REVEALING TOURIST HOTSPOTS IN YOGYAKARTA CITY BASED ON SOCIAL MEDIA DATA CLUSTERING

Totok Wahyu WIBOWO

Faculty of Geography, Universitas Gadjah Mada, Department of Geographic Information Science, Yogyakarta, Indonesia, e-mail: totok.wahyu@ugm.ac.id

Sigit Heru Murti Budi SANTOSA^{*}

Universitas Gadjah Mada, Faculty of Geography, Department of Geographic Information Science, Yogyakarta, Indonesia, e-mail: sigit.heru.m@ugm.ac.id

Bowo SUSILO Universitas Gadjah Mada, Faculty of Geography, Department of Geographic Information Science, Yogyakarta, Indonesia, e-mail: bowosusilo@ugm.ac.id

Taufik Hery PURWANTO

Universitas Gadjah Mada, Faculty of Geography, Department of Geographic Information Science, Yogyakarta, Indonesia, e-mail: taufik@ugm.ac.id

Citation: Wibowo, T.W., Santosa, S.H.M.B., Susilo, B., & Purwanto, T.H. (2021). REVEALING TOURIST HOTSPOTS IN YOGYAKARTA CITY BASED ON SOCIAL MEDIA DATA CLUSTERING. GeoJournal of Tourism and Geosites, 34(1), 218–225. https://doi.org/10.30892/gtg.34129-640

Abstract: Cities have a common characteristic in the form of land utilisation, which is dominated by built-up areas. Tourism is an essential aspect of city development because it can involve the identity of the city. Historical buildings, landmarks, shopping centres and museums are generally interesting places for tourists to visit. Yogyakarta, the research area, is synonymous as a city of culture and of students. Knowledge of the spatial clustering patterns of tourists can be one of the references for urban development. Social media data were used in the study as an alternative to direct data collection, which requires considerable resources. Flickr and Twitter were used as proxies to determine the distribution of tourists, and the DBSCAN and HDBSCAN clustering algorithms were used to determine the centres of tourist activity. Further more, Flickr data were analysed temporally to determine the impact of the COVID-19 pandemic on tourism in Yogyakarta City. The clustering of social media data results shows that there are several city hotspots, besides the already well-known tourist attractions. Apart from city landmarks, several other tourist hotspots were revealed through the clustering process, such as accommodation, shopping centres, entertainment venues and souvenir shops, which also support tourism activities. The impact of COVID-19 on tourism in Yogyakarta City can be reflected through the number of uploaded photos by tourists on Flickr, which has decreased since March 2020.

Key words: Social Media, Flickr, Twitter, HDBSCAN, COVID-19

* * * * * *

INTRODUCTION

Tourism in urban areas has certain characteristics because most of the land use is generally built-up areas. Various human-made landmarks, such as buildings, shopping centres, recreation zones, or iconic city locations can be beautiful places for tourists to visit (Hu et al., 2015). This condition is different from tourist attractions clearly defined as having a particular space and function. Because each city is unique, tourists can visit exciting places there. The problem is that in public areas, it is likely that tourism officers will not record the number of visitors. Several previous studies have called this tourist perception the 'vague place' concept (Montello et al., 2014; Montello et al., 2003). Urban tourism can generate much diversification of tourist choices, so it is not easy to decide where to visit when on short visits (Jansen-Verbeke, 1986; Salas-Olmedo et al., 2018). Apart from visits to exciting places, tourists also need adequate supporting infrastructure. Some of this may also become tourist destinations in its own right, because it offers absolute uniqueness. For example, in terms of hotel selection, tourists tend to choose hotels close to their destination, generally within walking distance (Shoval et al., 2011). This makes a tourist destination attractive because determination of a tourist centre is not only based on officers' counts at officially managed tourist sites. Knowledge of tourists' favourite locations is required for more effective city management, especially if there are new locations that offer opportunities for further exploration (Devkota et al., 2015).

