IMPLEMENTATION OF K-MEANS ALGORITHM AS A CLUSTERING METHOD FOR SELECTING ACHIEVEMENT STUDENTS BASED ON ACADEMIC GRADE

Indah Dwijayanthi Nirmala [1]; Prima Dina Atika [2];

Program Studi Informatika
Universitas Bhayangkara Jakarta Raya
www.ubharajaya.ac.id
indah.dwijayanthi@dsn.ubharajaya.ac.id [1]; prima.dina@dsn.ubharajaya.ac.id [2]

Abstract— Increased student success and low student failure rates are a reflection of good quality in the field of education. Awareness of the education importance determines the quality of utilizing existing resources, including human resources, facilities, and infrastructure as well as technological resources. A large number of students in school as well as the variety of different abilities and academic qualifications for each student, so that it is difficult for the school to select the excellent students based on academic scores. Therefore, the purpose of this study is to formulate and analyze problems, where these problems can be solved by using the K-Means Clustering Algorithm method based on grouping data on assignment scores, test scores, and student practice scores as variables that will be supporting values in selecting outstanding students. By applying the k-Means clustering algorithm, it can be proposed as an approach that can be used to predict the selection of excellent students. With this system, it can minimize the occurrence of errors in selecting excellent students based on academic scores.

Keywords: *K-Means, Clustering, System, Education*

Intisari— Meningkatnya keberhasilan siswa dan rendahnya tingkat kegagalan siswa merupakan cerminan dari kualitas yang baik dalam bidang Pendidikan. Kesadaran akan arti pentingnya Pendidikan menentukan kualitas dalam memanfaatkan Source daya yang dimiliki, termasuk Source daya manusia, sarana dan pra sarana juga Source daya teknologi. Banyaknya jumlah siswa didik disekolah serta beragamnya kemampuan dan kualifikasi akademik yang berbeda di setiap siswa didik, sehingga kesulitan bagi pihak sekolah dalam memudahkan pencarian untuk pemilihan siswa berprestasi berdasarkan nilai akademik. Maka dari itu tujuan penelitian ini merumuskan dan menganalisa permasalahan, dimana masalah tersebut dapat diatasi dengan menggunakan metode Algoritma K-Means Clustering berdasarkan pengelompokkan data nilai tugas, nilai ujian, dan nilai praktek siswa sebagai variabel yang akan menjadi nilai pendukung dalam pemilihan siswa berprestasi. Dengan penerapan algoritma K-Means clustering dapat diusulkan sebagai salah satu pendekatan yang dapat dilakukan untuk memprediksi dalam pemilihan siswa berprestasi. Dengan adanya sistem ini dapat meminimalisir terjadinya suatu kesalahan dalam pemilihan siswa berprestasi berdasarkan nilai akademik.

Kata Kunci: K-Means, Clustering, Sistem, Pendidikan

INTRODUCTION

Education is a process of learning knowledge that is passed down or taught from one generation to the next through teaching, training, or research that aims to develop one's own or individual abilities (Dodi 2016). According to Ki Hajar Dewantara, the meaning of education is the process of guiding all the natural forces that exist in the children of students, so that they as human beings and as members of society can achieve the highest safety and happiness (Dewantara 1961). The high level of student success and the low rate of student failure is a reflection of the good quality of education (Isnaini, Kusuma, & Noviani, 2015). The world of education today is demanded by the importance of having educational resources to prepare agents of change who strong, superior, participatory, and competitive. (Ningrum, 2009), (Rezky, Sutarto, Prihatin, Yulianto, & Haidar, 2020). Educational resources are anything that is used in the implementation of education, which includes educational personnel, the community, funds, facilities, and infrastructure for human resources. (Ningrum 2009)

Along with the increasing number of student data each year, the amount of student data continues to increase and the variety of abilities and academic qualifications is different for each

student. So that it is difficult for the school to facilitate the search for the selection of achievement students based on academic scores. (Irawan 2017)

Therefore it is necessary to do this data so that it can be processed into information and knowledge as a grouping of achievement students from assignment scores, test scores, and student practice scores as variables that will be supporting values in the selection of achievement students.

