Bulletin of Informatics and Data Science
Vol. 2 No. 2, November 2023
ISSN 2580-8389 (Media Online)
Page 86−95
https://ejurnal.pdsi.or.id/index.php/bids/index
DOI 10.61944/bids.v2i2.78

The Process of Grouping Elementary School Students Receiving PIP Assistance uses the K-Means Algorithm

Jen-Peng Huang[1], Pai-Chou Wang[1], Ridha Maya Faza Lubis[2,*]

[1] Department of Information Management, Southern Taiwan University of Science and Technology, Taiwan
[2] Department of Business and Management, Southern Taiwan University of Science and Technology, Taiwan
Email: [1]jehuang@stust.edu.tw, [2]pwang@stust.edu.tw, [3,*]db01g208@stust.edu.tw,
Coressponding Author: db01g208@stust.edu.tw

Abstract

As part of receiving support from the Smart Indonesia Program (PIP), this study intends to analyze and apply the K-Means algorithm in the process of grouping elementary school students. PIP is a government initiative that attempts to give money to elementary school pupils from disadvantaged or weaker homes. The effective and fair distribution of aid monies depends on the proper grouping of the students. The K-Means approach was selected because it can cluster data, allowing the grouping of pupils based on pertinent traits. Numerous characteristics that can affect kids' financial needs are included in the data utilized in this study, including family income, parental education level, proximity to the school, and other social and economic issues. This study makes use of empirical data from a PIP-affiliated elementary school in an urban setting. The data includes a large number of pertinent features and thousands of pupils. Based on how similar their characteristics are, pupils are divided into numerous clusters using the K-Means technique. The findings of this study will help us better identify the traits of students who are eligible for PIP support. By doing this, the government can allocate funds more wisely and guarantee that aid is given where it is most needed. The PIP program can benefit children in need more by streamlining the process of grouping the students. In addition, this research has broader implications for social aid and education policy. To guarantee effectiveness and equity in resource allocation, the K-Means algorithm can be used in a variety of additional aid initiatives. Data mining-based strategies, like those employed in this study, are becoming more crucial to boost the effectiveness of aid programs like PIP. The findings of this study can help the government and educational institutions improve the efficacy of aid initiatives designed to boost Indonesian children's education.

Keywords: Data Mining; K-Means Algorithm; Smart Indonesia Program; Receiving PIP

1. INTRODUCTION

One of the key foundational elements in a nation's development is education. The focus of the government's efforts in Indonesia to ensure that all children have equitable access to education is on the provision of basic education, particularly at the elementary school level. The Smart Indonesia Initiative Assistance (PIP) initiative is one way the government promotes access to education. PIP is a program that intends to give money to elementary school pupils from disadvantaged or weaker families. This initiative serves as a tool to eliminate social and economic barriers to educational opportunity [1]–[3].

The PIP initiative, however, faces some difficulties throughout its implementation. How to more effectively identify kids who qualify for this help is one of the primary challenges. This identification procedure takes into account several variables, including family income, parental education, proximity to the school, and other social and economic considerations. In this situation, technology and data analysis can significantly improve the speed and precision of the student eligibility identification process.

The clustering technique K-Means is employed in data analysis and data grouping. The objective is to organize the data into clusters (groups) that share common traits. By locating the cluster center (centroid) and assigning each piece of data to the cluster with the closest center, this technique organizes the data. The PIP tool can more precisely identify which pupils need help by categorizing them based on their traits. Additionally, the K-Means algorithm can be used in the identification process automatically to boost impartiality and eliminate bias thanks to information technology and computer systems. Decision-making in the PIP program can become more informed and evidence-based by utilizing the data that is currently accessible, such as family economic data and student education data [4]–[6].

The author used several earlier studies as references to complete the research while doing this study. "Fammaldo et al. conducted this study in 2019. The K-Means Clustering algorithm's use in the Smart Indonesia Program to categorize family welfare levels is discussed in this study. According to the study's findings, the K-Means approach can identify families that are classified as being poor, middle-class, or rich and in need of financial aid with an accuracy rate of 69%. Due to the volume and diversity of asset data, this decreased level of accuracy results in some families not meeting the preset group requirements [7].

Furthermore, previous research conducted by Usanto S in 2023 discussed the implementation of the K-Means algorithm in grouping prospective KIP recipient students. The study found that possible KIP scholarship winners might be suggested by applying clustering strategies based on the K-Means method. A total of 100 student data were examined for this study. From the 100 data, 52 students succeeded in meeting the requirements to receive a scholarship, 32 students were identified as potential candidates for receiving a scholarship at the next stage, and 16 students did not meet the requirements for receiving a scholarship [8].

Furthermore, research conducted by Derman Janner Lubis in 2022 discusses the application of the K-Means algorithm in grouping student scholarships at the Miftahul Huda Islamic Boarding School, Bogor. The research concluded that the results were tested by cluster evaluation utilizing the Silhouette Index method via the MATLAB program after the K-Means Clustering Algorithm method was applied to the student grouping system to get scholarships. This evaluation's output produced a rating of 0.7030, indicating a good structural basis for the cluster's grouping. An average value of 1 represents the degree of clustering within these clusters that is tight [9].