The number of incoming tourists is a good indicator to assess the competitiveness of regional tourism. There are three sources of incoming tourist number data, namely: statistical records, travel records, and social media (Mou et al., 2019). Statistical records can provide accurate incoming tourist data but are unable to show the spatial distribution of tourists. Meanwhile, the use of travel notes has two main challenges, namely the existence of incomplete data and an inaccurate report by the author. Social media provide opportunities for passive tourist data acquisition due to its increasing popularity (Girardin et al., 2008; Mou et al., 2019; Önder et al., 2014). The digital footprint left by tourists through social media content opens up opportunities for spatial tourist studies. In general, studies related to tourist concentration can be conducted using census techniques or surveys based on samples (Salas-Olmedo et al., 2018). However, both methods require considerable resources and do not necessarily cover a broad enough spatial or temporal scope. Big data offers new tourism research opportunities by providing high levels of spatial and temporal data to analyse large volumes of tourist spatio-temporal patterns (Goodchild, 2007). Big data for tourism studies can be obtained from three sources: users, devices and operations (Li et al., 2018). These three big data sources for tourism are influenced by developments in social media platforms, the Internet of Things (IoT), and services for tourism-related operations (e.g., web searches, web page visits and online ticket purchases). Based on big data from these three primary sources, tourist and tourism market behaviour can be better explored and understood by academia and industry. For example, big data on a large scale make it possible to overcome the limitations of survey data user sample size problems and provide new ways to understand tourist behaviour (Yang et al., 2015). Big data analysis is also known to provide sufficient data without sample bias to understand such behaviour (Li et al., 2017). The use of geotagged photos from photo-sharing services or other social media can enrich tourism data sources and can be used in tourism

^{*} Corresponding author

planning and management. The density of geotagged photos shows the distribution of the tourist presence across a city, and is easier to perform than direct measurements (García-Palomares et al., 2015). The direct benefit of density mapping is that it can be used to determine spatiotemporal clusters (Hu et al., 2015; Kisilevich et al., 2013) and identify urban morphology (Kádár, 2014).

MATERIAL AND METHODS

Study area

Yogyakarta is known as a city of culture and education, with a rich historical heritage in every corner of the city. It receives privileges from the Government of Indonesia through Law Number 13 of 2012 concerning the Privileges of the Special Region of Yogyakarta. This law further strengthens the cultural elements found in Yogyakarta. The city is located in the middle of Yogyakarta Special Region, with an area of 32.5 km².



Figure 1. Tourist attraction in Yogyakarta City



Figure 2. Philosophical axis of Yogyakarta (Tugu Jogja Monument – Yogyakarta Palace – Panggung Krapyak)

24 tourist attractions were recorded in the Government Office Tourism report (Statistik's, 2018), of which the majority were located in the centre of the city (Figure 1), which is filled with locations that are attractive to tourists, such as the Yogyakarta Palace (Keraton Ngayogyakarta Hadiningrat), Presidential Palace and several museums.

Apart from the tourist attractions listed in Figure 1, Yogyakarta City has other landmarks that are no less attractive to tourists, such as the Yogyakarta Monument, Malioboro Street, North Square (alunalun lor), South Square (alun-alun kidul), and Tugu Station (Figure 2).

Malioboro Street and the

buildings in it are even part of the identity of the city, given its long history (Septirina et al., 2016). Yogyakarta City is also famous for the philosophical axis formed between the Tugu Jogja monument, Yogyakarta Palace and Panggung Krapyak, which is a manifestation of the journey of human life from birth to death (Syahbudin et al., 2018). Because these locations are in the form of public spaces, it is somewhat challenging to record tourist visits with certainty, compared to attractions that have doors/gates for entry and exit. Information about tourist hotspot clusters is therefore needed, especially in the current era, as many tourist sites prioritize their photographic quality. Photogenic attractions have the potential to gain popularity through information dissemination on social media (Wibowo et al., 2019).

Data acquisistion and processing

The primary data used in the study were obtained from Flickr, in the form of photo-sharing services. Flickr allows users to post photos accompanied by a title and a description, and also supports the geotagging function, by which photos uploaded by users can be referenced to their location on the earth's surface. SmugMug, the owner of Flickr, is committed to providing data for free and open-access, for both commercial and noncommercial users (Flickr, 2018). Data collection is written using the Python programming language, which is used to communicate with the Flickr server via the Application Programming Interface (API). In general, the script used consists of four parts, namely function calls, providing access and authentication, commands for streaming data, and search limits

