Matrik: Jurnal Manajemen, Teknik Informatika, dan Rekayasa Komputer
Vol. 22, No. 1, November 2022, pp. 259∼272
ISSN: 2476-9843, accredited by Kemenristekdikti, Decree No: 200/M/KPT/2020
DOI: 10.30812/matrik.v22i1.2337

Implementation of Single Linked on Machine Learning for Clustering Student Scientific Fields

Saiful Nur Arif[1], Muhammad Dahria[2], Sarjon Defit[3], Dicky Nofriansyah[4], Ali Ikhwan[5]
[1,2,4]STMIK Triguna Dharma, Medan, Indonesia
[3]Universitas Putra, Padang, Indonesia
[5]Universitas Islam Negeri Sumatra Utara, Medan, Indonesia

Article Info

Article history:

Received August 27, 2022
Revised September 15, 2022
Accepted November 27, 2022

Keywords:

Clustering,
Euclidean,
Machine Learning,
Single Linkage,
Scientific Fields,

ABSTRACT

The problem with data mining methods is that they are limited in finding patterns. In contrast, machine learning methods are more well-tested in shaping training data and making predictions using new data sets. Therefore machine learning can be used to analyze quite a lot of data related to current research problems. It is not easy to map the scientific fields that students will use in submitting titles so that the results of the thesis are made less than optimal. For this reason, it is necessary to map this concentration to assist students in completing their thesis through specialization classes. One of the Mechanical Learning techniques used in solving this problem is the Single Linkage Technique. Testing the method begins with determining the raw data used and then looking for the proximity value using Euclidean to get cluster results later from scientific field mapping. From the Single Linkage Technique process that has been carried out, several cluster results will be obtained, namely clusters that map groups of STMIK Triguna Dharma students who have competence and clusters that map groups of STMIK Triguna Dharma students who lack competence. The results of this grouping produce 2 clusters based on manual calculations validated by the program. The institution will make specialization classes according to the resulting clusters. They are creating specialization classes that are by the competencies possessed by STMIK Triguna Dharma students.

Corresponding Author:

Saiful Nur Arif
Department Sistem Informasi,
STMIK Triguna Dharma, Medan. Indonesia,
Email: saiful.nurarief@gmail.com

How to Cite: S. Arif, M. Dahria, S. Defit, D. Novriansyah, and A. Ikhwan, Implementation of Single Linked on Machine Learning for Clustering Student Scientific Fields, MATRIK : Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer, vol. 22, no. 1, pp. 253-264, Dec. 2022.

1. INTRODUCTION

Mapping of scientific fields is one way to improve competence and produce quality human resources in accordance with their competencies. Clustering is grouping records, observing, or paying attention and forming classes of objects that have similarities. Cluster is a collection of records that have similarities with one another and have dissimilarities with records in other clusters [1]. Clustering is different from classification in that there is no target variable in the clustering. Clustering does not attempt to classify, estimate, or predict the value of the target variable. However, the clustering algorithm tries to divide the entire data into groups that have similarities (homogeneous), where the similarity of records in one group will be valuable [2].

Based on the observations that have been made, there are problems regarding the difficulty of classifying students' scientific fields according to their competencies. Currently at STMIK Triguna Dharma still using the concept of selecting a concentration class in the field of science by choosing according to the wishes of students, this has resulted in problems, including the incompatibility of competencies possessed with the chosen scientific field and less than optimal in the preparation of the thesis. The concept of choosing a concentration class is considered ineffective because the selection system does not consider other variables or factors.

Previous research still used the data mining method to determine patterns with the Association Rule technique by applying the a priori algorithm to search for patterns.[3–5] There were still many areas for improvement in it. The final results could have been more precise in testing because there was still a human influence on the results obtained. At the same time, research conducted using machine learning techniques shows the results of direct testing with manual calculations or validation with existing programs or tools in machine learning

Based on these problems, a technology is needed that can be used to map students' scientific fields based on the competencies possessed by each student. One of the technologies that can be used is Machine Learning. Machine Learning is a technology that is used to analyze data to be used as knowledge [6]. Another opinion suggests that Machine Learning is a branch of artificial intelligence that adapts human abilities to learn [7]. The use of Machine Learning has been tested in solving existing problems [8]. Such as classifying fast food outlets [9], then predicting an increase in Covid-19 cases [10].

The Mechine Learning method that will be used to cluster students' scientific fields is the Single Linkage Technique. The Single Linkage technique is a method of hierarchical clustering that is used to group data into knowledge [11]. In addition, the Single Linkage Technique is known as a method that is able to explore knowledge from data into groupings based on the value of the closest distance [12]. This method was chosen because it has been widely used in solving data clustering problems. Such as grouping data on patients with liver disease [13]. Then grouping customer data for telemarketing [14].

In the grouping of student scientific fields there are 15 categories of fields of science including: artificial neural networks, image processing, cryptography, steganography, machine learning, expert systems, decision support systems, management information systems, animation, data mining, simulations, robotics, deep learning, systems distributed and computer networks. The student's scientific field is obtained from the results of the institution's decision regarding the determination of the concentration class for the preparation of the thesis. The first process to be carried out is to standardize the data based on existing variables, then calculate the closest distance using the Euclidean concept so that later the results of the grouping of the data will be known.

With this research, it can be an innovation in data grouping which has been carried out unstructured and ignores the variables that support the clustering. The results of the grouping that have been obtained will later be used for the formation of specialization or concentration classes that are intended for students. This class will later assist students in preparing their thesis through a mentoring program. In addition, the results of this clustering can also be used as a reference by the institution to determine the level of student competence, so that it becomes a reference in making policies to improve the quality of students.

