tessellation cell. Data calculation process uses QGIS software with ‘count point in polygon’ tool. Calculation for the amount of data points in each cell affects the data visualization. Data visualization will be different if the data value of each cell is different which also concerns with the information obtained. Result of tourist data statistic transformation is also estimated using hexagonal tessellation to analyze visual graphic.

3.3 Visual Graphic Analysis

Visual graphic is a comparison graphic between the result of extraction big data Instagram and tourist statistic data. The comparison is based on each amount of tourist attribute on tessellation cell between extraction of Instagram data and tourist statistic data. Statistic data has a time range of monthly hence Instagram data is matched according to the month duration with a similar range. Based on the visual graphic, information obtained shows the average value of tourist statistic data (\bar{x}) and average value of Instagram data (and average value of Instagram data (\bar{y})). Therefore, based on both value, tourist attraction centroid data with coordinate (\bar{x}, \bar{y}) is obtained. This point is used as a base of data division to produce four quadrants. Figure 3.2 presents the division of characteristic quadrant in each tessellation cell based on the centroid data.

Quadrant I shows the most popular tourist attractions based on the Instagram and tourist statistic data. Quadrant II shows the most popular tourist attraction based on Instagram data and least popular based on tourist statistic data. Quadrant III shows the least popular tourist attraction based on Instagram and tourist statistic data. Quadrant IV shows the most popular tourist attraction based on tourist statistic data but the least popular ones based on Instagram data. As a result, through visual graphic, the characteristics of tourist attractions distributions in Central Java Province and Special Region of Yogyakarta can be known.

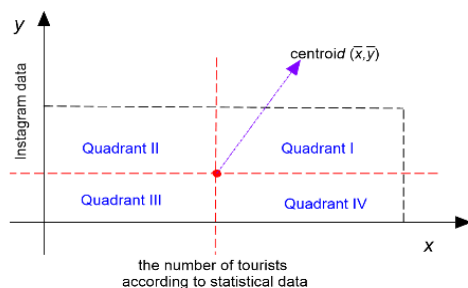


Figure 3.2 visual quadrant graphic

Hexagonal tessellation geovisualization uses big data Instagram for tourist requires testing of tessellation size to produce a more effective geovisualization. One of the ways to test the size of tessellation is using visual graphic, which is a comparison between tessellation Instagram data and tourist statistic data. Visual graphic has centroid data which can be used as a reference for division between data into 4 quadrants. Characteristic of each quadrant can be seen in table 3.1. The key in interpreting visual graphic is in quadrant I and III. Quadrant I is the most popular tourist attraction based on Instagram data and statistic data whereas quadrant III is the least popular based on Instagram data and statistic data. Thus, it can be assumed that the most effective tessellation size has a huge amount in quadrant I. This condition shows that the data can be represented well enough, which means the condition of statistic data is considered as more accurate and can be held responsible as well as represented by data social media Instagram.

Table 3.1 Visual graphic quadrant characteristic

No	Quadrant	Karakteristik
1	Quadrant I	Popular tourist attraction according to Instagram and statistic
2	Quadrant II	Popular tourist attraction according to Instagram and not popular according to statistic
3	Quadrant III	Not popular according to Instagram and statistic
4	Quadrant IV	Popular according to statistic and not popular according to Instagram

4. RESULT AND DISCUSSIONS

Result of Instagram data extraction that has been collected using Netlytic has an amount of 28 dataset with total of raw data with 277.638 post. Each dataset has a maximal of 10.000 post hence the total data maximum is 280.000 post. However, due to the data obtained has a gap of 2.362 post because on 11th December 2018, Instagram changed its policy to close API access. Result of raw data extraction shows there are 12 column data which are id, guid, link, medialink, title, like, count, filter, code, author, pubdate and coordinate. Id column has data identity which is in the numerical value from 1 to 10.000. Calculation of these numbers happens for every dataset thus the total of all dataset will have the same id value. In consequence, data analysis requires author as the identity of each data post. Author column is the information that shows the identity of the owner who shared the data. Pubdate column shows the information of date and time which can be used in Instagram data analysis.

