

Clustering Performance Between K-means and Bisecting K-means for Students Interest in Senior High School

Erni Seniwati^{1,*}, Acihmah Sidauruk², Haryoko³, Achmad Lukman⁴

^{1,2}Computer Science, Information System, Universitas Amikom Yogyakarta, Yogyakarta, Indonesia

³Computer Science, Information Technology, Universitas Amikom Yogyakarta, Yogyakarta, Indonesia

⁴Informatics, Information Technology, Telkom University, Bandung, Indonesia

Email: ^{1,*}erni.s@amikom.ac.id, ²acihmah@amikom.ac.id, ³haryoko@amikom.ac.id, ⁴alukman@telkomuniversity.ac.id

Correspondence Author Email: erni.s@amikom.ac.id

Submitted: 12/06/2023; Accepted: 29/06/2023; Published: 29/06/2023

Abstract–The interest of high school students is an important thing to do to see the talents of each student based on the academic scores obtained in the first and second semesters. There are two majors of interest in this case study, namely natural and social studies with criteria for natural studies scores including mathematics, chemistry, biology and physics. Meanwhile, the social studies criteria include history, economics, geography and sociology. This research propose comparing of clustering time and accuracy based on manual data from school as a reference of clustering in SMAN 1 Wonosari for 2011/2012 academic year using two clustering methods namely K-means and Bisecting K-Means. The results of this research compare to manual results interest from class teacher, so this work can demonstrate the run time comparison and accuracy of this study. The accuracy result shows 87.5% for both methods but different run times. For bisecting k-means got 0.0229849 seconds to complete the clustering process faster than k-means only got 0.0929448 seconds.

Keywords: Student Interest; K-Means; Bisecting K-Means

1. INTRODUCTION

Determination of interest in high school (SMA) is very important to classify students' tendencies in exploring the subjects they are interested in. There are many criteria that schools use to classify their students in the process of selecting science or social studies majors, the criteria used generally refer to subjects related to science or social studies majors including for science majors the criteria taken are mathematics, physics, biology and chemistry. Meanwhile, the Social Studies major has criteria for historical, economic, geography and sociology values. In addition to the general criteria for science and social studies that have been mentioned, the other research [1] adds IQ scores criteria as supporters in student majors. Another article [16] that related to this paper proposed a calculation using a t-Test to identify the effect of watching English tv series on youtube channels to improve the vocabulary size of senior high school at SMAN 2 Maros. The results show that the pre-test (49.97) was lower than the post-test (90.40), which indicated that it had a significant effect on students. Afika et al. [17] proposed a clustering technique to see the effect of increasing writing skills in descriptive text. The results show that the mean pre-test score was 40.33, and the post-test was 71.00, which increased significantly. This technique helps more students to improve their writing skill. Erdi and Yessica [21] proposed clustering scores of students to define the high grades of students. They used k-means clustering to categorize the student based on grades and attendance, and then they combined the results with Simple Additive Weighting (SAW) method to find out the top rank of the student. Sembiring et al. [22] used the K-means method to find out the clustering of the highest grades of students in their school. This research aims to achieve better placement of students suitable for their abilities. They got 29 students considered, five students proper in the highest grades, and 30 students were not considered. Sulistiyawati and Supriyanto [23] proposed a way of clustering grades of students using K-means clustering to define the better grades of students in a special student group. The results of the six classes in clustering showed 96 students in the highest grades group, and 96 students did not members of the highest grades group. Butsianto and Saepudin [24] proposed the implementation of Data mining to define the interests of students in math subjects. This research used Davies Boulding Index to measure the testing of k-means clustering. Dacwanda and Nataliani [25] proposed a k-means method implementation to analyze the academics grade of students according to their knowledge and skills. The results in skills showed 72 students in the smart group, 92 students in the average group, and 103 students in the low group. While in knowledge showed 71 students in the smart group, 87 students in the average group, and 107 students in a low group. Hutagalung et al. [26] proposed mapping the highest grades student used k-means clustering algorithm. They implemented k-means to cluster 120 students' data based on the reports of grading. The results of clustering such as 10 students in android programming, 62 students in web programming, 48 students in desktop programming.

