Developing a hybrid clustering model for Covid-19 vaccine distribution in Bangkok metropolis, Thailand

Thitima Wonginta[1], Anirut Kantasa-ard[2*]

Abstract

The Covid-19 cases are rapidly and continuously increasing in many areas recently, particularly in the Asia and Pacific region. Bangkok, as one of the biggest capitals in Southeast Asia, has a high prevalence rate of Covid-19 patients. Furthermore, the number of deaths from this pandemic is constantly rising. Therefore, the central government put an effort to control the Covid-19 situation by increasing the vaccination rate in many endemic areas, including Bangkok. Despite that Covid-19 vaccine has been prepared since the 1st quarter of 2021, vaccine distribution in this city is still limited. One of the main problems is insufficient vaccination centers serving all populations in Bangkok and metropolitan areas. This paper proposes a novel perspective of vaccine distribution in the metropolis, using a hybrid clustering model. This model is the integration between the concepts of partitional clustering and center-of-gravity. Bangkok is chosen as a case study to verify the performance of this model.

The results showed that K-Means is outperformed other benchmark via silhouette width value. In addition, this model suggests increasing the number of vaccination centers in some areas of Bangkok. This model suggests extra six vaccination centers apart from 25 current centers established by the central government, for a total of 31 centers. This can be a good case study for all involved organizations to consider increasing vaccination centers in different areas for the future period.

Keywords: Partitional Clustering, Vaccine distribution, Covid-19, K-Means, Center-of-gravity

---

[1] Assistance Professor, Faculty of Logistics, Burapha University, Chonburi Thailand
[2*] Lecturer, Faculty of Logistics, Burapha University, Chonburi Thailand
[2*] Corresponding author, email: anirutka@gmail.com

การพัฒนาแบบจำลองคลัสเตอร์แบบผสมสำหรับการจัดสรรวัคซีน Covid-19 ในพื้นที่กรุงเทพมหานคร

ฐิติมา วงศ์อินตา[1], อนิรุทธ์ ขันธสะอาด[2*]

บทคัดย่อ

จากสถานการณ์การแพร่ระบาดของเชื้อไวรัสโควิด-19 ทั่วโลกส่งผลให้มีจำนวนผู้ติดเชื้อโควิด-19 เพิ่มขึ้นอย่างรวดเร็วและต่อเนื่องโดยเฉพาะในภูมิภาคเอเชียและแปซิฟิก กรุงเทพมหานครเป็นหนึ่งในเมืองหลวงที่ใหญ่ที่สุดในเอเชียตะวันออกเฉียงใต้ที่มีจำนวนผู้ติดเชื้อไวรัสโควิด-19 เพิ่มขึ้นในอัตราที่สูง อีกทั้งจำนวนผู้เสียชีวิตจากการติดเชื้อไวรัสโควิด-19 ยังมีอัตราเพิ่มขึ้นอย่างต่อเนื่อง จากสถานการณ์ดังกล่าวรัฐบาลจึงเร่งดำเนินมาตรการในการลดการแพร่ระบาดของเชื้อไวรัสโควิด-19 โดยการเพิ่มอัตราการฉีดวัคซีนให้กับประชาชนโดยเฉพาะในพื้นที่กรุงเทพมหานครและปริมณฑลซึ่งเป็นพื้นที่ที่มีจำนวนผู้ติดเชื้อสูงสุดในประเทศ อย่างไรก็ดีแม้ว่าจะรัฐบาลจะมีการเตรียมการฉีดวัคซีนให้กับบุคลากรทางการแพทย์ตั้งแต่ช่วงไตรมาสที่ 1 ของปี 2021 แต่ยังพบว่าปัจจุบันการกระจายวัคซีนยังไม่สามารถกระจายให้กับประชาชนทั่วไปได้อย่างทั่วถึง หนึ่งในปัญหาที่สำคัญคือจำนวนวัคซีนไม่เพียงพอ ทำให้การกำหนดจุดให้บริการฉีดวัคซีนให้กับประชาชนทั่วไปยังไม่ครอบคลุมทั่วพื้นที่กรุงเทพมหานครและปริมณฑล ดังนั้นงานวิจัยนี้จึงมีแนวคิดในการพัฒนาแบบจำลองคลัสเตอร์แบบผสมโดยการผสมผสานระหว่างแนวคิดการจัดกลุ่มแบบ Partitional Clustering และวิธีจุดศูนย์กลางแรงโน้มถ่วง เพื่อวิเคราะห์หาจุดที่เหมาะสมสำหรับการจัดสรรวัคซีนโควิด-19 ให้กับประชาชนทั่วไป โดยงานวิจัยนี้ได้เลือกจังหวัดกรุงเทพมหานครเป็นกรณีศึกษาในครั้งนี้

ผลการวิเคราะห์พบว่าเทคนิคในการจัดกลุ่มด้วย วิธี K-mean มีประสิทธิภาพกว่าดีกว่าวิธีอื่นๆ เมื่อพิจารณาจากค่า silhouette จากการใช้วิธี K-Means ในการพัฒนาแบบจำลองพบว่าแบบจำลองแนะนำให้เพิ่มจำนวนศูนย์ฉีดวัคซีนเพิ่มในบางพื้นที่ของกรุงเทพมหานครอีกจำนวน 6 แห่ง จากเดิมที่รัฐบาลได้มีการกำหนดจุดฉีดวัคซีนไว้แล้วจำนวน 25 แห่ง รวมเป็น 31 แห่ง ซึ่งสามารถให้บริการฉีดวัคซีนให้กับประชาชนได้ครอบคลุมพื้นที่กรุงเทพมหานคร ซึ่งแนวคิดในการวิเคราะห์การกำหนดจุดฉีดวัคซีนของกรุงเทพมหานครในครั้งนี้นี้สามารถนำใช้เป็นแนวทางในการวิเคราะห์จุดที่เหมาะสมในการจัดสรรวัคซีนโควิด-19 ในพื้นที่อื่นๆได้

คำสำคัญ: การจัดกลุ่มแบบ Partitional Clustering, การกระจายวัคซีน, Covid-19, K-Means, วิธีจุดศูนย์กลางแรงโน้มถ่วง

---

[1] ผู้ช่วยศาสตราจารย์, คณะโลจิสติกส์ มหาวิทยาลัยบูรพา ชลบุรี

[2*] อาจารย์, คณะโลจิสติกส์ มหาวิทยาลัยบูรพา ชลบุรี

[2*] Corresponding author, email: anirutka@gmail.com

1. Introduction

Today`s covid-19 situation is worse than it was last year during the same time period.The number of patients and deaths are increasing continuously from time to time. In addition, the delta branch mutation is quickly spreading over Europe, North and Latin America, and the Asia Pacific regions. The number of infections was increasing in Thailand since April 2021. The trend of Covid-19 new cases in Thailand is shown in Figure 1.

