

# Hotspot Identification Through Pick-Up and Drop-Off Analysis of Ride-Hailing Transport Service

Ragil Saputra<sup>1</sup>, Suprpto<sup>2\*</sup>, Agus Sihabudin<sup>3</sup>

Department of Computer Science, Universitas Diponegoro, Jl. Prof. Soedarto SH Tembalang Semarang Indonesia<sup>1</sup>  
Department of Computer Science and Electronics, Universitas Gadjah Mada, Bulaksumur Yogyakarta Indonesia<sup>1,2,3</sup>

**Abstract**—It is important to extract hotspots in urban traffic networks to improve driver route efficiency. This research aims to identify hotspot pick-up and drop-off (PUDO) areas in ride-hailing transportation services using a clustering approach. However, there are challenges in applying clustering algorithms to trajectory data in the coordinates of the Global Positioning System (GPS). So this research proposes modifications to the Density-Based Spatial Clustering of Applications with Noise (DBSCAN) algorithm by considering the radius from the center of the cluster to determine the presence of amenities around the cluster. We used a dataset containing 55,988 trip trajectories of Grab drivers over a two-week period in Jakarta. A preliminary statistical analysis was carried out to understand the distribution of trips. Next, we identify the PUDO point of each trip for use in the clustering analysis. The research explores the various parameters and settings of the clustering method and their impact on the results. The study found that the results obtained from the clustering method are sensitive to parameter selection, including epsilon radius and minimum number of points needed to form a cluster. The optimal cluster with the best parameters (eps: 0.25, minpts: 100) in the pick-up (PU) location analysis produced 17 clusters with the silhouette coefficient of 0.752, while in the drop-off (DO) location there are 18 clusters with a silhouette coefficient of 0.694. Overall, the research highlights the potential of the clustering analysis method for ride-hailing transportation.

**Keywords**—Hotspot identification; ride-hailing; transportation; PUDO location; clustering analysis

## I. INTRODUCTION

### A. Background

The rise of application-based transportation services, such as ride-hailing has revolutionized the way people move around in modern cities [1]. In the context of transportation services, "pick-up" often denotes the time when a vehicle (such as a taxi, or ride-hailing service) arrives at a designated location to collect passengers or packages. While "drop off" can refer to the time when a vehicle arrives at the destination location, and passengers or packages are left at that location.

The exponential growth in the use of these services has created enormous opportunities to analyze and understand urban mobility patterns [2]. In the study of Zhang et al [3], mobility-based data analysis has become the main focus in uncovering complex urban movement patterns. There have been many studies that use taxi data to examine urban mobility, such as Veloso et al [4] use of taxi trajectory data from Lisbon to discuss the spatiotemporal variation of taxi services, correlations between pick-up and drop-off (PUDO) sites, and

driver behavior, and Liu et al [5] use of taxi trajectory data from Shanghai to explore human movement patterns. The taxi traces were also mined for characteristics and behaviors of the vehicular network, including anomalous and driver behavior, dynamics of mobility patterns, interactions between vehicles, and the relationship between gender and mobility [6]. Additionally, Keler et al [7] examined where automobile routes connect at specific rush hours in urban locations.

By identifying origin and destination (OD) flow clusters in urban travel data, it is possible to determine prospective routes for public transportation service settings [8]. In order to locate the taxi OD hotspot, the available OD pairs from empirical mobility traces are first grouped [9]. A deep understanding of hotspots, namely areas with heavy travel activity, has great potential for shaping more efficient transportation planning and better traffic management [10]. Although there have been previous studies exploring hotspot analysis in urban mobility, Dutta et al [11], use of more sophisticated and comprehensive clustering methods is still an important area of exploration.

Exploring urban mobility patterns has become a main focus in recent years, with previous research analyzing trajectory data [12]. Therefore, in this context, it is essential to define several key terms and assumptions that will form the basis of this study before presenting the main theorem. We define a hotspot as a location with a high concentration of PUDO points are adapted from [13]. We assume that the dataset used in this study is representative of the overall ride-hailing transportation system in the study area. Additionally, we assume that variables like the time of day, day of the week, and the location of well-known destinations have an impact on drivers' PUDO patterns.

### B. Motivation

The majority of previous research on ride-hailing has been on the routing method, with little emphasis paid to optimizing PUDO locations that are sensitive to the spatial and temporal need distribution [14]. When taking a shared trip with multiple riders, the PUDO optimization is vital to prevent pointless detours. Whereas in the conventional system, the vehicle is frequently obliged to pick up passengers at certain locations [15]. Detours are made in order to pick up additional riders frequently result in longer travel times and higher costs. The PUDO position, where all potential PUDO points must therefore be optimized urgently.

Shen et al [2], suggests a cluster analysis approach for identifying various metropolitan online ride-hailing operation trends. Based on the suggested intensity and stability indicators

of ride-hailing vehicle operational characteristics, k-means++ clustering technique is applied. The results show that there are three distinct operating patterns for online ride-hailing services. In the study of Zhang et al [3], the goal is to identify the distribution of areas with high travel demand as well as the relationship between travel demand and Point of Interest (POIs). Tang [16], utilizes taxi GPS trajectory data to analyze urban human activity and mobility in Harbin city. The researchers employ the DBSCAN algorithm for PUDO location clustering and develop four spatial interaction models to understand pick-up location searching behavior.

