EKSAKTA
Journal of Sciences and Data Analysis

E-ISSN: 2720-9326
P-ISSN: 2716-0459
DOI: 10.20885/EKSAKTA.vol3.iss1.art4

Research Article

Analysis of Hotels Spatial Clustering in Bali: Density-Based Spatial Clustering of Application Noise (DBSCAN) Algorithm Approach

Achmad Fauzan[1,*], Afdelia Novianti[1], Raden Rara Mentari Ayu Ramadhani[1], Marcelinus Alfafisurya Setya Adhiwibawa[2]

[1] Statistics Department, Faculty of Mathematics and Natural Science Universitas Islam Indonesia Jl.Kaliurang KM 14.5, Sleman-Yogyakarta, Indonesia
[2] Ma Chung Research Center for Photosynthetic Pigments, Universitas Ma Chung

[*] *Corresponding author: achmadfauzan@uii.ac.id.*

Received: 10 November 2021; Accepted: 4 December 2021; Published: 7 March 2022

Abstract: Bali is one of the hearts of tourism in Indonesia. The existence of the Covid-19 pandemic has made this tourist paradise also affected the wheels of the economy. Based on this, this study aims to determine the density clustering of one of the economic supporters in Bali, namely hospitality. The study began with the quadrant method and Ripley's K-Function to measure the distribution pattern of hospitality. From the results of the two methods, the distribution pattern of hotels in Bali is more towards clusters than random or regular distribution. If the point distribution pattern is more towards the cluster, it is continued with the Density-Based Spatial Clustering of Application Noise (DBSCAN) algorithm to form spatial clustering. In the DBSCAN algorithm, a combination of parameters, namely minimum points (MinPts) and epsilon (Eps), is carried out with evaluation using the silhouette average width value. From the results of the DBSCAN algorithm, the clustering results show that the distribution of hotels in Bali forms clusters and tends to approach the surrounding tourist attractions, such as near the beach, city market, and mountainous areas. It can help policymakers if they want to prioritize economic recovery after the Covid-19 pandemic.

Keywords: Quadrant Method, Ripley's K-Function, Density-Based Spatial Clustering of Application Noise (DBSCAN), Silhoutte Average Width.

Introduction

The Covid-19 pandemic is a pandemic caused by Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) or better known as the Coronavirus. The impact of the Covid-19 pandemic is so broad, not only on the health sector but also on various sectors, such as education, social and economic sectors. The crisis caused by the Covid-19 pandemic is also different from the crisis from previous years, both in terms of causes, scope, and severity [1]. The impact of the Covid-19 pandemic is likely to be more profound on mental health and well-being over the long term [2], [3]. One of the significant impacts is the economic side, including the economy in Indonesia, especially from the tourism sector.

As one of the biggest tourist attractions in Indonesia, Bali is also feeling the impact of the Covid-19 pandemic. This is due to a decrease in tourists (international and domestic) in Bali [4]. Even worse, not a few hotel industries in Bali are also stopped [5]. Table 1 presents the level of hotel occupancy in Bali, which is one indicator of economic development level.

Table 1. Room occupancy rates of hotels (%), in 2021 [6]

Hotel classification	January	February	March	April	May	June	July	August
5 star hotel	9.63	5.84	8.20	10.29	12.43	22.61	5.67	4.92
4 star hotel	11.92	9.77	11.20	10.30	9.05	14.75	4.77	4.26
3 star hotel	11.62	11.05	10.95	9.36	10.01	12.91	5.09	4.80
2 star hotel	11.76	12.51	10.46	11.09	10.14	11.77	6.50	7.51
1 star hotel	-	1.19	7.48	6.88	7.77	10.82	1.99	6.92
All star average	11.15	8.99	10.24	10.09	10.35	16.68	5.23	4.77

Based on Table 1, the average hotel occupancy is still low compared to before the Covid-19 pandemic, which reached 50-70% [6].

A year and a half of the pandemic have passed, and now gradually, there is a decrease in Covid-19 cases in Indonesia [7]. This is triggered by various factors, including the acceleration of vaccination and public awareness of health protocols. Presently is the time to begin the transition phase for economic recovery in Bali. One form of economic recovery is the determination of appropriate policies, including deciding the supporting profile of the economy in Bali. Based on this, this study focuses on the distribution of hospitality as one of the financial supports in Bali.

Analysis of the distribution of hotels plays an essential role in supporting government policies later when the pandemic is still ongoing or post-pandemic later. One of the distribution analyses that can be used is spatial clustering. The main objective is to find out the clustering of hotels in Bali so that it is expected to be an input or consideration for policymakers. This research focuses on density clustering or known as density based-cluster [8]. Density-Based Clustering is an unsupervised learning method for finding distinct groups/clusters in data. It is founded on the idea that in a data space, a cluster is a contiguous region of high point density separated from other clusters by contiguous low point density regions [9]. The spatial clustering used in this research is the Density-Based Spatial Clustering of Application Noise (DBSCAN) algorithm. The DBSCAN algorithm is an algorithm that develops areas with high enough density into clusters and finds clusters in arbitrary shapes in a spatial database containing noise [10].

Several previous studies related to the DBSCAN method include research Sari & Primajaya [11] which applies the DBSCAN method for agriculture, the problem in this research is how to analyze the characteristics of rice production in Karawang district using the DBSCAN algorithm. Harjanto *et al* [12] implement this method in determining the priority scale for handling stunting under-fives. This study was due to the stunting problem in Lebong Regency. The data variables of health facilities were used in the number of doctors, the number of nurses, the number of nutritionists, the number of integrated service health centers (Posyandu), the number of children under five who received exclusive breastfeeding. X. Han *et al.* [13] performed DBSCAN on the marine trajectory. The DBSCAN algorithm can be applied to historical or real-time Automatic Identification System (AIS) data so that vessel routes can be modeled, and the trajectories' anomalies can be detected, so does Sheridan *et al.*, who uses the DBSCAN method in-flight anomalies [14].

Scitovski & Sabo [15] researched various density data with the purpose of modification of the well-known DBSCAN algorithm. Fan *et al*. [16] conducted consumer clustering with geo-tagged social networks. In addition, this research method is also often compared with other clustering methods, such as the K-Means method in research Isnarwaty & Irhamah [17] research in 2019, Mustika *et al*. [18], Adha *et al*. [19], and also K-Medoids methods such as research by Mustakim *et al*. [20]. Indeed further discussing DBSCAN with nonlinear autoregressive (NAR) Neural Network [21], in this research, NAR is used to improve the accuracy of spatial load forecasting in the planning stage of the power grid.

Meanwhile, previous research related to the distribution pattern of hotels and tourism, including Hartawan's identifying the distribution pattern of hotels in the Ubud area [22]. Some of the findings in this research, among others, the tendency of the distribution pattern of star hotels in the Ubud Tourism Area varies depending on the class. The factors that affect the distribution pattern of star hotels are access and the potential for natural scenery. Barnad *et al*. [23] related to the spatial pattern of the distribution of objects and tourism supporting facilities in West Bali. Based on the study results, the spread of tourism support facilities forms a clustering pattern forming five groups of tourism support facilities in the West Bali National Park (TNBB) area with the highest density in the Gilimanuk Village area. The development of each cluster point is centered on Gilimanuk Village, with the direction of growth following the main path corridor that stretches along TNBB. Setyaningsih & Alam [24] described investigation on the impact and handling of the pandemic on sharia hotels. This research presents a comprehensive review of the impact of the covid-19 epidemic in Sharia hotels, as well as mitigation techniques.