(coordinates). The Flickr data used in this study were limited to those from January 2000 to August 2020, with 34,704 data items obtained. Twitter is a micro-blogging social media platform, which provides flexibility for users to upload multimedia information, mostly text, photos and videos. Because of its highly flexible functionality, data on Twitter have high variability. Twitter provides an option for users to include the location where a tweet was created, so geotagged tweets can be created. There are various ways to obtain Twitter data via the Twitter API (Twitter, 2020). In this study, the GET statuses/sample method was used, which provides 1% real-time random tweet data. Even though this is only a sample, it cannot be denied that the existence of Twitter data has added a new data source to mapping, as an output from location-based social media (Thatcher, 2014). Twitter data were collected from January 2019 to August 2020, with a total of 62,980 tweets. In terms of popularity, Twitter is more popular than Flickr; data published by Hootsuite in 2019 showed that Twitter occupied sixth position on a list of the most active social media in Indonesia (Hootsuite, 2019). However, Flickr is also widely used in various Volunteered Geographic Information (VGI)-based research in tourism (Stock, 2018). This is because photos are a product of tourism activities, and a photo-sharing services platform is able to optimise them. The distribution of Flickr data appears to be more widespread than Twitter data, especially in the areas around the Presidential Palace and Yogyakarta Palace (Figure 3).

The difference in the time duration between the two data sets affects the difference in data density. A striking similarity between the two sets of raw data is the density along a straight line parallel to Jalan Malioboro. The large number of posting activities made adjacent to roads makes the patterns seen in the data similar to that of the road network. Raw data processing was performed by means of filtering based on text and location. Keywords were used for the text-based filtering; in this case, posts that had words related to tourism activities were retained. Some of these keywords were holiday, tourism, long weekend and scene. Location-based filtering was performed by eliminating data that had identical coordinate points. This can occur in the case of a user making repeated uploads from the same location; for example, by an advertiser (Frias-Martinez et al., 2012).



Figure 3. Raw data from a) Flickr and b) Twitter of Yogyakarta City

After filtering the Twitter data, 7,224 tweets or about 11% of the raw data were left. Of these, there were 4,340 unique users, so the average user makes just below two tweets. This further strengthens the Twitter data, which has high variability. On the other hand, Flickr data comprised 12,371 data items, or 35% of raw data. There were 2,323 unique users, with an average of five posts per user. Based on these records, it can be seen that Flickr users are more loyal to the platform, because it produces a higher average post.

Clusterization method

Tourists in Yogyakarta City stay on average for two days, so they need to determine the places to visit. This behaviour can produce specific patterns that can be recognised through social media (Devkota et al., 2019). For this study, geotagged Flickr and Twitter data were analysed using a clustering algorithm to determine the centre of activity. Grouping is a data mining technique to identify patterns in data, which is generally used for exploratory data analysis. Grouping techniques are widely used to uncover hotspots of dense data points.

There are various kinds of clustering algorithm, which are generally divided into two types, partitioning and hierarchial (Ester et al., 1996). Partitioning works by dividing the data into predetermined clusters, which in this case requires domain knowledge. An example of a popular partitioning algorithm is the k-means clustering algorithm (Shaji et al., 2020). The popularity of this algorithm is supported by various types of GIS software that provide it. Density-based spatial clustering of applications with noise (DBSCAN) is another example of density-based partitioning algorithms with better efficiency (Ester et al., 1996). Hierarchies work by composing a hierarchical decomposition of data, which can generally be represented using a dendrogram. Hierarchial DBSCAN (HDBSCAN) is an improved version of DBSCAN by creating a simplified hierarchy composed only of the most significant clusters, which can be easily extracted (Campello et al., 2013).

The DBSCAN algorithm has two user-defined parameters, namely the search distance radius/epsilon (*eps*) and the minimum data count within the search radius (*minPts*). These two parameters need to be determined carefully in order to produce useful clustering. The search distance radius (*eps*) is strongly influenced by the area of the study area and the scale of the mapping; the greater the value of *eps*, the more general the cluster, and vice versa. Meanwhile, *minPts* affects the minimum number of points that can be considered as clusters. A low *minPts* value can result in a large number of clusters, but with the risk of considerable noise entering the cluster. Conversely, a higher value of *minPts* could cause cluster merging, meaning the level of detail could be reduced. On the other hand, the HDBSCAN algorithm, which is a development of the DBSCAN algorithm, applies DBSCAN to various epsilon values and integrates the results to find the grouping that provides the best stability (Campello et al., 2013). HDBSCAN also has the ability to find cultivars at varying densities, and it is also more robust in terms of parameter selection (McInnes et al., 2017). Assessment of clusterization results can be seen objectively from the value of the ratio of between to total sum of squares, with a value close to 1 indicating a good cluster. In this study, assessment of the clustering results was also made by conducting a visual assessment of the clusters formed, both from the DBSCAN and HDBSCAN algorithms.