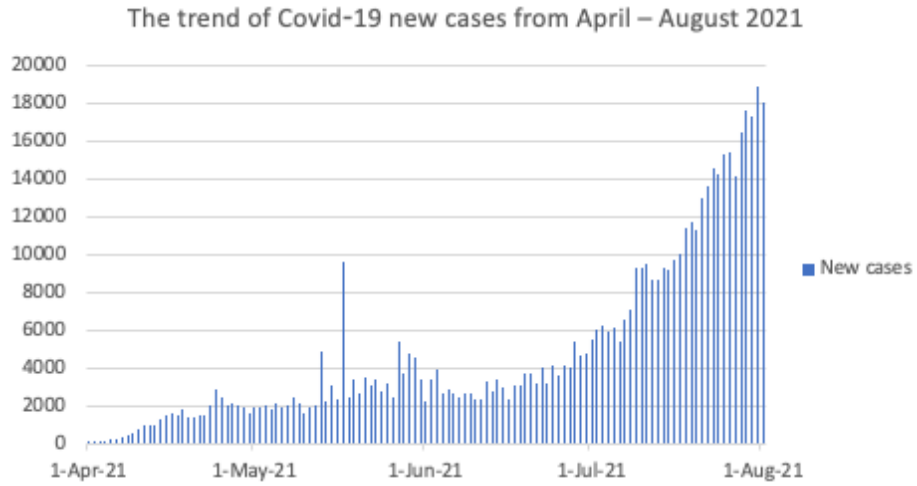


Figure 1 The trend of Covid-19 new cases from April – August 2021 (Dong et al., 2020)

Furthermore, as compared to other countries in the same region, the quantity of vaccines is insignificant. The number of Covid-19 first doses are around 14.2 million doses (20% of the total population), and the number of complete doses are around 3.9 million doses (5.6% of the total population)(Hannah et al., 2020). The total vaccination rate, as shown in Figure 2, is evaluated until 2 August 2021.

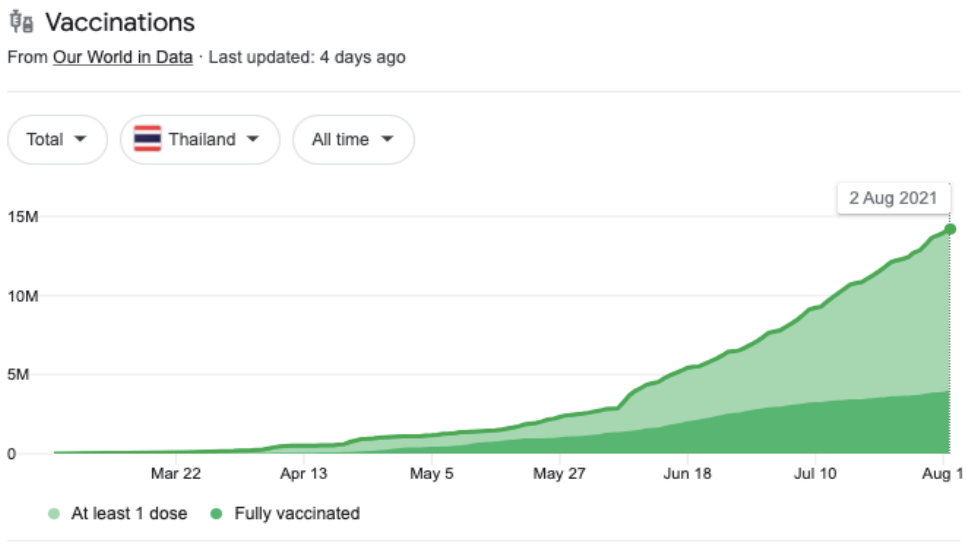


Figure 2 The Covid-19 vaccination rates in Thailand from April – August 2021(Hannah et al., 2020)

Regarding the information, people waiting for the vaccine are approximately 80% for the first dose and 94. 4% for the second dose. After considering the number of patients from many cities in Thailand, Bangkok has the highest rate of Covid-19 patients(Department of Disease Control, 2021). The details are shown in Figure 3.

Since Bangkok is Thailand`s capital and one of the country`s most important economic centers, it is crucial to keep Covid-19 infection under control. The public health system, economy and scoiety will be affected if the government does not lower Covid-19 infection rates in the economic sector. Furthermore, the rates of economic and social progress will be also slowed.

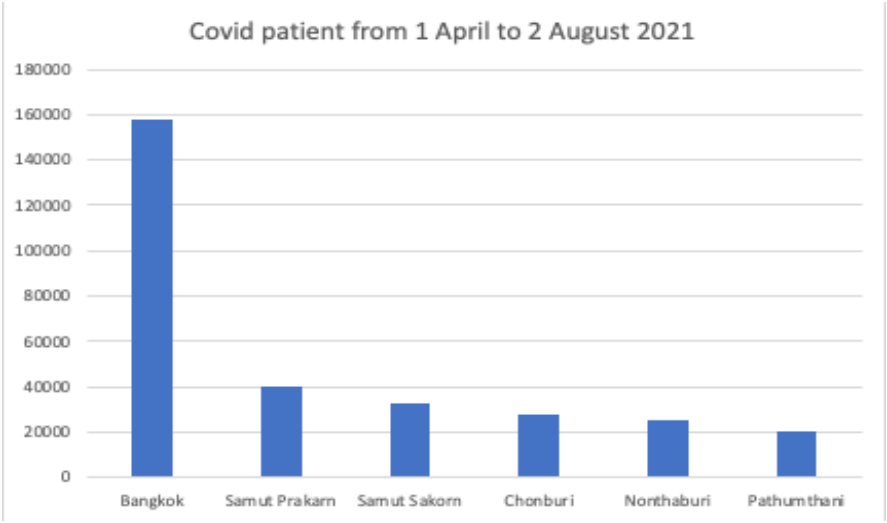


Figure 3 The top 6 provinces of Covid-19 patients in Thailand from April – August 2021(Department of Disease Control, 2021)

