

Application of Data Mining Using the K-Means Clustering Algorithm for Opening Industrial Classes in Vocational High Schools

¹Aan Rosydiana, ²Dian Sediana, ³Christina Juliane

^{1,2,3}Study Program of Information Systems, College of Informatics and Computer Management LIKMI

Email: ¹aanrosydiana@gmail.com, ²diansediana@gmail.com ³christina.juliane@likmi.ac.id

Article Info

Article history:

Received Aug 20th, 2022

Revised Sep 28th, 2022

Accepted Oct 20th, 2022

Keyword:

Clustering

Data Mining

Industrial Class

K-Means Algorithm

Vocational High Schools

ABSTRACT

Vocational High School has a goal to enter the world of work, meaning that it must have a skill program to be relevant to the industrial world. However, adapting to the industrial world is difficult, one of the things that is happening between industries is increasing. Various efforts continue to be made, among others, by establishing an industrial class, the formation of an industrial class is expected to produce students who have competencies in accordance with the standards required by the collaborating industries. The formation of an industrial class can be done by applying data mining methods, in order to form the right industrial class and in accordance with predetermined criteria. This study aims to classify new student registration data at State Vocational Schools at the Regional Education Office XIII Branch of West Java Province in 2022 and the results of the grouping are used to form industrial classes. The clustering process is carried out using the K-Means algorithm and cluster analysis is carried out with the help of RapidMiner software. The results showed that the data clustering was formed into 4 clusters. The cluster that has the highest number is cluster 1 and the cluster that has the lowest number is cluster 0. There are variables used for data grouping, including school variables and expertise programs, from these variables it is obtained that the schools selected by students are based on the highest order and have expertise programs contained in their clusters, which need to be considered when opening industrial classes.

Copyright © 2022 Puzzle Research Data Technology

Corresponding Author:

Aan Rosydiana,

Study Program of Information Systems,

College of Informatics and Computer Management LIKMI,

Jl. Ir. H. Juanda No. 96, Lebakgede, Coblong District, Bandung City, West Java, Indonesia.

Email: aanrosydiana@gmail.com

DOI: <http://dx.doi.org/10.24014/ijaidm.v2i2.19172>

1. INTRODUCTION

In the learning process at Vocational High Schools students are prepared to have competencies or skills that match the needs in the world of work or industry. This is so that after graduation students can immediately work and get the right job. However, the reality on the ground shows that the current existence of Vocational High Schools is considered to still lack a workforce that is ready to work for its graduates [1]. Various efforts continue to be made so that this can be realized, one of which is by involving the industrial world in the learning process in accordance with the majors in the Vocational High School or better known as the industrial class. The school's relationship with industry is very large for the sake of coaching, moral support, materials, and the use of the industrial environment as a learning resource [2].

Industrial class is a learning activity which is a collaboration between school and industry [2]. The formation of industrial classes in Vocational High Schools is expected to produce students who have competencies in accordance with the standards required by industry who understand [3]. State Vocational

Schools that have opened industrial classes at the Regional Education Office XIII Branch of West Java Province are 1 State Vocational School and 1 concentration of expertise. The number of industrial classes opened is still small and does not match the number of State Vocational Schools owned, namely 18 with a total of 93 concentration of expertise. The reason is, the school does not conduct an analysis of students who have the potential to enter the industrial class, so there are doubts that the industrial class will be opened.

Data processing of students who apply to State Vocational High Schools needs to be done, to find out important information about the number of prospective students who have the potential to enter the industrial class. Utilization of data mining to analyze student data can produce new discoveries [4] to determine patterns or useful information based on selected data using certain methods [5]. Data mining is also a technique that combines data analysis techniques and finds patterns in data [6]. The pattern obtained can be used for a knowledge base that supports decision making [7]. Decision making by grouping students quickly and precisely, results in quality and more optimal learning [8].

