

IDENTIFYING AREA HOTSPOTS AND TAXI PICKUP TIMES USING SPATIAL DENSITY-BASED CLUSTERING

Mulia Dea Lestari^{*1}, Lizda Iswari²

^{1,2}Informatics, Faculty of Industrial Engineering, Universitas Islam Indonesia, Indonesia
Email: ¹19523119@students.uii.ac.id, ²lizdaiswari@uii.ac.id

(Article received: April 10, 2023; Revision: May 04, 2023; published: October 15, 2023)

Abstract

Taxis are one of the competitive sectors of transportation and are recognized as convenient and easy means of transportation to meet individual needs. However, in the operation of a taxi there are some problems that would make the taxi service less optimal, such as the difficulty with finding a taxi at specific hours, the imbalance between demand and taxi supplies, and the length of passengers waiting for a taxi. Therefore, to optimize taxi service, a knowledge base is needed for strategic management decision making. In the study, data of exploration taxis uses a DBSCAN algorithm aimed at identifying and clustering pickup hotspots based on time during weekday and weekend time from Queens, New York City. As for the features used which are pickup latitude and pickup longitude. Accuracy scores for modeling use coefficients to achieve accuracy scores of 0.80 on weekdays and 0.77 on weekends where the accuracy score falls into the accurate category in modeling. Results show that there are three areas of taxi pickup centers based on high taxi demand in January 2016, where they are at LaGuardia airport, John f. Kennedy international, and the area around Steinway Street.

Keywords: Clustering, DBSCAN, Taxi, Taxi Service.

1. INTRODUCTION

In Indonesia one competitive business is the transport services industry. Transport efforts are not only moving goods or people from one place to another with static conditions, but also needed for repairs and improvements to technological development [1]. According to the study [2], taxis are one of the competitive sectors in certain transportation modes. According to the UITP (International Association of Public Transport) the flexibility of taxi services encourages global growth and popularity in the industry [3] so that taxi revenues by 2023 according to the Statista data portal reach US\$332.50 billion with revenue growth of 20.3% [4]. Taxis are recognized as comfortable mode of transportation [5], easy and free in the needs of individual transport [6].

In taxi operations, there are some problems from taxi service that make the service less optimal, such as how difficult it is to find a taxi at peak hours [7], a discrepancy between demand and supply taxi [8], and how long passengers wait for a taxi. Customer satisfaction with taxi service affects the company's image so that it is important for the company to improve the quality of the taxi service. Therefore, to optimize taxi service requires a basic knowledge base of the location and time of taxi operations for strategic management decision making.

Technology plays a crucial role in addressing business needs in a rapidly technological age [2]. GPS technology has been so widely used in taxis that the

real time data collection [7]. The raw data trajectory of the data gathering can be processed and analyzed to produce useful knowledge [9]. Thus, the importance of how to mine data pathways to improve taxi service [10] and understand the travel point patterns of data provide good opportunities for insights into taxi mobility [11].

The study will apply a density-based clustering method using a DBSCAN algorithm (Density-Based Spatial Clustering of Applications with Noise). Clustering is a grouping process of a number of data into several classes according to the characteristics of each class, and an efficient clustering algorithm for determining clusters in data density is a DBSCAN algorithm [12]. DBSCAN algorithm is a location-based grouping approach used to find connections and patterns in geographic data [13]. Time-based analysis helps to understand patterns or characteristics of changes over time. The results of the modeling will be analyzed based on each hour of the day and compared during weekdays and weekends based on high taxi demand and knowing the times when traffic is packed according to travel speeds. The study focuses only on taxis located in the Queens region of New York City with the goal of identifying and recommending pickup areas based on the result of time series analysis. Thus, it is hoped that the study will be used as basis for management strategies that can be used to boost taxi service.

2. RESEARCH METHODS

The research method consists of several stages of EDA (Exploratory Data Analysis), preprocessing, clustering process, and results, shown on figure 1 are the stages of the research method.