Additionally, Agustin Ely Rahayu et al. conducted research in 2019. Their study covered the use of K-Means to identify student Bidikmisi scholarship clustering. According to Peacock's research, the four clusters of potential scholarship recipients—very deserving, less worthy, considered, and not worthy of getting a Bidikmisi award—can be divided using the K-Means method. When it comes to accepting Bidikmisi scholarships for students, the clustering results should be viewed as a suggestion that aids in the decision-making process but is not the ultimate say. Before the selection of potential scholarship recipients, a more thorough review and verification process must still be completed. The K-Means method is only one tool used to offer a preliminary understanding of the procedure [10].

Additionally, a study done in 2020 by Nur Afriani Manihuruk et al. covered how to sort potential scholarship applicants using the K-Means method. According to the study's findings, the data were analyzed to identify potential scholarship applicants at Muhammadiyah 54 Kerasaan Private Middle School and classify them according to preset criteria. To determine the centroid values of the three clusters that had formed—high-level clusters, medium-level clusters, and low-level clusters—data processing was done using Microsoft Excel tools. According to data processing results, there are 73 students in the high-level cluster, 30 students in the medium-level cluster, and 25 students in the low-level cluster [11].

The purpose of this study is to investigate and apply the K-Means method in the process of classifying elementary school pupils in the context of getting PIP support, based on the justification provided above. With this strategy, it is believed that the PIP program would become more effective at selecting aid recipients, guarantee that aid is given to students who require it, and eventually help Indonesian children realize their right to an education.

2. RESEARCH METHODOLOGY

2.1 Research Framework.

Various steps need to be completed to prepare this research, including:

a. Problem-Solving
The purpose of problem analysis is to thoroughly describe the issue, pinpoint its underlying causes, and establish a precise research question or objective.

b. Gathering Data
During this phase, pertinent data and information must be gathered for the previously identified research issues or goals.

c. Literature Analysis
Reviewing pertinent material or references that have been published in the past about the research issue or topic is known as a literature study.

d. Implementation of the algorithm.
This phase entails the creation and application of algorithms or techniques used to address issues or accomplish research goals.

e. Conclusion
At this point, judgments are being drawn in light of the outcomes of the data analysis, literature review, and algorithm implementation.

The stages above can be seen in Figure 1 below:

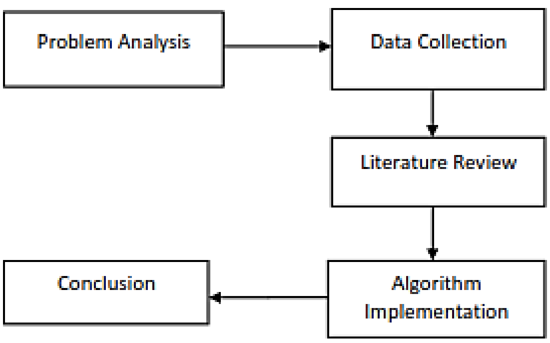

Figure 1. Research Framework

2.2 Data Mining

With the use of sophisticated computational and statistical analytic tools, valuable and meaningful knowledge may be extracted from several datasets or big data sets through the process of data mining [12]. Finding patterns, relationships, trends, or hidden information in complex and varied data is the core objective of data mining. Data processing, statistical

Bulletin of Informatics and Data Science
Vol. 2 No. 2, November 2023
ISSN 2580-8389 (Media Online)
Page 86−95
https://ejurnal.pdsi.or.id/index.php/bids/index
DOI 10.61944/bids.v2i2.78

modeling, machine learning, predictive analysis, and association rule mining are a few of the processes and methods used in data mining [2], [13]. In the process, data mining can disclose insights that are not readily apparent manually and assist users in receiving more detailed information and better decisions. Data mining has a wide range of applications that can be employed in a variety of industries, including business, science, finance, health, marketing, and many more. Customer behavior analysis, fraud detection, market trend prediction, data clustering, and many other applications that extract additional value from existing data are examples of applications where data mining is used. Data mining is developing and becoming an important tool to support decision-making and research across a variety of industries because of technological advancements and improvements in processing power [14]–[16].