There is a clustering method in data mining that can be used to group data [4] for new students for the 2022/2023 academic year. Clustering is an activity that aims to group data located between data and other data into clusters or groups of data in clusters that have a maximum level, and data between clusters has a minimum level [9]. The clustering process is carried out using the K-Means algorithm, the advantages of applying the K-Means algorithm are that it can group large objects [7] and the performance of the clustering process is good for large data sets [8]. The K-Means algorithm aims to divide the data into several groups [7].

Previous research related to the application of the K-Means algorithm, was able to determine the grouping of superior class students by inputting the centroid value and looking for the closest value [8]. The clustering method with the K-Means algorithm can determine the grouping of student achievement consisting of high, medium and sufficient. The results showed that the accuracy of the data on the problems that occurred was 70% so that it could recognize student achievement data in 10 data samples [10]. The clustering method can make it easier for schools to group student achievements quickly and accurately using the K-Means algorithm [11].

Various previous studies have discussed the clustering method with the K-Means algorithm. However, no one has discussed clustering for opening industrial classes in Vocational High Schools with the K-Means algorithm. This study aims to classify the registration data for new student admissions for the 2022/2023 academic year at the State Vocational High School Branch of the Regional XIII Education Office of West Java Province. The results of the research are expected to help make decisions and be able to analyze the opening of the industrial class, so that there is no doubt at the time of its opening.

2. RESEARCH METHOD

2.1 Research Stages

In carrying out the research there are stages that are carried out so that the research is in accordance with the objectives that have been determined. The grouping is carried out on the registration data of new students for the 2022/2023 academic year at the State Vocational High School Branch of the Regional XIII Education Office of West Java Province. The stages in this study are shown in Figure 1.

The source of data collection in this study was from the West Java Province new student admissions committee for the 2022/2023 academic year. The data used is the registrant data for the State Vocational High School Branch of the Regional XIII Education Office which includes Ciamis Regency, Banjar City and Pangandaran Regency. The number of registrants is 2,750 and some samples of student data can be seen in Table 1.

From the student data, 4 variables were determined based on the need for data analysis, each variable was defined by (X1) and so on until the last one. The variables (attributes) used in this study include: (1) Expertise Program; (2) Gender; (3) Average Report Score; (4) School Choices. Data cleaning is carried out if the information is not clear and incomplete and complete information data is selected for each attribute. The next process is to enter the selected data into RapidMiner for the data clustering process.

2.2 Data Mining

Data mining is a term used to describe the discovery of knowledge in databases [6], then used from various perspectives and concluded as important information [5] which is used for decision-making purposes [7]. Data mining can be defined as the process of extracting unknown data into useful information [8] and related knowledge from various large databases [9]. Data mining is also the process of finding useful new correlations, patterns and trends [10] extracted from within data chunks by managing large amounts of data stores [8].

Data mining is a science that studies the collection, maintenance, processing, analysis, and obtaining useful knowledge from data [4]. Data mining is the process of finding interesting patterns and knowledge

from big data [12]. Big data is a large amount of data that cannot be used with ordinary traditional tools and must use new ways and tools to extract value from data [13].