The process of grouping students for the selection of majors is generally done manually. So, that the school or specifically the teacher classes takes time to process the student majors. There are many approaches taken in previous research to solve this problem, including the K-Means[5][2] method which is commonly used, this method is quite simple in conducting the data grouping process, then the next method is to use Fuzzy C-Means, this method used by [6][7][2], who conducted research on the grouping of students' majors at the SMAN 1 Wonosari by grouping 2 clusters, namely science and social studies which is also a collection of data used in this paper.

In this study, we using one of the developments of the K-means algorithm, namely Bisecting K-Means [3][8].

This method is an innovation from the K-means method combined with the bisection method, which is the process of dividing each data in two, and then calculating the error of the two data clusters, then the data that has a large error (SSE) will divide in two cluster and so on. So, that it reaches a value of specified K.

2. RESEARCH METHODOLOGY

In this paper, we compare the performance of two clustering algorithms in terms of time, namely K-means and Bisecting K-means in grouping the interests of SMAN 1 Wonosari students. The research steps in this paper can be shown as follows.

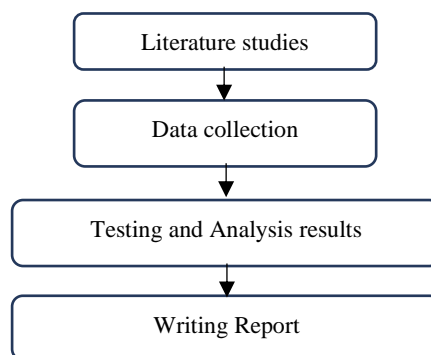


Figure 1. The block diagram of the research

2.1 Literature studies

In this part, we begin with literature review for several journals related to the research topic. We explore two clustering method to solve the high school student interest problem.

2.2 Data Collection.

We are collecting data from [2] that contains 189 students with 8 criteria namely math, physics, biology, chemist, history, geography, economic, and sociology. Then, cleaning the data by removing a several data outliers and then processing them into testing and analysis.

2.3 Testing and analysis results

In this part, we design two algorithms, such as K-means and Bisecting K-means. Next, we use both of these algorithms to process the data and achieve the results.

2.3.1 K-Means Algorithm

The traditional clustering[9][10] algorithm is K-means[2][11][12], this algorithm works by grouping data $D = \{x_1, x_2, \dots, x_n\}$ into several clusters based on the selected K value. The value of K is the initialization of the cluster center from random data that is used as the center of the cluster. To determine the members of the cluster center by calculating the distance between the cluster center and all existing data values using the SSE (sum of square error) or distance function [13]:

$$SSE = \sum_{i=1}^n (x - \bar{x})^2 \tag{1}$$

Pseudocode Algorithm 1: K-means

Input :

$D = \{x_1, x_2, \dots, x_n\}$

K number of Cluster

Output:

K

Assign initial Values for p_1, p_2, \dots, p_k

Repeat

Assign each data point x_i to the center of clusters with closest mean (eq. 1);

Calculate new mean for each cluster;

Until convergence

2.3.2 Bisecting Algorithm

Bisecting K-means[8][14][15][16][17][18][19][20] is a combined method of the K-means[5] algorithm with the bisection method[4], previously selecting the number of K, where data $D = \{x_1, x_2, \dots, x_n\}$ divided by two by determining two centers (C1) and (C1), then calculate the SSE (sum of square error) for each cluster, then choose

the SSE with the largest value between the two clusters which is then divided by two again until the number of clusters is equal to the number of K.

Pseudocode Algorithm 2 : Bisecting K-means

Input :
 $D = \{x_1, x_2, \dots, x_n\}$
K number of Cluster
Output:
 K
 For 1 to K
 If $SSE1 > SSE2$:
 Select Cluster
 Divide Data(D) as two cluster
 Calculate SSE for each cluster
 Until K

Furthermore, the work process flow of this research can be seen in Figure 2.