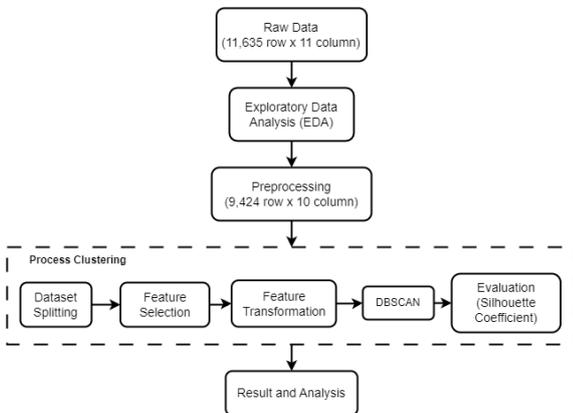


Figure 1. Flow of Method

2.1. Dataset

The dataset used in the study is a dataset of a New York City taxi taken from Kaggle web site accessible using links: <https://www.kaggle.com/competitions/nyc-taxi-trip-duration>. Dataset taken from January 1 to January 31, 2016, comprising 229,707 lines and 11 columns of data. New York City has five major boroughs of Bronx, Brooklyn, Manhattan, Queens, and Staten Island. However, the study focused only on the Queens area, which was therefore ran a data filtering, resulting in 11.635 lines of data. The following are variables and descriptions of the dataset on table 1:

Table 1. Description Dataset

Variable Name	Definition
id	A unique identifier for each trip
Vendor id	A code indicating the provider associated with the trip record
Pickup datetime	Date and time when the meter was engaged
Dropoff datetime	Date and time when the meter was disengaged
Passenger count	The number of passengers in the vehicle (driver entered value)
Pickup longitude	The longitude where the meter was engaged
Pickup latitude	The latitude where the meter was engaged
Dropoff longitude	The longitude where the meter was disengaged
Dropoff latitude	The latitude where the meter was disengaged
Store and fwd flag	This flag indicates whether the trip record was held in vehicle memory before sending to the vendor because the vehicle did not have a connection to the server
Trip duration	Duration of the trip in seconds

2.2. Exploratory Data Analysis (EDA)

This process is useful for improving understanding about the data and seeing the data

quality beforehand before moving on to the pre-processing stage. This process involves using statistical descriptive and graphic representation by looking at or discovering analysis patterns of each variable or correlation between variable.

2.3. Pre-processing

Preprocessing data is used to prepare raw data in order to get clean and quality data. Here are the stages of preprocessing data:

a. Feature Engineering

Feature engineering is useful in the process of selecting, modifying, creating, and manipulating raw data to produce new variables with the purpose of analysis and enhancing model accuracy. One step of feature engineering is feature creation. Feature creation process makes new features by extracting existing variables into new features, converts from numerical variables into category or called discretization, and produces new features from the results of the division between two variables

b. Data Cleaning

Data Cleaning is done by removing an outlier, removing unnecessary attributes, and deleting irrelevant data. In this case, dispose of over three hours of trips and any trips that ended outside of the Queens and Manhattan areas. These measures remove approximately 19% of the data collection, with 9424 lines of data remaining.

2.4. Clustering

The study implemented a DBSCAN algorithm for the clustering process. The clustering process aims to form a cluster according to each individual characteristic. The following are a sequence of clustering processes:

a. Dataset Splitting

The study splits the dataset into two groups by day, weekdays and weekends. This is done in order to find and understand the different patterns the taxi and the residents need in working days and weekends.

b. Feature Selection

In this process select relevant features to do modeling. In this case, it's only taken two features which are pickup latitude and pickup longitude.

c. Feature Transformation

The data used is spatial in their makeup, and because of the earth being spherical their statements harmonize with the documentation [14] metric Haversine Distance is a good measure to use in this case because it gives a good estimate of the distance between two points on the earth's surface. Next, the features that have been selected are transformed from degrees to radians. This is done to clustering by means of haversine distance meters and is useful in improving accuracy in models.

d. DBSCAN

Next, a DBSCAN algorithm is applied to modeling based on the chosen variables. DBSCAN

can determine the number of clusters itself, so it is not necessary for us to determine the number of clusters, but it requires two other input parameters [15]:

- a The epsilon is the maximum distance value that constitutes the neighborhood boundary
- b MinPts is the minimum number of dots in the epsilon radius

The parameters mentioned could be modified according to wishes, in which case experiments determine the Epsilon and MinPts parameters to choose which are better based on the metrics of the evaluation. Computing from DBSCAN algorithm as follows [16]:

1. Determine the parameters of MinPts and Epsilon to use
2. Select initial or c data at random
3. In this process, researchers are applying meters haversine distance. Calculate the distance between data c for all points of density reachable using a metric Haversine Distance in equations (1):

$$D = 2R \cdot \sin^{-1} \sqrt{\sin^2\left(\frac{y_2 - y_1}{2}\right) + \cos(y_1) \cdot \cos(y_2) \cdot \sin^2\left(\frac{x_2 - x_1}{2}\right)} \quad (1)$$

In the equation (1) known d for distance, then R represents the earth's radius of approximately 6371 km (6371 mi), x1, x2 is longitude and y1, y2 is latitude.

4. If that point meets the epsilon more than the minimum number (MinPts) then point c as the core points and cluster is formed.
5. If c is border point, and there is no point that density reachable with c, then the process will continue to another point.
6. Repeat steps 3-4 to all points processed.
- e. Evaluation

At the evaluation stage it is a stage for measuring performance on a model. Metric silhouette covaries, is used in order to determine the degree of accuracy in the model to the cluster that is formed.

Silhouette coefficient is stated in the equation (2) where Si is Silhouette Coefficient, ac is the nearest cluster distance for each sample and bc is the distance between sample and nearby cluster other than the sample.

$$S_c = \frac{b_c - a_c}{\max(a_c, b_c)} \quad (2)$$

According to [17] interpretation of silhouette coefficient are shown in table 2, can be concluded that if the accuracy score gets closer to 1 then the accuracy value can be categorized as accurate or good, conversely closer to 0 then the accuracy value is said to be bad or inaccurate.

Table 2. Interpretation of the Silhouette Coefficient Value

Silhouette Coefficient	Interpretation
0.71 - 1.00	A strong structure
0.51 - 0.70	A good structure
0.26 - 0.50	A weak structure
0.00 - 0.25	A bad structure

2.5. Result Analysis

After an experiment and knowing performance on a model, the model is chosen with the best accuracy. Later, the cluster that forms is used to identify the center of the cluster and is carried out an analysis process to identify the extraction area. Additionally, an in-depth analysis of taxi data profiles could be obtained from further feature engineering and exploratory data analysis using statistics and visualization.

3. RESULT AND DISCUSSION

3.1. Raw Dataset

The dataset used in this study consisted of 229707 rows and 11 columns in January. Shown in table 3 is an example of the raw dataset.

Table 3. Raw Dataset

no	id	vendor_id	pickup_datetime	dropoff_datetime	passenger count	pickup longitude	pickup latitude	dropoff longitude	dropoff latitude	store and fwd flag	trip duration
1	id3858529	2	2016-01-19 11:35:24	2016-01-19 12:10:48	1	-73.979027	40.763939	-74.005333	40.710087	N	2124
2	id0801584	2	2016-01-30 22:01:40	2016-01-30 22:09:03	6	-73.982857	40.742195	-73.992081	40.749184	N	443
3	id3970352	1	2016-01-09 12:55:35	2016-01-09 13:07:22	1	-74.008995	40.712589	-73.998634	40.722809	N	707
...
229706	id1049543	1	2016-01-10 07:35:15	2016-01-10 07:46:10	1	-74.000946	40.747379	-73.970184	40.796547	N	655
229707	id2714485	1	2016-01-05 15:56:26	2016-01-05 16:02:39	1	-73.982079	40.749062	-73.974632	40.757107	N	373

3.2. Result of Data Pre-Processing

In the preprocessing stage we create new features, data cleaning, and data selection, then obtained as many as 9424 rows and 10 columns. The resulting data can be viewed at table 4.