However, with the rise of ride-hailing services, there is a need to develop new methods for analyzing PUDO patterns of drivers. Gunawan and Susilawati [17], addresses limitations in current ride-hailing PUDO location selection practices, which often prioritize spatial distribution and company interests over passenger needs. Research using neural networks was carried out by [18], seeks to examine the integration of clustering models and deep learning techniques. The model, which can concurrently capture the spatial and temporal fluctuations of taxi hotspots, was proposed for taxi hotspot prediction.

There have been several studies interested in discussing the applications for the clustering analysis method in transportation systems. Zhang et al [19], used DBSCAN to cluster PUDO data from a ride-hailing service in China. A pick-up points recommendation model (PPRM) is introduced, utilizing DBSCAN to cluster historical orders. This clustering enables finding contextually relevant candidate PU points. The other research by Wang and Ren [20], introduces a two-level divide approach and enhances the K-means++ algorithm to refine the clustering of taxi passenger hot spots based on GPS location data. The method is validated using a week of New York City's green taxi data, demonstrating superior accuracy and comparable time efficiency when compared to traditional K-means and DBSCAN methods.

Based on previous research, clustering analysis method has become an increasingly popular approach in the field of data mining and machine learning. Rafiq and McNally [21], use clustering these data points, ride-hailing companies can gain insights into traffic patterns and usage trends of their customers, which can help them optimize their operations and improve their overall service quality.

To build on prior research that examined trajectory data to investigate urban mobility patterns, this study employs a clustering approach to identify hotspot PUDO areas for ride-hailing transportation services. The DBSCAN algorithm was

chosen due to its ability to identify clusters of varying shapes and sizes. However, it has limitations in handling spatial datasets, therefore, this study proposes modifying the DBSCAN algorithm to take into account the radius from the cluster's center to determine the presence of facilities in the cluster's vicinity. Our method identifies concentrations of PUDO locations that can be used to determine potential hotspots. While this study does not provide a direct comparison with previous methods, the proposed method is superior to previous approaches because it considers the radius from the center of the cluster to determine the presence of amenities around the cluster, uses a large dataset of ride-hailing trajectories.

The paper is structured as follows. Section II describes data processing, and proposed method. The findings and results gleaned from the methodology are discussed in Section III. Section IV concludes by summarizing the work and offering recommendations for the future.

## II. METHODOLOGY

### A. Data Collection

The dataset is derived from Grab's food delivery and logistics in Jakarta, and it includes 4,000 daily trajectories collected from 2019-04-08 to 2019-04-21 (inclusive, UTC/ Universal Time Coordinated). The trajectories were gathered from drivers' phones while they were on the road. The total number of GPS pings in the collection is 61.549.964; each GPS ping has values for its trajectory ID, latitude, longitude, timestamp, accuracy level, bearing, and speed [22]. The raw data sample is shown in Table I.

The names of the fields in Table I are covered below.

- 1) *Trajectory ID*: A number used to identify different GPS mobility trajectories.
- 2) *Latitude*: A GPS location's latitude coordinate.
- 3) *Longitude*: A GPS location's longitude coordinate.
- 4) *Timestamp*: Time the GPS locations were recorded is shown by the timestamp. The UTC standard is used in the format. One second is the quickest sample interval for each GPS point.
- 5) *Accuracy level*: Shows the radius of the circle that, with a certain probability, contains the real location.
- 6) *Bearing*: The degrees relative to true north.
- 7) *Speed*: The immediate speed is expressed in meters per second.

TABLE I. TEST RAW DATA

Trajectory ID	Latitude	Longitude	Timestamp	Accuracy Level	Bearing	Speed
4820	-6.591032	106.834468	09/04/19 10.33	4	194	20.42
46324	-6.247526	106.977663	10/04/19 05.22	10.72	101	4.22
15007	-6.267404	107.036016	19/04/19 02.57	3.149	108	1.99
4239	-6.293342	106.820029	10/04/19 01.31	10	288	1.01

### B. Data Preprocessing

The next step is to pre-process the data to create trajectory, the data includes the point of trajectory of captured trips. The trips were divided based on the trajectory ID, and the minimum and maximum trajectory times recorded for each trip were used to determine the PUDO information. The result of dataset contains in total 55,988 trajectories with the distribution of number of trips is show in Fig. 1.

The daily distribution of trips is depicted in Fig. 1. We simply display the relative frequency of travel requests discussed in this article in order to maintain data confidentiality. As seen in Fig. 1(a) pick-up (PU) time, the temporal travel demand is typically distributed over the normal working day. Between 6 P.M and 12 noon, it is relatively high, but after that, the tendency begins to decline until 9 P.M. The trend then went upward once more until 10 P.M. The same trend also occurs in drop-off (DO) time, presented in Fig. 2(b).