Materials and Methods

Materials

The population in this study is the location of all hotels in Bali, while the sample used is the location of the hotel obtained using google maps in 2021. The sample is obtained using the web scraping technique. Table 2 is an illustration of the data used in the study.

Table 2. Illustration of the data

id	Hotel Name	Address	Website	Rating	Reviews	Longitude	Latitude
1	Handara Golf & Resort Bali	Desa Singaraja-Denpasar Pancasari Sukasada, Pancasari, Kec. Buleleng, Kabupaten Buleleng, Bali 81161	handaragolfresort.com	4.4	1683	115.157	-8.2485
2	Grand Zuri Hotels Kuta - Bali	Jl. Raya Kuta No.81, Kuta, Kabupaten Badung, Bali 80361	zhmhotels.com	4.6	3272	115.1815	-8.71168
3	The Tusita Hotel	Jl. Kartika Plaza No.9x, Kuta, Kabupaten Badung, Bali 80361	tusitabali.com	4	1153	115.1671	-8.7365
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
399	Clove Tree Hill - Bali Eco Retreat	Village, Jl. Kanciana, Gunung Salak, East Selemadeg, Tabanan Regency, Bali 82162	clovetreehillvillas.com	4.6	31	115.064	-8.43452

Based on Table 1, longitude and latitude data are used to find the density of the distribution of hotel locations. Longitude and latitude show the geographical coordinates of the location of each hotel. The longitude line measures the distance of a point's site from the prime meridian, while the latitude line measures the length of a point from the equator [25]. Based on Table 1, the longitude line values for all points are positive, which means that all locations are located to the right of the prime meridian. Meanwhile, the weight on the latitude line is negative because all sites are south/below the equator.

Methods

After obtaining data relating to the location of hotels in Bali, the data processing stage is continued. Figure 1 shows the research flow chart used in this study.

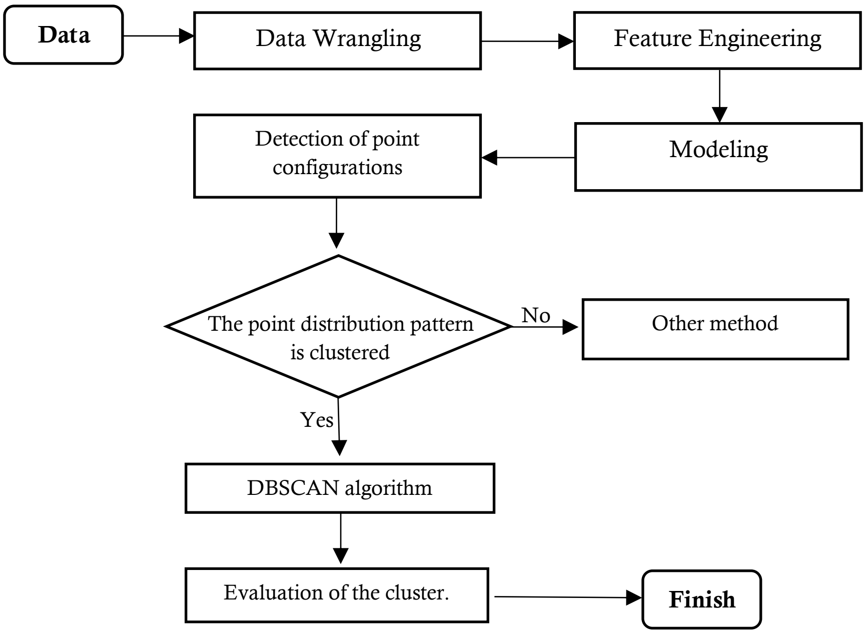

Figure 1. Research flow chart

The initial stage is data preparation/ data wrangling/ data preprocessing. The idea of data preparation is to execute/organize raw data into quality data [26]. At the data preparation stage, data profiling, data cleansing, and data anomalies are carried out. The second stage is Exploratory Data Analysis (EDA). EDA is used to convey initial investigations on the data to find patterns contained in the data to facilitate the analysis process. EDA techniques can be either graphical or non-graphical representations [27]. Then proceed with the feature engineering step.

The fourth stage is modeling. Two (2) main analyzes are used in this stage, namely: (1) detection of point configurations in space and (2) DBSCAN algorithm. Configure points in space using the quadrant method and Ripley's K-Function. This point configuration proposes to determine whether the point distribution pattern is random, regular, or clustered. The calculation phase of the quadrant method begins by dividing the area into m cells of almost the equivalent size. Then count the total occurrences in that area, say n. The next step is to calculate the average number of events per cell ($\bar{x}$), the value of the variance of the number of events per cell (s^2), and calculate the value of Variance-Mean Ratio (VMR), which is written in Equation 1 [28].

$$VMR = \frac{s^2}{\bar{x}} \tag{1}$$

$$\bar{x} = \frac{n}{m} \quad \text{and} \quad s^2 = \frac{\sum_{i=1}^{m}(x_i - \bar{x})}{m-1}$$

m: number of cells/grid, n: total occurrence, x_i: a number of events in cell i. The VMR value is the ratio of variance and the mean. If the value of $VMR = 0$ indicates the configuration of points in space is uniform or perfect regular. If the value of $VMR = 1$, it means that the structure of points in space is random. Likewise, if the value of $VMR < 1$ (the variance value is smaller than the average value), it indicates that the configuration of points in space is more inclined to the regular form. Meanwhile, $VMR > 1$ shows that points in space are more clustered than random.

The number of quadrants whose quadrant is less than 30, then $(m-1)\ VMR$ spreads following the Chi-Square distribution with degrees of freedom $m-1$. Chi-Square test statistics are presented in Equation 2.

$$\chi^2 = \frac{\sum_{i=1}^{m}(x_i - \bar{x})^2}{\bar{x}} \tag{2}$$

The null hypothesis (H_0) used is that the configuration of points in space is random, with statistic calculation = $(m-1)$VMR. Reject H_0 if the value $\chi^2 > \chi^2_{table}$. Meanwhile, if $m > 30$, then $(m-1)$ VMR spreads following the normal distribution $(m-1, 2(m-1))$ [29].

Besides the quadrant method, Ripley's K-Function is also used, which is also a method for analyzing the point distribution of the spatial structure [30]. The concept of Ripley's K-Function is to distinguish the distribution of point observations with a Poisson random model for all different radii [31]. Empirical Ripley's K-Function $\widehat{K}(r)$ results then visualized with a theoretical K-Function on the Poisson process $K_{pois}(r) = \pi r^2$, to obtain the plot presented in Figure 2.