Table 4. Description Dataset of Result Pre-processing

Variable Name	Definition
Pickup date	Date when the meter was engaged
Pickup longitude	The longitude where the meter was engaged
Pickup latitude	The latitude where the meter was engaged
Pickup Time	Time when the meter was engaged

Pickup GroupofDay	Grouping by day (Weekdays and Weekends)
Distance	The distance from pick-up point to the drop-off point
Duration Minutes	Taxi ride duration in minutes
Average Speed	Average taxi ride speed
Suburbs	One of the boroughs of New York City (Queens)

3.3. Defining the Cluster Parameters

Before implementing a DBSCAN algorithm, it needs to determine the optimal value of the two parameters, namely Epsilon and MinPts. This process is carried out in several tests to select optimal parameters. Parameters selected based on the results of the highest coefficient value-assessments. As shown on the table 5 the highest silhouette coefficient is a 0.80 with an epsilon 0.5 and MinPts 17 on weekday. At the table 6, the highest silhouette coefficient are 0.779 with epsilon 0.8 and MinPts 15 on the weekend.

Table 5. Silhouette Coefficient of Several Trials on Weekdays

MinPts	Eps				
	0.4	0.5	0.6	0.7	0.8
13	0.630	0.773	0.661	0.665	0.770
14	0.630	0.766	0.659	0.665	0.770
15	0.628	0.766	0.659	0.663	0.770
16	0.643	0.767	0.547	0.661	0.768
17	0.642	0.801	0.547	0.661	0.664

Table 6. Silhouette Coefficient of Several Trials on Weekends

MinPts	Eps				
	0.4	0.5	0.6	0.7	0.8
13	0.658	0.668	0.770	0.775	0.776
14	0.691	0.769	0.770	0.636	0.776
15	0.642	0.615	0.770	0.772	0.779
16	0.642	0.705	0.770	0.773	0.778
17	0.641	0.705	0.770	0.773	0.778

3.4. Clustering with DBSCAN

The optimum value parameters have been determined, which is on a workday with Eps 0.5 and MinPts 17, and for a weekend with eps 0.8 and MinPts 15. Applying algorithm DBSCAN work day has rows of 6869, and a weekend of 2555 rows of data. Clustering results on work days and weekends can be seen at table 7.

Table 7. Result of Clustering on Weekday and Weekend

	Weekdays	Weekends
Epsilon	0.5	0.8
MinPts	17	15
Number of Cluster	3 (3424, 1963, 1268)	3 (1038, 773, 670)
Outlier	214	74
Accuracy	0.80142	0.77902

Results show that both weekday and weekend groups have a strong or very good category of accuracy. Thus, this clustering modelling is already able to do a process of analysis of the cluster being formed. The clustering comparison results between weekdays and weekends are shown in figure 2.

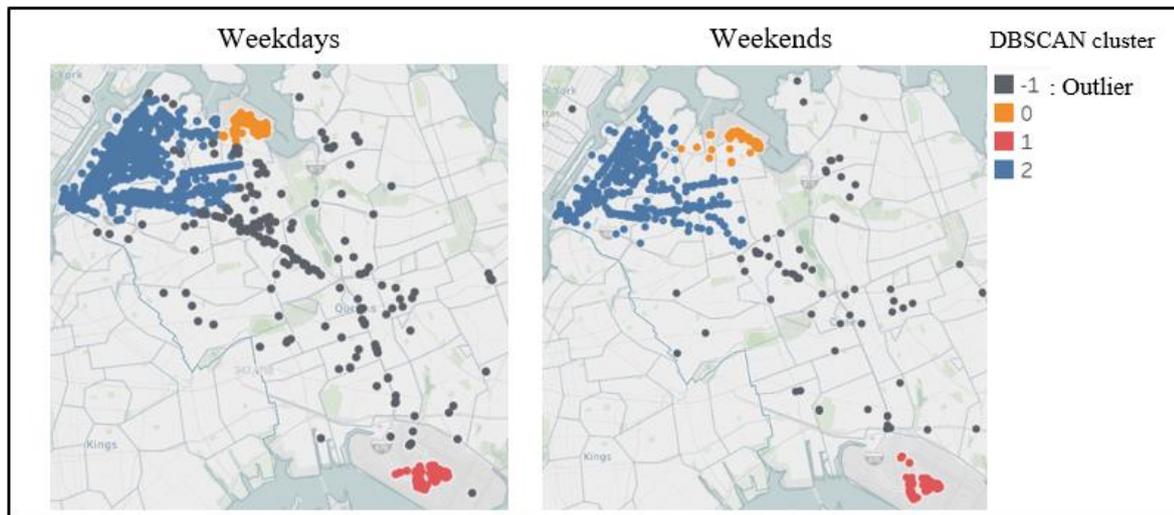


Figure 2. Comparison of Clustering Results Between Weekdays and Weekends

To analyze the data, researchers only focus on the data included in the cluster and delete the outlier. As Seen in figure 3 clustering results produce three clusters with the same location on both weekday and weekend.