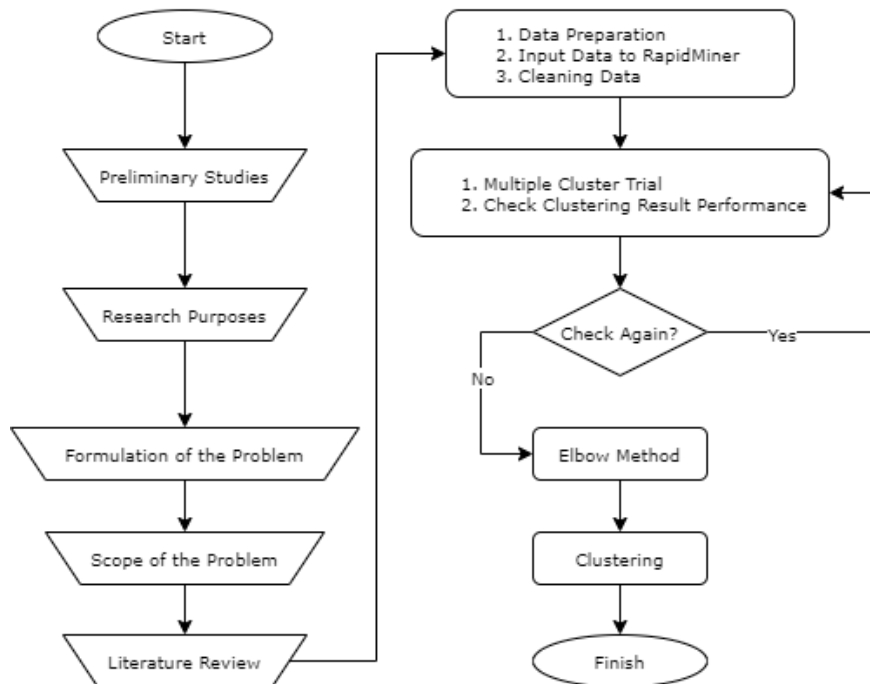


Figure 1. Research Stages

Table 1. Sample of Student Data (Source: New Student Admissions for West Java Province in 2022)

No	Expertise Program	Gender	Report Average Score	School Choice
1.	Teknik Otomotif	Male	85	SMKN 2 Ciamis
2.	Teknik Jaringan Komputer dan Telekomunikasi	Female	86	SMKN 1 Kawali
3.	Manajemen Perkantoran dan Layanan Bisnis	Male	87	SMKN 1 Ciamis
4.	Akuntansi dan Keuangan Lembaga	Female	85	SMKN 1 Panjalu
5.	Desain Komunikasi Visual	Male	80	SMKN 1 Panumbangan
6.	Pengembangan Perangkat Lunak dan Gim	Female	88	SMKN 1 Padaherang
7.	Desain Pemodelan dan Informasi Bangunan	Male	82	SMKN 2 Kalipucang
8.	Teknik Elektronika	Female	83	SMKN 2 Banjar
9.	Agriteknologi Pengolahan Hasil Pertanian	Male	81	SMKN 2 Pangandaran
10.	Teknik Konstruksi dan Perumahan	Female	90	SMKN 2 Ciamis

2.3 Clustering

Clustering is the process of grouping data into certain data groups (clusters) [9], to maximize the amount of data in a cluster and the availability of data items in the entire cluster [7]. Clustering is a method in data mining that functions to group data based on the similarity of data characteristics [9], then perform cluster clusters so that the analysis of large and complex data sets can be reached more clearly [8].

Clustering is a method to overcome the problem of data processing into an information source strategy, to support decision making [14]. Clustering is an important approach to find similarities in data and place similar data into groups [15]. Clustering is also useful for finding distribution patterns in data sets which help in the data analysis process [12].

2.4 K-Means Algorithm

The K-Means algorithm is a clustering algorithm that is classified as unsupervised learning group [12] which is easy to implement and customize, and provides relatively fast performance [11] for grouping

data into groups using a partition system [12]. The K-Means algorithm is a non-hierarchical cluster analysis method that groups data based on its characteristics into one or more clusters [14].

In grouping with the K-Means algorithm, K means the number of desired cluster constants and Means is the average value of a group of data (clusters) [5]. The steps in the K-Means Algorithm are as follows: (1) Prepare a data set; (2) Determine the number of clusters (K); (3) Choose a center point (centroid) randomly; (4) Grouping the data to form K clusters in the middle of each cluster; (5) Update center point value; (6) Repeat steps 3-5 until the centroid no longer changes [6]. The measurement of the distance between the center point (centroid) and the data can use the Euclidean distance. The Euclidean distance formula is as follows:

$$d_{ik} = \sqrt{\sum_{j=1}^m (x_{ij} - c_{kj})^2} \quad (1)$$

Where d_{ik} is the distance between the i -th cluster centers, m is the number of attributes, x_{ij} is the i -th data, c_k is the k -th cluster center data [16].

2.5 RapidMiner

RapidMiner is open-source software, for performing data mining analysis, text mining, and predictive analysis. RapidMiner uses a variety of descriptive and predictive techniques to provide users with insights so that they can make the best decisions [10]. RapidMiner is used for business and commercial, but also used for research, education, training, rapid prototyping. RapidMiner also supports machine learning processes including data preparation, result visualization, validation, and optimization [17].