After validating the situation of Covid- 19 in Bangkok, there are still some limitations such as insufficient vaccines, lack of outdoor vaccination centers, and difficulty to get the vaccination queue(Department of Disease Control, 2021). Therefore, this paper proposes a hybrid clustering model to improve the vaccine distribution during this epidemic time. This model will suggest how to allocate vaccination centers in each area by clustering method. Also, this model will propose new positions of vaccination centers to serve all populations in some districts of Bangkok and metropolitan areas.

This paper is divided into six sections. First section introduces the paper. Section 2 reviews the literature on clustering methods and typical location decisions. Section 3 describes the problem statements and objectives. Section 4 details the methodology and the implementation of the hybrid clustering model. Section 5 demonstrates the proposed model's results and compares the results between the proposed model and the existing solution. Section 6 contains the conclusion as well as potential future study directions.

2. Literature Reviews

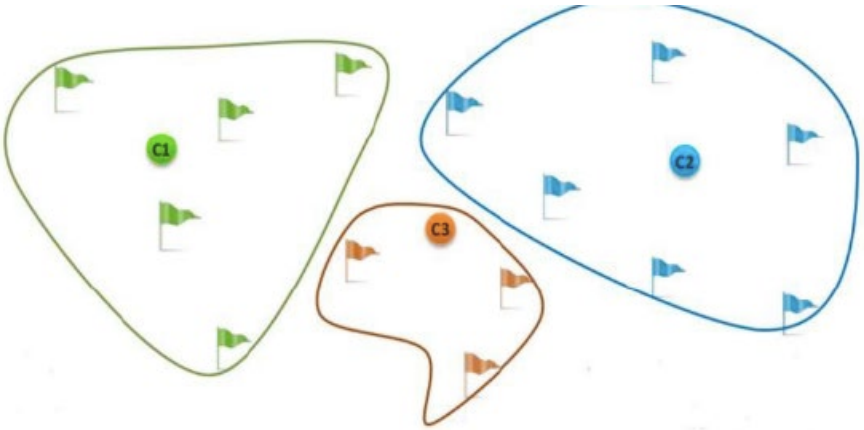
The literature review is structured as follows. Firstly, some clustering models are presented with their advantages and limitations. Secondly, the concept of location decisions and how to choose appropriate locations are reviewed. These concepts will be the starting point to construct the hybrid clustering model in this study.

2.1 Clustering method

The clustering method is one of the unsupervised learning techniques and frequently implemented to forecast data in various domains(Flynt & Dean, 2016; MacQueen, 1967). The main concept of this method is to cluster data into different groups before training and forecasting. The existing data will be clustered based on the similarity of input variables(MacQueen, 1967). Nowadays, the clustering method reduces the input and output variables' dimensions, particularly in big data(Nananukul, 2013). There are many clustering methods such as Hierarchical Clustering, Partitional Clustering, and Fuzzy Clustering. However, "partitional clustering" is one of the most often used clustering algorithms for grouping data.

Partitional clustering is easy to implement and appropriate to cluster data from small to large datasets with various parameters(Murray et al., 2015). Also, this clustering method has less computational time than other clustering methods. For the implementation, the partitional clustering will divide data into k groups regarding input parameters, both quantitative and qualitative types(MacQueen, 1967). Each k group will be composed of two main components: a center point and various data points. The example of partitional clustering is shown in Figure 4. In this figure, c1, c2, c3 stand for a center point of delivery locations in each cluster. The remaining flags are presented as various delivery locations.

There are many clustering methods for partitional clustering. However, two clustering methods, which are K-Means and K-Medoid, are frequently used to cluster data. Both K-Means and K-medoid are similar in structures. The data is clustered into k groups. Conversely, there are some different minor points between these two methods. K-Means groups data points into k groups and constructs a center point inside each group using the mean value (Kantasa-Ard et al., 2019). However, K-Medoid constructs a center point inside each group using the median value. The performance of these two methods is comparable. But, K-Means will be sensitive to anomaly data and the number of clusters are frequently changed (Kassambara, 2018).



(Gunawardena, 2016)

Moreover, some partitional clustering studies are focused on epidemic cases. For example, the first research(Dzimińska et al., 2021) work implemented K-Means to analyze the water usage rate of inhabitants in the northern region of Poland. The research also considered the fluctuation of usage rate during the epidemic of Covid-19. In addition, the second study (Davahli et al., 2021)used the Self-Organizing Map (SOM), an unsupervised learning approach, to cluster the number of states depending on Covid-19 infection rates. The states that have similar behavior will be clustered in the same group. In addition, this second work forecasted the number of infection rates using a

recurrent neural network. Another research work (Mitra et al., 2020) also used K-Means to cluster and analyze the children and teenagers' physical behavior in Canada. The research analyzed their behaviors before and during the epidemic of Covid-19. The main reason is to manage the risk of public health for all urban inhabitants. In addition, some research works in the medical domain have implemented the concept of hybrid clustering. For instance, the research works (Simić et al., 2018, 2021) proposed combining the fuzzy partition method and maximum likelihood estimation clustering for diagnosing medical diseases. Another work (Guhathakurata et al., 2021) visualized the death rates of different countries, including South Asian countries, using the hierarchical clustering-based machine learning approach.

Due to the current research above, the clustering method is frequently implemented to cluster and analyze data during epidemic time. Nevertheless, this technique still limits implementation to manage vaccine distribution. Also, the concept of hybrid clustering was implemented in few works for the medical domain. Therefore, it will be an exciting aspect to implement this technique to improve vaccine distribution in urban areas, especially in the metropolis. This research does not focus only on the cluster of vaccine distribution. It also mentions how to increase the number of new vaccination centers in the urban area. Then, the concept of typical location decisions will describe more details in this perspective.