### C. Proposed Method

The main analysis involves applying the Density-Based Spatial Clustering of Applications with Noise (DBSCAN) method to the preprocessed data to identify clusters of PUDO locations. DBSCAN was introduced by Ester et al [23], which is a clustering algorithm commonly used in data analysis to identify clusters of data points based on their spatial density.

It's particularly useful when dealing with data where clusters might have irregular shapes and varying densities. The algorithm categorizes data points into three main types [24]: core points, border points, and noise points. A core point is a data point that has at least a specified number of neighbor points (*minpts*) within a certain distance ( $\epsilon$ ). These points are at the heart of a cluster. A border point is a point that is within the  $\epsilon$  distance of a core point but doesn't have enough neighbors to be considered a core point itself. And finally, noise points are any points that are neither core nor boundary points.

The following steps make up the algorithm [25]:

- 1) Identify the core points or points visited by more than *minpts* neighbors by locating all neighbor points within  $\epsilon$ .
- 2) Make a new cluster for each core point if it hasn't already been done so.
- 3) Find all points connected to it by density recursively, then group them with the core point in the same cluster.
- 4) Points *a* and *b* are said to be density connected if point *c* has a significant number of points in its neighbors and both of those points are close to the  $\epsilon$ . Chains are used in this process. Inferring that *b* is a neighbor of *a*, if *b* is a neighbor of *c*, *c* is a neighbor of *d*, and *d* is a neighbor of *e*, then *b* is a neighbor of *a*.
- 5) Iteratively go over the remaining unexplored points in the dataset. All points that do not form a cluster are considered noise.

In this paper, we modified the DBSCAN algorithm from Kambe and Pe [25] to better suit the ride-hailing context by creating a function called DBSCAN\_FIT that takes three parameters: *x* (dataset),  $\epsilon$  (epsilon, the maximum distance between two points to be considered in the same cluster), and *minpts* (the minimum number of samples in a cluster). This function utilizes the DBSCAN algorithm to cluster the data and produces visualizations of the clusters, including those labeled as noise (points not belonging to any cluster). Additionally, the function calculates centroids for each cluster and assesses the amenity around each centroid. The pseudocode for the DBSCAN\_FIT is shown in Algorithm 1.

To obtain the amenity information with function *get\_amenity*, we used OpenStreetMap (OSM) data to identify the facilities or points of interest around each centroid. Specifically, we used the OSM API to query the database for the amenities within a certain radius of each centroid. We then used the OSM tags to extract the names of the amenities and assigned them as labels to the corresponding clusters. At a given average latitude on Earth, we use the number 111,320 as a conversion factor to translate differences in degrees of latitude or longitude into distances in kilometers.

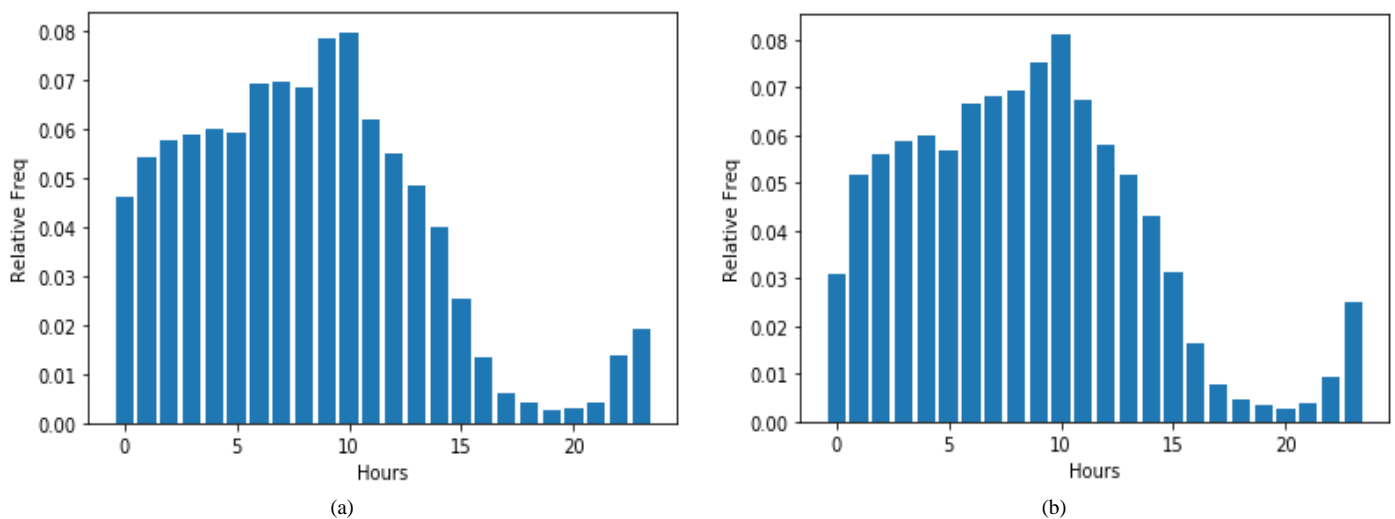


Fig. 1. Distribution average of (a) pick up time and (b) drop off time.