Based on several studies that have been conducted using clustering techniques in mapping student value data to predict student achievement (Sibuea and Safta 2017). Research (Ratih and Sonalitha Elta 2018) using the K-Means Clustering method is aimed at simplifying and accelerating the process of grouping students and producing the division of students into classes based on academic scores, gender, behavior, and names of students. Research (Metisen and Sari 2015) A hierarchical cluster process using the K-means method produces an information description of clustered or grouped sales.

Based on these problems, in this study, it was chosen to apply the K-means algorithm as a method of solving existing problems. So that the researchers conducted research entitled "Application of the K-Means Algorithm as a Clustering Method for Selection of Achievement Students based on Academic grades". So can assist and facilitate the school in process voter decision making an achievement student based on academic grades.

MATERIALS AND METHODS

A. Research Data

The material used in this research is data from school, in the form of school identity data and data of the school's students.

B. Research Method

The method in this research can be described in Figure 1 below:

Data Collection Method

The data used in this study are school identification data and student data of school. Data were taken in the study year 2019/2020.

Initial Data Processing Methods

The stages of the Data Processing process are as follows:

a. Data Integration, namely unifying storage areas. The school and student identity data obtained are put together in one storage medium.
b. Data reduction is getting data that has fewer attributes and records by reducing unnecessary/unfilled records.

Source: (Nirmala & Atika, 2020)

Gambar 1. Metode Penelitian

The Method Proposed

The method proposed in this research is the K-Means Clustering Algorithm to predict the selection of achievement students based on academic scores. The following is the K-means Clustering Algorithm process: a) Determine the number of clusters wanted, b) Determine the centroid value, c) Calculate the distance between the values and centroid values. The formula for calculating the distance between the values and centroid values:

Value range to centroid value to 1 :
$C1 = \sqrt{N1 - C1} + \sqrt{N2 - C1} + \sqrt{N3 - C1}$ .....(1)

Value range to centroid value to 2 :
$C2 = \sqrt{N1 - C2} + \sqrt{N2 - C2} + \sqrt{N3 - C2}$ .....(2)

Value range to centroid value to 3 :
$C3 = \sqrt{N1 - C3} + \sqrt{N2 - C3} + \sqrt{N3 - C3}$ .....(3)

Value range to centroid value to 4 :
$C4 = \sqrt{N1 - C4} + \sqrt{N2 - C4} + \sqrt{N3 - C4}$ .....(4)

Remarks:
N = student grade; C1 = 1st cluster value; C2 = 2st cluster value; C3 = 3st cluster value

Evaluation and Validation Result

The results of the evaluation and validation stage are expected to be able to observe and demonstrate the performance of the algorithms made.

RESULT AND DISCUSSION

This research uses a model testing experiment, the following stages are:

1. Prepare data for conducting experiments
2. Preprocessing data by reducing empty data
3. Experiment use the Clustering method with the K-Means Algorithm
4. Calculating the accuracy of the Clustering method using the K-Means Algorithm
5. Analyze the results of using the Clustering method using the K-Means Algorithm

The following is the calculation result of the K-Means algorithm:

1. Determining the Dataset

Table 1. Dataset

No	Nis Siswa	Nama Siswa	Nilai Tugas	Nilai Ujian	Nilai Praktek
1	0001295471	Moza Nur Afiyah	89	90	75
2	0001992007	Siti Zahra Chairunnisa	90	71	95
3	0002790331	Santoso Adi Kusumah	70	75	80
4	0003356318	Tri Utami Nurul Azizah	45	65	59
5	0003518994	Isri Mufidah	65	75	53
6	0004844053	Aldi Saputra	80	70	75
7	0007045603	Arya Dwi Pangga	90	85	81
8	0008020560	Dessy Lina Arnita Simanjuntak	70	70	73
9	0009498535	Amanda Putri Wijayanti Matrutti	96	93	85
10	00010491071	Cindy Fitriana	60	55	48

Source: (Nirmala & Atika, 2020)

2. Determine the number of clusters formed. The clusters that will be formed include:
   a. Cluster 1 (C1) = Verry Good
   b. Cluster 2 (C1) = Good
   c. Cluster 3 (C1) = Enough
   d. Cluster 4 (C1) = Less

