

IMPLEMENTATION OF K-MEANS CLUSTERING ALGORITHM TO ANALYZE THE FAMILIAL SENTIMENTS TOWARDS COVID-19 VACCINATION FOR ELEMENTARY SCHOOL STUDENTS IN KALAWAT DISTRICT

Indah Kairupan¹, Liza Wikarsa², Audreyvia G. C. A Kembuan³

^{1,2,3}Universitas Katolik De La Salle Manado

ikairupan@unikadelasalle.ac.id¹, lwikarsa@unikadelasalle.ac.id², 18013011@unikadelasalle.ac.id³

Abstract-- Due to the Ministry of Health's policy, the Indonesian government mandates the public to receive the COVID-19 vaccination as a form of immunity against the Coronavirus. This vaccination is not only for adults but also for children of a certain age. Regarding the provision of vaccination for elementary school students aged between 6 to 11 years, the families' responses to this predicament can cause significant barriers to those students being fully vaccinated. Thus, this research developed a web-based application that incorporated the K-means clustering method to group the sentiments of the families into three clusters, namely positive, neutral, and negative. The results show that the application can identify and cluster the different familial responses from 279 respondents in Kalawat District toward the administration of COVID-19 vaccination to their underage children. The most dominant familial sentiment is positive followed by neutral and negative sentiments with the number of respondents as many as 120 respondents (43%), 113 respondents (41%), and 46 respondents (16%) respectively. This research can help the Health Office in North Minahasa Regency to evaluate public sentiments about vaccination for elementary school students as well as look for better ways to encourage vaccine trust and confidence in this district.

Keywords: Clustering; COVID-19; K-Means; Vaccination.

I. INTRODUCTION

COVID-19 vaccination in Indonesia began on May 7, 2021, marked by the issuance of a regulation from the Minister of Health. Due to the rapid spread of the COVID-19 virus which can attack anyone, the government is trying to prevent the spread of the virus by vaccinating [1]. Vaccination is carried out by entering the inactive SARS-CoV-2 virus so that the immune system can recognize the virus and produce antibodies

that can prevent infection caused by the COVID-19 virus [2], [3]. With this vaccination effort, it is hoped that herd immunity will be formed so that the COVID-19 pandemic in Indonesia ends soon [1].

The Minister of Health again issued a regulation on December 13, 2021, regarding the provision of vaccinations to 26,5 million children aged 6 to 11 years [2]. With the start of the vaccination process for elementary school students, various responses emerged from families who have children aged 6 to 11 years [4]. There are concerns about the appropriateness of the vaccine to be used and the post-vaccination risks [5], [6]. Hence, the familial responses affect the administration of vaccination for these elementary school students and therefore it is necessary to know the familial sentiments.

Sentiment analysis is a computational detector, of opinion, sentiment, response, and subjectivity in a text [7]. Using sentiment analysis methods, several previous studies have tried to identify parents' responses to the mandatory COVID-19 vaccination of elementary school students, including the causes of COVID-19 vaccination and its implementation system. However, these studies only focused on one school so generalizations cannot be made about the acceptance of this vaccination on a larger scale [4-6]. In this regard, this research aims to identify the familial sentiments towards the administration of COVID-19 vaccination to students in 14 elementary schools in Kalawat district using the K-means clustering. By doing this research, it can

provide relevant perception and concerns of the families with regard to the matter in question.

K-means clustering is one of the data mining algorithms for unsupervised learning that has been widely used in various studies [8]-[11]. The study entitled Clustering of Vaccination Reception in Central Java Using the K-Means Method [12] classified vaccine acceptance in Central Java into 2 clusters. Cluster 1 was an area with low acceptance and Cluster 2 with high acceptance [12]. Cluster 1 consisted of 5 regions and Cluster 2 had 30 regions with the highest number of vaccination receipts. The second study was obtained from the journal Clinical Practice [9] which had 1,035 respondents. Yet, there were only 659 respondents who expressed their unwillingness to be vaccinated due to the confidence and trust issues regarding the administration of COVID-19 vaccinations to children [9]. The third study also used the K-means clustering algorithm to group the 900 recipients of the COVID-19 vaccination in Bengkulu city [11]. There are 2 types of vaccines given, namely AstraZeneca and Sinovac. The K-means clustering algorithm is used to classify the availability of vaccine stocks for the residents at the appointed health centers [13].

This research will group the responses from parents regarding the administration of COVID-19 vaccination to elementary students in 14 schools in Kalawat District with a population of 1,440 students. The number of respondents obtained using the Isaac and Michael sampling technique [11] with a 5% significance level was 279 respondents. The seven attributes used in this research are the enthusiasm for vaccines, the importance of vaccination, the level of knowledge, the environmental support, the school requirement for vaccinations, and the level of trust. Sentiments will be grouped into three, namely positive, neutral, and negative sentiments [4], [14] using the K-means clustering algorithm [10], [13]. Based on the description above, this study aims to determine the familial responses regarded the administration of COVID-19 vaccination to students aged between 6 and 11 years old and then group those sentiments into three clusters, namely positive, neutral, and negative sentiments. The research results can be used by the Health Office in North Minahasa

Regency as an insight to make better health and safety policies to further encourage vaccine trust and confidence for families with underage children.

