International Journal of Social Science and Human Research

ISSN (print): 2644-0679, ISSN (online): 2644-0695

Volume 08 Issue 05 May 2025

DOI: 10.47191/ijsshr/v8-i5-19, Impact factor- 8.007

Page No: 2785-2793

K-Means Clustering Based on Expertise on Applicant Data at Vocational High School Job Fairs Using the R Language Program

Andry Ananda Putra Tanggu Mara[1], Daniel Jesayanto Jaya[1,2], Ilma Zahriyatun Nadhiroh[1,3], Fakhrur Rozi[1,4], Hoirul Ummah[1,5]

[1]Universitas Stella Maris Sumba

[2,3,4,5]Universitas Negeri Yogyakarta

ABSTRACT: This research discusses the implementation of the K-Means Clustering method to group applicant data based on their expertise at the job fair in Vocational High Schools (SMK). The purpose of this research is to assist the job fair agenda in grouping applicants effectively so that the selection and job placement process can be carried out more efficiently. The clustering process is performed using the R program, which offers various statistical and data visualization functions to support the analysis. The applicant dataset was analyzed by considering variables such as field of expertise, basic skills, soft skills and academic ability. The results of this study show that the application of K-Means Clustering on job applicant data of SMK graduates can identify several groups with different characteristics. After conducting the analysis, it was found that there are three (3) main clusters formed based on the applicants' skills and job preferences. The first cluster consists of applicants who have high technical expertise, such as Light Vehicle Engineering, while the second cluster is dominated by applicants who have expertise in Computer, namely Network and Computer Engineering majors. The third cluster includes applicants who have an educational background in Multimedia, these results indicate a significant variation in expertise and preferences in the industrial world, which can help the industry in recruiting SMK graduates as needed.

KEYWORDS: K-Means Clustering, Skills, Job Market, R Program

INTRODUCTION

In the era of globalization and the rapid development of information technology, the world of education, especially vocational education, faces increasingly complex challenges. Vocational High Schools (SMK) are expected to produce graduates who are ready to work and have expertise in accordance with industry needs. However, there is often a gap between the competencies possessed by SMK graduates and labor market demand. According to Law of the Republic of Indonesia No. 20/2003 on the National Education System, vocational education aims to prepare students to have the skills, knowledge and attitudes needed in the world of work [1].

In this context, it is important to analyze the data of job applicants from SMK graduates. One method that can be used to analyze such data is K-Means Clustering. This method allows grouping data based on similar characteristics, so it can be used to understand patterns and trends among job applicants. K-Means Clustering has been widely applied in various fields, including education, to help make better decisions [2].

In a previous study, Wiharto and Suryani (2020) compared the K-Means algorithm with Fuzzy C- Means in the context of data clustering. The results show that K-Means is more efficient in terms of computation time, although Fuzzy C-Means provides more accurate results in some cases. This shows that the selection of an appropriate clustering algorithm depends largely on the type of data and the purpose of the analysis [3].

Along with technological developments, the use of data mining in education is increasing. Yunita (2018) explained that the application of the K-Means algorithm in new student admissions can help educational institutions in identifying prospective students who have high potential. Thus, institutions can be more selective in accepting students and adjusting study programs to market needs [4].

In addition, research by Agustina and Prihandoko (2018) shows that K-Means can be used to group employees based on discipline and performance levels. The results of this clustering can be used as a basis for decision making in human resource management. This shows that K-Means is not only useful in the context of education, but also in the industrial world [5]. Thus, this literature review shows that K-Means Clustering is a very useful tool in analyzing data, both in education and industry. This research will continue the use of this method

to analyze job applicant data of SMK graduates, hoping to provide deeper insights into the needs and characteristics of applicants in the job market.

This research aims to apply the K-Means Clustering algorithm in analyzing job applicant data of skill- based SMK graduates. By using the R language program, it is expected to produce relevant clustering and provide useful insights for schools and industries. In addition, this research will also explore how clustering results can be used to improve the quality of education and the relevance of the SMK curriculum to labor market needs. Through this research, it is expected to make a positive contribution to the development of vocational education in Indonesia, as well as assist SMK graduates in improving their chances of being accepted into the workforce. By understanding the characteristics of applicants and industry needs, it is expected that vocational education can be more focused and effective in preparing graduates who are ready to face global challenges.