Result of big data extraction through Netlytic software is not able to use directly for this research. This data has to undergo pre-processing process which is filtering or data filtering. Data filtering uses the approach of keywords text. Each post data has their own characteristics especially in texts structure. This information can be obtained in the description column which contains the description of the social media owner. Moreover, data filtering can use author column. In this research, data filtering which will be carried out are accounts who do iteration post and advertisements. The characteristic of accounts who do iteration post is identified via the keywords ‘Repost’. Repost accounts need to be filtered because it will disrupt the parameter temporal data. The characteristic of advertisements can be seen from several aspects which are account name, price and contact. The account name consist of several terms about tourism. The account name can be extracted in author column by using the keywords. This research uses the keywords ‘wisata’, ‘travel’, ‘sewa’ and ‘jasa’. An advertisement containing the price information of service cost for the service user. Filter in advertisement account data uses description column which contains price information such as service cost and promotions. The keywords in filtering are ‘Rp’, ‘,00’, ‘,-’ and ‘promo’. An advertisement also contains contact information so that the service user can easily contact if he/she is interested in the services offered. They keywords used in filtering contact information are ‘email’, ‘No. Hp’, ‘Whatsapp’, ‘Facebook’, ‘Twitter’, www., ‘BNI’, ‘BCA’, ‘Mandiri’ and ‘BRI’. They keywords chosen are due to people’s tendency to make an advertisement via social media Instagram. The total of data after going through selection and filtering is 12.616 post.

Important part in geovisualization analysis using big data Instagram is location defining. Some data obtained from big

data Instagram already contains information in the form of coordinate with projection system EPSG:4326. Not all data contains coordinate information. Coordinate data obtained is 56.525 post or 20,35% of total data. Few coordinate information percentage is due to several factors which are smartphone, internet connection and user behavior. Smartphone facilities in the process of uploading a data affects the availability of position information because not all smartphone has a good GPS. Smartphone needs internet connection to make coordinate information in photos or videos.

A few smartphone tends to be more light when the GPS is not turned on therefore some people does not use GPS to share data in social media. The last factor is user behavior who does not use location in sharing data to social media with several specific reasons. Based on seconder data, before it was transformed into tourist attractions, in Central Java Province has 579 attractions and Special Region of Yogyakarta has 226 tourist attractions thus the total of data is 805 attractions. Central Java Province has more attractions compared to Special Region of Yogyakarta because the administration area is wider. Result of data transformation of 524 tourist attractions or 65,17% of total data. This is based on Google My Map data which are not able to provide all spatial tourist attractions data. The building of Google My Map data uses the people participation hence the attractions which are not popular or recently built is not recorded in Google My Map.

Result of data extraction which has undergone selection process and filtering is still in the form of table. Therefore, these data has to go through transformation process to produce tourist spatial data based on big data Instagram. Figure 4.2 shows the distribution map of tourist attractions based on transformation results using Google My Map. The distribution comparison is seen from its patterns. Tourist attractions distribution tends to be equal compared to distribution made by big data Instagram. Tourist attractions are built based on the administration pattern of a region, which is different in tourists' data from big data Instagram which tends to be based on the more interesting area. Tourists' distribution visually is more on cities areas such as Yogyakarta and Semarang. This is due to the three aspects of citizens, infrastructure and tourist attractions. Citizens can be assumed that the more data is found in one location, then the more people is living in that area and has a higher chance in dividing the data in Instagram.