Algorithm 1. DBSCAN\_FIT

```
Input   : x (dataset), eps, minpts
Output  : labels, centroid, amenity
Function DBSCAN_FIT(x, eps, minpts){
    dbs = DBSCAN(eps,minpts)
    dbs.fit(X)
    labels = dbs.labels_
    label_unique = pd.Series(labels).value_counts()
    if label_unique.shape[0]>2 {
        cent = []
        amen = []
        for i in range(max(labels)+1){
            centroid = (X[labels==i].mean().values/111.320) + np.array([min_lat,min_lon])
            centroid += [centroid]
            amenity = get_amenity(centroid[0], centroid[1], eps*4*1000)
            amenity += [amenity]
        }
    }
    else
        return 0
return labels, centroid, amenity
}
```

D. Cluster Performance Measure

To evaluate the quality and coherence of clusters obtained through clustering algorithms, the Silhouette Coefficient (SC) was used. It quantifies the cohesion and separation between clusters. Ranging from -1 to 1, a higher SC indicates well-defined clusters where data points are closer to their own cluster members than to others. A score close to 0 suggests data points on cluster boundaries, while negative scores indicate potential misassignments [26].

A higher SC would indicate well-defined clusters of PUDO points, highlighting their coherence and separation from other clusters. This metric becomes essential in assessing the accuracy of clustering results and validating the effectiveness of algorithms. The silhouette score guides the determination of how accurately identified hotspots represent distinct patterns, thereby enhancing the credibility of the hotspot identification approach in ride-hailing transport services.

III. RESULT AND DISCUSSION

A. Experimental Result

Our research focuses on the complex interaction of epsilon ( $\epsilon$ ) values in the DBSCAN\_FIT algorithm in order to do a thorough study of clustering results. To achieve this goal, we conducted a preliminary analysis of the data and found that *minpts* values below 50 did not produce enough clusters to be useful for hotspot identification, while values above 100 resulted in too many clusters, making it difficult to identify meaningful patterns. Therefore, we chose to focus on *minpts* values of 50 and 100 in our analysis and visualization, as these values produced the most meaningful results for hotspot identification. In Fig. 2, the outcomes of this investigation are graphically summarized. Fig. 2(a) shows the silhouette value for pick up and Fig. 2(b) shows the silhouette value for drop-off.