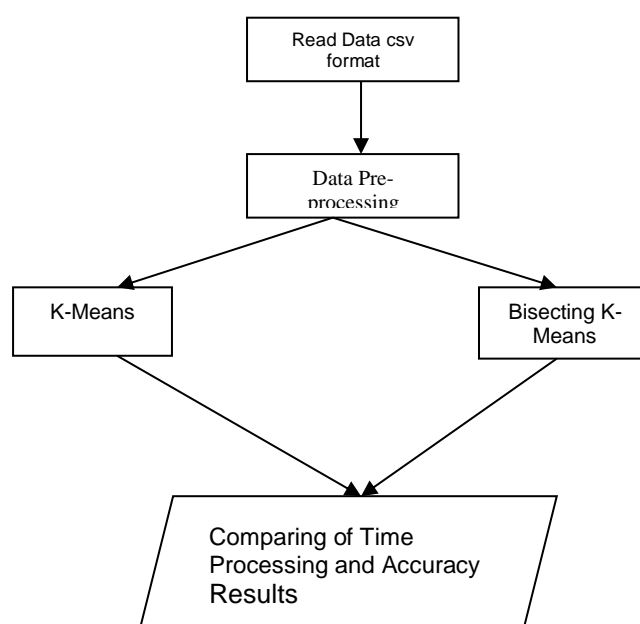


Figure 2. Flowchart process of comparing performance clustering algorithm

Figure 2 shows the working process of the K-means dividing algorithm for clustering. First, the 2011/2012 student grades of SMAN 1 Wonosari are collected in one file in CSV format, with several criteria as explained in more detail in section 3 of this paper. Interest criteria are filled in manually by the teacher assigned to make the selection so that they are not used in the grouping process but as a comparison to get the accuracy of the grouping results of the two methods used in this study. Then the values with these criteria are read as data points and assigned to K=2 because the target clustering is in two groups. Because this study uses two methods, namely traditional k-means clustering and bisecting k-means, the data read earlier is processed using these two methods, which both use the target two groups of IPA and IPS. The results of this clustering will be calculated using a confusion matrix.

Table 1. Confusion Matrix

		Prediction	
		Negative	Positive
Original Classes	Negative	a	b
	Positive	c	d

Accuracy is a comparison of cases with correct identification with all cases using the formula

$$AC = \frac{(a+d)}{(a+b+c+d)} \tag{2}$$

Error rate is the case that the identification is wrong with all cases using the formula

$$ER = \frac{(b+c)}{(a+b+c+d)} \tag{3}$$



2.4 Writing report

We resume and report all the experimental results in this paper.

3. RESULT AND DISCUSSION

3.1 The calculation manually

In this section, we show the calculation manually using 10 data sample for each method such as K-means and bisecting K-means algorithm.

3.1.1 K-means algorithm

As illustrate how traditional K-means works, we show the calculation of 10 data points as follows:

- a. We first prepare 10 data points in table 2.

Table 2. Contains of 10 data points examples

No	NIM	IPA scores				IPS scores			
		Math	Physics	Biology	Chemist	History	Geography	Economics	Sociology
1	10392	69	73	71	68	81	79	80	82
2	10393	70	70	72	76	77	76	80	85
3	10394	64	74	77	76	73	68	69	80
4	10400	75	75	68	73	76	77	80	80
5	10408	70	68	71	80	77	73	74	80
6	10425	63	71	74	72	82	72	79	76
7	10426	72	66	73	70	86	76	79	85
8	10430	67	76	68	68	76	76	75	83
9	10431	78	74	88	92	87	85	84	89
10	10439	87	81	90	88	88	87	90	89

- b. We define center point of clustering, we choose 2 point randomly from data point example such as :
75, 75, 68, 73, 76, 77, 80, 80 from data point 4 (C1) and 78, 74, 88, 92, 87, 85, 84,89 from data point 9 (C2)
- c. Calculate distance from center point using formula (1)

Table 3. The results of distance from center point

No	NIM	C1	C2
1	10392	107	1084
2	10393	102	805
3	10394	423	1364
4	10400	0	1053
5	10408	185	958
6	10425	275	1218
7	10426	251	932
8	10430	125	1420
9	10431	1053	0
10	10439	1314	191

- d. Define new cluster from Table 4 and we got the results of each new center point includes 68.75, 71.625, 71.75, 72.875, 78.5, 74.625, 77, 81.375 for C1 and 82.5, 77.5, 89, 90, 87.5, 86, 87, 89 for C2.
- e. Repeat steps C and D until the results do not change anymore.