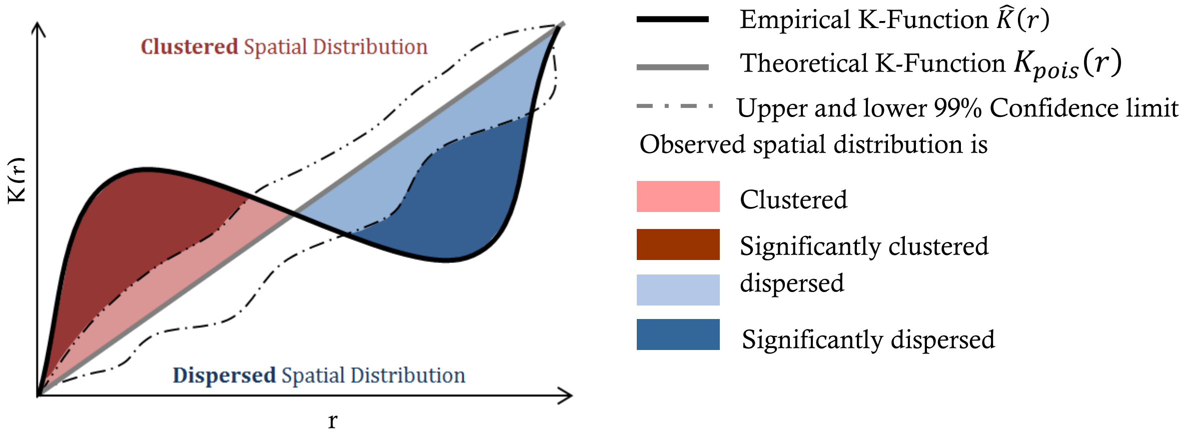

Figure 2. Interpret Ripley's K-Function [32]

Based on Figure 2, there are three (3) types of plots formed. The first condition is if the line in the empirical K-Function is below the theoretical K-Function ($\widehat{K}(r) < K_{pois}(r)$), this condition indicates that the type in this pattern has fewer neighbors than would be expected if the pattern is truly random. This pattern forms a regular pattern (points tend to avoid each other). If the empirical K-Function overlaps with K-Function theories ($\widehat{K}(r) = K_{pois}(r)$)) which means that the type of point in this pattern has the expected neighbors; it will enter independence (complete spatial randomness). The third condition is that the empirical K-Function is above the theoretical K-Function $\widehat{K}(r) > K_{pois}(r)$ which means that the type of point in this pattern has more neighbors than expected. This condition forms a group pattern. In the fourth stage, R software is used in statistical computing. This is because the R software has convenience, power, and other advantages (Rosadi, 2016). At the same time, the spatial visualization used Quantum Geographic Information System (QGIS) software. This is to make it easier to capture the general picture spatially.

If the distribution pattern of these points forms a cluster, then spatial clustering is continued using the DBSCAN Algorithm. The advantage of the DBSCAN algorithm is that it can detect outliers/noise and does not need to get input in the form of the number of clusters (k) as in the K-Means or K-Medoid methods. In addition, the wealth of this algorithm is that it can recognize difficult/irregular cluster [33].

DBSCAN illustrates a cluster as the maximum set of density-connected points. The cluster results from the DBSCAN algorithm will partition areas with high density and are considered clusters, while those with low density or are not included in the cluster are defined as noise [34]. DBSCAN settles the number of clusters generated, so there is no need to specify the desired number of clusters in advance. However, there are two main parameters in DBSCAN, namely: minimum points (MinPts) and epsilon (Eps). MinPts is defined as the minimum number of objects/ items/ points/ locations in a cluster. Eps is defined as the value for the distance between points which is the basis for forming the neighborhood of a point, or the radius that defines the boundary of the neighborhood from a certain point.

There are several terms in the DBSCAN algorithm, namely: core, border, density-reachable, directly density-reachable, density-connected, and noise. Suppose there is a neighborhood within the radius ($\in$), then it is called the neighborhood $\in$ of the object. If the neighborhood $\in$ of an object contains at least a minimum number, MinPts of an object, the object is called a core object [35]. In Figure 3, the core object is illustrated as a red point. The border is defined as an object that becomes a boundary in the core area (in Figure 3, it is illustrated as a blue point). The number of neighborhoods of the border points is less than the neighborhood of the core object [36].

An object is defined to be directly density-reachable from other objects if the distance between them is not more than the Eps value (within the radius of the distance). In Figure 3, the object is said to be directly density-reachable, namely in the illustration of the green point against the red point. An object is defined as density-reachable from another object if a path/chain connecting the two contains only objects directly density-reachable from the previous object. Object a is said to be density-connected to object b if there is an object o in the object set so that a and b are both density-reachable from o concerning Eps and MinPts (in Figure 3, they are illustrated as yellow points). In contrast, noise is defined as a set of objects that are not owned in any cluster (represented as a black dot). Figure 3 is an illustration of the border, core object, and density-reachable object.

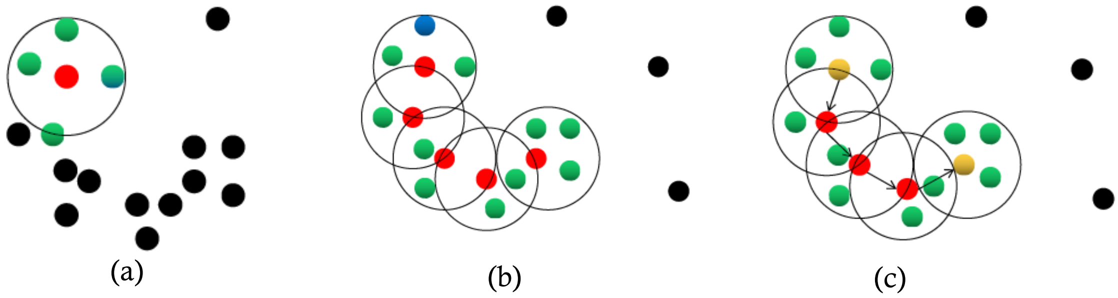

Figure 3. (a) a core object, (b) a border object, (c) a density-reachable object [37]

The essence of the DBSCAN algorithm is that for each object of a cluster, the neighborhood of a given radius contains at least a minimum number of objects; that is, the neighborhood's density must exceed the specified threshold [38]. The sequence of the DBSCAN algorithm generally includes six stages, namely: (1) choose a point p at random, (2) use Eps and MinPts to regain all points density-reachable from p, (3) if p is a core point, a cluster is formed, (4) if p is a border point, no points are density-reachable from p, and DBSCAN visits the succeeding point of the database, (5) remain the process until all of the points have been processed, (6) the result is independent of the order in which the points were [35]. The decision of MinPts and Eps is made iteratively. The best Eps determination is calculated from a distance around the elbow. After obtaining the optimal epsilon value, DBSCAN clustering was carried out so that clusters from each location were obtained based on their density and met MinPts and Eps, then visualized.

Later the modeling stage, the evaluation stage is continued. The evaluation stage of the cluster used is the average silhouette width. The estimate of the Silhouette method is presented in Equation 3 [39]. Suppose for data point $i \in C_i$ (data point i in the cluster C_i), let

$$s(i) = \frac{b(i) - a(i)}{\max\big(a(i), b(i)\big)} \quad (3)$$

$$a(i) = \frac{1}{|C_i| - 1} \sum_{j \in C_i, i \neq j} d(i,j) \quad \text{, and} \quad b(i) = \min_{k \neq i} \frac{1}{|C_k|} \sum_{j \in C_k} d(i,j)$$

$s(i)$ is the silhouette (value) at point i in a cluster. $b(i)$ the average distance of the i^{th} point to all data that is not in the same cluster with the i-th point. $a(i)$ the average distance of the i-th point to all data in one cluster. $|C_i|$ is the number of points belonging to cluster i. $d(i,j)$ is the distance between data object i and j in the cluster C_i. After obtaining the average silhouette width value, the average silhouette width value is interpreted based on Table 3 [40].