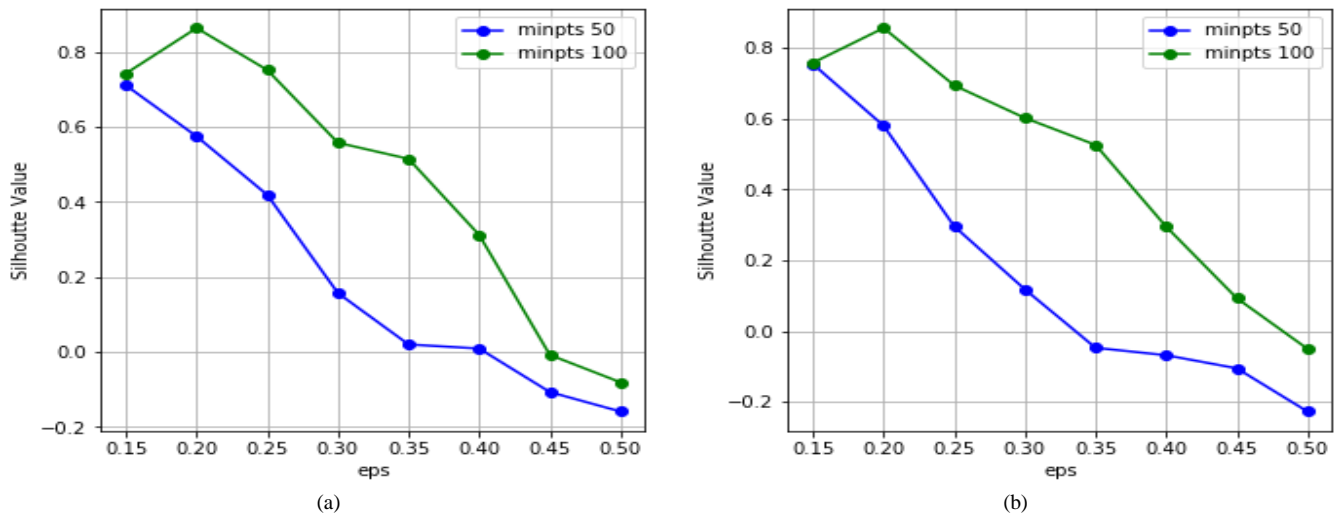


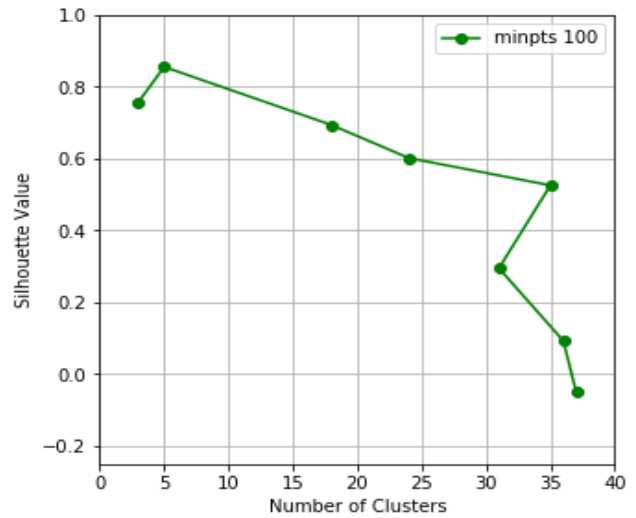
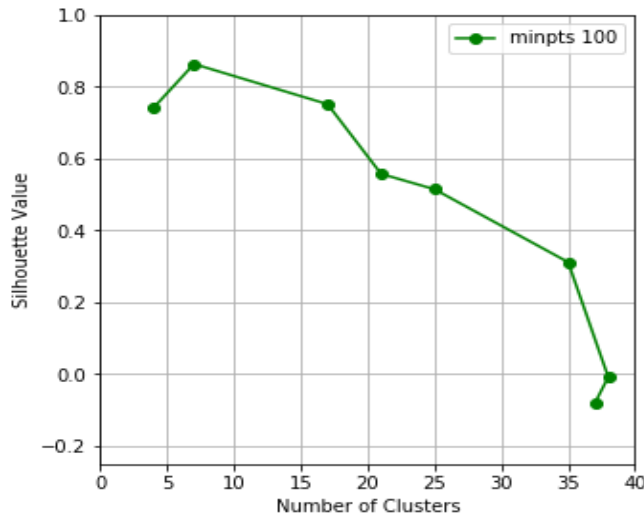
Fig. 2. Variation epsilon ( $\epsilon$ ) value for the (a) pick up and (b) drop off.

The graph on Fig. 2, the graph illustrates the variation in the silhouette coefficient with changing values of  $\epsilon$ . Two curves are depicted, one for a minimum number of points (*minpts*) set at 50, and the other for *minpts* set at 100. The results indicate that, for *minpts* 50, the silhouette coefficient (SC) tends to decrease as  $\epsilon$  increases, suggesting a negative impact on clustering quality. In contrast, for *minpts* 100, the SC exhibits fluctuations, with some  $\epsilon$  values resulting in higher coefficients. From the graph, it can be inferred that for *minpts* 100, the SC is more stable and potentially yields better clustering results compared to *minpts* 50 within a specific range of  $\epsilon$  values.

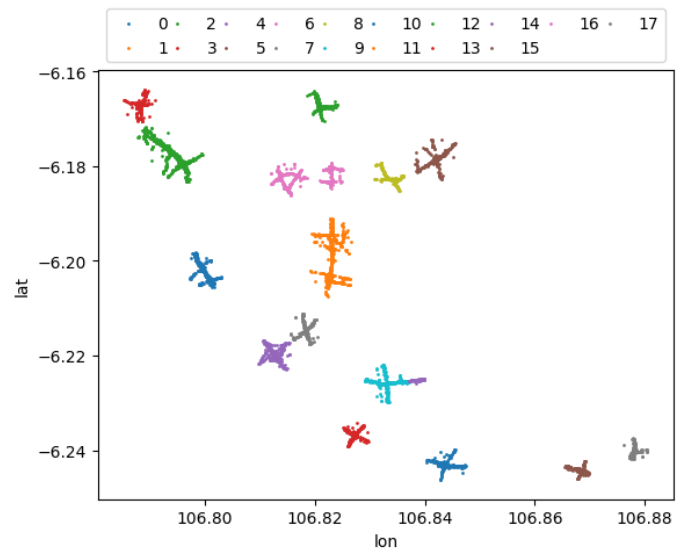
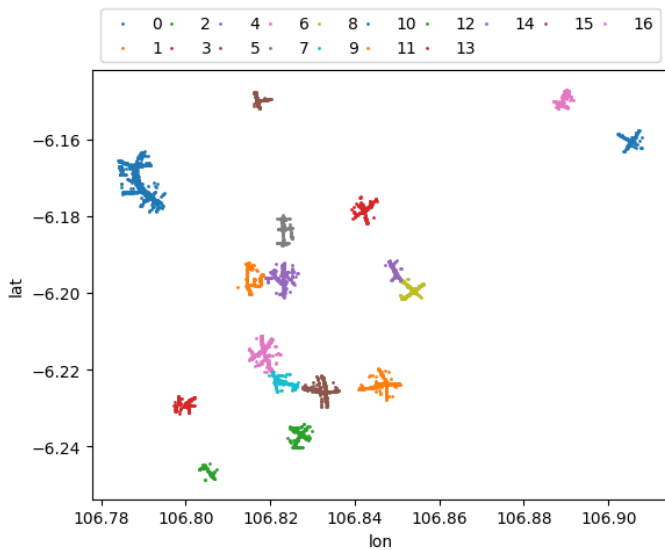
Furthermore, it is evident that the *minpts* value plays a pivotal role. The *minpts* 100 configuration consistently

outperforms *minpts* 50, yielding higher SC values. So in this study *minpts* 100 is used as a reference to find the optimal number of clusters. Compare the value of the SC with the number of clusters formed presented in the Fig. 3.