2.3 K-Means Clustering Algorithm

The K-Means algorithm is a data clustering technique that divides data into several groups or clusters based on certain shared criteria. This algorithm divides the data into several clusters, each of which has a centroid as its center [17]. One of the most popular and straightforward clustering methods is K-Means, but it has significant drawbacks, including sensitivity to initial initialization and the requirement to know the number of clusters in advance. Following are the steps for completing the K-Means Algorithm calculation:

a. When calculating the K value, one can predict how many clusters will develop.
b. Select some random points based on an established K value.
c. Determine the centroid that is closest to each input data by calculating the distance between each centroid and each input data using the Euclidean.
   Distance formula:

$$d(xi, \mu j)\sqrt{\sum (xi - \mu j)^2} \quad (1)$$

   Evidence :
   xi : $criteria\ data$
   $\mu j = the\ jth\ cluster's\ centroid$
d. using the centroid that is closest to each group of data and has the least distance.
e. Refresh the centroid value. The formula below is used to calculate the new centroid value from the average of the cluster in question:

$$\mu j(t+1) = \frac{1}{Nsj} \sum_{j \in Sj} xj \quad (2)$$

   Evidence :
   $\mu j(t+1)$: $the\ th\ iteration's\ new\ centroid$ $(t+1)$
   N_{sj} : many details about the cluster S_J
f. Till there are no more changes in the cluster members, which denotes that the grouping has converged and the cluster has not changed any further, repeatedly repeat steps 2 through 5 in that order.

2.4 Smart Indonesia Program (PIP)

The Indonesian government runs a financial aid program called Smart Indonesia Program aid (PIP), which is frequently abbreviated to PIP [18]. This program's primary goal is to give students at basic education levels (such elementary school or elementary school) who originate from underprivileged or economically disadvantaged families financial aid. PIP aid is intended to assist low-income families with the expenditures of their children's elementary education [19], [20].

3. RESULTS AND DISCUSSION

A review of the aid data that has been gathered is done at this point. The input will be processed by assigning weight values to each pertinent criterion as part of the calculation procedure. It had previously been decided to divide the data into two groups, the high-level group and the low-level group. The outcomes of the grouping will be used for additional analysis in this stage. Data about potential scholarship recipients will be processed first in the grouping process.

Table 1. Student Information

Number.	Student’s Name	Assessment Value	Parent’s Job	Number of Parents' Dependents
1	Adrian Putra	90	Farmer	3
2	Salim	90	Trader	3
3	Ari Juliano	85	Private sector employee	2
4	Alviansyah Santana	85	Farmer	2
5	Baim	80	Fisherman	2
6	Khairunnisa	85	Fisherman	2
7	May Randa	85	Farmer	1

Bulletin of Informatics and Data Science
Vol. 2 No. 2, November 2023
ISSN 2580-8389 (Media Online)
Page 86−95
https://ejurnal.pdsi.or.id/index.php/bids/index
DOI 10.61944/bids.v2i2.78

Number.	Student’s Name	Assessment Value	Parent’s Job	Number of Parents' Dependents
8	Julianus Fedinan Tarigan	80	Private sector employee	1
9	Muhammad Edy	80	Farmer	1
10	Arofi Arrahman	90	Farmer	4
11	Krisman Syah	90	Fisherman	4
12	Yuswardiman hiwa	85	Fisherman	3
13	Deri Putra Perdana	85	Trader	2
14	Debby Ayu	85	Trader	3
15	Nurhidayah Amel	85	Trader	3
16	Frans Farel	80	Private sector employee	3
17	Novelia Azzahra	80	Private sector employee	4
18	M. Dicky Putro	90	Farmer	4
19	Ramadani Purba	90	Farmer	5
20	Khairunnisa arrahmah	90	Private sector employee	4
21	May	85	Private sector employee	3
22	Julianus	85	Fisherman	3
23	Muhammad fahrozi	80	Fisherman	3
24	Arofi	80	Farmer	5
25	Krisman Syah Putra Duha	85	Farmer	5
26	Ridho Auliansyah Pohan	90	Farmer	5
27	Abdul Azizi	90	Fisherman	5
28	Dedi Hardian	85	Fisherman	4
29	Ade Agung Ichwansyah	85	Fisherman	3
30	Ripal Anwar Pohan	85	Fisherman	3
31	Desi Simanjuntak	85	Private sector employee	2
32	Muhammad Rizal	80	Private sector employee	3
33	Marihot Putra	85	Farmer	2
34	Siska Kristiana Enjel	85	Farmer	3
35	Shakya Suci Ramadani	85	Fisherman	4
36	Tarmizi	80	Trader	1
37	Juliana	80	Trader	1
38	Rayni Ayumi	85	Trader	2
39	Santi Susanti	85	Private sector employee	3
40	Reka Wulandari	85	Fisherman	4
41	Tohusokhi Giawa	80	Private sector employee	4
42	Teuku Ferdiansyah Putra	85	Private sector employee	5
43	Marta Nofita Sihotang	85	Fisherman	4
44	Efrin Edoardo Simatupang	85	Fisherman	4
45	Riska Rahayu Pasaribu	90	Trader	3
46	Sahrul Fahmi	90	Trader	4
47	Regina Oktaviana Br Purba	85	Fisherman	4
48	Rakhel Sitanggang	85	Fisherman	4
49	Dony Dharmawan	80	Trader	4
50	Ridho Auliansyah Pohan	80	Trader	3
51	Abdul Aziz	80	Trader	3
52	Dedi Hardiansyah	80	Fisherman	2
53	Ade Agung Ichwansyah Nst	85	Honorary Teacher	1
54	Ripal Anwar Pohan	85	Honorary Teacher	1
55	Desi Simanjuntak	85	Trader	1
56	Muhammad Rizal Harahap	80	Fisherman	4
57	Marihot	80	Honorary Teacher	3
58	Siska Kristiana Simanullang	85	Honorary Teacher	3
59	Shakya Suci Kamara	85	Trader	3
60	Tarmizi	85	Trader	3
61	Juliana Tamba	85	Fisherman	2
62	Syaaf Faathir	85	Fisherman	4
63	Santi Manullang	90	Trader	3
64	Reka Wulandari	90	Trader	2
65	Tohusokhi Giawa	85	Private sector employee	4
66	Teuku Ferdiansyah Putra	80	Private sector employee	5