The rest of this article is organized as follows. Section 1 will address the research background to understand the main context of this study. Section 2 discusses the related works and research methods. Section 3 shows the research results and discussion. Section 4 provides the conclusion and future work.

II. METHOD

The development framework used in the current study is divided into four stages, namely the data collection stage, the pre-processing stage, the processing stage, and the post-processing stage, as seen in Fig. 1.

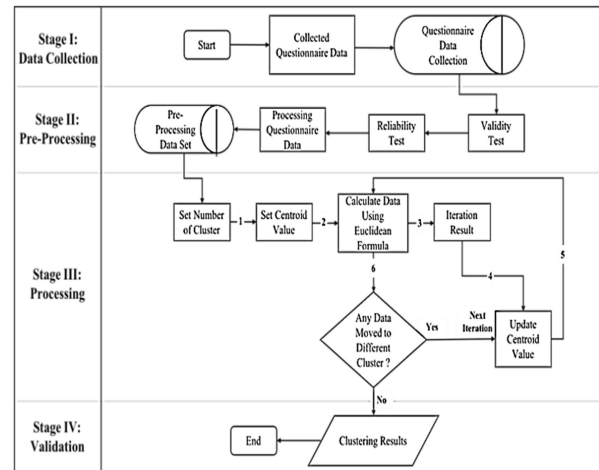


Fig. 1. Research Framework K-Means [13]

A. Stage I: Data Collection

This stage is the stage of distributing questionnaires to the families of elementary school students in the Kalawat district. The number of data or respondents will be determined by the formula for determining the number of samples, namely Table Isaac and Michael with a significance level of 0.05. The collection of questionnaires aims to obtain the value of each variable which will later be used in the application of the K-Means method [15], [16].

The data used were 279 parental responses to giving COVID-19 vaccinations to elementary school students with an age range of 6-11 years [17], [18]. Questionnaires were distributed twice in this study, where the first questionnaire was distributed to identify the independent variables. The second questionnaire to dig deeper related to the independent variables that have been obtained

from the first questionnaire. The results of the second questionnaire are indicators for each of these independent variables.

B. Stage II: Pre-processing

Several statistical tests will be applied to the questionnaire data collection. It is useful for testing the hypotheses involving population parameters. Pearson's formula will be used for data validity and correlation. In addition, the reliability test will use Cronbach's Alpha formula and the T-test.

C. Stage III: Processing

After passing through the pre-processing stage, data collection will be carried out. The K-means clustering algorithm is embedded in the application so that the clustering process is carried out through the application to be built. K-means is one of the data mining methods to perform the unsupervised learning data modeling process and uses the data grouping method of various partitions [10]. the objective of K-Means serves to minimize the function of objects that have been set in the grouping process. The following are the steps of the K-means clustering algorithm [10]:

1. The number of clusters is determined.
2. The initial centroid value is determined by taking the average value of each variable.
3. The Euclidean formula is used to calculate the distance for each data based on the shortest distance between the data and the centroid value.

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (1)$$

Note:

- d = distance between x and y
- x = data attribute
- y = cluster center data
- i = each data
- n = number of data
- x_i = data on each data to i
- y_i = data at the center of the cluster to i

4. Data that has the closest distance to the centroid value must be grouped.

$$\min \sum_k^i -a_{ik} = \sqrt{\sum_i^m (C_{ij} - C_{kj})^2} \quad (2)$$

5. A new centroid/center value is defined.

$$C_{kj} = \frac{\sum_k^i x_{ij}}{p} \quad (3)$$

6. Recalculate by repeating steps 3 to 5.
7. Grouping will be declared complete if there is no more data moving clusters/groups. However, if it is still found that something is moving, the third step must be repeated until no more data is transferred.

D. Stage IV: Validation

The validation stage is the final stage which will display the results of grouping the three sentiment clusters namely positive, neutral, and negative. Visual displays in the form of tables and graphs will be presented to display the number of people in the cluster.

III. RESULT AND DISCUSSION

A. Analysis

This study adopted a system rapid application development with four stages that are analysis, design, implementation, and testing. This analysis contains data collection and processing, problem identification, target user analysis, requirements specifications, and K-means clustering algorithm analysis.

1. Data Collection

Data on the familial responses to the COVID-19 vaccination was collected to get the percentage of what actions were deemed necessary by the respondents. The first questionnaire was to identify the independent variables, the second questionnaire was to dig deeper into the independent variables that have been obtained from the first questionnaire. In this questionnaire, there were 7 variables used (shown in Table I), such as enthusiasm [19], how important is vaccination? knowledge level [20], support from the environment [21], [22], vaccinations to enter school [23], reduce COVID-19 cases [22], belief in vaccination [24], [25].

2. Pre-Processing

At this stage, validity, reliability, and t-test tests will be carried out on the questionnaire [23]. The test results can be seen in Table II.