1. First iteration

In this process we calculate the distance between data point and new center point as point out in step D.

Table 4. The results of the distances form new center poin

No	NIM	C1	C2
1	10392	61.0625	1199.75
2	10393	40.3125	972.75
3	10394	205.5625	1633.75
4	10400	87.3125	1135.75
5	10408	81.8125	1199.75
6	10425	91.3125	1430.75
7	10426	127.3125	1080.75



8	10430	74.8125	1579.75
9	10431	1008.0625	47.75
10	10439	1453.3125	47.75

According to table 4, rows 3, 9, and 10 belongs to C2 and another belongs to C1, it means that data point tends to change based on new center point. Because the initial iteration different results with the first iteration. Next, we define the new center point based on table 4 such as 69.428, 71.285, 71, 72.42857, 79.28571, 75.57143, 78.14286, 81. 571 for C1 and 76.333. 76.333, 85, 85.333, 82.666, 80, 81, 86 for C2. So, we continue the calculation to second iteration.

2. Second iteration

We repeat step C to continue calculate the distances based on new center point from the first iteration.

Table 5. The results of the second iteration

No	NIM	C1	C2
1	10392	41.0612245	582.111111
2	10393	36.3469388	386.444444
3	10394	268.489796	726.111111
4	10400	72.9183673	535.111111
5	10408	99.9183673	500.111111
6	10425	102.489796	673.444444
7	10426	102.204082	536.777778
8	10430	79.6326531	782.111111
9	10431	990.77551	123.444444
10	10439	1408.91837	335.111111

In table 5, the results show that only rows 9 and 10 belongs to C2. Because first and second iteration still different in its results, we continue calculate again in the third iteration that the results shown in table 6.

Table 6. The results of the third iteration

No	NIM	C1	C2
1	10392	61.0625	1199.75
2	10393	40.3125	972.75
3	10394	205.5625	1633.75
4	10400	87.3125	1135.75
5	10408	81.8125	1199.75
6	10425	91.3125	1430.75
7	10426	127.3125	1080.75
8	10430	74.8125	1579.75
9	10431	1008.0625	47.75
10	10439	1453.3125	47.75

In the table 6, shows that the results in rows 9 and 10 belongs to C2. This is the results same with in table 7, So, we stop the calculation because the distances between data point and center point did not change again.

Furthermore, we assign C1 and C2 as IPA and others as IPS. So, we describe the results on table 7 as follows.

Table 7. The Final results of clustering

No	NIM	Results
1	10392	IPA
2	10393	IPA
3	10394	IPA
4	10400	IPA
5	10408	IPA
6	10425	IPA
7	10426	IPA
8	10430	IPA
9	10431	IPS
10	10439	IPS

According to table 7. The results show 8 students in the group IPA and 2 students in the group IPS.

3.1.2 Bisecting K-means

In this section, we calculate bisecting K-means manually based on the same data from K-means above as follows.

- a. Preparing data, this data the same of K-means process above.



- b. Dividing the data to two groups as bisection method principles.

Table 8. Data group 1 (C1)

No	NIM	IPA scores				IPS scores			
		Math	Physics	Biology	Chemist	History	Geography	Economics	Sociology
1	10392	69	73	71	68	81	79	80	82
2	10393	70	70	72	76	77	76	80	85
3	10394	64	74	77	76	73	68	69	80
4	10400	75	75	68	73	76	77	80	80
5	10408	70	68	71	80	77	73	74	80
6	10425	63	71	74	72	82	72	79	76
7	10426	72	66	73	70	86	76	79	85
8	10430	67	76	68	68	76	76	75	83
9	10431	78	74	88	92	87	85	84	89
10	10439	87	81	90	88	88	87	90	89

- c. Defining center point randomly for each group.

We define center point includes 70,70, 72, 76, 77, 76, 80, 85 for C1 and 67, 76, 68, 68, 76, 76, 75, 83 for C2.