2.6 Industrial Class

Cooperation between vocational education and industry is important for the alignment or relevance of competencies in Vocational High Schools [2]. Vocational High School as one of the important role holders in preparing the workforce to always be able to follow the development of market needs [1]. Vocational High School aims to improve the ability of students to be able to develop themselves in line with the development of science, technology and art; and prepare students to enter the world of work and develop professional attitudes [18].

Industrial class is prepared as a preparation for students' practice which will later be used to work in the industrial world [2]. Industrial class is a program of collaboration between industry and vocational education units in integrating learning in schools with the industrial world. Industrial class is also a pattern of education carried out in vocational schools that combines the school education system with the existing system in industry, which is relevant to schools and industry [3].

3. RESULTS AND ANALYSIS

Data transformation is done to change the data, so that the data can be grouped using the K-Means Algorithm [5]. The K-Means algorithm is a distance-based clustering method that divides data into several clusters and only works on numeric attributes [12]. For data transformation, it can be done with the following steps: (1) Sort the data based on its occurrence; (2) Initialization of data starts from the highest data with a value of 1, then the next data is 2, 3 and so on [16]. An example of a skill program variable data transformation is shown in Table 2.

Table 2. Transformation of Expertise Program Variables

Expertise Program	Amount	Initialization
Teknik Otomotif	614	1
Teknik Jaringan Komputer dan Telekomunikasi	495	2
Manajemen Perkantoran dan Layanan Bisnis	346	3
Akuntansi dan Keuangan Lembaga	273	4
Desain Komunikasi Visual	166	5
Pengembangan Perangkat Lunak dan Gim	165	6
Kuliner	142	7
Pemasaran	106	8
Desain Pemodelan dan Informasi Bangunan	75	9
Teknik Elektronika	60	10
Busana	48	11
Agriteknologi Pengolahan Hasil Pertanian	44	12
Teknik Mesin	43	13
Broadcasting dan Perfilman	38	14

Expertise Program	Amount	Initialization
Agribisnis Tanaman	24	15
Perhotelan	23	16
Teknik Furnitur	22	17
Nautika Kapal Penangkapan Ikan	17	18
Teknik Geologi Pertambangan	12	19
Usaha Layanan Pariwisata	11	20
Teknik Konstruksi dan Perumahan	11	21
Teknika Kapal Penangkapan Ikan	8	22
Agribisnis Perikanan	4	23
Seni Pertunjukan	2	24
Agribisnis Ternak	1	25

Student data processing can be done after the transformation process, so that student data can be used for grouping data using the K-Means algorithm. The application of K-Means using RapidMiner aims to group data on new student admissions for opening new industrial classes in Vocational High Schools and is shown in Figure 2 for the selection of the K-Means algorithm in RapidMiner.

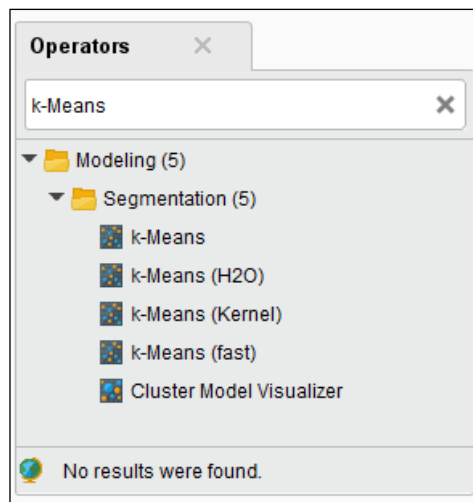


Figure 2. K-Means Operator in RapidMiner

The number of clusters is determined using the Elbow Method, where the most consistent average distance from each cluster center point is 3, after several iterations. The most consistent average distance can be defined as the number of clusters based on the most corner points and is shown in Figure 3 the result of iteration with the Elbow Method.