The number of clusters formed ranging from 3 to 37 presented by Fig. 3. In the context of hotspot identification, a higher number of clusters can provide a more detailed picture of variations in the distribution of hotspots. However, on the other hand, the SC value is also important because it determines the quality and coherence between clusters. So the optimal number of clusters was chosen as 17 clusters for the PU location and 18 clusters for the DO location. So that the map distribution of PUDO locations is illustrated in Fig. 4.



(a) (b)  
Fig. 3. Silhouette coefficient and number of clusters of (a) pick up and (b) drop off.



(a) (b)  
Fig. 4. The map distribution of (a) pick up location and (b) drop off location.

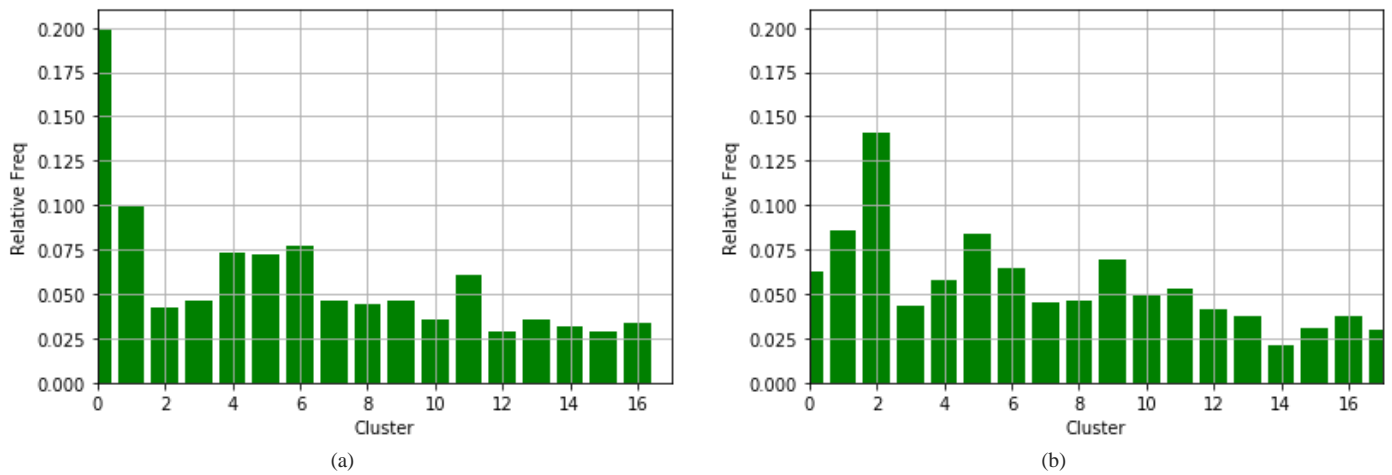


Fig. 5. Number of trips (relative frequency) in cluster (a) pick up and (b) drop off.

The distribution of PUDO locations inside each cluster is shown in the Fig. 5. The cluster number is represented on the x-axis, while the number of trips inside each cluster is represented on the y-axis, in relative frequency. This information is crucial in our effort to locate hotspots for application-based transportation services. These findings are directly related to our study's primary objective, which was to develop a special technique for identifying hotspots by closely analysing PUDO patterns. We have learned a lot from the clustering analysis. With the parameters  $\epsilon$ : 0.25 and *minpts*: 100, the ideal cluster structure specifically produced 17 clusters for the PU localization analysis, showing an outstanding SC of 0.752. On the other hand, the DO location analysis revealed 18 clusters with a silhouette coefficient of 0.694. These clusters have varied levels of activity, as seen by the bar chart in Fig. 5, indicating the existence of distinct hotspots throughout the transportation network.

Based on Fig. 4 and Fig. 5 show that the highest hotspots for both pick-up and drop-off sites. It follows that these places represent prospective areas that could be used to enhance driver trips. Using the methods we propose, most potential hotspots are presented in Table II along with their amenity labels.

**B. Discussion**

To assess the performance of clustering, we conducted a comparative analysis with two distinct inputs values: 50 and 100, while exploring the influence of the  $\epsilon$  parameter across a range from 0.15 to 0.5. The Silhouette Coefficient (SC) served as our evaluation metric [26]. Our findings are consistent with previous studies that have examined the impact of the  $\epsilon$  parameter on DBSCAN clustering [23] [24]. For example, [23] found that larger values of  $\epsilon$  can lead to the formation of overly large clusters, which can reduce the effectiveness of the clustering algorithm. Similarly, [24] found that larger values of  $\epsilon$  can lead to a decrease in the quality of the clustering results. Our study builds on these findings by examining the impact of both the  $\epsilon$  parameter and the *minpts* value on the performance of clustering in the context of ride-hailing services.

One important aspect of the study is the influence of algorithm parameters on the clustering results. The study found

that the results obtained from the clustering method are sensitive to parameter selection, including epsilon radius and minimum number of points needed to form a cluster. The study explored the influence of the  $\epsilon$  parameter across a range from 0.15 to 0.5 and found that for *minpts* 50, the SC tends to decrease as  $\epsilon$  increases, suggesting a negative impact on clustering quality. In contrast, for *minpts* 100, the SC exhibits fluctuations, with some eps values resulting in higher coefficients.