2. RESEARCH METHOD

In this study, data collection was carried out by determining student sample data obtained from the results of examinations conducted on students' scientific fields. The student's scientific field will be used as part of taking the thesis title. The scientific fields that will be used are artificial neural networks (V1), image processing (V2), cryptography (V3), steganography (V4), machine learning (V5), expert systems (V6), decision support systems (V7), information systems management(V8), animation(V9), data mining(V10), simulation(V11), robotics(V12), deep learning(V13), distributed system(V14) and computer network(V15). Stages of this research through the following stages, Identify the problem, namely finding problems with the problems on the Triguna Dharma campus in determining to map the scientific fields that students will use in submitting titles so that the results of the thesis are made less than optimal,Searching for literature, namely by collecting relevant references about the research that will be discussed by searching for national and international indexed journals as well as books, Data Collection stages in collecting data, both primary and secondary data, Data Cleaning is the stage in filtering data based on the attribute value restrictions used, Cluster Method Implementation, namely the stages in determining clusters to facilitate the stages in the search, Single Linkage at this stage is calculated manually using the formula by finding the closest point between the clusters that have been formed, Website Implementation Validation stages in testing validation based on tests that have been carried out manually and obtained the same results,The final decision is the final stage of the test results by obtaining 2 clusters in determining the final results.

2.1. Single Linkage

After collecting data, the next step is to apply the Single Linkage Technique in grouping student scientific data. The Single Linkage technique is one of the techniques in Mechine Learning in data clustering. Clustering is a way of analyzing data by grouping objects into groups based on a certain similarity. A cluster is a collection of objects that are combined together because of their similarities or proximity. Clustering (clustering) can also be interpreted as the process of grouping a set of objects into the same object classes. Clustering or clustering is also one of the many functions of the data mining process to find groups of objects that are almost the same [15]. Cluster analysis is an attempt to identify groups of similar objects and help find distribution patterns and relationship patterns in large data sets [16]. The important thing in the clustering process is declaring a set of patterns to the appropriate group that is useful for finding similarities and differences so that it can produce valuable conclusions [17].

$$\bar{X} = \frac{\Sigma_{i=1}^{n} X_i}{n} \tag{1}$$

From this equation, it can be seen that X is the average value of the variable, Xi is the value of the variable and N is the number or quantity of the object. Next is to find the value of the standard deviation of the variable with the equation [18].

$$std(x) = \sqrt{\frac{\Sigma_{i=1}^{n}(x_i - \bar{x})^2}{n-1}} \tag{2}$$

From this equation, it can be seen that std(X) is the average value of the variable, Xi is the value of the variable and N is the number or quantity of the object. Next is to find the standard zero value with the equation [19].

$$Z_i = \frac{X_i - \bar{X}}{Std(X)} \tag{3}$$

From this equation, it can be seen that Zi is the standard Zero value, std(X) is the average value of the variable, Xi is the variable value and N is the number or quantity of the object. Next is to find the value of the distance measurement using the equation [20].

$$d_{(x,y)} = \sqrt{(\Sigma_{i=0}^{k}(x_i - y_p)^2)} \tag{4}$$

From these equations, it can be seen that $d_((x,y))$ is a measure of the similarity or dissimilarity between the $x - th$ and $y - th$ objects, k is the number or quantity of objects, xi is The value of the data parameter r to 19, yp is the parameter value agent $r + 1$. Next is grouping using Euclidean Single Linkage. This technique uses the principle of minimum distance which begins with finding the two closest objects, and both form the first cluster with the equation [21].

$$Euclidean = Min(Zi; Zn) \tag{5}$$

After grouping the data, it can be seen that the results of student clustering are grouped based on scientific field variables.

3. RESULT AND ANALYSIS

Data obtained from student exams conducted by STMIK Triguna Dharma. The data includes test scores achieved by 15 students with 15 scientific fields. The following is the student exam data that was successfully obtained.

Table 1. Student Exam Score Data

No	Nim	Name	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12	V13	V14	V15
1	9983689340	Afdillah (A)	80	88	89	87	90	89	91	85	80	85	88	85	82	80	85
2	9950021406	Ahmad Afrizal (B)	82	87	88	87	92	94	89	91	80	85	89	91	80	95	83
3	9986294859	Ahmad Fauzi (C)	81	80	86	87	85	89	87	90	80	96	86	89	80	82	80
4	9986050887	Anggi Arafah (D)	81	95	87	87	91	90	94	85	80	85	88	87	81	81	80
5	9990261330	Audriyana (E)	82	92	87	86	93	95	96	93	80	90	89	87	85	89	83
6	9000137537	Ayu Andira (F)	81	91	85	85	92	93	90	89	80	90	91	92	80	80	80
7	9956819174	Aryansyah (G)	82	92	90	91	93	96	91	90	80	96	86	93	80	96	80
8	9972100774	Budi Irawan (H)	80	92	88	91	92	90	87	90	80	90	86	90	84	81	83
9	9986175197	Cici Adniyaty (I)	81	95	96	87	95	95	91	96	80	96	90	93	83	86	84
10	9980641790	Debby Nadillah (J)	96	94	96	95	95	96	95	96	95	96	95	95	85	95	95
11	9987371019	Dermilan Harahap (K)	91	93	96	96	94	96	95	90	84	96	92	92	86	96	90
12	9000137499	Halawiyah (L)	91	96	94	95	93	96	95	95	85	96	91	93	86	92	86
13	9986294826	Hari Topan Basroh (M)	84	87	90	87	88	90	90	90	80	85	85	90	80	89	81
14	9000137502	Irpan Pohan (N)	82	85	87	85	85	90	85	90	80	85	85	88	80	82	80
15	9000137505	M. Fakhriza (O)	80	83	85	85	86	89	85	91	80	85	86	85	81	80	82

STMIK Triguna Dharma

In using the Single Linkage Technique for grouping students' scientific fields, the results obtained from the stages carried out. The following are the results of the Single Linkage Technique stages.