Infrastructure can be assumed that the more complete it is, the more data points can be found such as internet connection or a good access. This problems are shown in figure 4.1 which presents tourist attractions has a uniform pattern in south beach and Central Java. Meanwhile, in figure 4.2 is seen to not be uniformed or focused in certain points. The building of tower in side beach is usually avoided by service providers of internet connection because of the chances of corrosion. Side of the beach has water vapor which contains salt that allows corrosion to be quicker. Corrosion can cause damage in parts of the tower components. Therefore, there are many beach areas which has minimal internet connection especially in areas which are further away from urban areas. The last aspect is tourist attractions. Tourist attractions makes the tourists attracted to visit these places. Visual appearances of three aspects can be seen in figure 4.3. Infrastructure and population is associated with city that has a huge amount of population and a complete infrastructure. This phenomenon is seen in figure 4.3 part a, which shows the buildup of tourists from big data Instagram focusing on Yogyakarta City, Surakarta City and Semarang City.

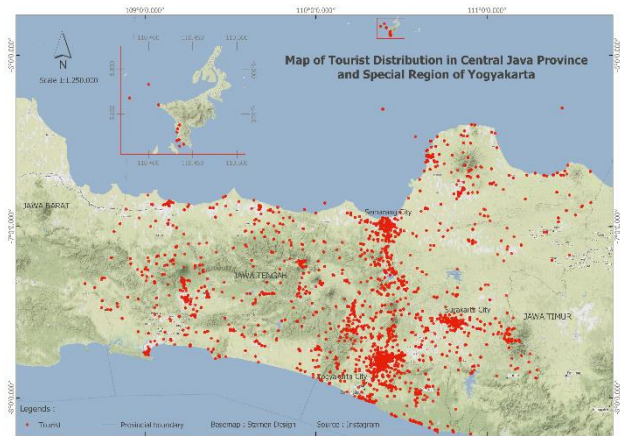


Figure 4.1 Map of tourist distribution using big data instagram

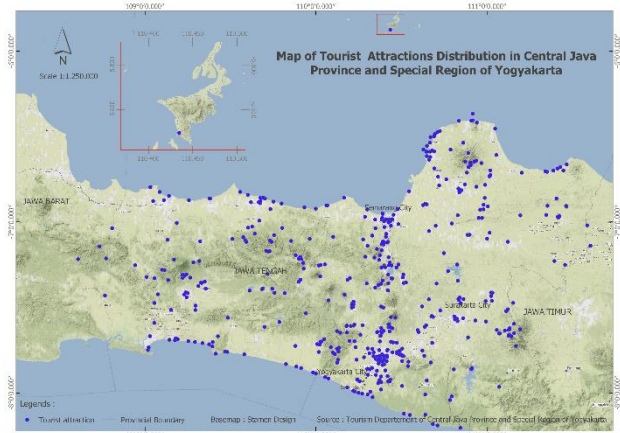


Figure 4.2 Map of tourist attractions distribution

The characteristic of Yogyakarta has a good infrastructure and known to the populated areas, as well as supported with the presence of the best university. Semarang is known for its industry and culture as well as a good infrastructure. Whereas Surakarta is known for its culture and a good infrastructure. Figure 4.3 part b, shows tourists distribution from big data Instagram which focuses on areas that has many distribution of tourist attractions such as in Yogyakarta and Semarang.

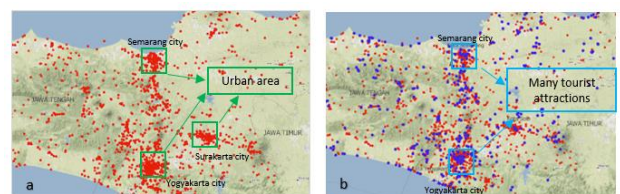


Figure 4.3. (a) Comparison of tourist distribution from big data Instagram with urban area; (b) Comparison of tourist distribution from big data Instagram with tourist attractions

The issues in geovisualizing points using big data becomes a recurring issues together with the development of big data which is constantly increasing like big data Instagram. Hexagonal tessellation geovisualization becomes one of the solutions in geovisualizing big data Instagram to find out the tourists distribution in an area. However, hexagonal tessellation geovisualization is discreet hence tessellation size on specific scale becomes the main focus. Figure 4.4 shows the difference in size of hexagonal tessellation which makes a difference in calculation of tourists in each cell. In the red circle, there is an equal amount of tourists but is in different cells. 1km size has 3

cells with value formation of 2, 1 and 1. 1,25km size has 3 cells with value formation of 1, 1 and 1. 2km size has 3 cells with value formation of 3, 2 and 0. Formation difference in value of each cell will affect data visualization. When using colour gradation, the higher value will has a bold colour. This proves the importance on paying attention to tessellation size in hexagonal tessellation geovisualization using big data Instagram.