3. Determining the central cluster value is:
   max = 90; mean = 80; min = 70; mn = 60; l = 4; m = 3; n = 2; o = 1

4. Calculation of Data Allocation to cluster distance. To result the distance to the cluster is obtained from the calculation with the formula:

$d(x_i,x_j) = \sqrt{\sum_k^n = 1 (x_{ik} - x_{jk})^2}$ .....(5)
$d(x_i,x_j) = Jarak\ dari\ x_i\ dan\ x_j$ .....(6)

Description of the Index Formula:
Assignment scores (p), test scores (q), practice score (r), max (a), mean (b), min(c), mn(d), l, m, n, dan o.

$C1 = \sqrt{(p - a)^2 + (q - a)^2 + (r - l)^2}$.....(7)

$C2 = \sqrt{(p - b)^2 + (q - b)^2 + (r - m)^2}$ .....(8)

$C3 = \sqrt{(p - c)^2 + (q - c)^2 + (r - c)^2}$ .....(9)

$C4 = \sqrt{(p - d)^2 + (q - d)^2 + (r - d)^2}$ .....(10)

Table 2. Results of the Calculation of the Clustering Algorithm K-Means Method

NO	Nis Siswa	Nama Siswa	C1	C2	C3	C4	Hasil
1	0001295471	Moza Nur Afiyah	71.00704	73.24616	78.03845	84.95293	Sangat baik
2	0001992007	Siti Zahra Chairunnisa	92.96236	92.97849	95.13149	99.28243	Sangat baik
3	0002790331	Santoso Adi Kusumah	80.00625	77.80745	78.16009	81.03086	Baik
4	0003356318	Tri Utami Nurul Azizah	75.3326	67.72001	62.44197	60.11655	Kurang
5	0003518994	Isri Mufidah	57.01754	52.44044	51.48786	54.35071	Cukup
6	0004844053	Aldi Saputra	74.43789	72.69113	73.68175	77.30459	Baik
7	0007045603	Arya Dwi Pangga	77.16217	78.79721	82.86133	89.02247	Sangat baik
8	0008020560	Dessy Lina Arnita Simanjuntak	74.57211	71.41428	71	73.37575	Cukup
9	0009498535	Amanda Putri Wijayanti Matrutti	81.2773	84.55176	89.96666	97.16481	Sangat baik
10	00010491071	Cindy Fitriana	63.72598	55.22681	49.40648	47.26521	Kurang

Source: (Nirmala & Atika, 2020)

Algorithm Implementation

Figure 2, source-code algorithm used by System for selecting achievement student.

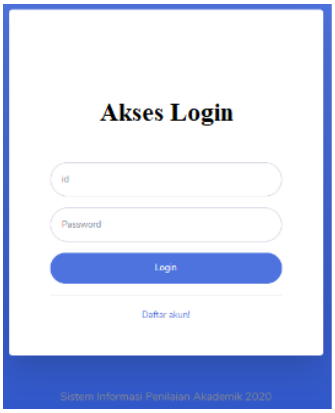
```
<!-- algoritma kmeans clustering -->
<?php
foreach($data as $d):

$jum = $no;


    $c1a = sqrt(pow(($d['n1']-$max),2)+pow(($d['n2']-$max),2)+pow(($d['n3']-$l),2));
    if ($c1a <= 0) { $c1a = 0;}
    $c1 = ($c1a);

    $c2a = sqrt(pow(($d['n1']-$mean),2)+pow(($d['n2']-$mean),2)+pow(($d['n3']-$m),2));
    if ($c2a <= 0) { $c2a = 0;}
    $c2 = ($c2a);

    $c3a = sqrt(pow(($d['n1']-$min),2)+pow(($d['n2']-$min),2)+pow(($d['n3']-$n),2));
    if ($c3a <= 0) { $c3a = 0;}
    $c3 = ($c3a);

    $c4a = sqrt(pow(($d['n1']-$mn),2)+pow(($d['n2']-$mn),2)+pow(($d['n3']-$o),2));
    if ($c4a <= 0) { $c4a = 0;}
    $c4 = ($c4a);

    if (($c1 < $c2) AND ($c1 < $c3) AND ($c1 < $c4)) {
    $result = "Sangat baik";
    } else if (($c2 < $c1) AND ($c2 < $c3) AND ($c2 < $c4)) {
    $result = "Baik";
    } else if (($c3 < $c2) AND ($c3 < $c1) AND ($c3 < $c4)){
     $result = "Cukup";
    }else if (($c4 < $c1) AND ($c4 < $c2) AND ($c4 < $c3)){
    $result = "Kurang";
    }

?>
```