Temporal Visualization

The COVID-19 pandemic has harmed the tourism industry in many countries. For example, Poland has experienced a decrease in the

number of visits and accommodation occupancy due to restrictions on international movements (Korinth and Ranasinghe, 2020). The magnitude of the pandemic is also expected to affect the pattern of tourism, accommodation services and commercial flights in South Africa (Rogerson and Rogerson, 2020). Indonesia, including the Yogyakarta City, cannot avoid these impacts because at the beginning of the pandemic it imposed large-scale social restrictions (PSBB). To determine the impact of social restrictions on tourism activities in the study area, a comparison of tourist patterns was conducted over the past five years. In this case, only Flickr data were used because Twitter's data coverage collected in this study was only until 2018.

RESULTS AND DISCUSSION

Data Clustering

Determination of the parameters for the DBSCAN algorithm is an important step in obtaining the appropriate clusters. An experiment was conducted by adjusting *minPts* values between 5-25 in multiples of 5, and *eps* values between 50 - 150 in multiples of 25. The 25 experimental results were then assessed visually and statistically. The most optimal clustering was achieved at the *minPts* setting of 20 and *eps* of 100 meters (Figure 4). A total of 34 clusters were generated from the Flickr data, while the Twitter data resulted in 39 clusters. Following the clustering theory, a low *minPts* value will result in a larger number of clusters, with a tendency to be overcluttered (Campello et al., 2015; McInnes et al., 2017). Tuning the value of *eps* to the same *minPts* value can reduce the number of clusters. This effect can be clearly seen in experiments 1-5; at the same *minPts* value of 50, the number of clusters decreased from 181 to 27 with the Flickr data and from 268 to 14 with the Twitter data. Too few clusters are more influenced by the choice of a too high *eps* value; this can be seen on the right-hand side of the graph in Figure 3. Changing the *minPts* value does not change the number of clusters produced if the *eps* value is too high. In this condition, the cluster distribution tends to be under clustered because there will be clusters that are too dominant. In this study, it was found that the eps value of 100 meters was an excellent value to represent the movement of tourists in tourist areas. This value may differ if applied to other types of travel or other applications.



Figure 4. Experiment result of eps and minPts applied on the data (Parameters used in HDBSCAN algorithm)

The clustered data show a massive cluster in the middle of the city, which stretches vertically from north to south. This cluster can be found from the results of the Flickr and Twitter data clustering. The location of the cluster starts from Pangeran Mangkubumi Street, Malioboro Street, Ahmad Yani Street and reaches the area around the Yogyakarta Palace. The largest cluster generated from the Flickr data has a wide enough coverage area to intersect with 16 tourist attractions (Figure 5a). Meanwhile, the largest cluster from the Twitter data has an elongated shape and includes only ten attractions (Figure 5b). In a more detailed finding, there are many compelling attractions contained in the large cluster, which will be discussed in the following section. On the outskirts of the city, clusters are sporadically encountered. The eastern part of the city of Yogyakarta is where there are the least clusters, because there are not many tourist attractions. In comparison, the Flickr data show seven clusters, while those from Twitter show five. The HDBSCAN algorithm was executed using the same *minPts* value setting. In order to obtain good clustering results, the minimum sample value was tuned for each set of data. The minimum sample values used were 80 and 40 for the Flickr and Twitter data respectively. A difference in the minimum sample value was made because the essential characteristics of the two types of data are different, with Flickr data tending to be clustered. In contrast, Twitter data tends to spread out more (Figure 3). The high minimum sample value was intended as a conservative effort to remove noise from the cluster. In general, the clustering generated by the HDBSCAN algorithm has the characteristics of narrow and not very large clusters, especially those generated from the Flickr data (Figure 6). The clusters from the cluster from the eastern city of Yogyakarta, which appear to be more spread out and quite large.

A longitudinal vertical pattern across Jalan Malioboro can be seen, which is broken down by several clusters. On the one hand, this narrower cluster intersects with fewer, only 14, tourist attractions. On the other hand, many new points of interest emerge, beside the tourist attractions that are defined in Figures 1 and 2. Visually, the clusters generated using Flickr data are more compact than those from Twitter data. This is reinforced by the value of the ratio between the total sum of the squares obtained, with the results from the clustering of the Flickr and Twitter data resulting in values of 0.997 and 0.988 respectively. One of the reasons for the low ratio value for the Twitter data is the presence of clusters that are relatively spread out in the eastern part of the city, with very high within-cluster sum of squares values (127.12). In perfect clustering, members in one cluster will tend to be uniform, but between clusters have significant differences. An indication of this is the low within-cluster sum of squares and the high between-cluster sum of squares. In the following section, clusterization data from Flickr will be used, because visually and statistically it is superior to Twitter data clustering.