TABLE I
Research Variables and Indicators

Variable	Indicator	Initial (%)	Final (%)	Variable Weight (%)
Knowledge	Enthusiastic	74.3	14.4	43
	How important is Vaccination?	73.3	14.3	
	Knowledge level	73.5	14.3	
Environment	Support from the Environment	73	14.2	42.8
	Vaccinations to Enter School	74.6	14.4	
	Reduce COVID-19 Cases	72.7	14.2	
Trust	Belief in Vaccination	72.7	14.2	14.2

TABLE II
Questionnaire Validity and Correlation Test

Item Code	r _{count}	r _{table}	Significance	Validity
Q1	0,880**	0,113	<0,001	Valid
Q2	0,914*	0,113	0,018	Valid
Q3	0,910**	0,113	0,002	Valid
Q4	0,902**	0,113	<0,001	Valid
Q5	0,927**	0,113	0,001	Valid
Q6	0,936**	0,113	<0,001	Valid
Q7	0,937**	0,113	<0,001	Valid

The data will be considered valid if the value of r_{count} exceeds the number of r_{table}. The number of respondents will determine the value of the r_{table}. The validity test in Table III shows the calculated r-value of the seven items in the questionnaire where the significance value is bigger than 0.001. The calculated r-value is greater than r_{table} for each item, namely 0.113, therefore it is known that the correlation value of each variable is connected, and it can be concluded that the tested variables correlate. The reliability test for the research instrument is shown in Table III.

TABLE III
Reliability Test

Cronbach's Alpha	Items Quantity
0.968	7

The Cronbach's alpha value of the reliability test on the questionnaires was 0.98 which is considered as best internal consistency [22]. In statistics, the t-test is part of a hypothesis test to determine whether the independent variable can partially affect the dependent variable.

TABLE IV
T-test

Item Code	T _{Result}	T _{Table 0.05}	Acceptance
Knowledge Level	3.72	1.968565	Accepted
Reduce COVID-19 Cases	3.67	1.968565	Accepted
Enthusiastic	3.68	1.968565	Accepted
How important is Vaccination?	3.66	1.968565	Accepted
Belief in Vaccination	3.73	1.968565	Accepted
Support from the Environment	3.64	1.968565	Accepted
Vaccinations to Enter School	3.65	1.968565	Accepted

3. Processing

The following is an analysis of the application of the K-means method to cluster the familial responses toward the administration of COVID-19 vaccinations to elementary school students in the Kalawat district. The training data used 5 data and the iteration process obtained as many as 2 iterations. Also, one needed to prepare the dataset before use as elaborated below.

a. Scores were obtained from the collection of questionnaires about the administration of giving COVID-19 vaccinations to elementary school students aged between 6-11 years.

TABLE V
Scores for Questionnaires

Respondent	Q1	Q2	Q3	Q4	Q5	Q6	Q7
A1	5	4	4	3	3	4	4
A2	4	5	4	5	4	5	4
A3	3	3	3	3	3	3	3
A4	3	3	3	3	4	3	3
A5	1	1	2	2	2	1	2

b. In Table I, the level of importance of the action will be multiplied by the value of the action obtained from the respondents.

c. In Table VI, the average value is obtained from the sum of the items included in the same

variable or type of indicator which is then divided by three.

TABLE VI

Example of the Processed Data for 1 Respondent

Id Respondent	001
Age	10
Gender	Male
Knowledge level	5
Reduce COVID-19 Cases	4
Enthusiastic	5
How important is Vaccination?	4
Belief in Vaccination	5
Support from the Environment	4
Vaccinations to Enter School	5

d. The amount of data processed is 279 data. Applications can only read input data in the form of a CSV extension, therefore, after it has been successfully processed, the input data will be stored in the CSV extension type so that the application can read it.

e. The implementation of K-means followed the steps:

1) Determine the number of clusters.

There are 3 sentiment clusters such as positive (C1), neutral (C2), and negative (C3).

2) Determine the initial centroid/center value randomly.

The initial centroid has an important role in determining the final result of the grouping. This stage will calculate the average value of the overall score of the questionnaire, the average value will be used as a reference to code labels on each respondent's data.

TABLE VII

Percentages for Indicators Scoring

	Result		Result
Q1.1	2,645	Q2.1	2,645
Q1.2	2,828	Q2.2	0
Q1.3	6,633	Q2.3	7,874
Q3.1	4,358	Q4.3	4,358
Q3.2	0	Q5.1	8,0622
Q3.3	4	Q5.2	4
Q4.1	4	Q5.3	0
Q4.2	1		

3) The Euclidean distance formula is used to calculate the distance for each data group based on the closest distance to the centroid value.

4) Data that is close to the centroid value will be grouped.

5) The new centroid value is determined.

6) Repeat steps C to D to calculate the distance from each data.

7) Grouping will be declared complete if there are no more data-moving clusters. However, if there is still data that has moved, then step C will be repeated until you get the same data grouping results as the previous iteration.