1. Average of each variable
   The average result of each variable is obtained through equation 6 with the following stages of completion.

$$\bar{v1} = \frac{V1_1 + V_2 + V_3 + V_4 + V_5 + V_6 + V_7 + V_8 + V_9 + V_10 + V_11 + V_12 + V_13 + V_14 + V_15}{N}$$

$$\bar{v1} = \frac{80 + 82 + 81 + 81 + 82 + 81 + 82 + 80 + 81 + 96 + 91 + 91 + 84 + 82 + 8}{15}$$

$$\bar{v1} = \frac{1245}{15} = 83,6$$

2. Variable Standard Deviation Value
   The results of the standard deviation of the variables are obtained through equation 2 with the following stages of completion.

$$std(\bar{v}1) = \frac{\sqrt{\begin{array}{l}(80-83,6)^2 + (82-83,6)^2 + (81-83,6)^2 + (81-83,6)^2 + (82-83,6)^2 + \\ (81-83,6)^2 + (82-83,6)^2 + (80-83,6)^2 + (81-83,6)^2 + (96-83,6)^2 + \\ (91-83,6)^2 + (91-83,6)^2 + (84-83,6)^2 + (82-83,6)^2 + (80-83,6)^2\end{array}}}{15-1}$$

3. Standard Zero Value (Zi) of each Object

The result of the standard zero value (zi) for each object is obtained through equation 3 with the completion stages.

$$
\begin{aligned}
Z(V1_1) &= \frac{v1_1 - \bar{v}1}{std(v1)} = \frac{80 - 83,6}{4,925} = -0,731 \\
Z(V1_2) &= \frac{v1_2 - \bar{v}1}{std(v1)} = \frac{82 - 83,6}{4,925} = -0,325 \\
Z(V1_3) &= \frac{v1_3 - \bar{v}1}{std(v1)} = \frac{81 - 83,6}{4,925} = -0,528 \\
Z(V1_4) &= \frac{v1_4 - \bar{v}1}{std(v1)} = \frac{81 - 83,6}{4,925} = -0,528 \\
Z(V1_5) &= \frac{v1_5 - \bar{v}1}{std(v1)} = \frac{82 - 83,6}{4,925} = -0,325 \\
Z(V1_6) &= \frac{v1_6 - \bar{v}1}{std(v1)} = \frac{81 - 83,6}{4,925} = -0,528 \\
Z(V1_7) &= \frac{v1_7 - \bar{v}1}{std(v1)} = \frac{82 - 83,6}{4,925} = -0,325 \\
Z(V1_8) &= \frac{v1_8 - \bar{v}1}{std(v1)} = \frac{80 - 83,6}{4,925} = -0,731 \\
Z(V1_9) &= \frac{v1_9 - \bar{v}1}{std(v1)} = \frac{81 - 83,6}{4,925} = -0,528 \\
Z(V1_{10}) &= \frac{v1_{10} - \bar{v}1}{std(v1)} = \frac{96 - 83,6}{4,925} = -0,518
\end{aligned}
$$

$$
\begin{aligned}
Z(V1_{11}) &= \frac{v1_{11} - \bar{v}1}{std(v1)} = \frac{91 - 83,6}{4,925} = -0,502 \\
Z(V1_{12}) &= \frac{v1_{12} - \bar{v}1}{std(v1)} = \frac{91 - 83,6}{4,925} = -0,502 \\
Z(V1_{13}) &= \frac{v1_{13} - \bar{v}1}{std(v1)} = \frac{84 - 83,6}{4,925} = -0,081 \\
Z(V1_{14}) &= \frac{v1_{14} - \bar{v}1}{std(v1)} = \frac{82 - 83,6}{4,925} = -0,325 \\
Z(V1_{15}) &= \frac{v1_{15} - \bar{v}1}{std(v1)} = \frac{80 - 83,6}{4,925} = -0,731
\end{aligned}
$$

This process is carried out for all variables and data objects in table 2 which is the overall result of the default value of zero (zi) for each object.

Table 2. Overall Standard Zero Score

NO	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12	V13	V14	V15
A	-0,731	-0,418	-0,150	-0,449	-0,274	-1,179	0,072	-1,746	-0,397	-1,060	-0,160	-1,616	-0,083	-1,065	0,360
B	-0,325	-0,627	-0,400	-0,449	0,313	0,489	-0,469	0,081	-0,397	-1,060	0,182	0,323	-0,918	1,240	-0,110
C	-0,528	-2,092	-0,901	-0,449	-1,739	-1,179	-1,011	-0,223	-0,397	1,099	-0,843	-0,323	-0,918	-0,758	-0,814
D	-0,528	1,046	-0,651	-0,449	0,020	-0,845	0,885	-1,746	-0,397	-1,060	-0,160	-0,970	-0,501	-0,912	-0,814
E	-0,325	0,418	-0,651	-0,708	0,606	0,823	1,426	0,690	-0,397	-0,078	0,182	-0,970	1,168	0,318	-0,110
F	-0,528	0,209	-1,151	-0,966	0,313	0,156	-0,199	-0,528	-0,397	-0,078	0,866	0,646	-0,918	-1,065	-0,814
G	-0,325	0,418	0,100	0,587	0,606	1,157	0,072	-0,223	-0,397	1,099	-0,843	0,970	-0,918	1,393	-0,814
H	-0,731	0,418	-0,400	0,587	0,313	-0,845	-1,011	-0,223	-0,397	-0,078	-0,843	0,000	0,751	-0,912	-0,110
I	-0,528	1,046	1,601	-0,449	1,192	0,823	0,072	1,604	-0,397	1,099	0,524	0,970	0,334	-0,143	0,125
J	2,518	0,837	1,601	1,622	1,192	1,157	1,155	1,604	3,323	1,099	2,234	1,616	1,168	1,240	2,709
K	1,502	0,627	1,601	1,881	0,899	1,157	1,155	-0,223	0,595	1,099	1,208	0,646	1,586	1,393	1,535
L	1,502	1,255	1,101	1,622	0,606	1,157	1,155	1,299	0,843	1,099	0,866	0,970	1,586	0,779	0,595
M	0,081	-0,627	0,100	-0,449	-0,860	-0,845	-0,199	-0,223	-0,397	-1,060	-1,185	0,000	-0,918	0,318	-0,579
N	-0,325	-1,046	-0,651	-0,966	-1,739	-0,845	-1,552	-0,223	-0,397	-1,060	-1,185	-0,646	-0,918	-0,758	-0,814
O	-0,731	-1,464	-1,151	-0,966	-1,446	-1,179	-1,552	0,081	-0,397	-1,060	-0,843	-1,616	-0,501	-1,065	-0,345