The fact is that there are three equal area recommendations of each other the LaGuardia airport, JFK airport, and the around Steinway Street. As shown in figure 3, visualization on Steinway

Street has more points, but the fact that it has the most points LaGuardia airport followed JFK airport. This is because the Steinway Street cluster has low density or gap enough distance to form one cluster large enough, while the airport's LaGuardia cluster and JFK airport has a high density between one dot and another, thus creating a cluster with dense, visible visualization.

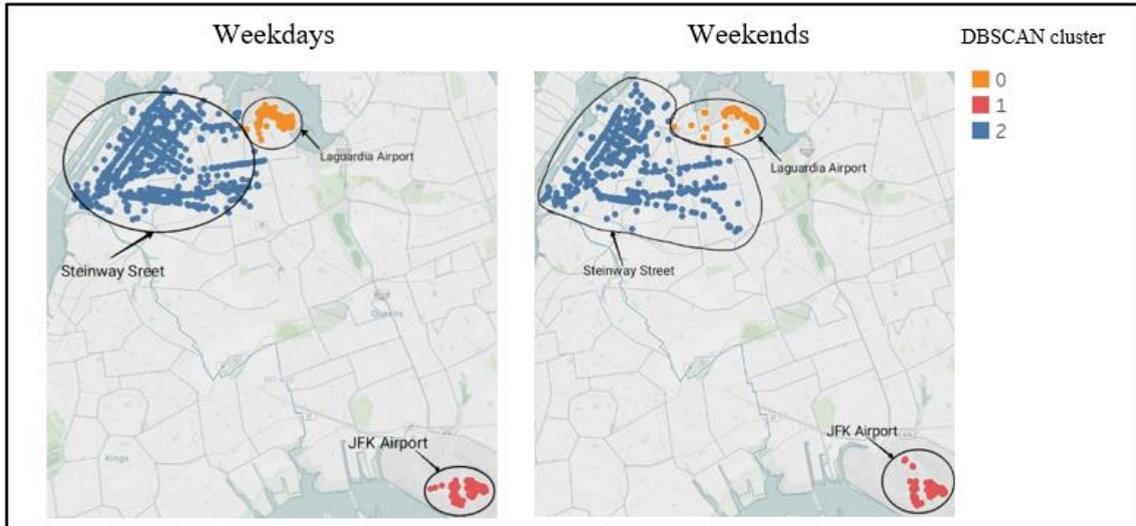


Figure 3. Comparison of Clustering Results Between Weekdays and Weekends Without Outlier

3.5. Data Taxi Profile

Before identifying the area where the taxi will be required to know the traffic situation at all hours,

so that the taxi can minimize the staying out of traffic and impact the reduced passenger wait time.

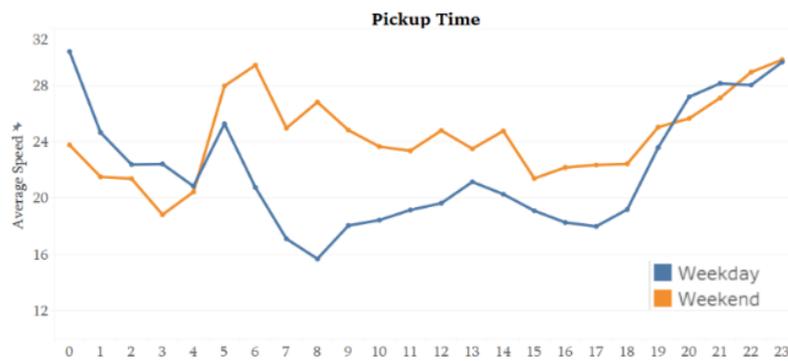


Figure 4. Comparison of Weekdays and Weekends based on Average Speed in Every Hour