2.2 Typical Location Decisions

This section focuses on two parts. The first part will describe the principle of typical location and relevant parameters to decide an appropriate location. The second part will mention how to choose the appropriate location based on different location decision methods. All details are described below.

2.2.1 The principle of typical location and relevant parameters

There are many definitions to explain the meaning of location recently. In the logistics aspect, location is a place to serve all operations, such as production and

distribution processes and to satisfy customers in the supply chain(Apipratchayasakul, 2010). All business units establish different location strategies based on their business nature, cost, profit, and customer relations. Therefore, the location strategy is essential to all business units nowadays.

Another important thing before choosing an appropriate location is considering all relevant parameters that affect the location(Chaiyawong, 2017). Various parameters are needed to be considered. However, some parameters are frequently used in the location decision, such as raw materials, labor, transportation cost, end-customers, funding, regulations, and all involved facilities. For example, the location of the manufacturing or assembly process should be nearby to the labor community. Another example, the location of consumable goods distribution should be established near to raw materials. It is also important to think about how to choose the right location because location is so important for all type of business. The following section will describe how to choose appropriate locations via the location decision method.

2.2.2 Location decision method

Four location decision methods are frequently implemented in various business units(Siriliang & Jarupan, 2019).

1. Center-of-Gravity method:

This method is one of the most popular methods to decide the appropriate location with the lowest distribution cost. This method is constructed based on the position of relevant locations on the map and product quantities. The mathematical formula is shown in Equations (1)-(2).

$$\bar{X} = \frac{\sum Q_i X_i}{\sum Q_i} \quad (1)$$

$$\bar{Y} = \frac{\sum Q_i Y_i}{\sum Q_i} \quad (2)$$

Where:

Q_i is the quantity product at destination i

X_i is the position X at destination i

Y_i is the position Y at destination i

$\bar{X}$ is the center point of position X

$\bar{Y}$ is the center point of position Y

2. Factor-Rating method:

This method chooses the appropriate location by considering relevant factors such as labor cost, raw materials supply, customer segment, and marketing. These factors can be both quantitative and qualitative data. Several steps will explain how to choose the appropriate location in each area.

a. Preparing all relevant factors that affect the location decisions
b. Setting a weight and score levels of each factor
c. Putting appropriate score inside each factor
d. Calculating the summation between weight and score of each factor
e. Choosing the location that has the highest score

3. Load-Distance method:

This method determines the appropriate location by calculating the summation of multiplication between each location's total distance and transportation rate. The formula of distance calculation is shown in Equation (3) and Figure 5.

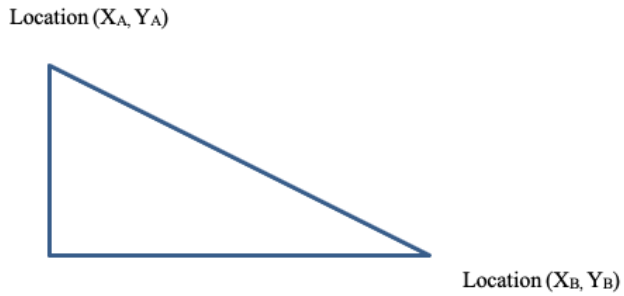


Figure 5 The example of distance calculation between locations A and B

$$d_{AB}=\sqrt{(X_A - X_B)^2 + (Y_A - Y_B)^2} \quad (3)$$

Where:

d_{AB} is the distance between locations A and B

X_A is the position X at location A

Y_A is the position Y at destination A

X_B is the position X at location B

Y_B is the position Y at destination B

Also, the multiplication of Load-Distance is shown in Equation (4).

$$LD = \sum l_{AB} * d_{AB} \quad (4)$$

Where:

LD is the total Load-Distance

l_{AB} is the transportation rate between location A and B

4 Linear Programming method:

This method implements the concept of solving-optimization problem to select the optimal location via the Linear Programming model. The objective is to minimize cost or maximize profit for product distribution. Three main components are considered in this method

a. Decision Variable: This component will define the decision of each variable.
b. Objective Function: This component will calculate an optimal solution.
c. Constraints: This component will provide the list of relevant constraints to support optimal solution.

Moreover, some relevant works focus on location decision methods. For instance, the first research(Panyasan & Klankamsorn, 2012) investigated how relevant factors affect the location selection of Automotive centers in Bangkok and metropolitan.

The research also implemented the center-of-gravity method to determine the appropriate location of the new Tire service center. Another research(Pluempirom & Ngamsaard, 2020) proposed the feasibility study of location decisions in the frozen food company. The center-of gravity was implemented to reduce the transportation cost of the frozen food distribution. This method was also integrated with AHP and factor-rating methods. The third research(Siriliang & Jarupan, 2019) demonstrated how to select the new distribution center using the center-of-gravity. The research also formulated the problem using Mathematical programming and run the model via CPLEX. These current works show that the center-of-gravity is frequently used to determine the appropriate location in different domains.

In this study, the center-of-gravity is chosen to investigate the location decision of vaccine distribution in Bangkok. This method also combines with the partitional clustering method. The objective is to balance the number of vaccination centers and target inhabitants in Bangkok. Inhabitants who stay in Bangkok will also have higher chances to get the vaccine after establishing new vaccination centers. The problem statements and objectives of this study will be described in the next section.

3. Problem Statements and Objectives

3.1 Problem statements

Regarding the vaccination situation in Bangkok and metropolitan areas, there are two main problems affecting the vaccination rate recently.

3.1.1 A large number of Covid-19 patients

The Covid-19 situation, as mentioned earlier (Department of Disease Control, 2021), demonstrated that the accumulated number of patients are higher than last year with the same period, particularly in Bangkok and metropolitan areas. One of the main problems is an inadequate vaccine to serve all populations during this epidemic time. In addition, many people become infected with delta variance, one of the most rapid variances that can easily be spread to many people within a short period.

3.1.2 The limitation of vaccination centers

Nowadays, there are approximately 25 vaccination centers in Bangkok to serve all populations in Bangkok and metropolitan areas (https://www.bangkokbiznews.com). However, some districts still do not have vaccination centers to serve populations. Besides, some vaccination centers are a bit far from the village or inhabitants' area.