4. Distance Measurement Value
   The results of the distance measurement values are obtained through equation 4 with the following stages of completion.

$$d_{(A,B)} = \sqrt{\begin{array}{l}(v1_1 - v1_2)^2 + (v2_1 - v2_2)^2 + (v3_1 - v3_2)^2 + (v4_1 - v4_2)^2 + (v5_1 - v5_2)^2 + \\ (v6_1 - v6_2)^2 + (v7_1 - v7_2)^2 + (v8_1 - v8_2)^2 + (v9_1 - v9_2)^2 + (v10_1 - v10_2)^2 + \\ (v11_1 - v11_2)^2 + (v12_1 - v12_2)^2 + (v13_1 - v13_2)^2 + (v14_1 - v14_2)^2 + (v15_1 - v15_2)^2\end{array}}$$

$$d_{(A,B)} = \sqrt{\begin{array}{l}(-0,731 - (-0325))^2 + (-0,418 - (-0,627))^2 + (-0,150 - (-0,400))^2 + \\ (-0,449 - (-0,449))^2 + (-0,274 - 0,313)^2 + (-1,179 - 0,489)^2 + (0,072 - (-0,469))^2 + \\ (-1,746 - 0,081)^2 + (-0,397 - (-0,397))^2 + (-1,060 - (-1,060))^2 + (-0,160 - 0,182)^2 + \\ (-616 - 0,323)^2 + (-0,083 - (-0,918)^2 + (-1,065 - 1,240)^2 + (0,360 - (-0,110)^2)\end{array}}$$

$$= 4,140$$

The same steps are carried out on all student data, so that at the end of the calculation the matrix in table 3 is obtained.

Table 3. Distance Measurement Matrix

No	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
A	0,000	4,140	4,244	2,299	4,351	3,821	5,521	3,281	5,960	9,304	6,910	7,106	3,282	3,530	3,295
B	4,140	0,000	4,520	4,127	3,636	3,078	3,160	3,746	4,367	7,834	5,802	5,669	2,605	3,881	4,371
C	4,244	4,520	0,000	4,976	5,623	4,209	5,181	4,097	6,292	9,915	7,929	7,638	3,340	2,598	2,839
D	2,299	4,127	4,976	0,000	3,966	3,129	4,876	3,464	5,672	9,279	6,876	6,740	3,425	4,170	4,419
E	4,351	3,636	5,623	3,966	0,000	3,860	4,150	3,902	3,849	7,419	5,131	4,596	4,318	5,399	5,355
F	3,821	3,078	4,209	3,129	3,860	0,000	3,972	3,386	4,392	8,498	6,601	6,159	3,596	3,993	4,395
G	5,521	3,160	5,181	4,876	4,150	3,972	0,000	4,094	3,773	7,547	5,065	4,764	3,936	5,411	6,154
H	3,281	3,746	4,097	3,464	3,902	3,386	4,094	0,000	4,313	8,287	5,956	5,564	3,272	3,754	3,915
I	5,960	4,367	6,292	5,672	3,849	4,392	3,773	4,313	0,000	6,431	4,680	3,836	5,160	6,335	6,715
J	9,304	7,834	9,915	9,279	7,419	8,498	7,547	8,287	6,431	0,000	3,947	3,897	8,766	10,136	10,394
K	6,910	5,802	7,929	6,876	5,131	6,601	5,065	5,956	4,680	3,947	0,000	2,159	6,625	8,190	8,505
L	7,106	5,669	7,638	6,740	4,596	6,159	4,764	5,564	3,836	3,897	2,159	0,000	6,354	7,823	8,190
M	3,282	2,605	3,34	3,425	4,318	3,596	3,936	3,272	5,160	8,766	6,625	6,354	0,000	2,325	3,232
N	**3,530**	**3,881**	**2,598**	**4,170**	**5,399**	**3,993**	**5,411**	**3,754**	**6,335**	**10,136**	**8,190**	**7,823**	**2,325**	**0,000**	**1,558**
O	**3,295**	**4,371**	**2,839**	**4,419**	**5,355**	**4,395**	**6,154**	**3,915**	**6,715**	**10,394**	**8,505**	**8,190**	**3,232**	**1,558**	**0,000**

5. Euclidean Single Linkage Grouping
The results of the single linkage Euclidean grouping are obtained through equation 4. N and O have the smallest value, which is 1.558, then the N and O objects are combined into one cluster, so that the distance between the NO clusters and other objects will be calculated. Following are the steps for calculating the Euclidean single linkage:

$$
\begin{aligned}
d(NO)A &= mindNA, dOA = dOA = 3,295 \\
d(NO)B &= mindNB, dOB = dNB = 3,881 \\
d(NO)C &= mindNC, dOC = dNC = 2,598 \\
d(NO)D &= mindND, dOD = dND = 4,170 \\
d(NO)E &= mindNE, dOE = dOE = 5,355 \\
d(NO)F &= mindNF, dOF = dNF = 3,993 \\
d(NO)G &= mindNG, dOG = dNG = 5,411 \\
d(NO)H &= mindNH, dOH = dNH = 3,754 \\
d(NO)I &= mindNI, dOI = dNI = 6,335 \\
d(NO)J &= mindNJ, dOJ = dNJ = 10,136 \\
d(NO)K &= mindNK, dOK = dNK = 8,190 \\
d(NO)L &= mindNL, dOL = dNL = 7,823 \\
d(NO)M &= mindNM, dOM = dNM = 2,325
\end{aligned}
$$

Thus, a new first derivative distance matrix is formed which becomes a cluster between N and O in table 4.