Bulletin of Informatics and Data Science
Vol. 2 No. 2, November 2023
ISSN 2580-8389 (Media Online)
Page 86−95
https://ejurnal.pdsi.or.id/index.php/bids/index
DOI 10.61944/bids.v2i2.78

Number.	Student's Name	Assessment Value	Parent's Job	Number of Parents' Dependents
67	Marta Nofita Sihotang	85	Farmer	5
68	Efrin Edoardo Simatupang	85	Farmer	4
69	Riska Rahayu Pasaribu	85	Fisherman	3
70	Sahrul Fahmi	80	Farmer	6
71	Widya Yanti Tumanggor	80	Farmer	5
72	Siti Anzani	85	Trader	5
73	M. Iqbal Marpaung	85	Trader	6
74	Silvia Arditha	85	Trader	6
75	Elisa Hasibuan	90	Fisherman	5
76	Muhammad Firhan Fahmi	90	Fisherman	5
77	Hansen Gabriel Siregar	90	Private sector employee	4
78	Yantonius Gulo	85	Private sector employee	3
79	Taufik Ramadoni Sirait	85	Trader	2
80	Widya Yanti Tumanggor	85	Fisherman	2

The school's final grades are then evaluated using the preliminary data, as indicated in Table 2 below.

Table 2. Data on Parental Employment

Mark	Information
Private sector employee	5
Fisherman	4
Farmer	3
Honorary Teacher	2
Trader	1

Table 3 below shows the normalized registrant data table based on predefined values:

Table 3. Information Normalization

Number.	Student's Name	Assessment Value	Parent's Job	Number of Parents' Dependents
1	Adrian Putra	90	3	3
2	Salim	90	1	3
3	Ari Juliono	85	5	2
4	Alviansyah Santana	85	3	2
5	Baim	80	4	2
6	Khairunnisa	85	4	2
7	May Randa	85	3	1
8	Julianus Fedinan Tarigan	80	5	1
9	Muhammad Edy	80	3	1
10	Arofi Arrahman	90	3	4
11	Krisman Syah	90	4	4
12	Yuswardiman hiwa	85	4	3
13	Deri Putra Perdana	85	1	2
14	Debby Ayu	85	1	3
15	Nurhidayah Amel	85	1	3
16	Frans Farel	80	5	3
17	Novelia Azzahra	80	5	4
18	M. Dicky Putro	90	3	4
19	Ramadani Purba	90	3	5
20	Khairunnisa arrahmah	90	5	4
21	May	85	5	3
22	Julianus	85	4	3
23	Muhammad fahrozi	80	4	3
24	Arofi	80	3	5
25	Krisman Syah Putra Duha	85	3	5
26	Ridho Auliansyah Pohan	90	3	5
27	Abdul Azizi	90	4	5
28	Dedi Hardian	85	4	4
29	Ade Agung Ichwansyah	85	4	3
30	Ripal Anwar Pohan	85	4	3

Bulletin of Informatics and Data Science
Vol. 2 No. 2, November 2023
ISSN 2580-8389 (Media Online)
Page 86−95
https://ejurnal.pdsi.or.id/index.php/bids/index
DOI 10.61944/bids.v2i2.78

Number.	Student's Name	Assessment Value	Parent's Job	Number of Parents' Dependents
...	....	....	....	....
80	Widya Yanti Tumanggor	85	4	2

3.1 Implementation of K-Means Algorithm

Follow these steps to finish the K-Means algorithm:
first revision

a. Counting the cluster values in a cluster (K)
b. As many times as the K value, choose the starting center point.
To group or categorize each piece of data in this study, the centroid center's first determination is made. Since the initial centroid is generated at random, it is unrelated to any particular data. Table 4 below provides more details regarding the initial centroids:

Table 4. First Centroid Information

Centroid	Data Number	Assessment Value	Parent's Job	Number of Parents' Dependents
Centroid 1	2	90	1	3
Centroid 2	4	85	3	2

c. the shortest route to the centroid is determined. (In regard to clusters 1 and 2). Cluster 1 values are calculated using the first centroid data 1, and cluster 2 values are calculated using the second initial centroid.
1. The separation between data point 1 and the centroid is 1.

$$= \sqrt{(90-90)^2 + (3-1)^2 + (3-3)^2} = 2$$

2. Distance between centroid point 2 and data point 1

$$= \sqrt{(90-90)^2 + (1-1)^2 + (3-3)^2} = 0$$

3. the separation between data point 1 and centroid 3

$$= \sqrt{(85-90)^2 + (5-1)^2 + (2-2)^2} = 7.4031242$$

Then, you can look at Table 5 below to see the outcomes of computing the minimum distance from data number 4 to data number 80. Utilize the same calculations as those detailed for Data Point 1.