Exploring Tourist Hotspot

Although several clusters do not represent tourists' position at tourist attraction sites, they are still closely related to tourism activities,

since their locations offer views of the city, hotels, stations or other attractions. In comprehensive tourism management, as pects of tourist comfort should also be considered. Based on the clustering results from the Flickr data, 29 hotspot clusters can be categorized into six groups: tourist attractions, accommodation, city scenery, shopping centres, souvenir shops, and entertainment (Figure 7).





Figure 6. Clustering result of DBSCAN algorithm using a) Flickr and b) Twitter data

There are nine clusters of tourist attractions, most of which around are Jalan Malioboro, except for Gembira Loka Zoo and XT Square. As one of the special symbols of the Yogyakarta Special Region, Yogyakarta Palace has the highest number of photo spots of all the clusters, making it a popular destination for tourists. The Taman Sari complex, which is part of cultural tourism, also has many photo points. These two popular cultural tourism sites can strengthen the image of Yogyakarta as a city of culture. The three intersection points starting from Tugu Jogja Monument, Malioboro Street Sign and the city centre are also attractive tourist destinations. Taking pictures at these landmarks can be proof that tourists have visited the city. The renovation and renewal of the three landmarks in recent confirms that vears the authorities are paying attention to tourist area development.

The Malioboro Street area is now even more friendly to pedestrians, making it easier for tourists to explore from the north to the south at Yogyakarta Palace. This development further supports the preservation of Yogyakarta's philosophical axis and efforts to support Yogyakarta City as a World Heritage City (Syahbudin et al.,, 2018). The accommodation cluster is not too far from the city centre, with a radius of approximately two kilometers.

For a two-day stay, two kilometers from the city centre

would be an ideal location for exploring the city, either on foot or by car. The Prawirotaman area, south of Yogyakarta Palace, is the most popular accommodation destination. Even in this area, there are two clusters which are not too far away. Not only known by local tourists, the Prawirotaman area is also well known among foreign tourists as a low budget accommodation area (Borgersen and Rye, 2004). However, the relative proximity to the city centre does not guarantee that the accommodation industry will survive. Ideally, development of the facilities provided by the inns should be made to improve the user experience and help the tourism industry survive. Yogyakarta City also provides city view tours that are attractive for tourists to take pictures at the various locations.

The data from the clustering results can present information on which areas are of interest to tourists, and include at least six locations, namely Code River, Kotabaru, the railway, Kepatihan Office, Kleringan Kewek Bridge, and South Square. Because these locations are open facilities, it is difficult to record the number of visitors. However, through social media data, it can then be revealed that these locations are attractive to tourists if simply to take photos. In recent years, some of these locations have undergone improvements to make them more attractive to tourists; for example, the construction of a boulevard in Kotabaru. Kotagede, which is located in the southern part of Yogyakarta City, offers souvenirs for tourists, including various well-known silver crafts. Apart from souvenirs, Kotagede also has tourist attractions such as The Kings Tomb, the Great Mosque of Kotagede, Kedhaton, and the Javanese Traditional House. However, due to the spread of the data point positions, the cluster formed is large. There are three shopping centre clusters, all three of which are well known, namely Malioboro Mall and Galeria Mall, which are modern markets, and Beringharjo Market, a traditional market. Preservation of the Beringharjo market is critical because it is not only attractive to tourists but also to maintain the culture of traditional Javanese markets (Herliana, 2015). Meanwhile, cinemas and football fields are a means for locals and tourists to find entertainment.

In addition to coordinate data, Flickr photo metadata also stores data on the date of photo uploads. This data can be used to determine the frequency of monthly photo uploads in Yogyakarta City. Figure 8 shows a graph of the number of Flickr photo uploads in the city,

juxtaposed with data on the number of tourists recorded by the Yogyakarta Special Region Tourism Office in 2018. The photo upload to Flickr graphic (in blue) fluctuates quite strikingly, with April being the lowest point and August the highest. In contrast, the tourist data graph (in orange) does not show significant differences, except for December, when it rose sharply. In general, the gap is quite large between August and October, when the number of photo uploads jumps from the previous month. In Yogyakarta City, these three months are in summer, with a transition to the rainy season, so they are very suitable for relatively distraction-free photo taking.