- d. Calculating SSE for each group using formula (1).

In this process, we get SSE results for each group as shown in table 10 and 11.

Table 10. The SSE results from Group 1 (C1)

No	NIM	C1	SSE
1	10392	109	
2	10393	0	
3	10394	303	605
4	10400	102	
5	10408	91	

According to table 10, to get number of SSE, we sum all values in the column C1.

Table 11. The SSE results from Group 2 (C2)

No	NIM	C2	SSE
6	10425	210	
7	10426	274	
8	10430	0	
9	10431	1420	3739
10	10439	1835	

In table 11, we implement the same calculations to get SSE results, and we can see that the group C1 and C2 shown the different SSE results. According to bisecting K-means algorithm, we choose greatest SSE to divide by two.

- e. Repeat step C and D until the group cannot divide by two.

- f. The final results, as shown in table 12.

Table 12. The results of Bisecting K-means

No	NIM	Results
1	10392	IPA
2	10393	IPA
3	10394	IPA
4	10400	IPA
5	10408	IPA
6	10425	IPA
7	10426	IPA
8	10430	IPA
9	10431	IPS
10	10439	IPS

As we can see, Table 12 shows that the results the same with K-means results, but we can see how fast run time for each algorithm to processes the whole data. Next, we show the results using python implementation.

3.2 Programming Implementation

We use a i3 core with 1.70 GHz, with memory RAM 6GB under Windows 10 operating system and software python 3.7 with anaconda packages to implement this experiment.

3.2.1 Data Pre-processing

Based on Figure 2, the first process flow is to read the csv format data. The process shown in Figure 3 with the initial five data display.

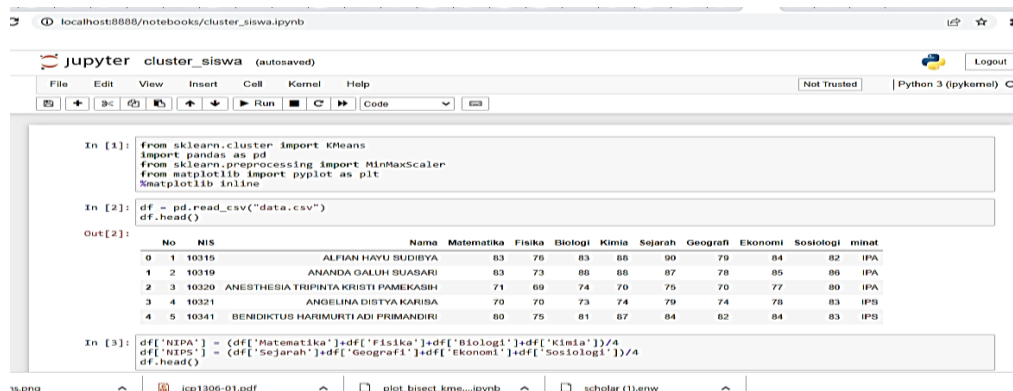


Figure 3. The first five row data samples of the dataset include 8 criterias

Data on Figure 3 are not all processed for clustering algorithm input, but only 8 criterias were selected, namely the value of mathematics, physics, biology, chemistry for the value of science (IPA) and such as history, economic geography and also sociology as social (IPS), while the interest column is filled in by the class teacher manually based on manual calculations. Furthermore, these criteria are used as representatives for IPA by adding up the criteria for each person then dividing by 4, as well as the criteria values for IPS as shown in Figure 4.

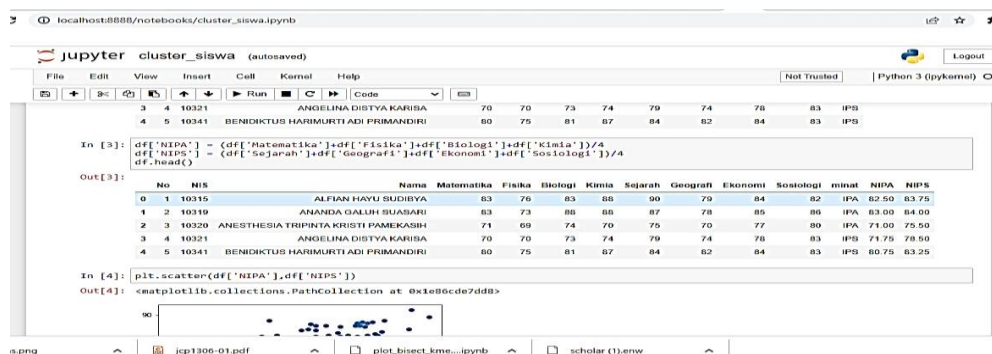


Figure 4. Criteria representation of natural science (NIPA) and social science (NIPS)