Below is an excerpt of the manual calculation.

$$d(1,1) = \sqrt{(5-5)^2 + (4-4)^2 + (4-5)^2 + (3-4)^2 + (3-5)^2 + (4-4)^2 + (4-5)^2} = 2,645$$

$$d(1,2) = \sqrt{(5-3)^2 + (4-3)^2 + (4-3)^2 + (3-3)^2 + (3-3)^2 + (4-3)^2 + (4-3)^2} = 2,828$$

$$d(1,3) = \sqrt{(5-1)^2 + (4-1)^2 + (4-2)^2 + (3-2)^2 + (3-2)^2 + (4-1)^2 + (4-2)^2} = 6,633$$

$$d(2,1) = \sqrt{(4-5)^2 + (5-4)^2 + (4-5)^2 + (5-4)^2 + (4-5)^2 + (5-4)^2 + (4-5)^2} = 2,645$$

$$d(2,2) = \sqrt{(4-3)^2 + (5-3)^2 + (4-3)^2 + (5-3)^2 + (4-3)^2 + (5-3)^2 + (4-3)^2} = 0$$

$$d(2,3) = \sqrt{(4-1)^2 + (5-1)^2 + (4-2)^2 + (5-2)^2 + (4-2)^2 + (5-1)^2 + (4-2)^2} = 7,874$$

$$d(3,1) = \sqrt{(3-5)^2 + (3-4)^2 + (3-5)^2 + (3-4)^2 + (3-5)^2 + (3-4)^2 + (3-5)^2} = 4,358$$

$$d(3,2) = \sqrt{(3-3)^2 + (3-3)^2 + (3-3)^2 + (3-3)^2 + (3-3)^2 + (3-3)^2 + (3-3)^2} = 0$$

$$d(3,3) = \sqrt{(3-1)^2 + (3-1)^2 + (3-2)^2 + (3-2)^2 + (3-2)^2 + (3-1)^2 + (3-2)^2} = 4$$

$$d(4,1) = \sqrt{(3-5)^2 + (3-4)^2 + (3-5)^2 + (3-4)^2 + (4-5)^2 + (3-4)^2 + (3-5)^2} = 4$$

$$d(4,2) = \sqrt{(3-3)^2 + (3-3)^2 + (3-3)^2 + (3-3)^2 + (4-3)^2 + (3-3)^2 + (3-3)^2} = 1$$

$$d(4,3) = \sqrt{(3-1)^2 + (3-1)^2 + (3-2)^2 + (3-2)^2 + (4-2)^2 + (3-1)^2 + (3-2)^2} = 4,358$$

$$d(5,1) = \sqrt{(1-5)^2 + (1-4)^2 + (2-5)^2 + (2-4)^2 + (2-5)^2 + (1-4)^2 + (2-5)^2} = 8,0622$$

$$d(5,2) = \sqrt{(1-3)^2 + (1-3)^2 + (2-3)^2 + (2-3)^2 + (2-3)^2 + (1-3)^2 + (2-3)^2} = 4$$

$$d(5,3) = \frac{\sqrt{(1-1)^2 + (1-1)^2 + (2-2)^2 + (2-2)^2 + (2-2)^2 + (1-1)^2 + (2-2)^2}}{=0}$$

TABLE VIII
First Iteration

	C1	C2	C3	Result
R1	2.645751	2.828427	6.63325	1
R2	2.645751		7.874008	2
R3	4.358899		4	2
R4	4	1	4.358899	2
R5	8.062258	4	0	3

Update the center of the cluster or centroid with the formula for the value of each variable from the grouping results divided by the total amount of data in a cluster. The results are as follows:

TABLE IX
New Centroid Values for the Last Iteration

	Q1	Q2	Q3	Q4	Q5	Q6	Q7
C1	5	4	5	4	5	4	5
C2	3.33	3.67	3.33	3.67	3.667	3.667	3.33
C3	1	1	2	2	2	1	2

The findings showed that no data moved to another cluster after the 2nd iteration. The results presented the responses of respondents regarding the response to COVID-19 vaccinations for elementary school students aged 6-11 years that have been carried out so far are presented in Table X.

TABLE X
The Results of Clustering

R*	C1	C2	C3	Result
1	10.34	2.48	34.51	2
2	11.78	2.48	7.87	2
3	7.93	1.63	4	2
4	8.36	1.63	4.35	2
5	4.35	5.62	0	3

*R=respondent

This section will implement the application design that has been made in the previous section by following the development stage of the RAD methodology.

Requirements Specification

- Application users can upload respondent data that has been processed through the pre-processing stage into the application in the form of a CSV file.
- On the home page, the application will display the respondent's data.

- In the application, the initial centroid can be processed randomly by taking the respondent's data that has been previously labeled.
- After the initial centroid is processed randomly, the application can perform data processing by calculating the distance of the respondent's data using the Euclidean distance formula.
- The new centroid at each iteration can be updated automatically according to the application of the K-means clustering method.
- To view the details of the iteration results, the user can select an iteration from the entire list of iterations displayed by the application through the grouping results of the K-means clustering method.
- The final results are presented in the form of 2 visualizations. The graphical form displays how many people from each cluster. While the tabular form presents information on the number of respondents who are members of each cluster and the new centroid used in iterations with the results of calculating the Euclidean distance.