Table3. Range of silhouette values average width

Range	Interpretation
0.71 – 1	Strong structure
0.51 – 0.70	Medium structure
0.26 – 0.50	Weak structure
<0.25	No substantial structure was found.

The last stage is deployment. In this study, the deployment step is presented in an offline graph following each cluster and then adjusted to the actual conditions on the data contained. Because the data used is hospitality, it can be analyzed more deeply related to the nearest tourist object from the clustering results received.

Results and Discussions

The first stage is data preparation. It starts from the data profiling stage. Data profiling is carried out by defining and adjusting the data obtained and accompanied by data cleansing, namely the process of eliminating hotels that are not included in the point being studied, namely in Bali. In addition, the detection of anomaly data is done by adjusting the coordinates that are not suitable.

In the EDA stage, a graphical representation is used with the guidance of the QGIS program. QGIS is used, one of the free GIS software but has almost the same capabilities as paid GIS software in mapping and spatial analysis [41]. A graphic representation of the point distribution of hotels in Bali is presented in Figure 4.

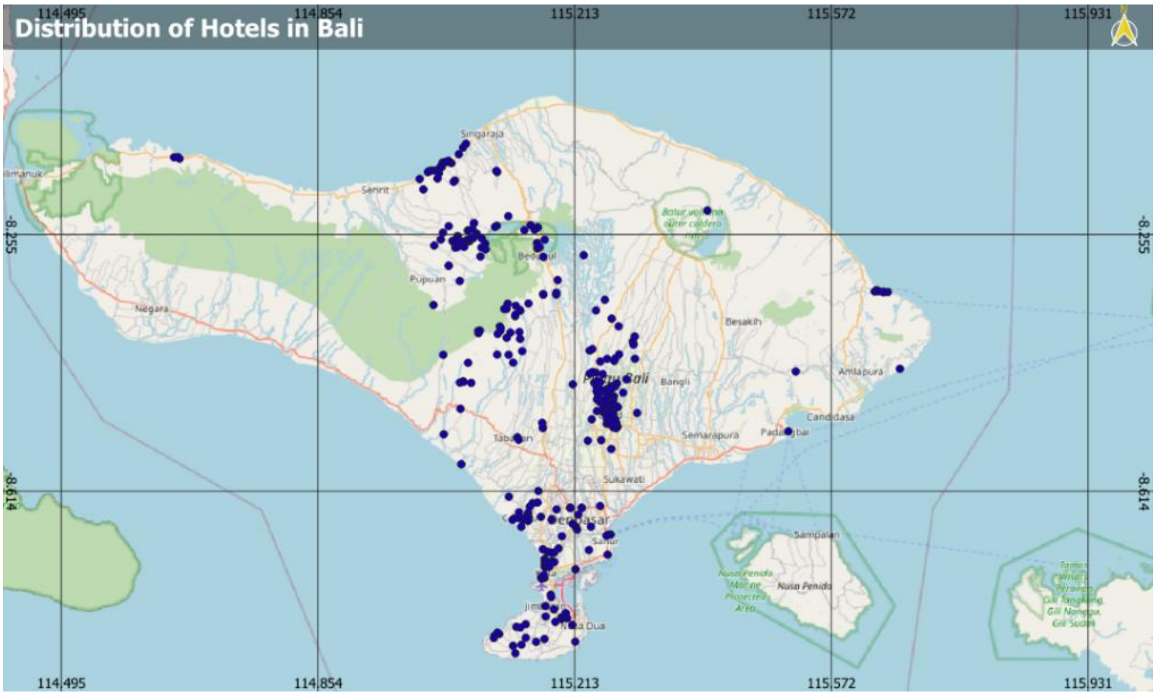

Figure 4. Distribution of hotels in Bali

In Figure 4, the blue point is the location of the hotel in Bali. At a glance, the distribution of hotels in Bali forms clusters in the northern, central, and southern parts of Bali. However, whether the hotel distribution includes a cluster or is random at the modeling stage will be proven.

Because this research focuses on the density pattern of hotels in Bali, at the feature engineering stage, coordinate variables (longitude and latitude) and hotel names are used from the data obtained. The coordinate system used is the geographic coordinate system. The distance of this coordinate variable will be calculated, which will later be used as one of the components that make up the DBSCAN algorithm.

Besides the coordinate variables, other variables such as an address, website, rating, and reviews in Table 1 are also used to support the results obtained.

Then the next stage is the modeling stage. Based on the above method, two main analyses were studied in this stage: (1) detection of point configurations in space and (2) DBSCAN algorithm. The space point configuration can use the quadrant method and Ripley's K test. Based on the results of the quadrant method by Equation 1, the value of $\chi^2 = 709.48 > 31.41 = \chi^2_{table}$indicates that the null hypothesis is rejected and is reinforced by the value of $VMR = 46.49 > 1$ which indicates the configuration of points in space is more clustered than random. Besides the quadrant method, this is supported from the visualization of Ripley's K-Function presented in Figure 5.

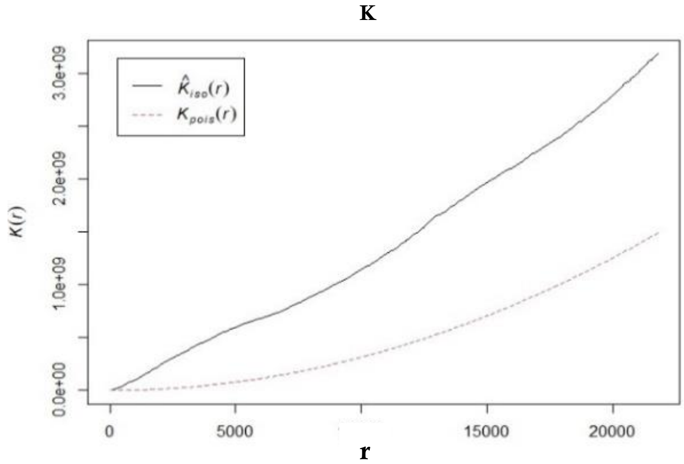

Figure 5. Ripley's K-Function

The empirical K-Function value ($\widehat{K}(r)$) is visualized as a black line, while the theoretical K-Function $K_{pois}(r)$ is visualized as a red dotted line. This is because of the value of $\widehat{K}(r) > K_{pois}(r)$. The next stage is the implementation of the DBSCAN algorithm. The DBSCAN algorithm uses two main parameters, namely MinPts and Eps. Before getting the two optimal parameters, k-Nearest Neighbor Distance (kNNdist) is calculated, which is defined as a quick calculation of the k-nearest neighbor distance in the points matrix. The results from kNNdist are used to make it easier to determine the appropriate eps and MinPts regions. The plot of kNNdist will later form a *knee* that can later be used to predict the optimal MinPts and Eps values. Figure 6 shows the kNNdist of the study with an illustration of the value of k=4.