3.2.2 K-means Results

The first process implement K-means algorithm with input values from NIPS and NIPA, the results shows in Figure 5 and table 13,

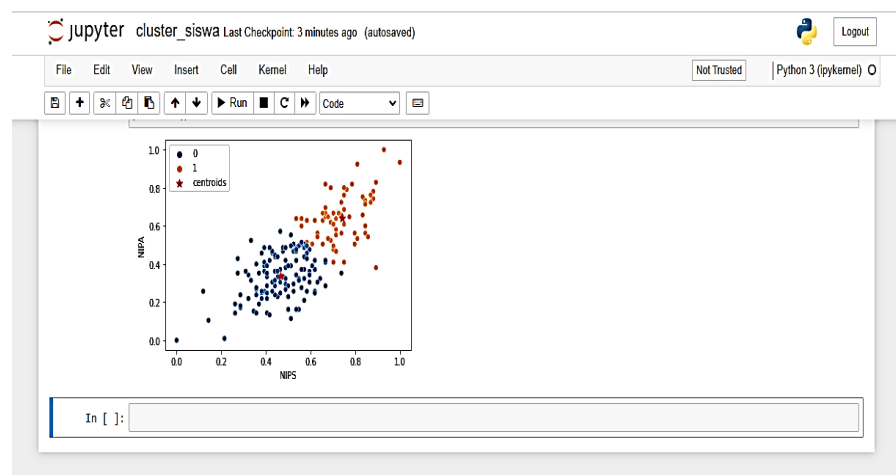


Figure 5. Clustering result of K-means

Table 13. The Accuracy Results of K-means

		Prediction	
		IPA	IPS
Original Classes	IPA	136	21
	IPS	0	12

$$ACC = \frac{(136 + 12)}{(136 + 21 + 0 + 12)} = 0.875 \times 100\% = 87.5\%$$

In Figure 5 shows the data is divided into two parts after the clustering process where the red part is identified as natural science while the blue part is social science, while the asterisk is the center point of each cluster.

3.2.3 Bisecting K-means Results

Next, the second process using the same pre-processing step with number of $K=2$ that represented of 2 interests such as IPA and IPS. In this section, after the initial data, we calculate using formula (1) to get SSE results for each group similarly manually step in section 3.1.2. this process taken 2 times iteration to get the results. As we can see the results in Figure 6 describe visualization from clustering data between IPA and IPS. The blue one represents IPA group and the other color represents IPS groups.