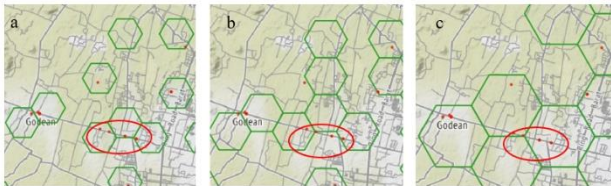


Figure 4.4 Tessellation size comparison (a) 1Km; (b) 1,25Km; and (c) 2Km of tourist distribution

Based on the comparison of tourists data results from extraction of big data Instagram with tourists data from statistic data is seen in visual graphic like figure 4.5. This graphic shows the distribution of tourist attractions which are symbolized with blue-coloured points. Three appearances of visual graphic shows the distribution pattern which are almost alike. This condition shows the data range in all of the tourist data from big data Instagram and statistic data.

Tourist data from big data Instagram has the value of hundreds, whereas statistic data has the value of thousands. Small value of data from each tessellation of big data Instagram is caused by tourists who does not always upload data to social media Instagram. Size determination of tessellation in the most effective way with map scale of 1:1.250.000 cannot be carried out visually because the three of them are almost alike. Therefore, calculation is done mathematically with the approach of centroid data which is used as a reference for data quadrant division.

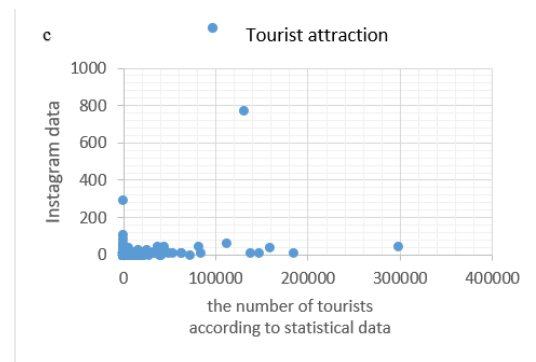
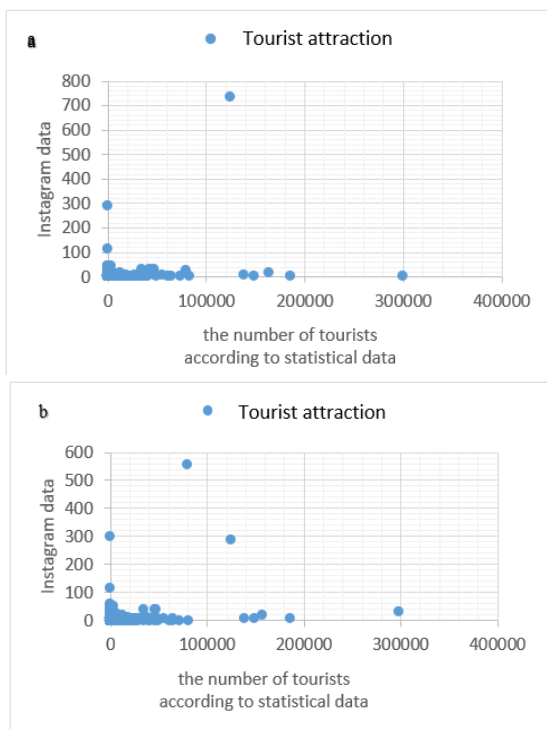