Source: (Nirmala & Atika, 2020)

Figure 2. Implementation of Code K-Means Clustering Algorithm

Implementation of Interface Design

a. Login Page Admin and Teacher

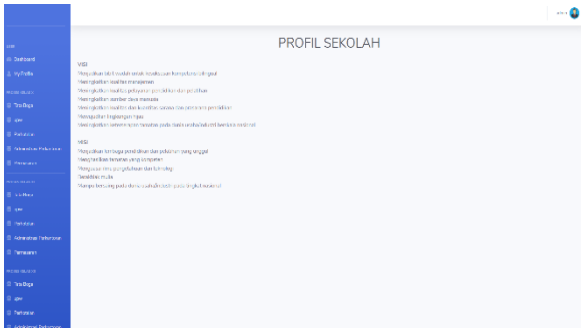


Source: (Nirmala & Atika, 2020)

Figure 4. Menu Login

Figure 4 The web page used to display system permissions for the user and admin.

b. Dashboard Page Admin and Teacher

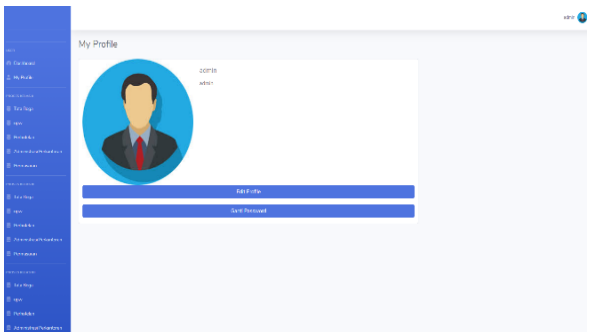


Source: (Nirmala & Atika, 2020)

Figure 5. Dashboard Menu

Figure 5, the web page used to display the admin and teacher dashboard pages.

c. Profile Admin

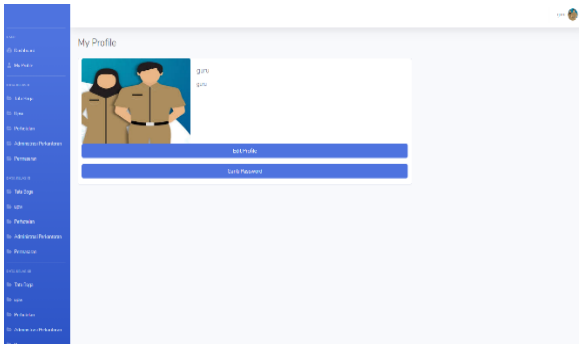


Source: (Nirmala & Atika, 2020)

Figure 6. Admin Profil Menu

Figure 6 Web page used to process admin profile information.

d. Profile Teacher

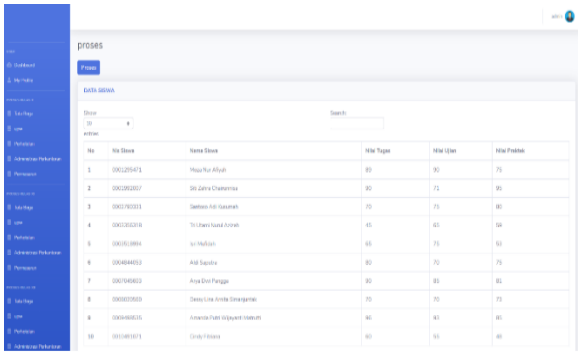


Source: (Nirmala & Atika, 2020)

Figure 7. Teacher Profil Menu

Figure 7 Web page used to process teacher profile information.