Figure 7. Tourist hotspot cluster in Yogyakarta City based on HDBSCAN clustering on Flickr data

COVID-19 Pandemic Impact

It is undeniable that the COVID-19 pandemic, that has taken place since the end of 2019, has greatly affected the tourism industry, including that of Indonesia. A few weeks after the announcement of the first case of COVID-19 in Jakarta in March 2020, many tourism industry players stopped their activities. Because of the highly contagious nature of the disease, crowds of tourists have the potential to become transmission clusters. Data on the distribution of clusters in the first half of the last four years in Yogyakarta City show a decreased number of tourist clusters, from 14 in 2017 to only 2 in 2020 (Figure 9). Naturally, the number of photos taken also fell. When viewed as a whole, the decline from year to year was consistent, at approximately the same rate. The clusters in 2017 were seen to be scattered from the Tugu Jogja Monument to Kotagede, while in 2018 they tended to be closer to the city centre.



There have been only two clusters in 2020, with the majority of uploads made from January to March. The two clusters are the area around Tugu Station, which is also adjacent to the Malioboro Street Sign, and the Yogyakarta Palace area. Both these areas and the city centre consist of clusters of the four sets of data used for clustering. The number of photos uploaded onto Flickr has decreased, which be an indicator that tourism in Yogyakarta has also been experiencing a decline. In 2017, the recorded number of tourists was 5,347,626. If in simulation the cluster points the represent the number of tourists, it is estimated that in early 2020 the number entering Yogyakarta was 800,000 to 1,000,000 visitors. With the closure of tourist sites, reduced hotel occupancy,



flight restrictions and government social restriction policies, the resulting impact could be even more severe as mentioned in other scholars findings (Korinth and Ranasinghe, 2020; Rogerson and Rogerson, 2020).



Figure 9. Clusters center for the first-half year in the last four years

CONCLUSION

Flickr and Twitter provide free access to large volumes of data, albeit with their respective limitations, which can be used for humanrelated research. The advantage of these data is that besides the coordinate option, a stamp team can be used for temporal analysis. In this study, both sources of data were used as proxies to determine tourist clusters. Flickr data has an advantage over Twitter data because its users only focus on uploading photos, making traveller recognition easier. On the other hand, Twitter data is rich, requiring delicate data pre-processing. However, the clustering results using the DBSCAN and HDBSCAN algorithms for both data sources resulted in the same number of clusters, with a similar spatial distribution. Determining clusters using the HDBSCAN algorithm is very flexible and can be applied in research based on two-dimensional point data to produce good clusters.

Determining parameters for execution can be done by adjusting the data used, mostly related to the study area and object. In this study, the HDBSCAN clustering results from the Flickr data show that cultural tourist attractions were still the main selling point of Yogyakarta City, withYogyakarta Palace and Taman Sari having the largest number of photos. Other locations that are no less popular are Yogyakarta city landmarks, such as the Tugu Jogja Monument, Malioboro Street Sign, and the city centre. Tourist accommodation options indicate that areas within a two-kilometer radius of the city centre are the preferred ones. The Prawirotaman area has even spawned two adjoining clusters. This clustering method can be duplicated in other cities to gain insight into tourism activities that are not officially recorded. It is undeniable that the COVID-19 pandemic has had a very negative impact on the tourism industry. Based on the results Flickr data clustering in the last first half of the last four years, it has been established that the number of tourist clusters has continued to decline. The prediction of the number of tourists visiting Yogyakarta City will also fall.

Acknowledgement

The author would like to thank the Directorate of Research at Universitas Gadjah Mada who gave the Final Project Recognition Grant to the author through assignment letter number 2488/UN1.P.III/DIT-LIT/PT/2020.

REFERENCES

Borgersen, T., & Rye, S.A. (2004). The dynamics of a low budget tourist area : The case of Prawirotaman. ASEAN Journal on Hospitality and Tourism, 3(1), 11–27. https://www.researchgate.net/publication/263968156_The_dynamics_of_a_low_budget_tourist_area_The_case_of_Prawirotaman

- Campello, R.J.G.B., Moulavi, D., & Sander, J. (2013). Density-Based Clustering Based on Hierarchical Density Estimates. In J. Pei, V. S. Tseng, L. Cao, H. Motoda, & G. Xu (Eds.), Advances in Knowledge Discovery and Data Mining. PAKDD 2013. Lecture Notes in Computer Science, 7819, 160–172. https://doi.org/10.1007/978-3-642-37456-2_14
- Campello, R.J.G.B., Moulavi, D., Zimek, A., & Sander, J. (2015). Hierarchical density estimates for data clustering, visualization, and outlier detection. ACM Transactions on Knowledge Discovery from Data, 10(1), 1–51. https://doi.org/10.1145/2733381