Table 5. Results of the first iteration's minimal distance calculation

Number.	Student's Name	Cluster 1	Cluster 2	Maximum Distance	Group Data
1	Adrian Putra	2	6	2	Cluster1
2	Salim	0	6,385165	0	Cluster1
3	Ari Juliono	7,4031242	2	2	Cluster2
4	Alviansyah Santana	6,3851648	0	0	Cluster2
5	Baim	11,440307	5,09902	5,099019514	Cluster2
6	Khairunnisa	6,8309519	1	1	Cluster2
7	May Randa	9,3851648	1	1	Cluster2
8	Julianus Fedinan Tarigan	14,77033	6,385165	6,385164807	Cluster2
9	Muhammad Edy	14,198039	6	6	Cluster2
10	Arofi Arrahman	3	9	3	Cluster1
11	Krisman Syah	4	9,09902	4	Cluster1
12	Yuswardiman hiwa	5,8309519	2	2	Cluster2
13	Deri Putra Perdana	6	2	2	Cluster2
14	Debby Ayu	5	3	3	Cluster2
15	Nurhidayah Amel	5	3	3	Cluster2
16	Frans Farel	10,77033	6,385165	6,385164807	Cluster2
17	Novelia Azzahra	11,77033	9,385165	9,385164807	Cluster2
18	M. Dicky Putro	3	9	3	Cluster1
19	Ramadani Purba	6	14	6	Cluster1
20	Khairunnisa arrahmah	5	9,385165	5	Cluster1
21	May	6,4031242	3	3	Cluster2
22	Julianus	5,8309519	2	2	Cluster2
23	Muhammad fahrozi	10,440307	6,09902	6,099019514	Cluster2
24	Arofi	14,198039	14	14	Cluster2
25	Krisman Syah Putra Duha	9,3851648	9	9	Cluster2
26	Ridho Auliansyah Pohan	6	14	6	Cluster1
27	Abdul Azizi	7	14,09902	7	Cluster1

Bulletin of Informatics and Data Science
Vol. 2 No. 2, November 2023
ISSN 2580-8389 (Media Online)
Page 86−95
https://ejurnal.pdsi.or.id/index.php/bids/index
DOI 10.61944/bids.v2i2.78

Number.	Student's Name	Cluster 1	Cluster 2	Maximum Distance	Group Data
28	Dedi Hardian	6,8309519	5	5	Cluster2
29	Ade Agung Ichwansyah	5,8309519	2	2	Cluster2
30	Ripal Anwar Pohan	5,8309519	2	2	Cluster2
....	....	....	....	....	....
80	Widya Yanti Tumanggor	6,8309519	1	1	Cluster2

Cluster 1 had 21 data while cluster 2 had 59 data in the first iteration's results.

d. The data are then sorted according to how close they are to the centroid with the smallest distance. Based on determining the minimal distance from each cluster, the data in Table 5 above has been divided into two clusters.
e. The centroid value is then updated by determining the average of the data in each relevant cluster. Iterations utilizing the revised centroid value are then performed after the cluster grouping results based on minimal distance are obtained. Table 6 results for the revised centroid values are displayed as follows:

Table 6. New Cluster Center/Centroid Iteration 2

Centroid 1	88,571	2,714	4,333
Centroid 2	84,915	3,288	2,983

Once the new cluster center has been established, use the new cluster center value to perform calculations, such as those used to determine the closest distance to the centroid in Table 5.

f. Then, repeat step 3 if the cluster center/centroid grouping findings change. Additionally, if the aggregated data stays put. The iterative computation procedure is then finished.

The iteration procedure ended after the third iteration as a result of the calculating process. where the cluster center/centroid results from the second and third iterations are identical (i.e., the centroid location has not moved). Tables 7 and 8 show the outcomes of distance computations for the second and third iterations.