Visible on figure 4, a chart of comparisons between work days and weekends based on average speed in every hour. By the time the workday started at 6:00 a.m., average speed began to decline, and it began to rise slowly again at 9:00. Next, significant traffic back at 9:00 a.m. to 07:00 p.m. can be seen at figure 4 on low average speed, and begins to rise at 8:00 a.m. to 00:00 a.m. Instead, at the end of the week, traffic was free from 5:00, and then slowly dropped from 2:00 p.m. to 6:00 p.m., and was back full from 12:00 p.m. to 4:00 a.m. This suggests relating to everyday life, when working days people travel in the morning to work or other activities all day long, and then home in the evening or evening, but instead on weekends, people tend to go from day to night to night. Thus, it is important to know the exact point of the pickup based on time in order to reduce passenger wait time because of overcrowding traffic or the distance between driver and passenger.

3.6. Result Recommended Area

On every graph in every recommended area, it is seen that weekends are always lower than

workdays, this is due to the number of different cab requests. By clustering our work day there are 6869 dots and 2555 taxi dots on the weekend, and produce the following three recommendations of the area picking the taxi:

a. LaGuardia Airport

It shows in figure 5 that taxi requests on weekdays and weekends are neither remotely different nor similar in much the same pattern, from 7:00 a.m. to 11:00 a.m. taxi requests fall into a high category. The result of this cluster is the airport's LaGuardia area, which accommodates limited domestic and international flight service and is the third busiest airport in the New York metropolitan area. Thus, this area can be a recommendation as a taxi base for making passenger pickup so that it can cut down on passenger waiting times because nearly every hour has multiple requests and is not down significantly.



Figure 5. Comparison of Weekdays and Weekend Pickups by Time (Hours)

b. John F. Kennedy International Airport



Figure 6. Comparison of Weekdays and Weekend Pickups by Time (Hours)

It shows on figure 6, that on weekdays and weekends it has a pretty similar pattern, where taxi vehicles have increased at 06:00 a.m. and then slowly decreased and again from 12:00 p.m. to late at night. The result is the John f. Kennedy international airport, the main international airport that serves New York

City and is the busiest airport in New York. Therefore, this area could be a cluster point for the morning pick-up, and continuing from evening to evening.

c. Steinway street and the vicinity

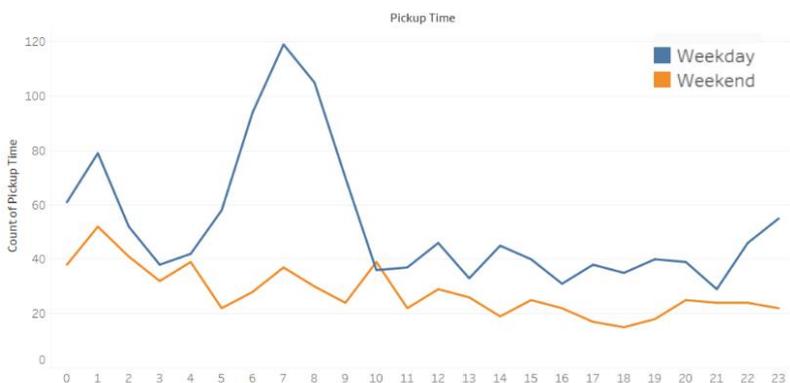


Figure 7. Comparison of Weekdays and Weekends Pickups by Time (Hours)

The cluster in figure 7 is the Steinway Street area, a main street in queens. This street is the main commercial district of the business improvement district, which involves the purchase and sale of goods and services that enumerate all industrial activities and relationships. All along this street are cafes, restaurants, fashion stores, sports centers, and

so on. Visible on weekdays, significant requests began at 05:00 a.m. to 07:00 a.m. then slowly declined. At the end of the week the taxi demand is quite stable at each hour from 05:00 a.m. to 12:00 p.m. and it rises significantly late at 00:00 p.m. to 4:00 a.m. The fact that the place is crowded on weekends, is because they spend time on vacation or

are free from work. Therefore, the pickup point can begin at 05:00 a.m. and continue at 08:00 p.m.