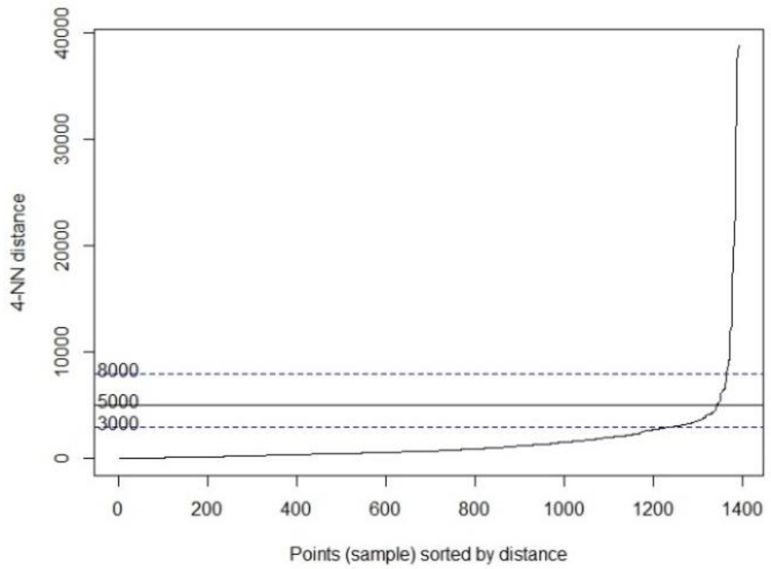

Figure 6. Plot kNN distance.

Based on Figure 6, the knee values are located in the 3000-8000 interval. Based on this, a combination of Eps values from 3000 to 8000 is carried out. In addition, a variety of MinPts from 3 to 8 MinPts is carried out. Evaluate the results using an average width silhouette by finding an enormous value of the MinPts and Eps combination pair. The test results are shown in Table 4.

Table 4. Evaluate the value of the silhouette average width

No	MinPts	Eps	Silhoutte Average Width	Number of cluster	Number of noise	Number of clustered points
1	3	3000	0.559	11	17	331
2	3	3050	0.559	11	17	331
3	3	3100	0.559	11	17	331
⋮	⋮	⋮	⋮	⋮	⋮	⋮
101	3	8000	0.503	4	4	344
102	4	3000	0.631	9	28	320
103	4	3050	0.631	9	28	320
⋮	⋮	⋮	⋮	⋮	⋮	⋮
412	7	3350	0.709	7	36	312
413	7	3400	0.707	7	35	313
414	7	3450	0.707	7	34	314
⋮	⋮	⋮	⋮	⋮	⋮	⋮
606	8	8000	0.52	2	16	332

From Table 4, the optimal average width silhouette value is obtained when the MinPts value = 7 and the Eps value = 3350, which forms 7 clusters with a silhouette average width value of 0.71 or is included in the category of a solid structure. The visualization of spatial clustering using the DBSCAN algorithm is presented in Figure 7.

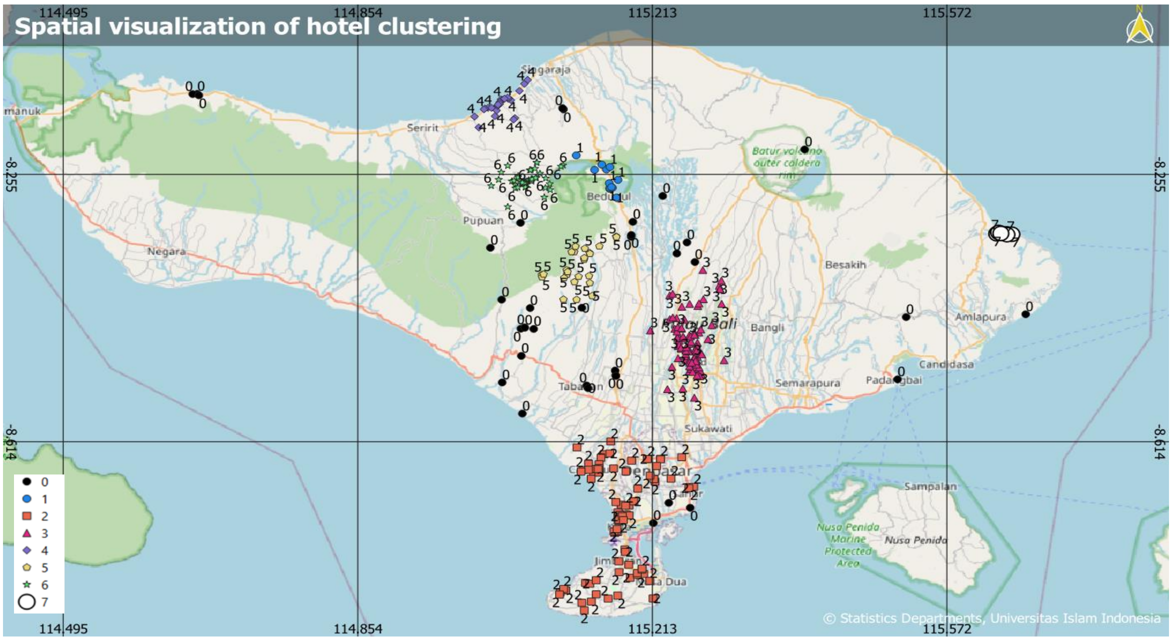

Figure 7. Spatial visualization of hotel clustering in Bali using the DBSCAN algorithm.

Based on Figure 7, the spatial clustering results of Bali hotels are presented, grouped into 7 clusters. Black points are noisy because they do not meet the specified MinPts and Eps. In comparison, the other points show the results of clusters in each region. The details of hotels from each cluster combined with several tourist attractions around the area are presented in Table 5.

Table 5. The results of the clustering of hotels and proximity to surrounding tourist objects.