These two problems affect the situation of vaccine distribution in Bangkok and metropolitan areas. Therefore, this study will propose a novel model to solve these problems with the following objectives below.

3.2 Objectives

Three main objectives are proposed to support this study.

1. Analyze how the factor correlation affects the cluster of vaccine distribution. This study considers some factors (e.g., the number of populations in each district, the number of accumulated patients, latitude-longitude of each district, and the number of current vaccination centers).
2. Develop a hybrid clustering model to suggest the proper number of vaccination centers inside each cluster.
3. Calculate extra vaccination centers to support some clusters that are missing vaccination centers.

4. Methodology

As shown in Figure 6, there are four steps to describe the construction of the hybrid clustering model. All steps are developed by R programming language(Flynt & Dean, 2016; Malika et al., 2014) in R studio.

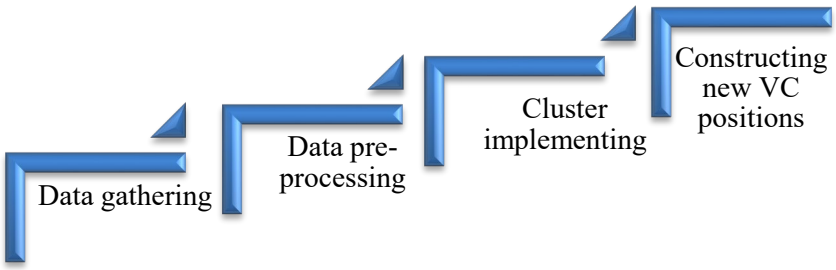


Figure 6 The steps of the hybrid clustering model

4.1 Data gathering

The input data is composed of the number of population in 50 districts of Bangkok(Bangkok, 2020). The number of population is updated up to 31 December 2020. Besides, the number of infections in each district from 1 April to 31 May 2021 is considered (Reference: Office of Health, Bangkok). The latitude-longitude positions of each district from Google Maps are also included. After gathering, all input data will be verified through the data pre-processing step.

4.2 Data pre-processing

In this step, data cleaning and data transformation are considered to pre-process data. Data cleaning can reduce the bias and make data applicable(Cadavid et al., 2019) before clustering. In this study, we check all missing and incomplete values for all input data as mentioned in the previous step. Then, we do the data transformation. The objective is to decrease errors and increase performance during training and forecasting (Cao et al., 2019; Zhang & Qi, 2005). Besides, data normalization using *MinMaxScaler* is considered to transform data in this study.

4.3 Cluster implementing

After completing the data pre-processing, the performance will be evaluated using Hopkins statistics(Banerjee & Rajesh N., 2004) before doing the clustering process. The data will be clustered using the partitional clustering. K-Means and K-Medoid are

considered for clustering data points. Also, all data points are clustered using Euclidean distance. NBClust is used to define appropriate k clusters for the data points in this study(Malika et al., 2014). Also, the cluster performance is measure using the Silhouette width value. In addition, there are two steps of the clustering process. Firstly, this study will cluster data to see the cluster behavior and trend. Each cluster will compose of the number of districts in Bangkok based on the similarity of input factors. Secondly, we will choose the cluster with the lowest ratio of vaccination centers per district to do sub-cluster. The objective is to validate which sub-cluster does not have any vaccination centers. For each sub-cluster, we assume that one vaccination center serves the population in three districts in case of no vaccination centers in the sub-cluster before.

4.4 Constructing new vaccination positions

When the list of sub-clusters is generated, we will see which cluster is missing the vaccination center. Then, the center-of-gravity is implemented to generate new positions of the vaccination center inside each cluster. The center-of-gravity formula was already mentioned in the literature reviews section. Besides, this study will compare the set of existing and new vaccination centers by displaying them on the Bangkok map. All details will be described in the next section.

5. Results analysis and discussion

5.1 Clustering performance

Based on the proposed dataset with relevant factors (e.g., the number of populations in each district, the number of accumulated patients, latitude-longitude of each district, and the number of current vaccination centers), the performance before cluster implementation is quite good with Hopkins statistic equal 0.64. In addition, the performance of partitional clustering is shown in Table 1. NBClust suggests the number of k clusters (2,3,4,10) in this experiment.

Table 1 The cluster performance of the 50 districts' dataset in Bangkok using the Silhouette width value

No. of K clusters	K-Mean	K-Medoid
K=2	0.26	0.24
K=3	0.31	0.25
K=4	0.27	0.23
K=10	0.23	0.22

The results in Table 1 show that K-Means with three clusters provide an optimal solution with the highest silhouette width value. Moreover, the data clustering is also visualized using the Principal Component Analysis (PCA), as shown in Figure 7. The total dimension 1 and 2 is equal to 63.3%

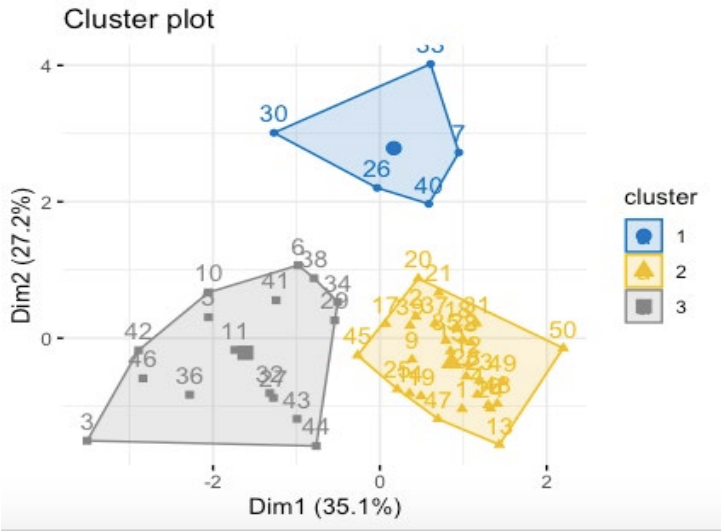


Figure 7 The cluster performance of 50 districts using K-Means

Furthermore, when investigated more details of three clusters, the results show that cluster 2 provides the lowest ratio of the vaccination center per district. All details are described in Table 2. Then, cluster 2 was chosen to do the sub-cluster in the second

step. The performance of location decisions via the sub-cluster method will be described in the next section.