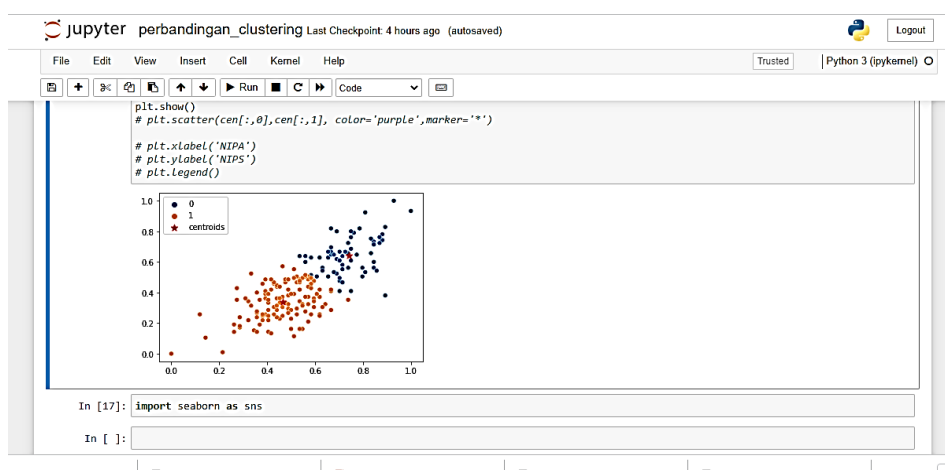


Figure 6. Clustering result of Bisecting K-means

Table 14. The Accuracy Results of Bisecting K-means

		Prediction	
		IPA	IPS
Original Classes	IPA	136	21
	IPS	0	12

$$ACC = \frac{(136 + 12)}{(136 + 21 + 0 + 12)} = 0.875 \times 100\% = 87.5\%$$

3.2.4 Comparison Clustering Time

We measure clustering time using time python library, and we show the results of then in table 15.

Table 15. Comparing clustering time

Algorithm	Run time (second)
<i>K-means</i>	0.0929448
<i>Bisecting K-means</i>	0.0229849

We can see in table 15, bisecting k-means faster than traditional K-means because they use principle of divide and conquer from computer science term. This term can break down the data into two or more sub-data.

4. CONCLUSION

In this paper we compare two clustering methods algorithms K-means and Bisecting K-means and the results tend to have the same results are 87.5%, but different run times, as shown in table 15. Bisecting K-means has a faster

clustering time than K-means because this algorithm can break down the process. thus speeding up the clustering process. Based on this experiment, we can conclude that the Bisecting K-means is superior in terms of time processing (0.0229849 seconds) than traditional K-means (0.0929448 seconds), and the results show that both methods can help senior high school teachers cluster their students on several majors faster than manual ones.