No	Cluster	Number of hotels	Hotel name	Average hotel rating[1]	Proximity to tourist spots
1	1	13	Handara Golf & Resort Bali, CLV Hotel & Villa, Handara Golf & Resort Bali, ECOcamp Bedugul, Wanagiri Cosmic Nature Villa, Strawberry Hill Hotel and Restaurant, ECOcamp 2 Bedugul, Dajan Buyan Homestay, Bedugul camping, Little Lake Villa, Enjung Beji Resort, PonDaNu, Villa Sinta.	4.51	Lake Beratan Bedugul, Wanagiri Peak, Cinta Waterfall, Bali Botanical Garden Cactus Garden, Bali Botanical Garden.
2	II	83	Grand Zuri Hotels Kuta, The Tusita Hotel, The Aromas of Bali Hotel & Residence, The Kubu Hotel, Beneyasa Beach Inn 1, Coco De Heaven Guest House, Pecatu Guest House, The Hide Hostel @ Canggu, Loft Legian Hotel, Kiki Residence Bali, RedDoorz near Stadion Kompyang Sujana Bali, Tirtha Canggu Suites, Agus Beach Inn, BB Hostel Canggu, RedDoorz Hostel near Lippo Mall Kuta, OYO 4012 Ari Beach Inn, OYO 672 Bali Radiance Canggu, Bali Manik Beach Inn, Green Roof Homestay, Casa Dasa Boutique Hotel, RedDoorz near Teuku Umar Barat, Berlian Inn, Kubu Container, Surfers Dorm House Bali, Villa Ricca Eco Lodge, Caroline Guest House, RedDoorz near Exit Toll Nusa Dua, OYO 1952 Hotel Dewata Indah, Waringin Homestay, Wave & Chill House, OYO 1406 I Love Grass, OYO 1445 Jimbaran 12 Residence, OYO 90067 Hotel Nuansa Indah, RedDoorz @ Uluwatu Bali, Gusti Homestay Bali, Draper Startup House for Entrepreneurs, Swandewi Homestay, Cempaka 2 Accomodation, OYO 3261 Hotel Ratu, Akatara Stay Jimbaran, Ratu Guest House, Pudi Hostel, Ayu Lili Garden Hotel, OYO 90082 Gana Inn Legian, RedDoorz @ Pecatu Ungasan, Indopurejoy House, RedDoorz near Pantai Sanur Bali, Cottage DenAyu Homestay, Betesda Guest House, RedDoorz @ Sanur Bali Beach, A&W guest house, OYO 2280 Wisma Guntur II, DSTAY Kost Bali, OYO 3244 Grand Chandra Hotel, Kampung Canggu, Nan Berlian Inn, OYO 2367 Kenanga Homestay, OYO 90104 Lanata Hotel, Twenty4Seven Bed & Breakfast, Rai House Sanur, Hotel Tambora, Belong Bunter Homestay, OYO 1927 Hotel Candra Adigraha, SPOT ON 2426 Hotel Aget Jaya Ii, Sri Krisna, Mugeri Homestay, OYO 90096 Hotel Tiana, Maybe Tonight Kimberly D 17, Hang Five Hostel, Rumah D'Soka Nusa Dua, Anandianta Guesthouse, KoolKost @ Bypass Nusa Dua, Twins Homestay, Padang padang sari homestay, Made Roejas Home Stay, Abian Dedari Mesare, Cempaka Losmen, Suji Bungalow Hotel, Gong Corner Homestay, Putra Dhadi Guest House, OYO 90076 Guesthouse Graha Pande Residence, Paranyogan Homestay 2, Made Roejas Family Homestay, Belong Bunter Homestay 2, Gong Corner Guest House II, Repag Wayan Guest House, RedDoorz Hostel near Trans Studio Mall, Safe House Bed & Breakfast, Wika Bali Beach House.	4.12	Kuta, Benoa bay, Garuda Wisnu Kencana Cultural park, Sanur, Nusa Dua beach, Jimbaran beach, Pandawa beach, Tegal Wangi beach, Balangan beach, Uluwatu temple, Bali exotic marine park, upside down world Bali, Bali fun adventures.
3	III	102	Jannata Resort and Spa, The Evitel Resort Ubud, Daun Lebar Villas, The Sankara Suites & Villas by Pramana, Bali Bohemia, Four Seasons Resort Bali At Sayan, Aksari Resort, Element by Westin Bali Ubud, Bayad Ubud Bali Villa, Puri Sunia Resort, Halaman Depan Hostel, Hotel Tjampuhan, Best Western Premier Agung Resort Ubud,	4.54	Joger Bali, Sangeh, Elephant cave, Gianyar, Bali Zoo, Bali Bird Park, Sukawati, Badung market.

[1] Skala rating 1 sampai 5

No	Cluster	Number of hotels	Hotel name	Average hotel rating[1]	Proximity to tourist spots
			Umah Dauh Homestay, Sila Urip Guest House, Swan inn ubud, Beji Ubud Resort, Leket house, The Kayon Resort Ubud by Pramana, Nandini Jungle Resort and Spa Bali, The Payogan Villa Resort and Spa, The Udaya Resorts & Spa, Padma Resort Ubud, Pondok Sebatu Villa, Puri Sebali Resort, Arjuna Homestay Ubud, Sari Villa Ubud, The Lokha Ubud Resort, Villas & Spa Villa Santun, Pita Maha Resort & Spa, Royal Kamuela Villas & Suites at Monkey Forest, Ubud, Ubud Wins Bungalow, Hati Padi Cottages, Pering Bungalow, Desa Visesa Ubud, Puri Gangga Resort Ubud, Ashoka Tree Resort Ubud, Kuwarasan A Pramana Experience, Kupu Kupu Barong Villas & Tree Spa by L'Occitane, Padi Bali Eco Villas, Sastra Ubud House, Mandapa, a Ritz-Carlton Reserve, Wapa di Ume Ubud, Pramana Watu Kurung, In Da Lodge, Natya Resort Ubud, Dedary Kriyamaha Ubud by Pramana, Parks Dewangga House, The Kayon Jungle Resort by Pramana, Dwaraka The Royal Villas – Ubud, Alam Dania Cottage, The Royal Pita Maha, Kabera Bungalow, Suarapura Resort and SPA, Komaneka at Rasa Sayang, Amora Ubud Boutique Villas, Rijasa Agung Resort Villas, Di pondok, Tegaltis Ubud Villas, Kenari Guest House, Degeg ubud homestay, Parthi Puri Ubud Bed & Breakfast, Black Penny Villas, FuramaXclusive Resort & Villas, Ubud Bali, Teba House, COMO Uma Ubud, Ayu Homestay, Bucu hidden guest house, Shindu Homestay, Ben's Homestay, Alila Ubud, Damuh Guest House, Depong House, Song Broek Jungle Resort, Chapung Se Bali Resort and Spa, The Samaya Ubud Bali, Ubud Padi Villas, Ulun Ubud Resort, Sulendra Bungalow, Pratama House, Puri Wulandari A Boutique Resort & Spa JagaRoni House, Tegal Sari, Market Hostel, Bunut Garden Luxury Private Villa, Taman Indah Homestay, Ode Hostel, Nyoman House Ubud, Nova Homestay, Liang Homestay and Hostel, Jero Sebali Villa, Villa Neyang, Citrus Tree Villas - Sulendra Ubud, Nau villa ubud, Sangeh Uma Dong Loka Villa, Komaneka at Bisma, OPLES Homestay, Dekwah Homestay, Pondok Gepokan Homestay, Bidadari Private Villas & Retreat, Samsara Ubud, Ayani Good House, Arjuna Hous I, Komaneka at Tanggayuda		
4	IV	24	Adirama Beach Hotel, Frangipani Beach Hotel, Aditya Beach Resort, The Lovina Bali Resort, Padmasari Resort Hotel, Hotel Puri Saron Lovina, Odika Lovina House and Villa, Puri Bagus Lovina Resort, Bali Taman Beach Resort and Spa-Lovina, Hotel Melamun, Puri Mangga Sea View Resort & Spa, Nugraha Lovina Seaview Resort & Spa, Villa Teman Lovina, Banyualit Spa 'n Resort Lovina Bali, The Hamsa Resort, Ju'Blu Hotel, Bagus Beach Resort Lovina, Lovina Life Room & Café, Aneka Lovina Villas & Spa, The Damai, Family Hostel, Santhiku Hotel, Villas, Yoga & SPA, Funky Place Lovina, Lovina Central Hostel	4.35	Kroya waterfall, Lovina beach, Tembok Barak waterfall, Penimbangan beach tourism, Harbour Tourist area of Buleleng, Blue Lagoon, Singaraja city park, Kerobokan beach
5	V	25	Padi Bali Cottages, Jatiluwih Home Stay Bali, Kubu d'ume Homestay, Yoga Above the Clouds, Village Above The Clouds, Alassari Plantation, Sang Giri Mountain Glamping Bali, Ti Amo Bali, Buana Asti Homestay, Mountains side Villa, Bhuana Agung, Saridev'i Ecolodge, Tikanadi Homestay, Sarinbuana Eco Lodge, Damara villa jatiluwih, Sang Tirta Farm Living Resort, Pondok Yeh Ho, Suweden Homestay, Carik Tangis Botique Homestay, Batukaru Mountain Farmstay, Batukaru Coffee Estate, Tepi Sawah	4.62	Blemantung waterfall, Sanda Desa wisata, Besi Kalung waterfall, Belulang hot spring, Penatahan Kaja hot spring, desa wisata Pinge, The SILA'S Agrotourism, Angseri hot spring