5.2 Location decision performance

This section demonstrates the results of the sub-cluster method. The results in Table 3 show that some sub-clusters (e.g., SC2-1 and SC2-3) have already established vaccination centers. However, some sub-clusters (e.g., SC2-2, SC2-4, and SC2-5), still do not have vaccination centers. Based on our assumption, the assignment of new vaccination centers is proposed. The center-of-gravity method is proposed to define appropriate average latitude and longitude positions from the google map for new vaccination centers. All details are described in Tables 4-5. Table 4 describes the details of latitude and longitude positions of each district for sub-clusters that require new vaccination centers. Table 5 describes the details of estimated new positions of vaccination centers, which are calculated using data from Table 4. We can see that each sub-cluster has estimated vaccination centers at least once. In addition, the visualization of the estimated position of vaccination centers is shown in Figures 8-9. The map shows in Figure 9 is reference from (https://maps-bangkok.com/bangkok-district-map).

Table 2 The details of each cluster from 50 districts' dataset in Bangkok

Cluster order	Avg. Population	Avg. Patient	No. of VC	No. of district	VC Ratio per district
Cluster 1	120128	1108	11	5	2
Cluster 2	**89395**	**512**	**6**	**29**	**0.2**
Cluster 3	149694	452	8	16	0.5

Table 3 The details of vaccination centers inside sub-clusters 2

Sub-cluster order	No. of district	No. of current VCs	No. of estimated new VCs
SC2-1	2	2	0
SC2-2	3	0	1
SC2-3	4	4	0
SC2-4	6	0	2
SC2-5	5	0	1
SC2-6	5	0	1
SC2-7	4	0	1

Table 4 The details of latitude and longitude positions of each district for sub-clusters that require new vaccination centers.

Sub-cluster order	District order	Latitude position	Longitude position	The number of populations
SC2-2	19	13.7769	100.4565	103,617
	25	13.7952	100.5051	89,417
	48	13.7732	100.3531	78,749
SC2-4.1	2	13.7774	100.5206	83,897
	8	13.7585	100.5130	41,524
	37	13.7605	100.5342	69,264
SC2-4.2	17	13.7776	100.5795	84,233
	39	13.7442	100.5856	81,623
	45	13.7644	100.6057	107,458
SC2-5	1	13.7666	100.4984	44,923
	4	13.7307	100.5235	45,757

Sub-cluster order	District order	Latitude position	Longitude position	The number of populations
	13	13.7314	100.5140	21,324
SC2-6	12	13.6967	100.5425	76,564
	24	13.6819	100.5057	78,687
	49	13.6117	100.5084	123,700
SC2-7	15	13.7251	100.4858	103,377
	22	13.7149	100.4369	124,318
	35	13.6776	100.4844	148,290

Table 5 The details of estimated new positions of vaccination centers

Sub-cluster order	No. of district	Avg. latitude position	Avg. longitude position	Estimated area of new VCs position
SC2-2	19,25,48	13.78185	100.44253	Promlikhit badminton, Revenue Office
SC2-4.1	2,8,37	13.76736	100.52382	Rama hospital, Faculty of Science MU
SC2-4.2	17,39,45	13.76244	100.59162	Parco Sukaparp Huai Kwan, Thai-Japan Ass School
SC2-5	1,4,13	13.74523	100.51162	BKK Revenue Office, Wat Pub Pla Chai
SC2-6	12,24,49	13.65483	100.51700	Market Thai Somboon Phrapadeang
SC2-7	15,22,35	13.70299	100.46908	Public Health Center 29