4. Validation

The 7-fold cross-validation (K=7) was used in this research by splitting the dataset into seven folds. This stage will show the results of the 3 clusters of sentiment which is positive, neutral, and negative. Each cluster will indicate how many respondents are in the cluster on several visual displays such as a table or chart.

B. Design

Table XI shows the flow of the program module design in pseudocode form which started from the process of determining the cluster and then determined the centroid, defined the Euclidean Distance value, and calculated the object distance of the Euclidean value. Also, there was a decision whether the object has moved, if there was not then the process is complete. However, if there were still moving objects, it would return to the process of determining the centroid.

TABLE XI
Module Program

```

Pseudocode: Calculate Euclidean Distance


---


Input:
D {atr1, atr2..n} // set of elements
K //number of desired clusters
Output:
K //Set of Clusters
K-Means Algorithm:
Assign initial values for C1, C2, C3
Repeat
Assign each item to the cluster which has the closest mean;
Calculate the new mean for each cluster;
Until convergence criteria are met;
    
```

Fig. 2 displays the database design for the application that contains the response values from the questionnaires in the previous stage. The centroid value at each iteration and the squared distance value using the Euclidean Distance formula are also included in this database.

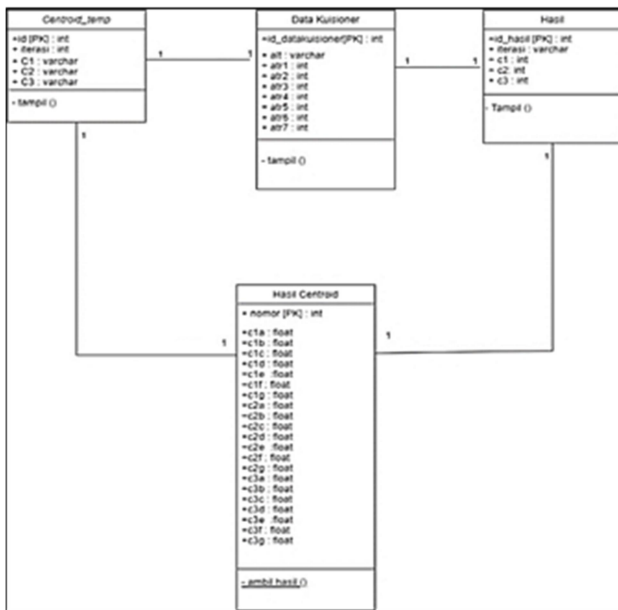


Fig. 2. Database Design

Data Collection

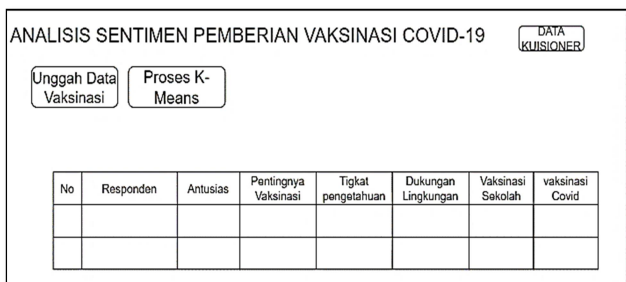


Fig. 3. Home Page

Fig. 3 contains 3 important sections. First, the process to include the CSV file. Second, is a

display of data files that have been successfully uploaded. Third, if you press the K-means Process button, then process the data for the grouping stage.

a. Grouping

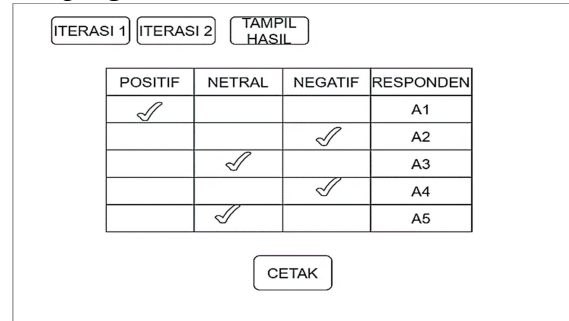


Fig. 4. A Page for Grouping

This page has 5 important sections, including the iteration 1 button and, the next iteration button. both are steps that need to be done before the Euclidean Distance formula is executed.

b. Processing and Result

One can press the "Acak Centroid" button to randomize the initial centroid.

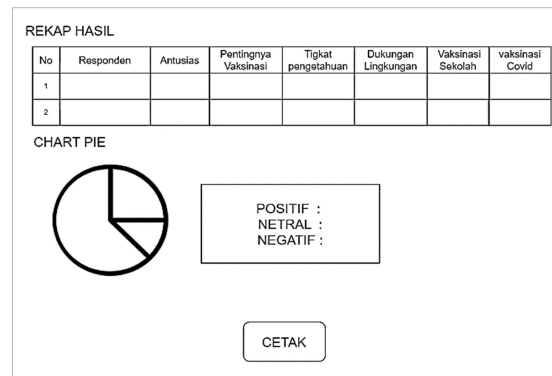


Fig. 5. Iteration Detail

This page displays the details of the iteration results of the grouping carried out. Data will be displayed in several visual forms such as tables and graphs. This page also shows the details of the centroid value and the results of calculating the distance from the Euclidean formula.