Figure 4.5 Visual graphic with tessellation size (a) 1 Km; (b) 1,25 Km; and (c) 2Km

Mathematical calculation of the visual graphic result produce different value of each tessellation size. Figure 4.3 shows the result of mathematic calculations of visual graphic of each tessellation size. The amount of tourist attractions in quadrant I class has a maximum of tessellation size of 2km with the value of 48 attractions. The amount of tourist attractions in quadrant II has a maximum of tessellation size of 2km with the value of 26 attractions. The amount of tourist attractions in quadrant III has a maximum of tessellation size of 1km with the value of 73 attractions. Consequently, the amount of tourist attractions in quadrant IV has a maximum of tessellation size of 1,25 Km with the value of 66 attractions. The total amount of tourist attractions has different values and not linear with the increase of tessellation size. This difference is caused by tessellation shift due to the difference of tessellation size.

Table 4.3 Result of mathematical calculation of tessellation

No	Type of Calculation	Tessellation size		
		1 Km	1,25 Km	2 Km
1	Centroid data (x,y)	(1885, 3)	(2139, 3)	(2665, 3)
2	Quadrant I Class	40	42	48
3	Quadrant II Class	19	20	26
4	Quadrant III Class	73	65	70
5	Quadrant IV Class	65	66	60

Determination of effective tessellation size is based on the assumption where Instagram data is able to represent tourist statistic data. On visual graphic has to show the high statistic data so that Instagram data is also high. Meaning that quadrant I has the most sum of total compared to all of the tessellation cell size. Based on table 4.3 can be seen on its quantity that the highest value from quadrant I is 48 with the tessellation size of 2 Km. This data shows that the most effective tessellation size for map scale of 1:1.250.000 is 2 Km. Hexagonal tessellation size 2Km is a theory of Shalton et al (2014) which states that the national scale uses 65 km and urban areas of 2 km. However, Shelton research is located in USA where it is not relevant to be applied in Indonesia. This issue is caused by the difference of area and characteristics of territory. In this research, the theory of Shelton et al (2014) regarding hexagonal tessellation size of 2 km for urban area is relevant to be applied in Indonesia especially Central Java Province and Special Region of Yogyakarta on the scale of 1:1.250.000.

Effective tessellation size is used to classify tessellation with visual graphic. Figure 4.5 presents the result of tourist attraction classification using visual graphic. Result of classification shows that the distribution of tourist attraction distribution is classified under quadrant III with white color symbol. In quadrant I, it shows that they are distributed in urban area and only few distributed in non-urban area such as hills, beaches

and mountains. An example of tourist attraction included in quadrant I and located in the hills is Tawang Mangu Karangayar. This place offers a beautiful waterfall with the combination of pine forest and a swimming pool with the water supply from the mountains. An example of tourist attraction located in the beach is Parangtritis Bantul Beach offers the beauty of south beach in Java and has sand in the color of black. An example of tourist attraction located in the mountains is Dieng. This place offers a scenic view of nature in the mountains and various local cultures. Karimun Jawa is a tourist attraction located further away from settlement and in a small island at the north of Java. This tourist attraction is included in the classification of quadrant I which is popular in statistic data and Instagram. Tourist attractions located further from settlement but classified in quadrant I cannot be compared with other attractions. This shows that Karimun Jawa as an attraction, has a strong potential to attract tourists. Karimun Jawa is known for its mesmerizing coast view especially under the sea.