Table 7. Iteration 2 of the minimum distance

Number.	Student's Name	Cluster 1	Cluster 2	Maximum Distance	Group Data
1	Adrian Putra	1,975	5,093	1,975	Cluster1
2	Salim	2,599	5,576	2,599	Cluster1
3	Ari Juliono	4,840	1,976	1,976	Cluster2
4	Alviansyah Santana	4,276	1,028	1,028	Cluster2
5	Baim	8,976	5,063	5,063	Cluster2
6	Khairunnisa	4,456	1,217	1,217	Cluster2
7	May Randa	4,894	2,006	2,006	Cluster2
8	Julianus Fedinan Tarigan	9,477	5,570	5,570	Cluster2
9	Muhammad Edy	9,201	5,308	5,308	Cluster2
10	Arofi Arrahman	1,495	5,193	1,495	Cluster1
11	Krisman Syah	1,951	5,234	1,951	Cluster1
12	Yuswardiman Hiwa	4,023	0,717	0,717	Cluster2
13	Deri Putra Perdana	4,598	2,492	2,492	Cluster2
14	Debby Ayu	4,180	2,290	2,290	Cluster2
15	Nurhidayah Amel	4,180	2,290	2,290	Cluster2
16	Frans Farel	8,971	5,205	5,205	Cluster2
17	Novelia Azzahra	8,877	5,303	5,303	Cluster2
18	M. Dicky Putro	1,495	5,193	1,495	Cluster1
19	Ramadani Purba	1,602	5,478	1,602	Cluster1
20	Khairunnisa Arrahmah	2,716	5,461	2,716	Cluster1
21	May	4,445	1,714	1,714	Cluster2
22	Julianus	4,023	0,717	0,717	Cluster2
23	Muhammad Fahrozi	8,769	4,967	4,967	Cluster2
24	Arofi	8,602	5,321	5,321	Cluster2
25	Krisman Syah Putra Duha	3,644	2,039	2,039	Cluster2
26	Ridho Auliansyah Pohan	1,602	5,478	1,602	Cluster1
27	Abdul Azizi	2,034	5,516	2,034	Cluster1
28	Dedi Hardian	3,810	1,244	1,244	Cluster2
29	Ade Agung Ichwansyah	4,023	0,717	0,717	Cluster2
30	Ripal Anwar Pohan	4,023	0,717	0,717	Cluster2
...	....	.....	....	....	......
80	Widya Yanti Tumanggor	4,456	1,217	1,217	Cluster2

Bulletin of Informatics and Data Science
Vol. 2 No. 2, November 2023
ISSN 2580-8389 (Media Online)
Page 86−95
https://ejurnal.pdsi.or.id/index.php/bids/index
DOI 10.61944/bids.v2i2.78

The results obtained for clusters 1 and 2 are shown in Table 7 above as 16 and 64 data, respectively. As can be seen from tables 6 and 7, which display the outcomes of various cluster data, the third iteration calculation was then repeated.

Table 8. Iteration 3 of the New Cluster Center/Centroid

Cluster 1	90,000	2,875	3,938
Cluster2	83,438	3,156	3,141

Table 9. Three iterations minimum

Number.	Student's Name	Cluster 1	Cluster 2	Maximum Distance	Group Data
1	Adrian Putra	1,975	5,093	1,975	Cluster1
2	Salim	2,599	5,576	2,599	Cluster1
3	Ari Juliono	4,840	1,976	1,976	Cluster2
4	Alviansyah Santana	4,276	1,028	1,028	Cluster2
5	Baim	8,976	5,063	5,063	Cluster2
6	Khairunnisa	4,456	1,217	1,217	Cluster2
7	May Randa	4,894	2,006	2,006	Cluster2
8	Julianus Fedinan Tarigan	9,477	5,570	5,570	Cluster2
9	Muhammad Edy	9,201	5,308	5,308	Cluster2
10	Arofi Arrahman	1,495	5,193	1,495	Cluster1
11	Krisman Syah	1,951	5,234	1,951	Cluster1
12	Yuswardiman hiwa	4,023	0,717	0,717	Cluster2
13	Deri Putra Perdana	4,598	2,492	2,492	Cluster2
14	Debby Ayu	4,180	2,290	2,290	Cluster2
15	Nurhidayah Amel	4,180	2,290	2,290	Cluster2
16	Frans Farel	8,971	5,205	5,205	Cluster2
17	Novelia Azzahra	8,877	5,303	5,303	Cluster2
18	M. Dicky Putro	1,495	5,193	1,495	Cluster1
19	Ramadani Purba	1,602	5,478	1,602	Cluster1
20	Khairunnisa arrahmah	2,716	5,461	2,716	Cluster1
21	May	4,445	1,714	1,714	Cluster2
22	Julianus	4,023	0,717	0,717	Cluster2
23	Muhammad fahrozi	8,769	4,967	4,967	Cluster2
24	Arofi	8,602	5,321	5,321	Cluster2
25	Krisman Syah Putra Duha	3,644	2,039	2,039	Cluster2
26	Ridho Auliansyah Pohan	1,602	5,478	1,602	Cluster1
27	Abdul Azizi	2,034	5,516	2,034	Cluster1
28	Dedi Hardian	3,810	1,244	1,244	Cluster2
29	Ade Agung Ichwansyah	4,023	0,717	0,717	Cluster2
30	Ripal Anwar Pohan	4,023	0,717	0,717	Cluster2
...	....	....	....	....	....
80	Widya Yanti Tumanggor	4,456	1,217	1,217	Cluster2

Table 9 above shows that iteration 3 results in cluster 1 with 16 data. likewise, cluster 2 generates 64 data. Clusters 2 and 3 both have the same findings. The following iteration phase is then deemed to have ended.