No	Cluster	Number of hotels	Hotel name	Average hotel rating[1]	Proximity to tourist spots
			Lodge, Tegal Jero Homestay, Teras Subak Homestay, Dina Home Stay		
6	VI	49	Puri Lumbung Cottages, Restaurant & Spa - Munduk Village, Ekommunity, Villa Sande, Sanak Retreat Bali, Munduk Menir Villas, Atres Villa, Umah De Madya & Rooftop De Madya, Terrasse du Lac Tamblingan, Puri Alam Bali Bungalows, Munduk Heaven Luxury Villas, Munduk Moding Plantation Nature Resort & Spa, Munduk Moding Plantation Nature Resort & Spa, Gesing panorama guest house, Bali Rahayu Homestay, Munduk Sari, Munduk farm house, Bali De Pardis Cottages, Atres Sari Resort, The GARUDA villa & Restaurant, Taman Sari Homestay, Melanting Cottages, Puri Sunset Homestay, Muntis homestay, Swar Bali Lodge, Lesong Hotel and Restaurant, Puri Sunny Cottage, OYO 3264 Bukit Kembar Ecotourism, Amartya Puri Green Cottages, Villa Dua Bintang, Nadira Bali Villa, Geriya Siena, Adila Warung and Homestay, One Homestay, Jojo's Homestay, Dong Paloh Hostel, Pondok Asri Homestay, Munduk Retreat-Pondok Pekak Lelut, Nadya Homestay, Made Oka warung and homestay, La MontaÃƒÂ±a, Puri Sunny Guest House, Warung Made And Aris Homestay, Edy Homestay, Arya Utama Garden Villa, Manik Bulan Hotel, Dangin Mangkalan Homestay, Puri Sunny Camping, Gosela Homestay, Shitala Villa	4.51	Sleeping Budha statue, Santhipala waterfall, Munduk waterfall, Bali jungle tracking, Melanting waterfall,
7	VII	7	Bali Fab Dive center, Bali Yogi, Backpackers Home Amed, Amed stop Inn, Cahaya Melasti, Kputra Homestay, Kampung hostel	4.26	Bali Edelweys park, Penglipuran village, Bukit putung, Savana Tianyar, Gubug hill, Belong hill.

In general, the rating from tourists is excellent (rated more than four on a scale of 1-5). This is what makes Bali a paradise for tourists.Based on Figure 7 and Table.5, there is a buildup of hotels in the southern part of Bali (cluster II). This is because there are many beach spots of tourism in the area that are the main tourist attractions such as Kuta, Benoa bay, Sanur, Nusa Dua beach, Jimbaran beach, Pandawa beach, and Tegal Wangi beach, Balangan beach, etc. This cluster is a cluster that prioritizes tourist objects close to beaches in Bali. In addition, the arrangement of hotels in Bali also adjusts the five (5) priority programs of the Tourism Office in Bali, which include development that is patterned and integrated throughout Bali and development that is developed according to the potential of each Regency/City area [42] .

Cluster I and cluster VI show areas close to the Beratan Bedugul lake tourist attraction. The spread of this hotel offers a natural atmosphere that is not much related to mountainous regions and natural panoramas. Meanwhile, the areas in clusters III and V are central to the economy and trade or the city center of Bali. The area in cluster IV is a tourist area around Singaraja. Cluster VII also focuses on nearshore areas as well as cluster II. While in Figure 6, the black dot area is noise because it is not included in any cluster.

Based on Table 5, it can also provide input for stakeholders if they want to be initiated for economic recovery, which can be initiated from regions that are included in clusters II and III. This is because clusters II and III areas are the most tourist-intensive areas, both in terms of tourist objects and tourism supporting facilities. These results are in line with research on the distribution of hotels based on Hartawan's study [22] which suggests that there are factors that influence the distribution pattern of star hotels, namely access and the potential for natural scenery. In this context are hotels. As for furthermore, it can be continued with areas in other clusters. Apart from the density of points, recovery priorities can also be carried out based on the characteristics of tourist objects such as beaches, mountains, and shopping.

Conclusion

The results show that the distribution of hotels in Bali is more clustered than random. Based on the combination of MinPts and Eps values, the number of clusters formed in seven using the DBSCAN algorithm. From the evaluation results, the silhouette average width value is 0.709 or is included in the category of a strong structure in this spatial clustering. Each cluster shows the characteristics of each tourist attraction that is the main attraction for the spread of the hotel. It can help policymakers if they want to prioritize economic recovery after the Covid-19 pandemic.

Acknowledgement

The author thanks to Statistic Department, Faculty of Mathematics and Natural Science, Universitas Islam Indonesia which has assisted in the research process.