Table 4. First Derivative Distance Matrix

	NO	A	B	C	D	E	F	G	H	I	J	K	L	M
NO	0,000	3,295	3,881	2,598	4,170	5,355	3,993	5,411	3,754	6,335	10,136	8,190	7,823	2,325
A	3,295	0,000	4,140	4,244	2,299	4,351	3,821	5,521	3,281	5,960	9,304	6,910	7,106	3,282
B	3,881	4,140	0,000	4,520	4,127	3,636	3,078	3,160	3,746	4,367	7,834	5,802	5,669	2,605
C	2,598	4,244	4,520	0,000	4,976	5,623	4,209	5,181	4,097	6,292	9,915	7,929	7,638	3,340
D	4,170	2,299	4,127	4,976	0,000	3,966	3,129	4,876	3,464	5,672	9,279	6,876	6,740	3,425
E	5,355	4,351	3,636	5,623	3,966	0,000	3,860	4,150	3,902	3,849	7,419	5,131	4,596	4,318
F	3,993	3,821	3,078	4,209	3,129	3,860	0,000	3,972	3,386	4,392	8,498	6,601	6,159	3,596
G	5,411	5,521	3,160	5,181	4,876	4,150	3,972	0,000	4,094	3,773	7,547	5,065	4,764	3,936
H	3,754	3,281	3,746	4,097	3,464	3,902	3,386	4,094	0,000	4,313	8,287	5,956	5,564	3,272
I	6,335	5,960	4,367	6,292	5,672	3,849	4,392	3,773	4,313	0,000	6,431	4,680	3,836	5,160
J	10,136	9,304	7,834	9,915	9,279	7,419	8,498	7,547	8,287	6,431	0,000	3,947	3,897	8,766
K	8,190	6,910	5,802	7,929	6,876	5,131	6,601	5,065	5,956	4,680	3,947	0,000	2,159	6,625
L	7,823	7,106	5,669	7,638	6,740	4,596	6,159	4,764	5,564	3,836	3,897	2,159	0,000	6,354
M	2,325	3,282	2,605	3,340	3,425	4,318	3,596	3,936	3,272	5,160	8,766	6,625	6,354	0,000

Furthermore, the calculation of the Euclidean single linkage is still carried out until all of its derivative clusters are met or completed. The following is the result of the derived distance matrix obtained:

Table 5. Pembagian data untuk Training dan Testing

	NO	KL	A	B	C	D	E	F	G	H	I	J	M
NO	0,000	7,823	6,910	5,669	7,638	6,740	4,596	6,159	4,764	5,564	3,836	3,897	6,354
KL	7,823	0,000	3,295	3,881	2,598	4,170	5,355	3,993	5,411	3,754	6,335	10,136	2,325
A	6,910	3,295	0,000	4,140	4,244	2,299	4,351	3,821	5,521	3,281	5,960	9,304	3,282
B	5,669	3,881	4,140	0,000	4,520	4,127	3,636	3,078	3,160	3,746	4,367	7,834	2,605
C	7,638	2,598	4,244	4,520	0,000	4,976	5,623	4,209	5,181	4,097	6,292	9,915	3,340
D	6,740	4,170	2,299	4,127	4,976	0,000	3,966	3,129	4,876	3,464	5,672	9,279	3,425
E	4,596	5,355	4,351	3,636	5,623	3,966	0,000	3,860	4,150	3,902	3,849	7,419	4,318
F	6,159	3,993	3,821	3,078	4,209	3,129	3,860	0,000	3,972	3,386	4,392	8,498	3,596
G	4,764	5,411	5,521	3,160	5,181	4,876	4,150	3,972	0,000	4,094	3,773	7,547	3,936
H	5,564	3,754	3,281	3,746	4,097	3,464	3,902	3,386	4,094	0,000	4,313	8,287	3,272
I	3,836	6,335	5,960	4,367	6,292	5,672	3,849	4,392	3,773	4,313	0,000	6,431	5,160
J	3,897	10,136	9,304	7,834	9,915	9,279	7,419	8,498	7,547	8,287	6,431	0,000	8,766
M	6,354	2,325	3,282	2,605	3,340	3,425	4,318	3,596	3,936	3,272	5,160	8,766	0,000

In table 6, the calculation of the Third Derivative Distance Matrix has been formed, from the results of the Single Linkage.

Table 6. Third Derivative Distance Matrix

	AD	KL	NO	B	C	E	F	G	H	I	J	M
AD	0,000	6,740	3,295	4,127	4,244	3,966	3,129	4,876	3,281	5,672	9,279	3,282
KL	6,740	0,000	7,823	5,669	7,638	4,596	6,159	4,764	5,564	3,836	3,897	6,354
NO	3,295	7,823	0,000	3,881	2,598	5,355	3,993	5,411	3,754	6,335	10,136	2,325
B	4,127	5,669	3,881	0,000	4,520	3,636	3,078	3,160	3,746	4,367	7,834	2,605
C	4,244	7,638	2,598	4,520	0,000	5,623	4,209	5,181	4,097	6,292	9,915	3,340
E	3,966	4,596	5,355	3,636	5,623	0,000	3,860	4,150	3,902	3,849	7,419	4,318
F	3,129	6,159	3,993	3,078	4,209	3,860	0,000	3,972	3,386	4,392	8,498	3,596
G	4,876	4,764	5,411	3,160	5,181	4,150	3,972	0,000	4,094	3,773	7,547	3,936
H	3,281	5,564	3,754	3,746	4,097	3,902	3,386	4,094	0,000	4,313	8,287	3,272
I	5,672	3,836	6,335	4,367	6,292	3,849	4,392	3,773	4,313	0,000	6,431	5,160
J	9,279	3,897	10,136	7,834	9,915	7,419	8,498	7,547	8,287	6,431	0,000	8,766
M	3,282	6,354	2,325	2,605	3,340	4,318	3,596	3,936	3,272	5,160	8,766	0,000

In table 7, the calculation of the Fourth Derivative Distance Matrix has been formed from the Single Linkage results, namely the merger between NO clusters and M Clusters.