C. Implementation

This section will implement the application design that has been made in the previous section by following the development stage of the RAD methodology.

TABLE XII
Development Environment

Device Type	Resource
Hardware	Minimum Laptop with specifications: CPU Intel Celeron, RAM 4GB, Windows 11.
Software	<ol style="list-style-type: none"> 1. PHP7 2. Php MyAdminversi 3. Visual Studio Code version 1.70.2 4. Google Chrome version 104.0.5112.102

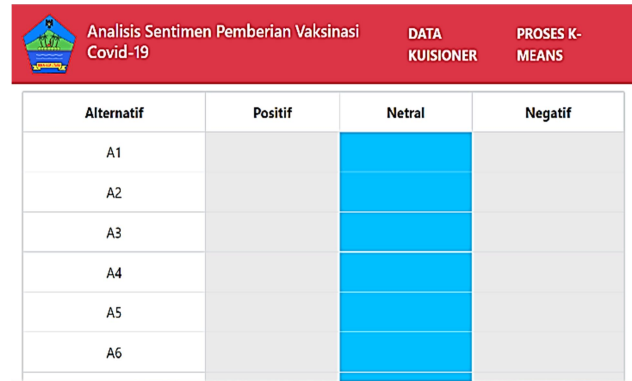


Fig. 9. Grouping Result

Fig. 6. Uploading CSV File

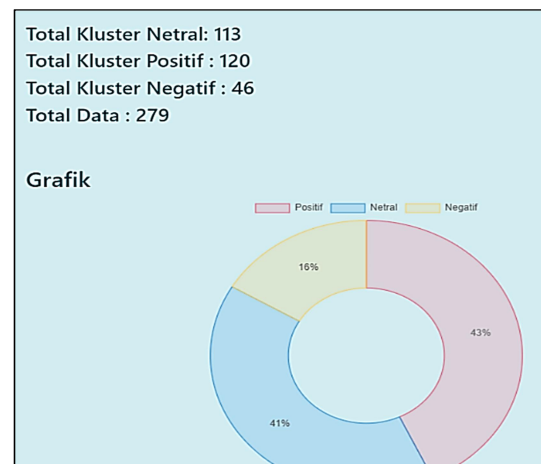


Fig. 10. Result in Chart

Fig. 7. The 1st Iteration

Fig. 8. The 2nd Iteration

D. Testing

The dataset was divided into 7 parts using 7-fold cross-validation (K=7). Each validation has a different initial centroid to get the desired grouping results. The results of this validation can be seen in Table XIII.

TABLE XIII
Test Result of 7-Fold Cross-Validation

Testing	Positive	Neutral	Negative
1	105	124	49
2	49	124	105
3	24	49	105
4	44	44	191
5	30	22	226
6	105	128	45
7	105	124	49

Based on Table XIII, the moderate neutral cluster has the highest amount of data in 7 tests. The following conclusions can be made.

1. All the features in the application, such as the CSV file upload feature, the random initial centroid, and the iteration results can run well.
2. The cluster with the least amount of data is the high addiction cluster where in experiments 1

and 3 and 2 and 4 have the same values.

3. The application can provide stable grouping results as well as describe the dominant conditions in the processed data.

IV. CONCLUSION

The K-means method is relatively simple to implement. The number of K clusters must be determined at the beginning of the algorithm. At the processing stage, 3 sentiment clusters were determined, namely positive (C1), neutral (C2), and negative (C3). It is seen that when the amount of data is not so much, the initial clustering will affect the cluster significantly. It will also hold back from predicting the value of K and knowing the actual cluster. Using the same data, if entered in a different order, may result in different clusters when the amount of data is small. In other words, the K-means method is sensitive to the initial conditions because the results will be greatly affected. K-means clustering often produces varying results in different algorithmic processes. Random selection of cluster patterns results in different clustering results which can lead to inconsistencies. This study used a random value for the initial centroid used for 7-fold cross-validation [26] and the results show consistency despite the initial centroid value.

Based on the tests carried out by 279 respondents, the familial responses toward the administration of COVID-19 vaccination to elementary school students consisted of 120 respondents who gave positive sentiments (43%), 113 respondents with neutral sentiments (41%), and 46 respondents had negative sentiments (16%). The majority of the families gave positive responses due to the high level of understanding and knowledge about COVID-19 vaccination for small children. Yet, there are a small number of families who do not fully agree with this policy due to the lack of knowledge and information regarding the administration procedure of COVID-19 vaccination for elementary school children among other reasons.