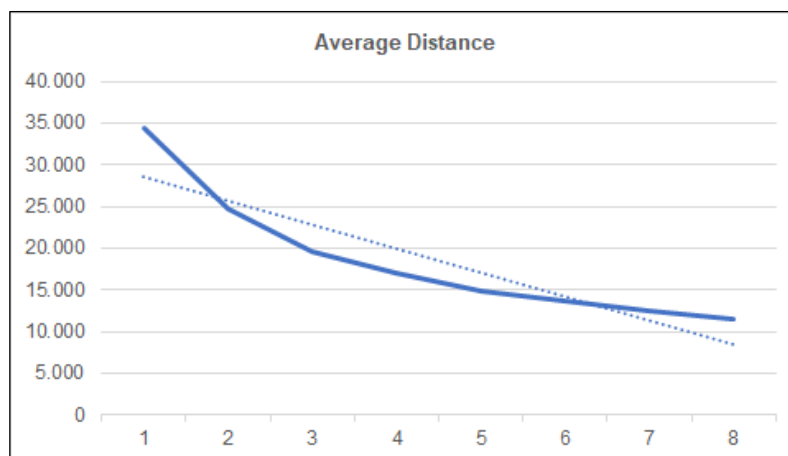


Figure 3. Graph of the Elbow Method

The results of the modeling using the K-Means algorithm as shown in Figure 3 with the number of data are 2,750 and the initialization of the number of clusters is 3, according to the definition of the value of the k parameter. The number of each cluster differs from one cluster to another as shown in Figure 4.

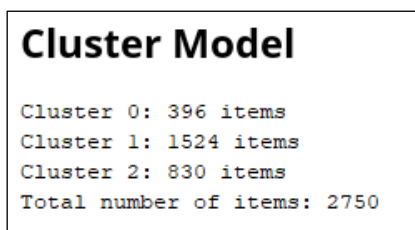


Figure 4. Cluster Model

Based on the cluster model formed, the cluster with the highest number is cluster 1, which is 1524 out of 2,750 data and the cluster with the least number is cluster 0, which is 396 out of 2,750 data. The distribution of each cluster is shown in Figure 5.

Row No. ↑	id	cluster	Expertise Program	Gender	School Choice	Report Average Score
1	1	cluster_1	2	1	4	76
2	2	cluster_1	1	1	4	85
3	3	cluster_1	2	2	4	80
4	4	cluster_1	1	1	4	77
5	5	cluster_1	1	1	4	81
6	6	cluster_1	1	1	4	77
7	7	cluster_1	1	1	4	77
8	8	cluster_1	1	1	4	78
9	9	cluster_0	21	1	4	76
10	10	cluster_1	2	2	4	90
11	11	cluster_1	1	1	4	83
12	12	cluster_0	13	1	4	80
13	13	cluster_0	13	1	4	79
14	14	cluster_0	13	1	4	79
15	15	cluster_0	13	1	4	79

ExampleSet (2,750 examples, 2 special attributes, 4 regular attributes)

Figure 5. Data Distribution of Each Cluster

The K-Means algorithm is used to determine the pattern of industrial class grouping in the State Vocational High School Branch of the Regional XIII Education Office of West Java Province based on the Distance Performance value of the proposed variable. Distance performance in the clustering method can be calculated using equation (2), if the value of each variable has a numeric type. In this method the data calculation is carried out in several stages, namely: (1) Determining the number of clusters, in this process 3 clusters are determined (k value); (2) Determining the centroid value, Table 3 is the centroid value of each independent variable in each data group.

Table 3. Centroid Value of Each Cluster

Cluster	X1	X2	X3	X4
Cluster 0	13.694	1.386	5.869	80.432
Cluster 1	3.223	1.492	3.624	81.940
Cluster 2	3.507	1.381	10.581	80.334

By using the K-Means algorithm, 3 clusters are formed for 4 variables (attributes), namely: (1) Expertise Program (X1); (2) Gender (X2); (3) School Choice (X3); (4) Report Average Score (X4). The grouping of data for each cluster is based on the distance between the center point (centroid) and the data on each attribute and the results of the analysis for each cluster are as follows:

Table 4. Details of Clustering Results

Cluster	Expertise Program	Gender		School Choice	Report Average Score
		Male	Female		
Cluster 0	<ul style="list-style-type: none"> ▪ Pemasaran = 2; ▪ Desain Pemodelan dan 	243	153	<ul style="list-style-type: none"> ▪ SMKN 2 Banjar = 59; ▪ SMKN 1 Ciamis = 18; 	80

Cluster	Expertise Program	Gender		School Choice	Report Average Score
		Male	Female		
	<ul style="list-style-type: none"> Informasi Bangunan = 28; ▪ Teknik Elektronika = 58; ▪ Busana = 48; ▪ Agriteknologi Pengolahan Hasil Pertanian = 44; ▪ Teknik Mesin = 43; ▪ Broadcasting dan Perfilman = 38; ▪ Agribisnis Tanaman = 24; ▪ Perhotelan = 23; ▪ Teknik Furnitur = 22; ▪ Nautika Kapal Penangkapan Ikan = 17; ▪ Teknik Geologi Pertambangan = 12; ▪ Usaha Layanan Pariwisata = 11; ▪ Teknik Konstruksi dan Perumahan = 11; ▪ Teknik Kapal Penangkapan Ikan = 8; ▪ Agribisnis Perikanan = 4; ▪ Seni Pertunjukan = 2; ▪ Agribisnis Ternak = 1; 			<ul style="list-style-type: none"> ▪ SMKN 1 Banjar = 56; ▪ SMKN 2 Ciamis = 64; ▪ SMKN 1 Pangandaran = 50; ▪ SMKN 1 Kawali = 7; ▪ SMKN 1 Padaherang = 51; ▪ SMKN 1 Rajadesa = 36; ▪ SMKN 3 Banjar = 23; ▪ SMKN 1 Cipaku = 25; ▪ SMKN 2 Pangandaran = 7; 	
Cluster 1	<ul style="list-style-type: none"> ▪ Teknik Otomotif = 369; ▪ Teknik Jaringan Komputer dan Telekomunikasi = 330; ▪ Manajemen Perkantoran dan Layanan Bisnis = 300; ▪ Akuntansi dan Keuangan Lembaga = 187; ▪ Desain Komunikasi Visual = 75; ▪ Pengembangan Perangkat Lunak dan Gim = 94; ▪ Kuliner = 79; ▪ Pemasaran = 67; ▪ Desain Pemodelan dan Informasi Bangunan = 23; 	774	750	<ul style="list-style-type: none"> ▪ SMKN 2 Banjar = 272; ▪ SMKN 1 Ciamis = 298; ▪ SMKN 1 Banjar = 232; ▪ SMKN 2 Ciamis = 169; ▪ SMKN 1 Cijulang = 225; ▪ SMKN 1 Pangandaran = 165; ▪ SMKN 1 Kawali = 136; ▪ SMKN 1 Padaherang = 27; 	82
Cluster 2	<ul style="list-style-type: none"> ▪ Teknik Otomotif = 245; ▪ Teknik Jaringan Komputer dan Telekomunikasi = 165; ▪ Manajemen Perkantoran dan Layanan Bisnis = 46; ▪ Akuntansi dan Keuangan Lembaga = 86; ▪ Desain Komunikasi Visual = 91; ▪ Pengembangan Perangkat Lunak dan Gim = 71; ▪ Kuliner = 63; ▪ Pemasaran = 37; ▪ Desain Pemodelan dan Informasi Bangunan = 24; ▪ Teknik Elektronika = 2; 	514	316	<ul style="list-style-type: none"> ▪ SMKN 1 Pangandaran = 5; ▪ SMKN 1 Kawali = 67; ▪ SMKN 1 Padaherang = 95; ▪ SMKN 1 Rajadesa = 121; ▪ SMKN 3 Banjar = 130; ▪ SMKN 4 Banjar = 147; ▪ SMKN 1 Rancah = 103; ▪ SMKN 1 Kalipucang = 82; ▪ SMKN 1 Panumbangan = 41; ▪ SMKN 1 Panjalu = 23; ▪ SMKN 2 Pangandaran = 15; ▪ SMKN 1 Cimerak = 1; 	80