Table 7. Fourth Derivative Distance Matrix

	NO-M	AD	KL	B	C	E	F	G	H	I	J
NOM	0,000	3,282	6,354	2,605	2,598	4,318	3,596	3,936	3,272	5,160	8,766
AD	3,282	0,000	6,740	4,127	4,244	3,966	3,129	4,876	4,150	5,672	9,279
KL	6,354	6,740	0,000	5,669	7,638	4,596	6,159	4,764	3,972	3,836	3,897
B	2,605	4,127	5,669	0,000	4,520	3,636	3,078	3,160	3,746	4,367	7,834
C	2,598	4,244	7,638	4,520	0,000	5,623	4,209	5,181	4,097	6,292	9,915
E	4,318	3,966	4,596	3,636	5,623	0,000	3,860	4,150	3,902	3,849	7,419
F	3,596	3,129	6,159	3,078	4,209	3,860	0,000	3,972	3,386	4,392	8,498
G	3,936	4,876	4,764	3,160	5,181	4,150	3,972	0,000	4,094	3,773	7,547
H	3,272	3,281	5,564	3,746	4,097	3,902	3,386	4,094	0,000	4,313	8,287
I	5,160	5,672	3,836	4,367	6,292	3,849	4,392	3,773	4,313	0,000	6,431
J	8,766	9,279	3,897	7,834	9,915	7,419	8,498	7,547	8,287	6,431	0,000

In table 8, the calculation of the Fifth Derivative Distance Matrix has been formed from the Single Linkage results, namely the merger between NOM clusters and C Clusters.

Table 8. Fifth Derivative Distance Matrix

	NOM-C	AD	KL	B	E	F	G	H	I	J
NOMC	0,000	3,282	6,354	2,605	4,318	3,596	3,936	3,272	5,160	8,766
AD	3,282	0,000	6,740	4,127	3,966	3,129	4,876	4,150	5,672	9,279
KL	6,354	6,740	0,000	5,669	4,596	6,159	4,764	3,972	3,836	3,897
B	2,605	4,127	5,669	0,000	3,636	3,078	3,160	3,746	4,367	7,834
E	4,318	3,966	4,596	3,636	0,000	3,860	4,150	3,902	3,849	7,419
F	3,596	3,129	6,159	3,078	3,860	0,000	3,972	3,386	4,392	8,498
G	3,936	4,876	4,764	3,160	4,150	3,972	0,000	4,094	3,773	7,547
H	3,272	4,150	3,972	3,746	3,902	3,386	4,094	0,000	4,313	8,287
I	5,160	5,672	3,836	4,367	3,849	4,392	3,773	4,313	0,000	6,431
J	8,766	9,279	3,897	7,834	7,419	8,498	7,547	8,287	6,431	0,000

In table 9, the calculation of the Sixth Derivative Distance Matrix has been formed from the Single Linkage results, namely the merger between NOMC clusters and B Clusters.

Table 9. Sixth Derivative Distance Matrix

	NOMC-B	AD	KL	E	F	G	H	I	J
NOMCB	0,000	3,282	5,669	3,636	3,078	3,160	3,272	4,367	7,834
AD	3,282	0,000	6,740	3,966	3,129	4,876	4,150	5,672	9,279
KL	5,669	6,740	0,000	4,596	6,159	4,764	3,972	3,836	3,897
E	3,636	3,966	4,596	0,000	3,860	4,150	3,902	3,849	7,419
F	3,078	3,129	6,159	3,860	0,000	3,972	3,386	4,392	8,498
G	3,160	4,876	4,764	4,150	3,972	0,000	4,094	3,773	7,547
H	3,272	4,150	3,972	3,902	3,386	4,094	0,000	4,313	8,287
I	4,367	5,672	3,836	3,849	4,392	3,773	4,313	0,000	6,431
J	7,834	9,279	3,897	7,419	8,498	7,547	8,287	6,431	0,000

In table 10, the calculation of the Seventh Derivative Distance Matrix has been formed from the Single Linkage results, namely the merger between NOMCB clusters and F Clusters.

Table 10. Seventh Derivative Distance Matrix

	NOMCB-F	AD	KL	E	G	H	I	J
NOMCBG	0,000	3,129	5,669	3,636	3,160	3,272	4,367	7,834
AD	3,129	0,000	6,740	3,966	4,876	4,150	5,672	9,279
KL	5,669	6,740	0,000	4,596	4,764	3,972	3,836	3,897
E	3,636	3,966	4,596	0,000	4,150	3,902	3,849	7,419
G	3,160	4,876	4,764	4,150	0,000	3,386	4,392	8,498
H	3,272	4,150	3,972	3,902	3,386	0,000	4,313	8,287
I	4,367	5,672	3,836	3,849	4,392	4,313	0,000	6,431
J	7,834	9,279	3,897	7,419	8,498	8,287	6,431	0,000

In table 11, the calculation of the Eighth Derivative Distance Matrix has been formed from the Single Linkage results, namely the merger between NOMCBF clusters and AD Clusters.