Devkota, B., Miyazaki, H., & Witayangkurn, A. (2019). Using Volunteered Geographic Information and Nighttime Light Remote Sensing Data to Identify Tourism Areas of Interest. Sustainability, Switzerland, 11(17), 1–29. https://doi.org/10.3390/su11174718

Ester, M., Kriegel, H.P., Sander, J., & Xu, X. (1996). Density-Based Clustering Methods. *KDD-96 Proceedings*, 226–231. https://doi.org/10.1016/B978-044452701-1.00067-3 Frias-Martinez, V., Soto, V., Hohwald, H., & Frias-Martinez, E. (2012). *Characterizing urban landscapes using geolocated tweets*. 2012 International Conference

on Privacy, Security, Risk and Trust and 2012 International Conference on Social Computing, 239–248. https://doi.org/10.1109/SocialCom-PASSAT.2012.19 García-Palomares, J.C., Gutiérrez, J., & Mínguez, C. (2015). Identification of tourist hot spots based on social networks: A comparative analysis of European

metropolises using photo-sharing services and GIS. *Applied Geography*, 63, 408–417. https://doi.org/10.1016/j.apgeog.2015.08.002 Girardin, F., Dal Fiore, F., Ratti, C., & Blat, J. (2008). Leveraging explicitly disclosed location information to understand tourist dynamics: A case study. *Journal of Location Based Services*, 2(1), 41–56. https://doi.org/10.1080/17489720802261138

Goodchild, M.F. (2007). Citizens as sensors: The world of volunteered geography. *GeoJournal*, 69(4), 211–221. https://doi.org/10.1007/s10708-007-9111-y

Herliana, E.T. (2015). Preserving Javanese Culture through Retail Activities in Pasar Beringharjo, Yogyakarta. Procedia - Social and Behavioral Sciences, 184(August 2014), 206–213. https://doi.org/10.1016/j.sbspro.2015.05.081

Hu, Y., Gao, S., Janowicz, K., Yu, B., Li, W., & Prasad, S. (2015). Extracting and understanding urban areas of interest using geotagged photos. *Computers, Environment and Urban Systems*, 54, 240–254. https://doi.org/10.1016/j.compenvurbsys.2015.09.001

Jansen-Verbeke, M. (1986). Inner-city tourism: Resources, tourists and promoters. Annals of Tourism Research, 13(1), 79–100. https://doi.org/10.1016/0160-7383(86)90058-7