REFERENCES

- [1] C. Purnamaningsih, 2013. "Pemanfaatan metode k-means clustering dalam penentuan penjurusan siswa SMA". JURNAL ITSMART Vol 3. No 1.
- [2] B.A. Hastuti, E. Utami, T.E. Luthfi, 2013, Implementasi Metode Fuzzy C-Means Dan Topsis Dalam Membangun Sistem Pendukung Keputusan Penentuan Jurusan SMA (STUDI KASUS : PENENTUAN JURUSAN DI SMA NEGERI 1 WONOSARI), Jurnal DASIS Vol. 14 No. 2
- [3] D. Napoleon, M. Praneesh, S. Sathya, & M. SivaSubramani, (2011, December). "An efficient numerical method for the prediction of clusters using k-means clustering algorithm with bisection method". In International Conference on Computing and Communication Systems (pp. 256-266). Springer, Berlin, Heidelberg.
- [4] A. Eiger, S. Kris, and S. Frank. "A bisection method for systems of nonlinear equations." ACM Transactions on Mathematical Software (TOMS) 10, no. 4 (1984): 367-377.
- [5] H. H. Bock, (2007). Clustering methods: a history of k-means algorithms. Selected contributions in data analysis and classification, 161-172.
- [6] R. Kohavi, and F. Provost. "On Applied Research in Machine Learning. In Editorial for the Special Issue on Applications of Machine Learning and the Knowledge Discovery Process", Columbia University, New York, volume 30. 1998.
- [7] R. L. Cannon, J. V. Dave, & J. C. Bezdek, (1986). Efficient implementation of the fuzzy c-means clustering algorithms. IEEE transactions on pattern analysis and machine intelligence, (2), 248-255.
- [8] Z. Zhou, A. Ran, S. Chen, X. Zhang, G., Li, B. Wei, & H. Sun, (2020). A fast screening framework for second-life batteries based on an improved bisecting K-means algorithm combined with fast pulse test. Journal of Energy Storage, 31, 101739.
- [9] J. Han, M. Kamber. (2001: pp34-39). Data Mining: Concepts and Techniques, The Morgan Kaufmann Series.
- [10] D. Nanjaya, (2005). Clustering Data Non-Numerik dengan Pendekatan Algoritma K-Means dan Hamming Distance Studi Kasus Biro Jodoh. Jurnal Ilmiah Teknologi Informasi, 46-53.
- [11] Y. Agusta, (2007). K-Means-Penerapan, Permasalahan dan Metode Terkait. Jurnal Sistem dan Informatika Vol.3 ,47-60.
- [12] J. Zhang, L. Zhuo, & Y. Zhu, (2015). The improvement and application of a K-means clustering algorithm. Application of Electronic Technology, 1.
- [13] R. Nainggolan, R. Perangin-angin, E. Simarmata, & A. F. Tarigan, (2019, November). Improved the performance of the K-means cluster using the sum of squared error (SSE) optimized by using the Elbow method. In Journal of Physics: Conference Series (Vol. 1361, No. 1, p. 012015). IOP Publishing.
- [14] Puspitasari, N., Widiyans, J. A., & Setiawan, N. B. (2020). Segmentasi pelanggan menggunakan algoritme bisecting k-means berdasarkan model recency, frequency, dan monetary (RFM). Jurnal Teknologi dan Sistem Komputer, 8(2), 78-83.
- [15] Savaresi, S. M., & Boley, D. L. (2004). A comparative analysis on the bisecting K-means and the PDDP clustering algorithms. Intelligent Data Analysis, 8(4), 345-362.
- [16] Asmawati, A., & Widyastuti, S. (2022). THE EFFECT OF 'LEARN ENGLISH WITH TV SERIES' IN INCREASING VOCABULARY SIZE FOR THE THIRD GRADE STUDENTS AT SMAN 2 MAROS. English Language, Linguistics, and Culture International Journal, 2(2), 134-144.
- [17] Afika, N., Abubakar, M., & Nawir, M. S. (2022). THE EFFECT OF USING CLUSTERING TECHNIQUE ON THE STUDENTS' WRITING SKILLS IN DESCRIPTIVE TEXT AT SENIOR HIGH SCHOOL 10 MAKASSAR. English Language, Linguistics, and Culture International Journal, 2(1), 26-41.
- [18] Lailiyah, S., Yulsilviana, E., & Andrea, R. (2019). Clustering analysis of learning style on anggana high school student. TELKOMNIKA (Telecommunication Computing Electronics and Control), 17(3), 1409-1416.
- [19] Rahim, R., Santoso, J. T., Jumini, S., Bhawika, G., Susilo, D., & Wibowo, D. (2021). Unsupervised data mining technique for clustering library in Indonesia. Library Philosophy and Practice (e-journal), 4866.
- [20] Widiyaningtyas, T., Prabowo, M. I. W., & Pratama, M. A. M. (2017, September). Implementation of K-means clustering method to distribution of high school teachers. In 2017 4th International Conference on Electrical Engineering, Computer Science and Informatics (EECSI) (pp. 1-6). IEEE.
- [21] Saputra, E. A., & Nataliani, Y. (2021). Analisis Pengelompokan Data Nilai Siswa untuk Menentukan Siswa Berprestasi Menggunakan Metode Clustering K-Means. Journal of Information Systems and Informatics, 3(3), 424-439.
- [22] Sembiring, S. N. B., Winata, H., & Kusnasari, S. (2022). Pengelompokan Prestasi Siswa Menggunakan Algoritma K-Means. Jurnal Sistem Informasi Triguna Dharma (JURSI TGD), 1(1), 31-40.
- [23] Sulistiyawati, A., & Supriyanto, E. (2021). Implementasi Algoritma K-means Clustering dalam Penentuan Siswa Kelas Unggulan. Jurnal Tekno Kompak, 15(2), 25-36.
- [24] Butsianto, S., & Saepudin, N. (2020). Penerapan Data Mining Terhadap Minat Siswa Dalam Mata Pelajaran Matematika Dengan Metode K-Means. J. Nas. Komputasi dan Teknol. Inf, 3(1), 51-59.
- [25] Dacwanda, D. O., & Nataliani, Y. (2021). Implementasi k-Means Clustering Untuk Analisis Nilai Akademik Siswa Berdasarkan Nilai Pengetahuan dan Keterampilan. AITI, 18(2), 125-138.
- [26] Hutagalung, J. (2022). Pemetaan Siswa Kelas Unggulan Menggunakan Algoritma K-Means Clustering. JATISI (Jurnal Teknik Informatika dan Sistem Informasi), 9(1), 606-620.