There are several recommendations made for further improvements related to this research. Firstly, it is highly suggested to differentiate the type of responses towards COVID-19 vaccination for children that are cognitive response

(knowledge), affective response (behavior), and conative response (action). Secondly, there is a need to use different clustering algorithms as a comparison to see which algorithm gives better results, and also the research topic can be made broader. Secondly, adding more sentiment polarities can increase the accuracy of identifying and classifying the familial responses pertinent to the COVID-19 vaccination in all the elementary schools in the Kalawat district. Thirdly, the classifier built into our application has the flexibility to accommodate new data from other sources to determine the sentiments of respective participants about the administration of COVID-19 vaccination. Lastly, increasing the training data can certainly improve the capabilities of the classifier to curtail Euclidean distance between data points.

V. ACKNOWLEDMENT

We want to express our gratitude to Universitas Katolik De La Salle Manado for their ongoing assistance and support, which made it possible for us to carry out this research.

VI. REFERENCES

- [1] Kementerian Kesehatan Republik Indonesia, "Keputusan Menteri Kesehatan Republik Indonesia Nomor HK 01.07/MENKES/6688/2021 Tentang Pelaksanaan Vaksinasi Corona Virus Disease 2019 (COVID-19) Bagi Anak Usia 6 Sampai Dengan 11 Tahun," Kementerian Kesehatan Republik Indonesia, Jakarta, 2021.
- [2] Kementerian Kesehatan Republik Indonesia, "Pelaksanaan Vaksinasi Corona Virus Disease 2019 (COVID-19) Bagi Anak Usia 6 (Enam) Sampai Dengan 11 (Sebelas) Tahun," Kementerian Kesehatan, Jakarta, 2021.
- [3] J. G. Shim, K. H. Ryu, S. H. Lee, E. A. Cho, Y. J. Lee and J. H. Ahn, "Text Mining Approaches to Analyze Public Sentiment Changes Regarding COVID-19 Vaccines on Social Media in Korea," *Int. J. Environ. Res. Public Health*, vol. 18, pp. 1-9, 2021.
- [4] F. F. Rachman and S. Pramana, "Analisis Sentimen Pro Dan Kontra Masyarakat Indonesia Tentang Vaksin COVID-19 Pada Media Sosial Twitter," *Indonesian of Health Info. Manage. J.*, vol. 8, no. 3, pp. 100-109, 2020.
- [5] Zharifah, "Respon Orang Tua Terhadap Siswa Yang Diwajibkan Vaksinasi Covid-19 Di SD IT Mutiara Hati Payakumbuh," *JOM FISIP*, vol. 10, no. 1, pp. 1-13, 2023.
- [6] N. Kurniatillah, F. Hayat and N. Nurjaman, "Pengaruh Peran Orang Tua Terhadap Pemberian Vaksinasi