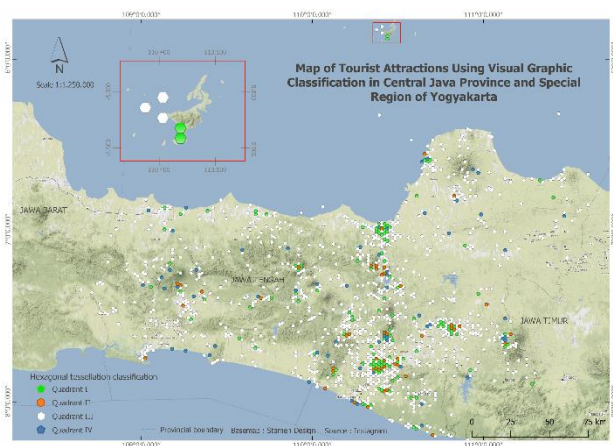


Figure 4.5 Map of tourist attractions using graphic classification

Hexagonal tessellation geovisualization uses visual graphic classification in the form of area. All of the tourist attractions are represented in each tessellation cell despite it being many. Henceforth, in justifying tourist attraction, it needs to be supported with statistic data of tourist attractions. Figure 4.5 shows 10 most popular tourist attractions justification using tourist attractions statistic data. The main source of data uses big data Instagram and only quadrant I is included in the decision of the most popular tourist attractions. Quadrant I is chosen because of its ability to represent the field condition.

Whereas for the name of tourist attractions, will be using tourist attractions data from Tourism Board in Central Java Province and Special Region of Yogyakarta. Study of extraction tourist attractions information can be obtained from Instagram, without statistic data. However, this needs a further comprehensive study and this research focuses on geovisualization.

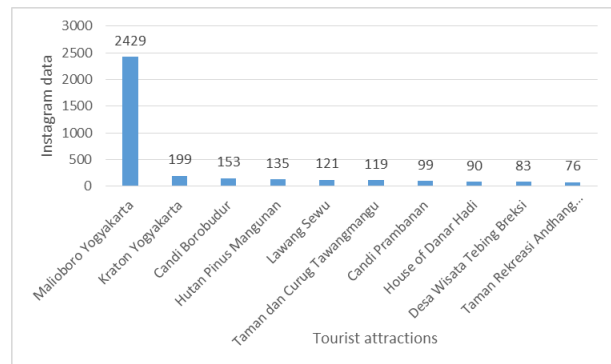


Figure 4.5 Graphic popular tourist attractions

Justification of the most popular tourist attractions with information of each tessellation cell, uses the local knowledge approach. Based on the analysis of the most popular tourist attraction using Instagram data, Malioboro is the most popular attraction. In the span of 3 month, Malioboro has been visited by 2,429 tourists, second is Kraton Yogyakarta with 199 tourists and third, Borobudur Temple with 153 tourists. In figure 4.5, it shows some tourist attractions recently built in the last few years is classified into to 10 most popular. These tourist attractions are Pine Forest Mangunan and Tebing Breksi. Pine Forest Mangunan is a tourist attractions with nuance of nature and the view of pine forests. Tebing Breksi is an area of breksi stone designed beautifully for tourists to take a photo.

5. CONCLUSION

Tourists data extraction using hashtag “wisata” in the span of 3 months, obtained a total of 277.638 posts. However, after going through pre-processing, the remaining is only 56.525 posts. This shows the variety of data in Central Java Province and Special Region of Yogyakarta. Knowledge extraction using big data Instagram for tourists can be applied using geovisualization. However, research for extraction quantitative of tourists needs a comprehensive research because the range data of tourists between Instagram and tourists statistic is huge. Result of tourists’ analysis using hexagonal tessellation using 2 Km is the most effective tessellation size in the scale of 1:1.250.000. Based on the density value, Malioboro Yogyakarta is the most popular tourist attraction with a total of 2,429 tourists

ACKNOWLEDGEMENTS

This research is supported by Universitas Gadjah Mada. The author would like to express his gratitude to an organization which has supported this research. The author would also like to thank all of the team members who has worked together in completing this research.

REFERENCES

- APJII. (2016). *Penetrasi dan Perilaku Pengguna Internet Indonesia 2016*. Jakarta: Teknopreneur.
- APJII. (2017). *Penetrasi dan Perilaku Pengguna Internet Indonesia 2017*. Jakarta: Teknopreneur.
- Baja, S. (2012). *Tata Guna Lahan dalam Pengembangan Wilayah*. Yogyakarta: Penerbit ANDI.