Although our study does not explicitly mention the most surprising results, we found that our approach of analyzing pick-up and drop-off (PUDO) patterns was effective in identifying hotspots in ride-hailing transport services. Our analysis revealed commuting patterns of users and different hotspots in the transportation network. In this study, several limitations were identified that require careful consideration. Firstly, the dataset utilized in this research was limited to a single ride-hailing service, which may limit the generalizability of the analysis results to other services. Secondly, this study only accounted for factors such as time and location in the PUDO analysis, while other factors such as weather or special events in certain areas may influence PUDO patterns and were not considered in this study.

TABLE II. THE POTENTIAL HOTSPOT

Cluster	Type	Center Latitude	Center Longitude	Amenity
0	PU	-6.2210668	106.824997	restaurant: Planet Hollywood Jakarta / cinema: Setiabudi 21 / fuel: SPBU / restaurant: Loewy / restaurant: Bakso Solo
2	DO	-6.1999132	106.824234	nightclub: DanceSignal / restaurant: Spot - Immigrant / parking_entrance: restaurant: Lanna Thai / cafe: nan / restaurant: Skye

**IV. CONCLUSIONS**

Our study reveals that DBSCAN\_FIT clustering with  $\epsilon$ : 0.25 *minpts*: 100 and yields 17 clusters for PU locations and 18 clusters for DO locations, demonstrating their potential as

hotspots in ride-hailing services. However, the limitation lies in the trade-off between cluster quantity and quality. A more comprehensive understanding of ride-hailing hotspots is achieved, emphasizing the need for a balance between cluster granularity and silhouette coefficients. Other researchers can use our method of analyzing hotspots through PUDO patterns to find hotspots in other transportation networks. This approach can help improve efficiency and help drivers optimize the routing of transportation services.

Future studies should consider utilizing datasets from multiple ride-hailing services to enhance the generalizability of the analysis results. And, additional factors such as weather or special events in certain areas should be taken into account in the PUDO analysis to provide a more comprehensive understanding of the factors that influence PUDO patterns.

#### ACKNOWLEDGMENT

This work was supported by Directorate General of Higher Education, Research, and Technology, the Ministry of Education, Culture, Research, and Technology Indonesia under Grant No. 1911/UN1/DITLIT/Dit-Lit/PT.01.03/2022.