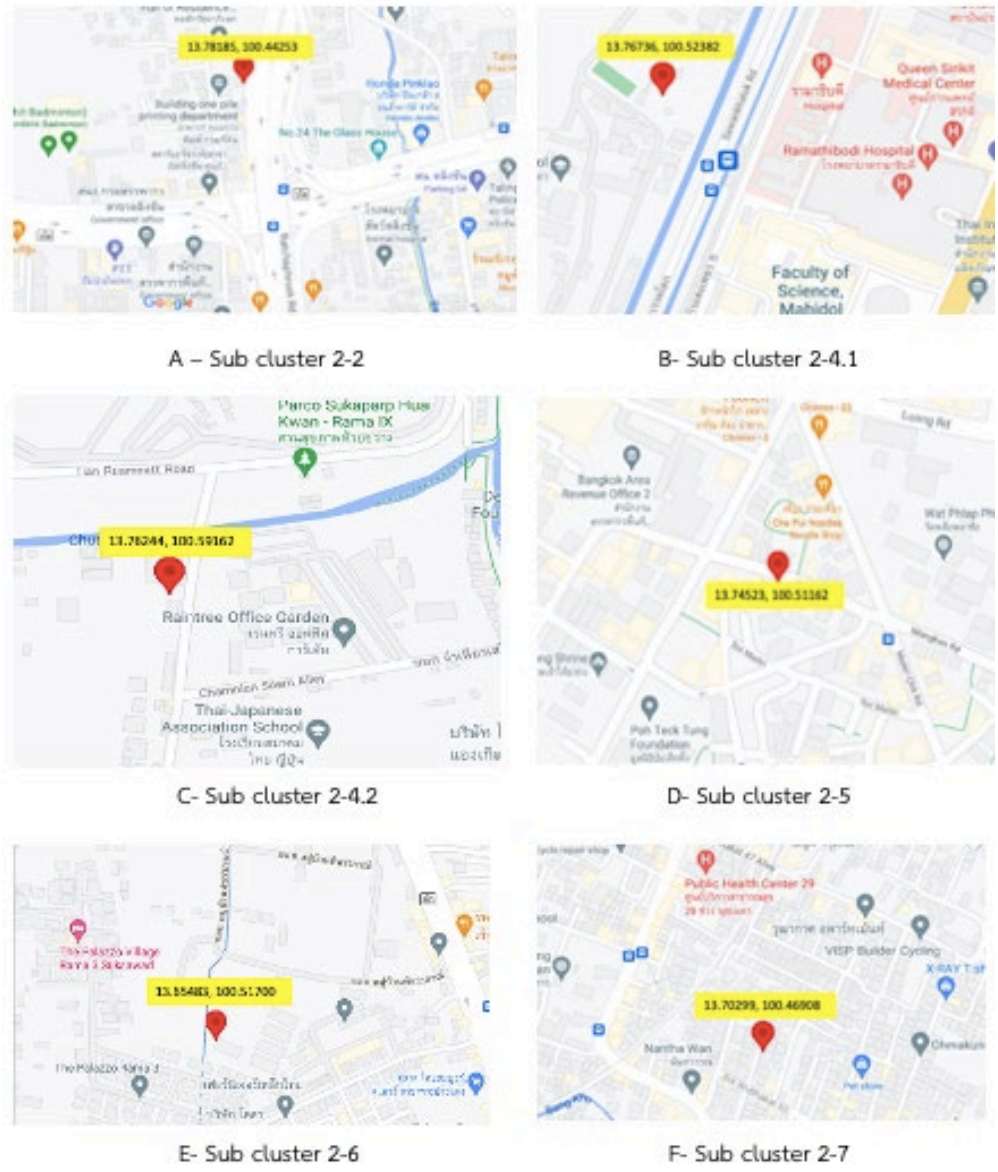


Figure 8 The visualization of the estimated vaccination centers in each cluster (A-F)

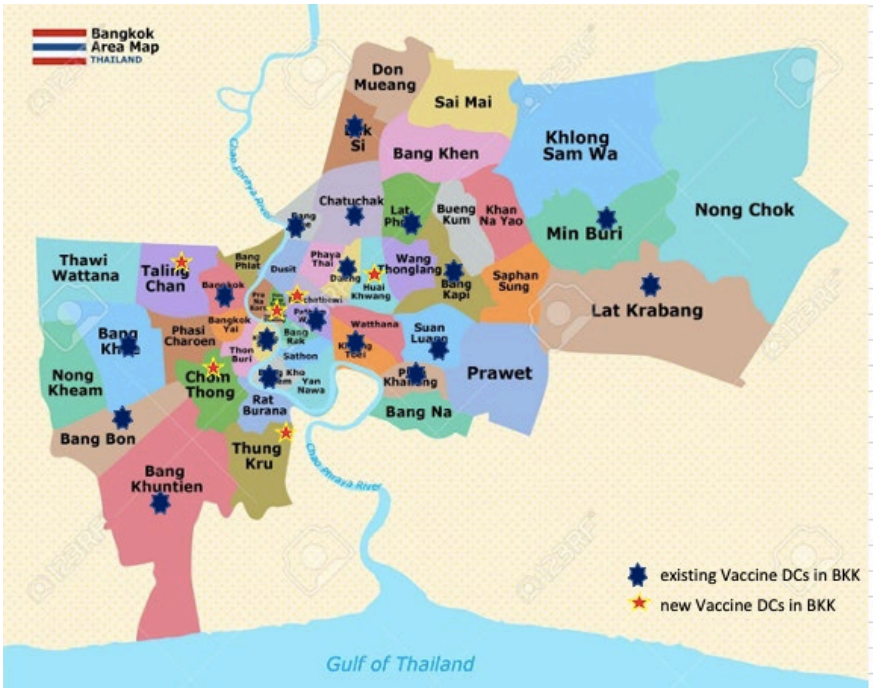


Figure 9 The visualization of exist and new vaccination centers in Bangkok

5.3 Discussion

Due to the proposed positions of new vaccination centers, we can see that each position is closed to many public organizations, such as the local market, hospitals, and revenue offices. Many people can access these places using public transportation easily. Also, the results and visualizations demonstrate the opportunity to the central government to establish some extra vaccination centers supporting all inhabitants who live in proposed areas or nearby. This strategy can reduce the density of people and avoid extra Covid-19 infections in current vaccination centers. In addition, the clustering method is not only implemented and provides good results in this research. It is also applied in different case studies. For example, the clustering method was already implemented to balance the number of distribution hubs (Kantasa-Ard et al., 2019), the number of taxi fleets (Chang et al., 2009), and the number of delivery points for pizza (Gunawardena, 2016). As seen from previous research above, we can confirm that this

method provides the appropriate number of data points in each cluster in different contexts.

6. Conclusion

This study proposes the hybrid clustering model, which combines both partitional clustering and the center-of-gravity methods. For partitional clustering, K-Means provides the best clustering performance with the highest silhouette width value. The first three clusters are suggested. Then, we implemented the sub-cluster method inside cluster 2 to see the trend and behavior of 50 districts in Bangkok. The results illustrated that some sub-clusters are still missing the vaccination center. Therefore, we propose the center-of-gravity method to define the estimated position of the vaccination center. This study also compares the number of new positions and existing positions. The results showed that we should add six more vaccination centers to serve the population far from 25 current vaccination centers, which are established by the central government. Then, the total vaccination centers are equal to 31 centers. Moreover, additional vaccination centers can serve more populations in some areas nearby, and many people can access extra vaccination centers easily.

Acknowledgement

This research was developed to be the best practice for vaccination distribution during the epidemic time. We would like to thank you with the data from Ministry of Public Health and Bangkok metropolis. This research was the grant from the internal scholarship from Burapha University.

References

Apipratchayasakul, K. (2010). *Inventory Management* (1st ed.). Focus Media And Publishing Company LTD. http://lib.neu.ac.th/ULIB/dublin.php?ID=16126

Banerjee, A., & Rajesh N., D. (2004). Validating clusters using the Hopkins statistic. *2004 IEEE International Conference on Fuzzy Systems (IEEE Cat. No.04CH37542)*, 149–153. https://doi.org/10.1109/FUZZY.2004.1375706

Bangkok, S. (2020). *Bangkok 2020 Statistics*. http://www.bangkok.go.th/upload/user/00000130/BMA_STATISTICS 2563/ebook 63.pdf

Cadavid, J. P. U., Lamouri, S., Grabot, B., & Fortin, A. (2019). Machine Learning in Production Planning and Control : A Review of Empirical Literature. *IFAC-PapersOnLine 52*, *13*, 385–390.