Table 11. Eighth Derivative Distance Matrix

	NOMCBF-AD	KL	E	G	H	I	J
NOMCBFAD	0,000	5,669	3,636	3,160	3,272	4,367	7,834
KL	5,669	0,000	4,596	4,764	3,972	3,836	3,897
E	3,636	4,596	0,000	4,150	3,902	3,849	7,419
G	3,160	4,764	4,150	0,000	3,386	4,392	8,498
H	3,272	3,972	3,902	3,386	0,000	4,313	8,287
I	4,367	3,836	3,849	4,392	4,313	0,000	6,431
J	7,834	3,897	7,419	8,498	8,287	6,431	0,000

In table 12, the calculation of the Ninth Derivative Distance Matrix has been formed from the Single Linkage results, namely the merger between NOMCBFAD clusters and G Clusters.

Table 12. Ninth Derivative Distance Matrix

	NOMCBFAD-G	KL	E	H	I	J
NOMCB FADG	0,000	4,764	3,636	3,272	4,392	7,834
KL	4,764	0,000	4,596	3,972	3,836	3,897
E	3,636	4,596	0,000	3,902	3,849	7,419
H	3,272	3,972	3,902	0,000	4,313	8,287
I	4,392	3,836	3,849	4,313	0,000	6,431
J	7,834	3,897	7,419	8,287	6,431	0,000

In table 13, the calculation of the Tenth Derivative Distance Matrix has been formed from the Single Linkage results, namely the merger between NOMCBFADG clusters and H Clusters.

Table 13. Tenth Derivative Distance Matrix

	NOMCBFADG-H	KL	E	I	J
NOMCBF ADGH	0,000	3,972	3,636	4,313	7,834
KL	3,972	0,000	4,596	3,836	3,897
E	3,636	4,596	0,000	3,849	7,419
I	4,313	3,836	3,849	0,000	6,431
J	7,834	3,897	7,419	6,431	0,000

In table 14, the calculation of the Eleventh Derivative Distance Matrix has been formed from the Single Linkage results, namely the merger between NOMCBFADGH clusters and E Clusters.

Table 14. Eleventh Derived Distance Matrix

	NOMCBFADGH-E	KL	I	J
NOMCBFADGHE	0,000	3,972	3,849	7,419
KL	3,972	0,000	3,836	3,897
I	3,849	3,836	0,000	6,431
J	7,419	3,897	6,431	0,000

In table 15, the calculation of the Twelfth Derivative Distance Matrix has been formed from the Single Linkage results, namely the merger between KL clusters and I Clusters.

Table 15. Twelfth Derivative Distance Matrix

	KL-I	NOMCBFADGHE	J
KLI	0,000	3,849	3,897
NOMCBFADGHE	3,849	0,000	7,419
J	3,897	7,419	0,000

In table 16, the calculation of the Thirteenth Derivative Distance Matrix has been formed from the Single Linkage results, namely the merger between KLI clusters and NOMCBFADGHE Clusters.

Table 16. Thirteenth Derivative Distance Matrix

	KLI-NOMCBFADGHE	J
KLINOMCBFADGHE-J	0	3,897
	3,897	0

In table 17, the calculation of the Fourteenth Derivative Distance Matrix has been formed from the Single Linkage results, namely the merger between KLINOMCBFADGHEclusters and J Clusters.

Table 17. Fourteenth Derivative Distance Matrix

0	KLINOMCBFADGHE-J
KLINOMCBFADGHE-J	0

From the Euclidean single linkage calculation process, 14 derived clusters are obtained so that the resulting cluster data.

Table 18. Cluster Results

Nomor Cluster	Nirm	Nama
1	9980641790	Debby Nadillah Harahap
	9987371019	Dermilan Harahap
	9000137499	Halawiyah Primadeny Yanuary
	9986175197	Cici Adniyaty
2	9000137502	Irpan Pohan
	9000137505	Muhammad Fakhriza Ananda Hasibuan
	9986294826	Hari Topan Basroh
	9986294859	Ahmad Fauzi Tanjung
	9950021406	Ahmad Afrizal
	9000137537	Ayu Andira
	9983689340	Afdillah
	9986050887	Anggi Arafah
	9956819174	Bonar Aryansyah
	9972100774	Budi Irawan
	9990261330	Audriyana

4. CONCLUSION

Based on the application of the Single Linkage Technique in solving problems regarding the grouping of student scientific fields that have been carried out, the results obtained are grouping 2 clusters out of 14 derived clusters. The first cluster contains four students categorized as a group of students with good competence. In comparison, the second cluster contains 12 students categorized as a group of less competent students. From the cluster results, it can be seen that the student specialization class will be divided into 2 class groups, and group 2 is the class that will get priority learning concentration because this group has poor grades compared to group 1. Hence, it needs to be more intensive and needs assistance. The data from the Cluster calculation results have been validated with the system created and have the same results as manual calculations. As for suggestions for future researchers, it can combine several of its algorithms to produce a more accurate analysis in providing recommendations for researchers.

5. ACKNOWLEDGEMENTS

As for this research, this is one of the most popular computer majors on our campus. Therefore, the researcher conducted this research and collaborated with several lecturers in the field of computers in the fields of data mining and data science in developing the lat research on the problems found.

6. DECLARATIONS

AUTHOR CONTIBUTION

In this study, the first and second authors have ideas about the problems in this journal, and the two authors collect data and test their theories. The third author is a supervisor in this research and an expert in data mining, and the fourth author is a party conducting testing and testing data. The fifth author is a translator and vocabulary improvement in this study

FUNDING STATEMENT

The Funding for this Research is a Party from an Institution that fully supports this Research activity in improving the quality of teaching staff in developing the world of education in the future

COMPETING INTEREST

the authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

REFERENCES

[1] A. O. Oyedele, A. O. Ajayi, and L. O. Oyedele, “Machine Learning with Applications Machine Learning Predictions for Lost Time Injuries in Power Transmission and Distribution Projects,” *Machine Learning with Applications*, vol. 6, no. September, pp. 1–12, 2021.