- Covid 19 di MAN I Kota Serang," *Journal of Baja Health*, vol. 2, no. 1, p. 18–23, 2022.
- [7] F. F. Mailo and L. Lazuardi, "Analisis Sentimen Data Twitter Menggunakan Metode Text Mining Tentang Masalah Obesitas Di Indonesia," *J. of Inf. Syst. for Public Health*, vol. 4, no. 1, p. 2019, 28-36.
- [8] L. Kartikawati, K. Kusri and E. T. Luthfi, "Algoritma K-Means Pada Pengelompokan Pembelajaran Tatap Muka Terbatas Sesudah Vaksinasi COVID-19," *J. Eksplora Informatika*, vol. 11, no. 1, pp. 20-28, 2021.
- [9] M. Yilmaz and M. K. Sahin, "Parents' Willingness and Attitudes Concerning the COVID-19 Vaccine: A Cross-Sectional Study," *The Int. J. of Clinical Health*, vol. 75, no. 9, pp. 1-11, 2021.
- [10] K. P. Sinaga and M. S. Yang, "Unsupervised K-Means Clustering Algorithm," *IEEE Access*, vol. 8, pp. 80716-80727, 2020.
- [11] R. L. Batu, N. I. Suryani, N. Septia and P. F. Sekaryahya, "Pengaruh Harga Dan Inovasi Layanan Aplikasi Terhadap Keputusan Penggunaan Jasa Taksi Express: Survei Pada Pengunduh Aplikasi Express Taxis," *J. Nasional Manajemen Pemasaran & SDM*, vol. 1, no. 1, pp. 1-22, 2020.
- [12] M. Z. Zhang, X. M. Li, S. B. Yue and L. Q. Yang, "An Empirical Study of TextRank for Keyword Extraction," *IEEE Access*, vol. 8, pp. 178849-178858, 2020.
- [13] S. Sarah and M. Mustakim, "Analisis Penerimaan Vaksin COVID-19 Berbasis Fuzzy Clustering Machine Learning Di Provinsi Riau," *JURIKOM (J. Riset Komputer)*, vol. 8, no. 6, p. Mustakim, 2021.
- [14] I. Ha, B. H. Back and B. C. Ahn, "MapReduce Functions to Analyze Sentiment Information from Social Big Data," *Int. J. of Distributed Sensor Netw.*, vol. 11, no. 6, pp. 2-11, 2015.
- [15] M. Robani and A. Widodo, "Algoritma K-Means Clustering Untuk Pengelompokan Ayat Al Quran Pada Terjemahan Bahasa Indonesia," *JSINBIS (J. Sistem Informasi Bisnis)*, vol. 6, no. 2, pp. 164-176, 2016.
- [16] Anggraini, Novita, E. S. N. Harahap and T. B. Kurniawan, "Text Mining-Text Analysis Related to COVID-19 Vaccination Issues," *J. Ilmu Pengetahuan Dan Teknologi Komunikasi*, vol. 23, no. 2, pp. 141 - 153, 2021.
- [17] E. Farshi, B. Kasmapur and A. Arad, "Investigation of Immune Cells on Elimination of Pulmonary-Infected COVID-19 and Important Role of Innate Immunity, Phagocytes," *Rev Med Virol*, vol. 31, no. 2, pp. 1-6, 2021.
- [18] S. Brandstetter, M. M. Böhmer, M. Pawellek, B. Seelbach-Göbel, M. Melter, M. Kabesch, C. Apfelbacher and KUNO-Kids study group, "Parents' Intention to Get Vaccinated and to Have Their Child Vaccinated against COVID-19: Cross-Sectional Analyses Using Data from the KUNO-Kids Health Study," *Eur J Pediatr*, vol. 180, no. 11, pp. 3405-3410, 2021.
- [19] S. Lukas and T. Triyani, "Penyuluhan Kesehatan Tentang: Pentingnya Vaksinasi COVID-19 Pada Lansia Di RW.01 Kelurahan Batu Ampar," *Berdikari*, vol. 3, no. 2, pp. 1-14, 2020.
- [20] M. A. Baloch, T. L. Zheng, X. F. Pei and S. Baloch, "The Coronavirus Disease 2019 (COVID-19) Pandemic," *Tohoku J. of Exp. Medicine*, vol. 250, pp. 271-278, 2020.
- [21] Iskak, M. Z. Rusydi, R. Hotauruk, S. Chakim and W. R. Ahmad, "Meningkatkan Kesadaran Masyarakat Tentang Pentingnya Vaksinasi Di Masjid Al – Ikhlas, Jakarta Barat," *Pengabdian Dharma Masyarakat (PADMA)*, vol. 1, no. 3, pp. 222-226, 2021.
- [22] T. Surtimanah, H. Hanifah, D. Alfianita, N. Nataria, S. S. Audia, P. Mulyawan and I. N. Sjamsuddin, "Penyuluhan Pencegahan COVID-19 Melalui Video Bagi Masyarakat Perdesaan Dan Perkotaan," *AS-SYIFA: J. Pengabdian dan Pemberdayaan Kesehatan Masyarakat*, vol. 2, no. 1, pp. 43-54, 2021.
- [23] N. Laili and W. Tanoto, "Model Kepercayaan Kesehatan (Health Belief Model) Masyarakat Pada Pelaksanaan Vaksin COVID-19," *J. Ilmu. Kesehatan Keperawatan*, vol. 17, no. 3, pp. 198-207, 2021.
- [24] A. Solichin and K. Khairunnisa, "Klusterisasi Persebaran Virus Corona (COVID-19) Di DKI Jakarta Menggunakan Metode K-Means," *Fountain of Inform.*, vol. 5, no. 2, pp. 52-59, 2020.
- [25] L. Maulida, "Penerapan Datamining Dalam Mengelompokkan Kunjungan Wisatawan Ke Objek Wisata Unggulan Di Prov. Dki Jakarta Dengan K-Means," *JISKA (J. Informatika Sunan Kalijag)*, vol. 2, no. 3, p. 167–174, 2018.
- [26] J. Wainer and G. Cawley, "Nested Cross-Validation When Selecting Classifiers Is Overzealous for Most Practical Applications," *Expert Syst. with Appl.*, vol. 182, pp. 1-9, 2021.
- [27] Q. M. Liu and L. J. Wang, "T-Test and ANOVA for Data with Ceiling and/or Floor Effects," *Behavior Research Methods*, vol. 53, p. 264–277, 2021.
- [28] J. BPK, "PMK No 10 Tahun 2021 Tentang Pelaksanaan Vaksinasi Dalam Rangka Penanggulangan Pandemi Corona Virus Disease 2019 (COVID-19)," Peraturan Menteri Kesehatan Nomor 10 Tahun 2021, Februari 2021. [Online]. Available: <https://peraturan.bpk.go.id/Details/169665/permenkes-no-10-tahun-2021>. [Accessed 10 Januari 2023].
- [29] G. Tabacchi, G. Battaglia, G. Messina, A. Paoli, A. Palma and M. Bellafiore, "Validity and Internal Consistency of the Preschool-Flat, a New Tool for the Assessment of Food Literacy in Young Children from the Training-to-Health Project," *Int J Environ Res Public Health*, vol. 17, no. 8, pp. 1-15, 2020.
- [30] B. D. Lund and J. X. Ma, "A Review of Cluster Analysis Techniques and Their Uses in Library and Information Science Research: K-Means and k-Medoids Clustering," *Performance Measurement and Metrics*, vol. 22, no. 3, pp. 161-173, 2021.