4. DISCUSSION

There were recent studies of data-mining trajectory techniques proposed to increase taxi service. In [18] research aims to identify taxi pick-up and drop-off points using the DBSCAN with the dataset of the city of Kunshan, China which is useful for taxi operation management and user travel pattern analysis. The results show that some pick-up and drop-off cluster center points are very close to each other, and there are also pick-up and drop-off clusters that are relatively far away from each other. This suggests that there is a relationship between people's daily travel characteristics and land use. Research [19] for the introduction and visualization of taxi hotspot using the DBSCAN+ method with the dataset of the city of Huai'an, China in hopes of providing important decisions for further city planning and traffic efficiency.

Research on the location is also done by [20] who aims to predict the taxi destination using the RNN method of using a dataset city of Porto, San Francisco, and Manhattan. The study is useful when demand is high, taxis will be near new passengers' pickup sites so they can distribute better resources, and then the results that are obtained provide relevant information to taxi company carriers. It is also studied by [21] with cab data in Chengdu, China using DeepFM methods.

In this study, the DBSCAN clustering method is applied, then the results of the cluster are analyzed time series based on weekdays and weekends at every hour. The purpose of this research is to recognize areas of high taxi pick-up demand at any time in the hope of knowing the right taxi demand area, and can be a recommendation for taxi stands.

5. CONCLUSION

In this study clustering models with a DBSCAN algorithm, and clustering results were analyzed in time series during weekdays and weekends on the taxi dataset in Queens area of New York City in January. The accuracy score derived from modeling is 0.80 on weekdays, and 0.77 on weekends, this values are included in the category were accurate for use. Thus, the results obtained are the patterns and characteristics of the three taxi pick-up areas, which are located at Laguardia Airport, John F. Kennedy International Airport, and the area around Steinway Street. The proposed method is expected to be useful in identifying the potentially high-altitude taxi pickup areas. For further research, a more specific modeling could be done with datasets such as at holidays, and could try using other machine learning algorithms.

REFERENCES

- [1] R. Rivai, "Identifikasi Perilaku Penggunaan

Dan Persepsi Pengguna Tentang Layanan Pemesanan Dan Pengiriman Makanan Dengan Transportasi Online," other, Universitas Komputer Indonesia, 2020. doi: 10/BAB%20IV%20-%20Unikom%20-%20Riyaldi%20Rivai%20-%2010615004.pdf.

- [2] A. S. Athoillah, M. Firdaus, and B. Sanim, "Competitive strategy of taxi company in facing environmental changes," *BISMA (Bisnis dan Manajemen)*, vol. 12, no. 1, Art. no. 1, Oct. 2019, doi: 10.26740/bisma.v12n1.p66-87.
- [3] UITP, "GLOBAL TAXI BENCHMARKING STUDY 2019," Nov. 2020. Accessed: Dec. 06, 2022. [Online]. Available: https://cms.uitp.org/wp/wp-content/uploads/2020/11/Statistics-Brief-TAXI-Benchmarking_NOV2020-web.pdf
- [4] "Ride-hailing & Taxi - Worldwide | Statista Market Forecast," *Statista*, Dec. 2022. <https://www.statista.com/outlook/mmo/shared-rides/ride-hailing-taxi/worldwide> (accessed Mar. 14, 2023).
- [5] M. B. Ulak, A. Yazici, and M. Aljarrah, "Value of convenience for taxi trips in New York City," *Transportation Research Part A: Policy and Practice*, vol. 142, pp. 85–100, Dec. 2020, doi: 10.1016/j.tra.2020.10.016.
- [6] H. Yao, Z. Huang, X. Ye, M. Grifoll, G. Liu, and P. Zheng, "Analysis of Taxi Travels during an Epidemic Period Using System Dynamics Method," *Sustainability*, vol. 14, no. 6, Art. no. 6, Jan. 2022, doi: 10.3390/su14063457.
- [7] X. Wang, H. Zhang, and L. Wang, "A Demand-Supply Oriented Taxi Recommendation System for Vehicular Social Networks," *IEEE Access*, vol. PP, pp. 1–1, Jul. 2018, doi: 10.1109/ACCESS.2018.2857002.
- [8] B. Hu, X. Xia, H. Sun, and X. Dong, "Understanding the imbalance of the taxi market: From the high-quality customer's perspective," *Physica A: Statistical Mechanics and its Applications*, vol. 535, p. 122297, Dec. 2019, doi: 10.1016/j.physa.2019.122297.
- [9] R. Ibrahim and M. O. Shafiq, "Detecting taxi movements using Random Swap clustering and sequential pattern mining," *J Big Data*, vol. 6, no. 1, Art. no. 1, Dec. 2019, doi: 10.1186/s40537-019-0203-6.
- [10] Z. Ning, X. Wang, X. Kong, and W. Hou, "A Social-Aware Group Formation Framework for Information Diffusion in Narrowband Internet of Things," *IEEE Internet of Things Journal*, vol. 5, no. 3, pp. 1527–1538, Jun.