Table 10. Data from Cluster 1 (C1) Students

Number.	Student's Name
1	Adrian Putra
2	Salim
10	Arofi Arrahman
11	Krisman Syah
18	M. Dicky Putro
19	Ramadani Purba
20	Khairunnisa arrahmah
26	Ridho Auliansyah Pohan
27	Abdul Azizi
45	Riska Rahayu Pasaribu
46	Sahrul Fahmi
63	Santi Manullang
64	Reka Wulandari
75	Elisa Hasibuan

Bulletin of Informatics and Data Science
Vol. 2 No. 2, November 2023
ISSN 2580-8389 (Media Online)
Page 86−95
https://ejurnal.pdsi.or.id/index.php/bids/index
DOI 10.61944/bids.v2i2.78

Number.	Student's Name
76	Muhammad Firhan Fahmi
77	Hansen Gabriel Siregar

Additionally, information from Cluster 2 pupils is included in Table 11

Table 11. Student Cluster 2 Data

Number	Student's Name	Mark
3	Ari Juliano	85
4	Alviansyah Santana	85
5	Baim	80
6	Khairunnisa	85
7	May Randa	85
8	Julianus Fedinan Tarigan	80
9	Muhammad Edy	80
12	Yuswardiman hiwa	85
13	Deri Putra Perdana	85
14	Debby Ayu	85
15	Nurhidayah Amel	85
16	Frans Farel	80
17	Novelia Azzahra	80
21	May	85
22	Julianus	85
23	Muhammad fahrozi	80
24	Arofi	80
25	Krisman Syah Putra Duha	85
28	Dedi Hardian	85
29	Ade Agung Ichwansyah	85
30	Ripal Anwar Pohan	85
31	Desi Simanjuntak	85
32	Muhammad Rizal	80
33	Marihot Putra	85
34	Siska Kristiana Enjel	85
35	Shakya Suci Ramadani	85
36	Tarmizi	80
37	Juliana	80
38	Rayni Ayumi	85
39	Santi Susanti	85
40	Reka Wulandari	85
41	Tohusokhi Giawa	80
42	Teuku Ferdiansyah Putra	85
43	Marta Nofita Sihotang	85
44	Efrin Edoardo Simatupang	85
47	Regina Oktaviana Br Purba	85
48	Rakhel Sitanggang	85
49	Dony Dharmawan	80
50	Ridho Auliansyah Pohan	80
51	Abdul Aziz	80
52	Dedi Hardiansyah	80
53	Ade Agung Ichwansyah Nst	85
54	Ripal Anwar Pohan	85
55	Desi Simanjuntak	85
56	Muhammad Rizal Harahap	80
57	Marihot	80
58	Siska Kristiana Simanullang	85
59	Shakya Suci Kamara	85
60	Tarmizi	85
61	Juliana Tamba	85
62	Syaaf Faathir	85
65	Tohusokhi Giawa	85
66	Teuku Ferdiansyah Putra	80
67	Marta Nofita Sihotang	85

Bulletin of Informatics and Data Science
Vol. 2 No. 2, November 2023
ISSN 2580-8389 (Media Online)
Page 86−95
https://ejurnal.pdsi.or.id/index.php/bids/index
DOI 10.61944/bids.v2i2.78

Number	Student's Name	Mark
68	Efrin Edoardo Simatupang	85
69	Riska Rahayu Pasaribu	85
70	Sahrul Fahmi	80
71	Widya Yanti Tumanggor	80
72	Siti Anzani	85
73	M. Iqbal Marpaung	85
74	Silvia Arditha	85
78	Yantonius Gulo	85
79	Taufik Ramadoni Sirait	85
80	Widya Yanti Tumanggor	85

4. CONCLUSION

The K-Means algorithm's results have made student grouping easier, ensured that PIP support is provided to those who need it most effectively, and optimized the distribution of educational resources. Thus, this article serves as an example of how technology and data analysis techniques like K-Means may help social initiatives like PIP to maximize societal advantages. K-means algorithm computations result in 16 student data from cluster 1 of the output. Cluster 2 also generates 64 data.