- Barudin, Fitriyani, I. A., & Indriati, D. (2016). *Statistik Profil Wisatawan Nusantara Tahun 2016*. Jakarta Pusat: BPS.
- Birch, C. P., Oom, S. P., & Beecham, J. A. (2007). Rectangular and hexagonal grids used for observation, experiment and simulation in ecology. *Elsevier*, 347-359.
- Bivand, R. S., Pebesma, E. J., & Gomez-Rubio, V. (2008). *Applied Spatial Data Analysis with R*. New York: Springer Science+Business Media.
- Boy, J. D., & Uitermark, J. (2017). Reassembling the city through Instagram. *Transactions of the Institute of British Geographers*, 612-624.
- BPS. (2018, Oktober 6). *Pariwisata*. Retrieved from Badan Pusat Statistik: <https://www.bps.go.id/dynamictable/2018/05/18/1329/proporsi-kontribusi-pariwisata-terhadap-pdb-2015.html>
- Dharmawan, R. D., Suharyadi, & Farda, N. M. (2017). Geovisualization Using Hexagonal Tessellation for Spatiotemporal Earthquake Data Analysis in Indonesia. *Soft Computing in Data Science*, 177-187.
- Dodge, M., McDerby, M., & Turner, M. (2008). Geographic Visualization : Concept, Tools, and Applications. In M. Dodge, M. McDerby, & M. Turner, *The Power of Geographical Visualization* (pp. 1-10). Chichester: JohnWiley & Sons Ltd.
- Han, J., Kamber, M., & Pei, J. (2012). *Data Mining : Concepts and Techniques*. Burlington: Elsevier.
- Harun, M., & Syam, H. M. (2018). Pengaruh Penggunaan Media Sosial Instagram terhadap Perubahan Sikap Tujuan Wisata Liburan Mahasiswa Universitas Syiah Kuala. *Jurnal Ilmiah Mahasiswa FISIP Unsyiah*, Vol 3, No 2.
- Instagram. (2018, Oktober 6). *Wisata*. Retrieved from Instagram: <https://www.instagram.com/explore/tags/wisata/>
- KataData. (2018, Oktober 11). *Pengguna Instagram*. Retrieved from <https://databoks.katadata.co.id/datapublish/2018/02/09/berapa-pengguna-instagram-dari-indonesia>
- Kerren, A., Ebert, A., & Meyer, J. (2007). *Human-Centered Visualization Environments*. Berlin: Springer.
- Kraak, M.-J., & Ormeling, F. (2007). *Kartografi Visualialisasi Data Geospasial*. Yogyakarta: UGM Press.
- Makhabel, B., Mishra, P., Danneman, N., & Heimann, R. (2017). *R: Mining Spatial, Text, Web, and Social Media Data*. Birmingham: Packt Publishing Ltd.
- Perumal, M., Velumani, B., Sadhasivam, A., & Ramaswary, K. (2015). Spatial Data Mining Approaches for GIS – A Brief Review. *Emerging ICT for Bridging the Future - Proceedings of the 49th Annual Convention of the Computer Society of India CSI* (pp. 579-592). Coimbatore: Springer.
- Robinson, A. C., Demsar, U., Moore, A. B., Buckley, A., Jiang, B., Field, K., . . . Sluter, C. R. (2017). Geospatial big data and cartography: research challenges and opportunities for making maps that matter. *International Journal of Carthography*, 2372 - 9341.
- Shelton , T., Poorthuis, A., Graham, M., & Zook, M. (2014). Mapping the data shadows of Hurricane Sandy: Uncovering the sociospatial dimensions of ‘big data’. *Geoforum*, 167–179.
- Vu, H. Q., Li, G., Law, R., & Ye, B. H. (2013). Exploring the travel behaviors of inbound tourists to Hong Kong using. *Tourism Management*, 222-232.

Revised August 2019