Cao, J., Li, Z., & Li, J. (2019). Financial time series forecasting model based on CEEMDAN and LSTM. *Physica A: Statistical Mechanics and Its Applications*, *519*, 127–139. https://doi.org/10.1016/j.physa.2018.11.061

Chaiyawong, K. (2017). *A guideline for selecting the location of super store* [Thammasat University]. http://ethesisarchive.library.tu.ac.th/thesis/2017/TU_2017_5716033187_3158_8131.pdf

Chang, H. wen, Tai, Y. chin, & Hsu, J. Y. jen. (2009). Context-aware taxi demand hotspots prediction. *International Journal of Business Intelligence and Data Mining*, *5*(1), 3. https://doi.org/10.1504/ijbidm.2010.030296

Davahli, M. R., Karwowski, W., & Fiok, K. (2021). Optimizing COVID-19 vaccine distribution across the United States using deterministic and stochastic recurrent neural networks. *PLoS ONE*, *16*(7 July), 1–14. https://doi.org/10.1371/journal.pone.0253925

Department of Disease Control, M. (2021). *Covid-19 statistic Thailand*. https://ddc.moph.go.th/

Dong, E., Du, H., & Gardner, L. (2020). *An interactive web-based dashboard to track COVID-19 in real time*. Lancet Inf Dis. https://doi.org/10.1016/S1473-3099(20)30120-1

Dzimiń ska, P., Drzewiecki, S., Ruman, M., Kosek, K., Mikołajewski, K., & Licznar, P. (2021). The use of cluster analysis to evaluate the impact of covid-19 pandemic on daily water demand patterns. *Sustainability (Switzerland)*, *13*(11). https://doi.org/10.3390/su13115772

Flynt, A., & Dean, N. (2016). A Survey of Popular R Packages for Cluster Analysis. *Journal of Educational and Behavioral Statistics*, *41*(2), 205–225. https://doi.org/10.3102/1076998616631743

Guhathakurata, S., Saha, S., Kundu, S., Chakraborty, A., & Banerjee, J. S. (2021). South Asian countries are less fatal concerning COVID-19: a hybrid approach using machine learning and M-AHP. In *Computational Intelligence Techniques for combating COVID-19* (pp. 1–26). Springer, Cham.

Gunawardena, T. (2016). *Algorithms□: Clustering*. https://www.slideshare.net/tilanigunawardena/hierachical-clustering?from_action=save

Hannah, R., Esteban, O.-O., Diana, B., Edouard, M., Joe, H., Bobbie, M., Charlie, G., Cameron, A., Lucas, R.-G., & Max, R. (2020). *Coronavirus Pandemic (COVID-19)*. OurWorldInData.Org. https://ourworldindata.org/coronavirus

Kantasa-Ard, A., Nouiri, M., Bekrar, A., Ait El Cadi, A., & Sallez, Y. (2019). Dynamic clustering of PI-hubs based on forecasting demand in physical internet context. In *Studies in Computational Intelligence* (Vol. 853, pp. 27–39). Springer Verlag. https://doi.org/10.1007/978-3-030-27477-1_3

Kassambara, A. (2018). *Type of clustering methods: overview and quick start R code*. https://www.datanovia.com/en/blog/types-of-clustering-methods-overview-and-quick-start-r-code/

MacQueen, J. (1967). Some Methods for classification and Analysis of Multivariate Observations. *5th Berkeley Symposium on Mathematical Statistics and Probability 1967*. https://doi.org/citeulike-article-id:6083430

Malika, C., Nadia, G., Boiteau, V., & Niknafs, A. (2014). NbClust: An R Package for Determining the Relevant Number of Clusters in a Data Set. *Journal of Statistical Software*, *61*(6), 1–36. http://www.jstatsoft.org/v61/i06/paper

Mitra, R., Moore, S. A., Gillespie, M., Faulkner, G., Vanderloo, L. M., Chulak-Bozzer, T.,

Rhodes, R. E., Brussoni, M., & Tremblay, M. S. (2020). Healthy movement behaviours in children and youth during the COVID-19 pandemic: Exploring the role of the neighbourhood environment. *Health and Place*, *65*(August), 102418. https://doi.org/10.1016/j.healthplace.2020.102418

Murray, P. W., Agard, B., & Barajas, M. A. (2015). Forecasting supply chain demand by clustering customers. *IFAC-PapersOnLine*, *28*(3), 1834–1839. https://doi.org/10.1016/j.ifacol.2015.06.353

Nananukul, N. (2013). Clustering model and algorithm for production inventory and distribution problem. *Applied Mathematical Modelling*, *37*(24), 9846–9857. https://doi.org/10.1016/j.apm.2013.05.029

Panyasan, C., & Klankamsorn, S. (2012). Site Selection of Automotive Changing Tire Center in Bangkok and Metropolitan. *IE Network Conference*, 489–491. http://www.dms.eng.su.ac.th/filebox/FileData/WPS012.pdf

Pluempirom, J., & Ngamsaard, W. (2020). Feasibility Study and Finding a Suitable Location for Frozen Food Distribution Center Investment Project: A Case Study of XYZ Company Limited. *The 15th RSU National Graduate Research Conference*, 1145–1156.

Simić, S., Banković, Z., Simić, D., & Simić, S. D. (2018). A Hybrid Clustering Approach for Diagnosing Medical Diseases. *Hybrid Artificial Intelligent Systems*, *10870*, 741–752. https://doi.org/https://doi.org/10.1007/978-3-319-92639-1_62

Simić, S., Banković, Z., Villar, J. R., Simić, D., & Simić, S. D. (2021). A hybrid fuzzy clustering approach for diagnosing primary headache disorder. *Logic Journal of the IGPL*, *29*(2), 220–235. https://doi.org/10.1093/jigpal/jzaa048

Siriliang, P., & Jarupan, P. (2019). *Selection of Location for Good Transportation Case Study of Delivery Service Company for Consumer Goods* [King Mongkut's University of Technology North Bangkok]. http://mhle.eng.kmutnb.ac.th/upload/student-abstract/d0147d95ff96ec10f1f3b59c77b9e43b.pdf

Zhang, G. P., & Qi, M. (2005). Neural network forecasting for seasonal and trend time series. *European Journal of Operational Research*, *160*(2), 501–514. https://doi.org/10.1016/j.ejor.2003.08.037