Table 4 is a breakdown of the results of clustering consisting of cluster 0, cluster 1 and cluster 2. Each cluster has proposed variables (attributes), namely: (1) Expertise Program; (2) Gender; (3) Average Report Score; (4) School Choices. From the results of the clustering in Table 4, it can be analyzed that the schools chosen by students based on the highest order and having expertise programs contained in their clusters can consider opening industrial classes. There are variables (attributes) of Gender and the Average Value of the Reports submitted in the clustering process. Cluster 0 consisted of 243 men, 153 women and an average report card score of 80. Cluster 1 consisted of 774 men, 750 women and an average report card score of 82. Cluster 2 consisted of 514 men, 316 women and a score of 80 for the average report card.

From the registration data for new student admissions in 2022, which amounted to 2,750, it has been identified and divided into 3 clusters. The clustering results show that the data mining process produces new information regarding the grouping pattern of industrial class openings at the State Vocational High School Branch of the Regional XIII Education Office of West Java Province. This can be used as a reference for State Vocational Schools, especially those in the Regional XIII Education Office Branch of West Java Province to disseminate information on the opening of industrial classes so that there is no doubt in their opening.

4. CONCLUSION

Based on the results of the research that has been done, it can be concluded that the clustering process produces 3 clusters, each of which consists of 396 for cluster 0, 1524 for cluster 2 and 830 for cluster 3 with a total of 2,750 data. The cluster that has the most number is cluster 1 and the cluster that has the least number is cluster 0. From the grouping of schools and expertise programs that have been formed, they can be taken into consideration at the time of the opening of the industrial class so that there is no doubt in its opening, especially for the State Vocational High School Branch of the Regional XIII Education Office of West Java Province. The results of the cluster analysis state that the schools chosen by students based on the highest order and having expertise programs contained in their clusters, may consider opening industrial classes.