REFERENCES

[1] S. A. Rahmah and J. Antares, "Klasterisasi Seleksi Mahasiswa Calon Penerima Beasiswa Yayasan Menggunakan K-Means Clustering," *INFORMATIKA*, vol. 13, no. 2, pp. 25–30, 2022.
[2] Baskoro, Sriyanto, and L. Setya Rini, "Prediksi Penerima Beasiswa dengan Menggunakan Teknik Data Mining di Universitas Muhammadiyah Pringsewu," *Semin. Nas. Has. Penelit. dan Pengabdi. Masy. Inst. Inform. dan Bisnis Darmajaya*, pp. 87–94, 2021.
[3] A. Salam, D. Adiatma, and J. Zeniarja, "Implementasi Algoritma K-Means Dalam Pengklasteran untuk Rekomendasi Penerima Beasiswa PPA di UDINUS," *JOINS (Journal Inf. Syst.*, vol. 5, no. 1, pp. 62–68, 2020.
[4] B. G. Sudarsono and S. P. Lestari, "Clustering Penerima Beasiswa Yayasan Untuk Mahasiswa Menggunakan Metode K-Means," *J. Media Inform. Budidarma*, vol. 5, no. 1, pp. 258–263, 2021.
[5] M. Z. Hossain, M. N. Akhtar, R. B. Ahmad, and M. Rahman, "A dynamic K-means clustering for data mining," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 13, no. 2, pp. 521–526, 2019.
[6] D. Aulia, M. Safii, and D. Suhendro, "Penerapan Algoritma K-Means dalam Proses Clustering Penilaian Kinerja Aparatur Sipil Negera di Sekretariat DPRD Pematangsiantar," *Jurasik (Jurnal Ris. Sist. Inf. dan Tek. Inform.*, vol. 6, no. 1, p. 47, 2021.
[7] E. Fammaldo and L. Hakim, "Penerapan Algoritma K-Means Clustering Untuk Pengelompokan Tingkat Kesejahteraan Keluarga Untuk Program Kartu Indonesia Pintar," *J. Ilm. Teknol. Infomasi Terap.*, vol. 5, no. 1, pp. 23–31, 2019.
[8] N. I. Danu, "Identifikasi Berita Hoax Menggunakan Kombinasi Metode K-Nearest Neigbor (KNN) dan TF-IDF Berbasis Web Dengan Menggunakan Framework Codeigniter," 2021.
[9] D. J. Lubis and M. B. Tamam, "Penerapan K-Means Untuk Pengelompokkan Beasiswa Santri di Pondok Pesantren Miftahul Huda Bogor," *Teknois J. Ilm. Teknol. Inf. dan Sains*, vol. 12, no. 1, pp. 7–20, 2022.
[10] A. E. Rahayu, K. Hikmah, N. Yustia, and A. C. Fauzan, "Penerapan K-Means Clustering Untuk Penentuan Klasterisasi Beasiswa Bidikmisi Mahasiswa," *Ilk. J. Comput. Sci. Appl. Informatics*, vol. 1, no. 2, pp. 82–86, 2019.
[11] N. A. Manihuruk, M. Zarlis, E. Irawan, H. S. Tambunan, and I. Irawan, "Penerapan Data Mining Dalam Mengelompokkan Calon Penerima Beasiswa Dengan Menggunakan Algoritma K-Means," *KOMIK (Konferensi Nas. Teknol. Inf. dan Komputer)*, vol. 4, no. 1, 2020.
[12] N. Agustina and M. Hermawati, "Implementasi Algoritma Naïve Bayes Classifier untuk Mendeteksi Berita Palsu pada Sosial Media," *Fakt. Exacta*, vol. 14, no. 4, pp. 1979–276, 2021.
[13] R. C. PRIHANDARI, "Data Mining: Konsep Dan Apikasi Menggunakan Rapidminer (Series: Supervised Learning Dan Unsupervised Learning)," 2022.
[14] D. Darlinda and J. N. Utamajaya, "Sistem Pendukung Keputusan Penerima Beasiswa Program Indonesia Pintar Menggunakan Metode Algoritma K-Means Clustering," *JURIKOM (Jurnal Ris. Komputer)*, vol. 9, no. 2, p. 167, 2022.
[15] A. W. Fathurrahman, M. Thoriqulhaq, and F. Arianto, "Penerapan Machine Learning untuk Pengklasifikasian Hoaks pada Platform Media Sosial," *Senada*, vol. 2022, no. Senada, pp. 66–68, 2022.
[16] Y. T. Handika, S. Defit, and G. W. Nurcahyo, "Text Mining Dalam Membandingkan Metode Naïve Bayes Dengan C.45 Dalam Mengidentifikasi Berita Hoax Pada Media Sosial," *Rang Tek. J.*, vol. 5, no. 1, pp. 116–123, 2022.
[17] R. K. Putri and M. Athoillah, "Identifikasi Berita Hoax Terkait Virus Corona Menggunakan Long Short-Term Memory," *Semin. Nas. Has. Ris. dan Pengabdi.*, no. April, pp. 506–513, 2022.
[18] M. H. Al Farisi, *Analisis Sentimen Komentar Masyarakat Terhadap Kebijakan Pemerintah Tentang Sistem Zonasi Sekolah Menggunakan Algoritma K-Means Dan Algoritma Levensthein Distance*. 2019.
[19] E. Edward, "Identifikasi Berita HOAX Berbasis Web Menggunakan Algoritma C4.5," *J. Ilmu Komput. dan Sist. Inf.*, vol. 9, no. 1, p. 53, 2021.
[20] Y. T. Handika, S. Defit, and G. W. Nurcahyo, "TEXT MINING DALAM MEMBANDINGKAN METODE NAÏVE BAYES DENGAN C. 45 DALAM MENGIDENTIFIKASI BERITA HOAX PADA MEDIA SOSIAL," *Rang Tek. J.*, vol. 5, no. 1, pp. 116–123, 2022.