[2] U. Maheshwera, R. Paturi, S. Cheruku, V. Phani, K. Pasunuri, S. Salike, N. S. Reddy, and S. Cheruku, “Machine Learning with Applications Machine Learning and Statistical Approach in Modeling and Optimization of Surface Roughness in Wire Electrical Discharge Machining,” *Machine Learning with Applications*, vol. 6, no. May, pp. 1–11, 2021.

[3] M. Aprilliyani and M. Y. Putra, “Algoritma Apriori untukMenentukan Pola Penjualan pada 212 Mart Cibitung,” *Information System for Educators and Professionals*, vol. 3, no. 2, pp. 215–226, 2019.

[4] S. Sunarti, F. Handayanna, and E. Irfiani, “Analisa Pola Penjualan Makanan Dengan Penerapan Algoritma Apriori,” *Techno.Com*, vol. 20, no. 4, pp. 478–488, 2021.

[5] S. Saefudin and S. DN, “Penerapan Data Mining dengan Metode Algoritma Apriori untuk Menentukan Pola Pembelian Ikan,” *JSiI (Jurnal Sistem Informasi)*, vol. 6, no. 2, pp. 110–114, 2019.

[6] M. R. Bachute, P. D. Electronics, I. L. Officer, and M. S. M. S. A. Electronics, “Machine Learning with Applications Autonomous Driving Architectures : Insights of Machine Learning and Deep Learning Algorithms,” *Machine Learning with Applications*, vol. 6, no. March, pp. 1–25, 2021.

[7] J. Chai, H. Zeng, A. Li, and E. W. T. Ngai, “Machine Learning with Applications Deep Learning in Computer Vision : A Critical Review of Emerging Techniques and Application Scenarios,” *Machine Learning with Applications*, vol. 6, no. March, pp. 1–13, 2021.

[8] L. Ganda, R. Putra, and A. Anggrawan, “Pengelompokan Penerima Bantuan Sosial Masyarakat dengan Metode K-Means,” *Matrik: Jurnal Manajemen, Teknik Informatika, dan Rekayasa Komputer*, vol. 21, no. 1, pp. 205–214, 2021.

[9] T. R. P. Bishop, S. V. Hinke, B. Hollingsworth, A. A. Lake, H. Brown, and T. Burgoine, “Machine Learning with Applications Automatic classification of Takeaway Food Outlet Cuisine Type using Machine (Deep) Learning,” *Machine Learning with Applications*, vol. 6, no. July, pp. 1–11, 2021.

[10] L. A. Bugnon, J. Raad, G. A. Merino, C. Yones, F. Ariel, D. H. Milone, and G. Stegmayer, "Machine Learning with Applications Deep Learning for The Discovery of New Pre-MiRNAs : Helping The Fight Against COVID-19," *Machine Learning with Applications*, vol. 6, no. March, pp. 1–8, 2021.

[11] I. F. Yuliati and P. R. Sihombing, "Penerapan Metode Machine Learning dalam Klasifikasi Risiko Kejadian Berat Badan Lahir Rendah di Indonesia Implementation of Machine Learning Method in Risk Classification on Low Birth weight in Indonesia," vol. 20, no. 2, pp. 417–426, 2021.

[12] S. Laqtib, K. E. Yassini, and M. L. Hasnaoui, "A Technical Review and Comparative Analysis of Machine Learning Techniques for Intrusion Detection Systems in MANET," *International Journal of Electrical and Computing Engineering (IJECE)*, vol. 10, no. 3, pp. 2701–2709, 2020.

[13] T. R. S. Mary and S. Sebastian, "Predicting Heart Ailment in Patients with Varying Number of Features Using Data Mining Techniques," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 9, no. 4, pp. 2675–2681, 2019.

[14] P. Subarkah, E. P. Pambudi, S. Oktaviani, and N. Hidayah, "Perbandingan Metode Klasifikasi Data Mining untuk Nasabah Bank Telemarketing," *MATRIK : Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer*, vol. 20, no. 1, pp. 139–148, 2020.

[15] R. R. Pratama, "Analisis Model Machine Learning Terhadap Pengenalan Aktifitas Manusia," *MATRIK : Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer*, vol. 19, no. 2, pp. 302–311, 2020.

[16] A. R. Riszky and M. Sadikin, "Data Mining Menggunakan Algoritma Apriori untuk Rekomendasi Produk bagi Pelanggan Data Mining using Apriori Algorithm for Product Recommendation for Customers," *Jurnal Teknologi dan Sistem Komputer*, vol. 7, no. 3, pp. 103–108, 2019.

[17] E. N. Osegi and E. F. Jumbo, "Machine Learning with Applications Comparative Analysis of Credit Card Fraud Detection in Simulated Annealing trained Artificial Neural Network and Hierarchical Temporal Memory," *Machine Learning with Applications*, vol. 6, no. June, pp. 1–10, 2021.

[18] K. Giri, T. Kr, and P. Sarkar, "Machine Learning with Applications ECR-DBSCAN : An Improved DBSCAN Based on Computational Geometry," *Machine Learning with Applications*, vol. 6, no. July, p. 100148, 2021.

[19] E. Nsugbe, O. Obajemu, O. William, and I. Sanusi, "Machine Learning with Applications Enhancing Care Strategies for Preterm Pregnancies by using a Prediction Machine to Aid Clinical Care Decisions," *Machine Learning with Applications*, vol. 6, no. July, pp. 1–11, 2021.

[20] M. Kempa and Y. Peng, "Machine Learning with Applications Machine Learning Algorithms for Fraud Prediction in Property Insurance : Empirical Evidence using Real-World Microdata," *Machine Learning with Applications*, vol. 5, no. July, pp. 1–14, 2021.

[21] P. C. Pendharkar, "Machine Learning with Applications Hybrid Radial Basis Function DEA and its Applications to Regression , Segmentation and Cluster Analysis Problems," *Machine Learning with Applications*, vol. 6, no. June, pp. 1–9, 2021.