REFERENCES

- [1] N. Wibowo, "Upaya memperkecil kesenjangan kompetensi lulusan sekolah menengah kejuruan dengan tuntutan dunia industri," *J. Pendidik. Teknol. dan Kejur.*, vol. 23, no. 1, pp. 45–59, 2016.
- [2] H. Achسانی, D. Kustono, and S. Suhartadi, "Model Kelas Industri pada Mitsubishi School Program di Sekolah Menengah Kejuruan," *J. Pendidik. Teor. Penelitian, dan Pengemb.*, vol. 5, no. 8, pp. 1078–1085, 2020.
- [3] S. S. Danutirta, "Pengelolaan Kelas Industri di SMK N 2 Klaten," *Hanata Widya*, vol. 7, no. 6, pp. 30–44, 2018.
- [4] A. Saifudin, "Metode data mining untuk seleksi calon mahasiswa pada penerimaan mahasiswa baru di Universitas Pamulang," *J. Teknol.*, vol. 10, no. 1, pp. 25–36, 2018.
- [5] F. Yunita, "Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Pada Penerimaan Mahasiswa Baru," *Sist. J. Sist. Inf.*, vol. 7, no. 3, pp. 238–249, 2018.
- [6] T. Purwanti, W. Ramdhan, and S. Santoso, "Penerapan Metode Klasterisasi K-Means untuk Strategi Promosi Pada SMK Tamansiswa Sukadama," *JUTSI (Jurnal Teknol. dan Sist. Informasi)*, vol. 1, no. 2, pp. 141–146, 2021.
- [7] M. Triandini, S. Defit, and G. W. Nurcahyo, "Data Mining dalam Mengukur Tingkat Keaktifan Siswa dalam Mengikuti Proses Belajar pada SMP IT Andalas Cendekia dengan Menggunakan Metode K-Means Clustering," *J. Inf. Dan Teknol.*, pp. 167–172, 2021.
- [8] J. Hutagalung, "Pemetaan Siswa Kelas Unggulan Menggunakan Algoritma K-Means Clustering," *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 9, no. 1, pp. 606–620, 2022.
- [9] A. Aditya, I. Jovian, and B. N. Sari, "Implementasi K-Means Clustering Ujian Nasional Sekolah Menengah Pertama di Indonesia Tahun 2018/2019," *J. Media Inform. Budidarma*, vol. 4, no. 1, pp. 51–58, 2020.
- [10] M. L. Sibuea and A. Safta, "Pemetaan Siswa Berprestasi Menggunakan Metode K-Means Clustering," *JURTEKSI (Jurnal Teknol. dan Sist. Informasi)*, vol. 4, no. 1, pp. 85–92, 2017.
- [11] S. N. B. Sembiring, H. Winata, and S. Kusnasari, "Pengelompokan Prestasi Siswa Menggunakan Algoritma K-Means," *J. Sist. Inf. Triguna Dharma (JURSI TGD)*, vol. 1, no. 1, pp. 31–40, 2022.
- [12] Z. Nabila, A. R. Isnain, P. Permata, and Z. Abidin, "Analisis Data Mining Untuk Clustering Kasus Covid-19 Di Provinsi Lampung Dengan Algoritma K-Means," *J. Teknol. Dan Sist. Inf.*, vol. 2, no. 2, pp. 100–108, 2021.
- [13] R. A. Indraputra and R. Fitriana, "K-Means Clustering Data COVID-19," *J. Tek. Ind.*, vol. 10, no. 3, pp. 275–282, 2020.
- [14] R. P. Primanda, A. Alwi, and D. Mustikasari, "Data Mining Seleksi Siswa Berprestasi Untuk Menentukan Kelas Unggulan Menggunakan Metode K-Means Clustering (Studi Kasus di MTS Darul Fikri)," *KOMPUTEK*, vol. 5, no. 1, pp. 88–100, 2021.
- [15] T. Syahputra, J. Halim, and E. P. Sintho, "Penerapan Data Mining Dalam Menentukan Pilihan Jurusan Bidang Studi Sma Menggunakan Metode Clustering Dengan Teknik Single Linkage," *JURTEKSI (Jurnal Teknol. Dan Sist. Informasi)*, vol. 4, no. 2, pp. 205–208, 2018.
- [16] A. Ali, "Klasterisasi Data Rekam Medis Pasien Menggunakan Metode K-Means Clustering di Rumah Sakit Anwar Medika Balong Bendo Sidoarjo," *MATRIK J. Manajemen, Tek. Inform. dan Rekayasa Komput.*, vol. 19, no. 1, pp. 186–195, 2019.
- [17] R. Nofitri and N. Irawati, "Analisis Data Hasil Keuntungan Menggunakan Software Rapidminer," *JURTEKSI (Jurnal Teknol. dan Sist. Informasi)*, vol. 5, no. 2, pp. 199–204, 2019.
- [18] W. Widiyanti, S. Solichin, and Y. Yoto, "Kerjasama Sekolah Menengah Kejuruan Dan Industri (Studi Kasus Pendidikan Kelas Industri Smk Nasional Malang Dengan Astra Honda Motor)," *Teknol. dan Kejur.* *J. Teknol. Kejur.* *dan Pengajarannya*, vol. 40, no. 2, pp. 181–192, 2017.

BIBLIOGRAPHY OF AUTHORS

Aan Rosydiana is currently as an active Student of the STMIK LIKMI Business Information Systems Masters Study Program and active teacher in State Vocational High School 2 Ciamis in Computer and Network Engineering Program. He completed his Bachelor Program at State University of Yogyakarta, With him expertise in Computer Network Engineering, Information Systems and data mining.



Dian Sediana is currently as an active Student of the STMIK LIKMI Business Information Systems Masters Study Program and active teacher in State Vocational High School 1 Kawali in Computer and Network Engineering Program. He completed his Bachelor Program at STMIK Tasikmalaya, with him expertise in Computer Network Engineering, Information Systems and data mining.



Christina Juliane is a lecturer at the Information Systems Masters Study Program, STMIK LIKMI. She completed her Masters and Doctoral Programs at STEI-ITB, with her expertise in the fields of Software and system engineering, Information Systems, and data mining.