References

[1] C. M. Reinhart, This time trully is different, 2020. [Online]. Available: https://www.project-syndicate.org/commentary/covid19-crisis-has-no-economic-precedent-by-carmen-reinhart-2020-03?barrier=accesspaylog [Accessed 1 October 2021].
[2] M. Hotopf, E. Bullmore, R. C. O'Connor, and E. A. Holmes, The scope of mental health research during the COVID-19 pandemic and its aftermath, Br. J. Psychiatry, 217 (4) (2020) 540–542.
[3] E. A. Holmes et al., Multidisciplinary research priorities for the COVID-19 pandemic: a call for action for mental health science, The Lancet Psychiatry, 7 (6) (2020) 547–560.
[4] I. N. Subadra and H. Hughes, Pandemic in paradise: Tourism pauses in Bali, Tour. Hosp. Res 2021.
[5] L. S. Wedaningsih, N. U. Vipriyanti, W. Maba, and I. G. Y. Partama, Mapping the employee layoff of star hotels in Denpasar City: an effort to reduce the impacts of the Covid-19 pandemic, SOSHUM J. Sos. dan Hum., 11 (1) (2021) 100–111.
[6] Bali Statistics Agency, Tingkat Penghunian Kamar (TPK) Hotel Bintang Menurut Kelas di Provinsi Bali (Persen), 2021, [Online]. Available: https://bali.bps.go.id/indicator/16/230/1/tingkat-penghunian-kamar-tpk-hotel-bintang-menurut-kelas-di-provinsi-bali.html [Accessed 1 October 2021].
[7] World Health Organization (WHO), WHO Coronavirus (COVID-19) Dashboard, 2021. [Online]. Available: https://covid19.who.int/table?tableDay=yesterday [Accessed 1 October 2021].
[8] K. Mumtaz, M. Studies, and T. Nadu, An analysis on Density Based Clustering of multi dimensional spatial data, Indian J. Comput. Sci. Eng., 1 (1) (2010) 8–12.
[9] J. Sander, Density-Based Clustering, in Encyclopedia of Machine Learning, C. Sammut and G. I. Webb, Eds. Boston, MA: Springer US (2010) 270–273.
[10] J. Sander, M. Ester, H. P. Kriegel, and X. Xu, Density-based clustering in spatial databases: The algorithm GDBSCAN and its applications, Data Min. Knowl. Discov. 2 (2) (1998) 169–194.
[11] B. N. Sari and A. Primajaya, Penerapan clustering DBSCAN untuk pertanian padi di kabupaten Karawang, J. Inform. dan Komput 4 (1) (2019) 28–34.
[12] T. D. Harjanto, A. Vatresia, and R. Faurina, Analisis penetapan skala prioritas penanganan balita stunting menggunakan metode DBSCAN clustering, J. Rekursif, 9 (1) (2021) 30–42.
[13] X. Han, C. Armenakis, and M. Jadidi, DBSCAN optimization for improving marine trajectory clustering and anomaly detection, Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. - ISPRS Arch 43 (B4) (2020) 455–461.
[14] K. Sheridan, T. G. Puranik, E. Mangortey, O. J. Pinon, M. Kirby, and D. N. Mavris, An application of DBSCAN clustering for flight anomaly detection during the approach phase, AIAA Scitech 2020 Forum, 1 (PartF) (2020) 1–20.
[15] R. Scitovski and K. Sabo, DBSCAN-like clustering method for various data densities, Pattern Anal. Appl 23 (2) (2020) 541–554.
[16] T. Fan, N. Guo, and Y. Ren, Consumer clusters detection with geo-tagged social network data using DBSCAN algorithm: a case study of the Pearl River Delta in China, GeoJournal 86 (1)

(2021) 317–337.
[17] D. P. Isnarwaty and Irhamah, Text clustering pada akun twitter layanan ekspedisi JNE , J&T, dan Pos Indonesia menggunakan metode Density-Based Spatial Clustering of Applications with Noise (DBSCAN), J. Sains dan Seni 8 (2) (2019) 2–9.
[18] M. P. M, C. Dewi, E. P. Siam, G. A. Wijayanti, N. Aulia, and R. Nooraerni, Comparison of DBSCAN and K-Means clustering for grouping the village status in Central Java 2020, J. Mat. Stat. dan Komputasi 17 (3) (2021) 394-404.
[19] R. Adha, N. Nurhaliza, U. Soleha, P. Studi, S. Informasi, and F. Sains, Perbandingan algoritma DBSCAN dan K-Means clustering untuk pengelompokan kasus Covid-19 di dunia, J. Sains, Teknol. dan Ind 18 (2) (2021) 206–211.
[20] Mustakim, M. Z. Fauzi, Mustafa, A. Abdullah, and Rohayati, Clustering of Public Opinion on Natural Disasters in Indonesia Using DBSCAN and K-Medoids Algorithms, J. Phys. Conf. Ser 1783 (1) (2021) 1-6.
[21] Z. Han, M. Cheng, F. Chen, Y. Wang, and Z. Deng, A spatial load forecasting method based on DBSCAN clustering and NAR neural network, J. Phys. Conf. Ser 1449 (1) (2020) 1-6.
[22] I. P. Hartawan, Pola persebaran hotel resort di kawasan pariwisata Ubud, Jurnal Anala 9 (2) (2021)1–22.
[23] T. O. Barnad, I. G. A. A. R. Asmiwyati, and N. N. A. Mayadewi, Pola ruang sebaran objek dan fasilitas penunjang wisata berbasis sistem informasi geografis di kawasan Taman Nasional Bali Barat, J. Arsit. Lansek 7 (1) (2021) 66-75.
[24] S. Setyaningsih and A. Alam, Impact of Covid-19 pandemic on sharia hotels and their handling strategies (a case in Indonesia), International Conference on Islamic Economics, Islamic Finance, & Islamic Law (ICIEIFIL) (2021) 26–54.
[25] The Environmental Data Standards Council (EDSC), Latitude/ Longitude data standard 2006.
[26] Z. S. Abdallah, L. Du, and G. I. Webb, Data preparation, in Encyclopedia of machine learning and data mining, Second Edi., New York: Springer References (2017) 318–325.
[27] M. Komorowski, D. C. Marshall, J. D. Salciccioli, and Y. Crutain, Exploratory data analysis, in MIT Critical Data: Secondary Analysis of Electronic Health Records, 2nd ed., USA: Springer Open, (2016) 185–203.
[28] M. N. Aidi, Konfigurasi Titik dalam Ruang. Bogor: Institut Pertanian Bogor 2013.
[29] H. Schabenberger, Spatial count regression Repository.John Wiley & Sons.CRAN, 2009.
[30] E. Marcon, S. Traissac, and G. Lang, A statistical test for Ripley's K function rejection of poisson null hypothesis, ISRN Ecol 2013 (March) 1–9, 2013.
[31] A. Dennett, Analysing spatial patterns III: point pattern analysis, in Geocomputation 2020-2021 Work Book, University College London 2021.
[32] J. A. Lentz, Developing a Geospatial Protocol for Coral Epizootiology (2012) 99-101.
[33] M. T. Furqon and L. Muflikhah, Clustering the potential risk of Tsunami using Density-Based Spatial Clustering of Application with Noise (DBSCAN), J. Enviromental Eng. Sustain. Technol 3 (1) (2016) 1–8.
[34] P. B. Nagpal and P. A. Mann, Comparative study of density-based clustering algorithms, Int. J. Comput. Appl 27 (11) (2011) 44–47.
[35] I. M. S. Putra, Algoritma DBSCAN (Density-Based Spatial Clustering of Application with Noise) dan contoh perhitungannya (2018) 1-37.
[36] M. Ester, H. P. Kriege, J. Sander, and X. Xu, A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise, Compr. Chemom (1996) 226–231.
[37] H. Cui, W. Wu, Z. Zhang, F. Han, and Z. Liu, Clustering and application of grain temperature statistical parameters based on the DBSCAN algorithm, J. Stored Prod. Res 93 (April) (2021) 101819.
[38] Q. Ye, W. Gao, and W. Zeng, Color image segmentation using Density-Based Clustering, in IEE Intertational Conference on Acoustics, Speech, & Signal Processing (2003) 401-404.
[39] S. Prabakaran, M. Wayland, and C. Penfold, An introduction to machine learning 2017.
[40] P. J. Rousseeuw, Silhouettes: A graphical aid to the interpretation and validation of cluster analysis, J. Comput. Appl. Math 20 (C) (1987) 53–65.
[41] J. M. Flenniken, S. Stuglik, and B. V. Iannone, Quantum GIS (QGIS): an introduction to a free alternative to more costly GIS platforms, edis 2020 (2) (2020) 1-7.
[42] Bali Government Tourism Office, "Program Prioritas Dinas Pariwisata Bali," 2021. [Online]. Available: https://disparda.baliprov.go.id/progam-prioritas/. [Accessed: 02-Dec-2021]