e. Admin Class Assessment Menu

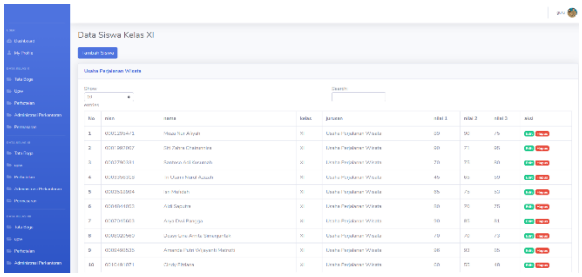


Source: (Nirmala & Atika, 2020)

Figure 8. Admin Class Assessment Menu

Figure 8 Web page used to display the information of student value data which has not been processed using the K – means Algorithm Clustering.

f. Teacher Class Assessment Menu

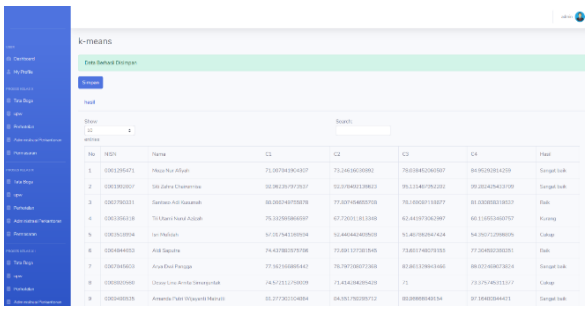


Source: (Nirmala & Atika, 2020)

Figure 9. Teacher Class Assessment Menu

Figure 9 Web page that is used to process the information (add data, edit data, and delete data) one of the data values of the student.

g. The K-Means Clustering Value Process Page

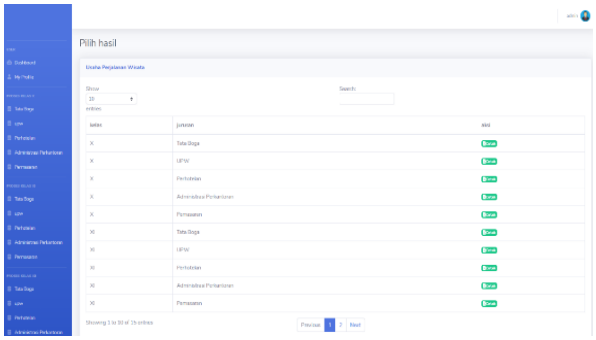


Source: (Nirmala & Atika, 2020)

Figure 10. Menu K-Means Process Clustering

Figure 10 Web page that is used to display the information about the calculation process of one of the data of the student processed using the K - means Algorithm Clustering.

h. Report Page

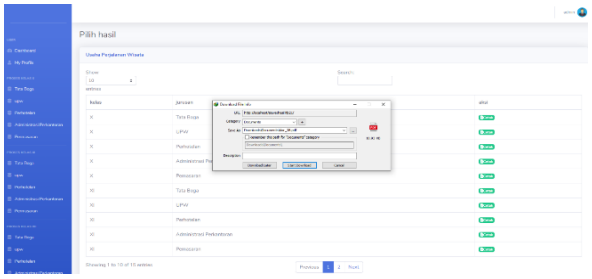


Source: (Nirmala & Atika, 2020)

Gambar 11. Menu Report Page

Figure 11 Web page used to print the results of the K-Means Algorithm Clustering from one of the student value data.

i. Response Download



Source: (Nirmala & Atika, 2020)

Figure 12. Download Report

Figure 12 Display response of the download processing of the print report.

j. Pdf Report Results

Data Pengelompokkan Siswa

Kelas : XI (sebelas)
Jurusan : Usaha Perjalanan Wisata

Nisn Siswa	Nama Siswa	Nilai Tugas	Nilai Ujian	Nilai Praktek	Hasil
0001295471	Moza Nur Afiyah	89	90	75	Sangat baik
0001992007	Siti Zahra Chairunnisa	90	71	95	Sangat baik
0002790331	Santoso Adi Kusumah	70	75	80	Baik
0003356318	Tri Utami Nurul Azizah	45	65	59	Kurang
0003518994	Isri Mufidah	65	75	53	Cukup
0004844053	Aldi Saputra	80	70	75	Baik
0007045603	Arya Dwi Pangga	90	85	81	Sangat baik
0008020560	Dessy Lina Arnita Simanjuntak	70	70	73	Cukup
0009498535	Amanda Putri Wijayanti Matrutti	96	93	85	Sangat baik
0010491071	Cindy Fitriana	60	55	48	Kurang

Figure 13. Report Data Sheet

Figure 13 Web page used to display the results of the print report data from one of the data value of the student have been processed using the K-Means Algorithm Clustering.