#### REFERENCES

- [1] X. Shen, L. Wang, Q. Pei, Y. Liu, and M. Li, "Location Privacy-Preserving in Online Taxi-Hailing Services," *Peer-to-Peer Netw. Appl.*, vol. 14, no. 2021, pp. 69–81, 2021.
- [2] Z. Xiong, Jian Li, and H. Wu, "Understanding Operation Patterns of Urban Online Ride-Hailing Services: A Case Study of Xiamen," *Transp. Policy*, vol. 101, no. August 2020, pp. 100–118, 2021, doi: 10.1016/j.tranpol.2020.12.008.
- [3] B. Zhang, S. Chen, Y. Ma, T. Li, and K. Tang, "Analysis on Spatiotemporal Urban Mobility Based on Online Car-Hailing Data," *J. Transp. Geogr.*, vol. 82, no. February 2019, p. 102568, 2019, doi: 10.1016/j.jtrangeo.2019.102568.
- [4] M. Veloso, S. Phithakkitnukoon, and C. Bento, "Sensing Urban Mobility with Taxi Flow," in *Proceedings of the 3rd ACM SIGSPATIAL International Workshop on Location-Based Social Networks*, 2011, pp. 41–44, doi: 10.1145/2063212.2063215.
- [5] Y. Liu, C. Kang, S. Gao, Y. Xiao, and Y. Tian, "Understanding Intra-Urban Trip Patterns From Taxi Trajectory Data," *J. Geogr. Syst.*, vol. 14, no. 4, pp. 463–483, 2012, doi: 10.1007/s10109-012-0166-z.
- [6] F. A. Silva, C. Celes, and A. A. F. Loureiro, "Filling the Gaps of Vehicular Mobility Traces Categories and Subject Descriptors," *Proc. 18th ACM Int. Conf. Model. Anal. Simul. Wirel. Mob. Syst. (MSWiM '15)*, pp. 47–54, 2015.
- [7] A. Keler, J. M. Krisp, and L. Ding, "Detecting Vehicle Traffic Patterns in Urban Environments Using Taxi Trajectory Intersection Points," *Geo-Spatial Inf. Sci.*, vol. 20, no. 4, pp. 333–344, 2017, doi: 10.1080/10095020.2017.1399672.
- [8] M. Fang, L. Tang, Z. Kan, X. Yang, T. Pei, and Q. Li, "An Adaptive Origin-Destination Flows Cluster-Detecting Method to Identify Urban Mobility Trends," pp. 1–17, 2021, [Online]. Available: <http://arxiv.org/abs/2106.05436>.
- [9] E. R. Magsino, A. J. Abello, and J. M. Lalusin, "Taxi Hotspots Identification through Origin and Destination Analysis of Taxi Trips using K-means Clustering and H-indexing," *J. Phys. Conf. Ser.*, vol. 1997, no. 1, 2021, doi: 10.1088/1742-6596/1997/1/012006.
- [10] S. Mohammed, A. H. Alkhereibi, A. Abulibdeh, R. N. Jawarneh, and P. Balakrishnan, "GIS-based spatiotemporal analysis for road traffic crashes; in support of sustainable transportation Planning," *Transp. Res. Interdiscip. Perspect.*, vol. 20, no. January, p. 100836, 2023, doi: 10.1016/j.trip.2023.100836.
- [11] S. Dutta, A. Das, and B. K. Patra, "CLUSTMOSA: Clustering For GPS Trajectory Data Based on Multi-Objective Simulated Annealing to Develop Mobility Application," *Appl. Soft Comput.*, vol. 130, p. 109655, 2022, doi: 10.1016/j.asoc.2022.109655.
- [12] S. Wang, Z. Bao, J. S. Culpepper, and G. Cong, "A Survey on Trajectory Data Management, Analytics, and Learning," *ACM Comput. Surv.*, vol. 54, no. 2, 2021, doi: 10.1145/3440207.
- [13] D. Zhou, R. Hong, and J. Xia, "Identification of Taxi Pick-Up and Drop-Off Hotspots Using The Density-Based Spatial Clustering Method," in *CICTP 2017: Transportation Reform and Change - Equity, Inclusiveness, Sharing, and Innovation - Proceedings of the 17th COTA International Conference of Transportation Professionals*, 2018, no. January, pp. 196–204, doi: 10.1061/9780784480915.020.
- [14] X. Qian, W. Zhang, S. V. Ukkusuri, and C. Yang, "Optimal Assignment and Incentive Design in The Taxi Group Ride Problem," *Transp. Res. Part B Methodol.*, vol. 103, pp. 208–226, 2017, doi: 10.1016/j.trb.2017.03.001.
- [15] A. Fielbaum, "Optimizing a Vehicle's Route in an On-Demand Ridesharing System in Which Users Might Walk," *J. Intell. Transp. Syst. Technol. Planning, Oper.*, vol. 26, no. 4, pp. 432–447, 2022, doi: 10.1080/15472450.2021.1901225.
- [16] J. Tang, *Urban Travel Mobility Exploring with Large-Scale Trajectory Data*. Elsevier Inc., 2018.
- [17] R. K. Gunawan and Susilawati, "A Study of Spatiotemporal Distribution of Mobility-On-Demand in Generating Pick-Up/Drop-Offs Location Placement," *Smart Cities*, vol. 4, no. 2, pp. 746–766, 2021, doi: 10.3390/smartcities4020038.
- [18] H. Yu, Z. Li, G. Zhang, P. Liu, and J. Wang, "Extracting and Predicting Taxi Hotspots in Spatiotemporal Dimensions Using Conditional Generative Adversarial Neural Networks," *IEEE Trans. Veh. Technol.*, vol. 69, no. 4, pp. 3680–3692, 2020, doi: 10.1109/TVT.2020.2978450.
- [19] L. Zhang, Z. He, X. Wang, Y. Zhang, J. Liang, and G. Wu, "Pick-Up Point Recommendation Using Users' Historical Ride-Hailing Orders," in *Wireless Algorithms, Systems, and Applications*, 2022, pp. 393–405.
- [20] Y. Wang and J. Ren, "Taxi Passenger Hot Spot Mining Based on a Refined K-Means++ Algorithm," *IEEE Access*, vol. 9, pp. 66587–66598, 2021, doi: 10.1109/ACCESS.2021.3075682.
- [21] R. Rafiq and M. G. McNally, "An Exploratory Analysis of Alternative Travel Behaviors of Ride-Hailing Users," *Transportation (Amst.)*, vol. 50, no. 2, pp. 571–605, 2023, doi: 10.1007/s11116-021-10254-9.
- [22] X. Huang, Y. Yin, S. Lim, G. Wang, B. Hu, and J. Varadarajan, "Grab-Posisi: An Extensive Real-Life GPS Trajectory Dataset in Southeast Asia," in *Proceedings of the 3rd ACM SIGSPATIAL International Workshop on Prediction of Human Mobility, PredictGIS 2019*, 2019, pp. 1–10, doi: 10.1145/3356995.3364536.
- [23] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, "A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise," in *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*, 1996, pp. 226–231.
- [24] D. Deng, "DBSCAN Clustering Algorithm Based on Density," *Proc. - 2020 7th Int. Forum Electr. Eng. Autom. IFEEA 2020*, pp. 949–953, 2020, doi: 10.1109/IFEEA51475.2020.00199.
- [25] J. Ha, M. Kambe, and J. Pe, *Data Mining: Concepts and Techniques*, 3rd ed. Elsevier Inc., 2011.
- [26] P. J. Rousseeuw, "Silhouettes: A Graphical Aid to The Interpretation and Validation of Cluster Analysis," *J. Comput. Appl. Math.*, vol. 20, no. C, pp. 53–65, 1987, doi: 10.1016/0377-0427(87)90125-7.