- 2018, doi: 10.1109/JIOT.2017.2777480.
- [11] Z. Zhou, J. Yu, Z. Guo, and Y. Liu, "Visual exploration of urban functions via spatio-temporal taxi OD data," *Journal of Visual Languages & Computing*, vol. 48, pp. 169–177, Oct. 2018, doi: 10.1016/j.jvlc.2018.08.009.
- [12] I. P. S. Almantara, N. W. S. Aryani, and I. B. A. Swamardika, "Spatial Data Analysis using DBSCAN Method and KNN classification," *International Journal of Engineering and Emerging Technology*, vol. 5, no. 2, pp. 77–80, Dec. 2020, doi: 10.24843/IJEET.2020.v05.i02.p013.
- [13] M. Amiruzzaman, R. Rahman, M. R. Islam, and R. M. Nor, "Logical analysis of built-in DBSCAN Functions in Popular Data Science Programming Languages." OSF Preprints, Jun. 27, 2022. doi: 10.31219/osf.io/ge654.
- [14] "sklearn.metrics.pairwise.haversine_distances — scikit-learn 1.2.2 documentation." Accessed: May 04, 2023. [Online]. Available: https://scikit-learn.org/stable/modules/generated/sklearn.metrics.pairwise.haversine_distances.html
- [15] S. T. I Made Suwija Putra, "ALGORITMA DBSCAN (DENSITY-BASED SPATIAL CLUSTERING OF APPLICATIONS WITH NOISE) DAN CONTOH PERHITUNGANNYA," Jul. 2018, Accessed: Dec. 18, 2022. [Online]. Available: <http://erepo.unud.ac.id/id/eprint/20097/>
- [16] A. Devi, I. Putra, and I. Sukarsa, "Implementasi Metode Clustering DBSCAN pada Proses Pengambilan Keputusan," *Lontar Komputer : Jurnal Ilmiah Teknologi Informasi*, p. 185, Dec. 2015, doi: 10.24843/LKJITI.2015.v06.i03.p05.
- [17] P. Rousseeuw, M. Hubert, and A. Struyf, "Clustering in an Object-Oriented Environment," *Journal of Statistical Software*, vol. 01, Feb. 1997, doi: 10.18637/jss.v001.i04.
- [18] D. Zhou, R. Hong, and J. Xia, *Identification of Taxi Pick-Up and Drop-Off Hotspots Using the Density-Based Spatial Clustering Method*. 2018, p. 204. doi: 10.1061/9780784480915.020.
- [19] Z. Huang, S. Gao, C. Cai, H. Zheng, Z. Pan, and W. Li, "A rapid density method for taxi passengers hot spot recognition and visualization based on DBSCAN+," *Sci Rep*, vol. 11, no. 1, Art. no. 1, May 2021, doi: 10.1038/s41598-021-88822-3.
- [20] A. Rossi, G. Barlacchi, M. Bianchini, and B. Lepri, "Modelling Taxi Drivers' Behaviour for the Next Destination Prediction," *undefined*, 2020, doi: 10.1109/TITS.2019.2922002.
- [21] X. Wang, Y. Liu, Z. Liao, and Y. Zhao, "DeepFM-Based Taxi Pick-Up Area Recommendation," in *Pattern Recognition. ICPR International Workshops and Challenges*, A. Del Bimbo, R. Cucchiara, S. Sclaroff, G. M. Farinella, T. Mei, M. Bertini, H. J. Escalante, and R. Vezzani, Eds., in Lecture Notes in Computer Science. Cham: Springer International Publishing, 2021, pp. 407–421. doi: 10.1007/978-3-030-68821-9_36