CONCLUSION

This system draft can assist in data processing academic grade of students in the selection of achievement students, Application of Algorithm K-Means clustering method can be used to grouping the data of students based on academic grades assignment, exam, and practice, that are very well, good, enough, and less. So it can be used to see the students that should be given an additional study to achieve a passing grade value, Clustering Method can help the school in the grouping of achievement students who have excellent academic grades.

APPRECIATION

Appreciation and thanks to Kementrian Pendidikan dan Kebudayaan that have financed the activities of Penelitian Dosen Pemula (PDP) in this research. Thanks also to Lembaga Penelitian dan Pengabdian Masyarakat Universitas Bhayangkara who has supported this research through Kopertis Region III Jakarta.

REFERENSI

Dewantara, K. H. (1961). *Karya Ki Hajar Dewantara bab I: Pendidikan*. Jakarta: Majelis Luhur Taman Siswa.

Dodi, N. (2016). Pentingnya Guru Untuk Mempelajari Psikologi Pendidikan. *Nusantara (Jurnal Ilmu Pengetahuan Sosial) Volume 1 Desember 2016*, *1*, 59–63.

Irawan, Y. (2017). Sistem Pendukung Keputusan Untuk Penilaian Prestasi Belajar Siswa pada Sekolah Dasar Negeri 167 Pekanbaru dengan Metode AHP. *Jurnal Ilmu Komputer*, *6*, 3–8.

Isnaini, M., Kusuma, D., & Noviani, L. (2015). PENGARUH KOMPETENSI DOSEN DAN FASILITAS BELAJAR TERHADAP KEPUASAN MAHASISWA PENDIDIKAN EKONOMI FKIP UNS. *BISE: Jurnal Pendidikan Bisnis Dan Ekonomi*, *1*(2), 1–20. https://doi.org/10.20961/BISE.V1I2.17968

Metisen, B. M., & Sari, H. L. (2015). Analisis clustering menggunakan metode K-Means dalam pengelompokkan penjualan produk pada Swalayan Fadhila. *Jurnal Media Infotama*, *11*(2), 110–118.

Ningrum, E. (2009). PENGEMBANGAN SOURCE DAYA MANUSIA BIDANG PENDIDIKAN. *Jurnal Geografi GEA*, *9*(1), 1–9. Retrieved from https://ejournal.upi.edu/index.php/gea/article/view/1681/

Nirmala, I. D., & Atika, P. D. (2020). *Laporan Akhir Hibah Penelitian: Penerapan Algoritma K-Means Sebagai Metode Clustering Untuk Pemilihan Siswa Berprestasi Berdasarkan Nilai Akademik*. Jakarta.

Ratih, S., & Sonalitha Elta. (2018). Aplikasi k-means clustering untuk pengelompokkan siswa ke dalam kelas berdasarkan nilai akademis, jenis kelamin, perilaku dan nama siswa di sma negeri 1 srengat. *Seminar Nasional Sistem Informasi*, 1179–1187.

Rezky, M. P., Sutarto, J., Prihatin, T., Yulianto, A., & Haidar, I. (2020). Generasi Milenial yang Siap Menghadapi Era Revolusi Digital (Society 5.0 dan Revolusi Industri 4.0) di Bidang Pendidikan Melalui Pengembangan Source Daya Manusia. *Seminar Nasional Pascasarjana (Prosnampas) 2019*, 1117–1125. Semarang: Universitas Negeri Semarang. Retrieved from https://proceeding.unnes.ac.id/index.php/snpasca/article/view/424

Sibuea, M. L., & Safta, A. (2017). Pemetaan Siswa Berprestasi Menggunakan Metode K-Means Clustring. *Jurteksi*, *4*(1), 85–92. https://doi.org/10.33330/jurteksi.v4i1.28