- Kádár, B. (2014). Measuring tourist activities in cities using geotagged photography. *Tourism Geographies*, 16(1), 88–104. https://doi.org/10.1080/14616688.2013.868029
- Kisilevich, S., Keim, D., Andrienko, N., & Andrienko, G. (2013). Geospatial Visualisation. In A. Moore & I. Drecki (Eds.), Geospatial Visualisation, Lecture Notes in Geoinformation and Cartography. https://doi.org/10.1007/978-3-642-12289-7
- Korinth, B., & Ranasinghe, R. (2020). Covid-19 Pandemic's Impact on Tourism in Poland in March 2020. GeoJournal of Tourism and Geosites, 31(3), 1004– 1009. https://doi.org/10.30892/gtg.31308-531
- Li, J., Xu, L., Tang, L., Wang, S., & Li, L. (2018). Big data in tourism research: A literature review. *Tourism Management*, 68, 301-323. https://doi.org/10.1016/j.tourman.2018.03.009
- Li, X., Pan, B., Law, R., & Huang, X. (2017). Forecasting tourism demand with composite search index. *Tourism Management*, 59, 57-66. https://doi.org/10.1016/j.tourman.2016.07.005
- McInnes, L., Healy, J., & Astels, S. (2017). hdbscan: Hierarchical density based clustering. The Journal of Open Source Software, 2(11), 1-2. https://doi.org/10.21105/joss.00205
- Montello, D.R., Friedman, A., & Phillips, D.W. (2014). Vague cognitive regions in geography and geographic information science. International Journal of Geographical Information Science, 28(9), 1802–1820. https://doi.org/10.1080/13658816.2014.900178
- Montello, D.R., Goodchild, M.F., Gottsegen, J., & Fohl, P. (2003). Where's downtown?: Behavioral methods for determining referents of vague spatial queries. Spatial Cognition and Computation, 3(2–3), 185–204. https://doi.org/10.1080/13875868.2003.9683761
- Mou, N., Yuan, R., Yang, T., Zhang, H., Tang, J., & Makkonen, T. (2019). Exploring spatio-temporal changes of city inbound tourism flow: The case of Shanghai, China. *Tourism Management*, 76(2020). https://doi.org/10.1016/j.tourman.2019.103955
- Önder, I., Koerbitz, W., & Hubmann-Haidvogel, A. (2014). Tracing Tourists by Their Digital Footprints: The Case of Austria. Journal of Travel Research, 55(5), 566–573. https://doi.org/10.1177/0047287514563985
- Rogerson, C.M., & Rogerson, J.M. (2020). Covid-19 Tourism Impacts In South Africa: Government and Industry Responses. *GeoJournal of Tourism and Geosites*, 31(3), 1004–1009. https://doi.org/10.30892/gtg.31321-544
- Salas-Olmedo, M.H., Moya-Gómez, B., García-Palomares, J.C., & Gutiérrez, J. (2018). Tourists' digital footprint in cities: Comparing Big Data sources. *Tourism Management*, 66, 13–25. https://doi.org/10.1016/j.tourman.2017.11.001
- Septirina, S.N., Takeo, O., & Satoru, K. (2016). Conservation of Historical Architecture in Malioboro Street, Yogyakarta City, Indonesia. Procedia Social and Behavioral Sciences, 225, 259–269. https://doi.org/10.1016/j.sbspro.2016.06.025
- Shaji, H. E., Tangirala, A. K., & Vanajakshi, L. (2020). Prediction Of Trends In Bus Travel Time Using Spatial Patterns. Transportation Research Proceedia, 48, 998–1007. https://doi.org/10.1016/j.trpro.2020.08.128
- Shoval, N., McKercher, B., Ng, E., & Birenboim, A. (2011). Hotel location and tourist activity in cities. Annals of Tourism Research, 38(4), 1594–1612. https://doi.org/10.1016/j.annals.2011.02.007
- Stock, K. (2018). Mining location from social media: A systematic review. Computers, Environment and Urban Systems, 71(March), 209-240. https://doi.org/10.1016/j.compenvurbsys.2018.05.007
- Syahbudin, A., Phenomenon, S.P., Meinata, A., Hanindita, A.S.H., & Mulyana, B. (2018). City of Philosophy: Evaluation of Tree Philosophy and Its Architecture in Yogyakarta Philosophical Axis Towards UNESCO World Heritage. In A. Sukartiko, T. Nuringtyas, S. Marliana, & A. Isnansetyo (Eds.), Proceeding of the 2nd International Conference on Tropical Agriculture, 249–274. https://doi.org/10.1007/978-3-319-97553-5_25
- Thatcher, J. (2014). Living on fumes: Digital footprints, data fumes, and the limitations of spatial big data. *International Journal of Communication*, 8(1), 1765–1783. https://ijoc.org/index.php/ijoc/article/view/2174
- Wibowo, T.W., Bustomi, A.F., & Sukamdi, A.V. (2019). Tourist Attraction Popularity Mapping based on Geotagged Tweets. Forum Geografi, 33(1), 82–100. https://doi.org/10.23917/forgeo.v33i1.8021
- Yang, X., Pan, B., Evans, J.A., & Lv, B. (2015). Forecasting Chinese tourist volume with search engine data. *Tourism Management*, 46, 386–397. https://doi.org/10.1016/j.tourman.2014.07.019
- Zhou, X., Xu, C., & Kimmons, B. (2015). Detecting tourism destinations using scalable geospatial analysis based on cloud computing platform. Computers, Environment and Urban Systems, 54, 144–153. https://doi.org/10.1016/j.compenvurbsys.2015.07.006
- *** Flickr (2018). Flickr APIs Terms of Use. Retrieved May 10, 2019. https://www.flickr.com/help/terms/api
- *** Hootsuite (2019). Digital 2019: Indonesia. In Global Digital Insights. https://datareportal.com/reports/digital-2020-indonesia
- *** Statistik Kepariwisataan DIY 2018. (2018). Bantul, Yogyakarta: Dinas Pariwisata Daerah Istimewa Yogyakarta. https://visitingjogja.com/19962/statistikpariwisata-diy-2018/

Accepted: 19.02.2021

Available online: 10.03.2021

*** Twitter (2020). Twitter API - Tap into what's happening. Retrieved April 1, 2020. https://developer.twitter.com/en/products/twitter-api

Article history:	Received: 12.10.2020	Revised: 22.01.2021	
------------------	----------------------	---------------------	--

225