LITERATURE ANALYSIS AND PROBLEM STATEMENT

K-Means Clustering is a clustering method that is often used in data analysis. This method works by dividing data into groups based on similar attributes. According to Fadhli (2017), this method is very effective in clustering data that has high dimensions, so it can help in visualizing and understanding complex data [6]. In the context of education, K-Means can be used to analyze student data, including major specialization and learning achievement [7].

K-Means Algorithm

K-means is a partitioning clustering method that separates data into different groups by partitioning iteratively, K-Means is able to minimize the average distance of each data to its cluster. The K-means algorithm is as follows:

- Determine the value of k as the number of clusters to be formed.
- Determining the centroid value

In determining the centroid value (cluster center point) for the beginning of the iteration, the initial value is done randomly. Meanwhile, if determining the centroid value which is the stage of iteration, the following formula is used.

$$\overline{v}_{ij} = \frac{1}{N_i} \sum_{k=0}^{N_i} x_{kj} \ ,$$

where:

- **Vij** is the centroid/mean of the Ith cluster for the jth variable.
- **Ni** is the amount of data that is a member of the i-th cluster
- **i,k** is the index of the cluster
- **j** is the index of the variable
- **Xkj** is the kth data value in the
- in the cluster for the jth variable[10].

Calculating the distance between the centroid and the point of each object, to calculate this distance, you can use the correlation formula between two objects, namely the Euclidean Distance.

$$D_e = \sqrt{(x_i - s_i)^2 + (y_i - t_i)^2}$$

where:

- **De** is the Euclidean Distance
- **i** is the number of objects,
- **(x,y)** is the object coordinate and (s,t) is the centroid coordinate[11].

Object Grouping: To determine cluster members is to take into account the minimum distance of the object. The value obtained in the data membership in the distance matrix is 0 or 1, where the value 1 is for data allocated to the cluster and the value 0 is f o r data allocated to another cluster. Back to step 2, iterate until the resulting centroid value is fixed and cluster members do not move to other clusters.

MATERIALS AND METHODS OF RESEARCH

This research methodology consists of several stages designed to collect and analyze data on job applicants from vocational schools.

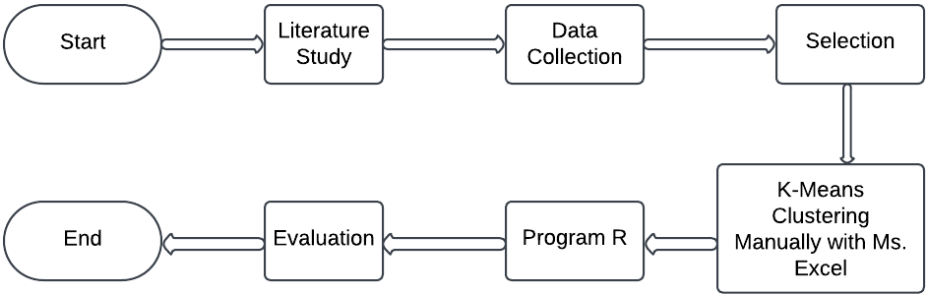

Figure 1. Research Stages

Based on Figure 2, it can be explained the stages carried out as follows: First, collecting information and studies on K-Means Clustering, R Language Program, Expertise and Job Exchange for SMK graduates from various sources such as Google Scholar, Previous Research, relevant books and reliable internet sources. Second, data collection from various sources, such as schools, training institutions, and job exchange portals. The data collected includes information on educational background, skills, work experience and job preferences. Data collection was conducted using questionnaires distributed to SMK graduates in several regions. After data collection, the next stage is Selection with data cleaning and processing methods. Incomplete or irrelevant data will be deleted to ensure that the analysis conducted is accurate and reliable. Proper data processing is essential in clustering analysis to produce valid results. Once the data is ready, the next step is to apply the K-Means Clustering algorithm. In this study, the number of clusters will be determined based on the elbow method, which is a common technique for determining the optimal number of clusters in K-Means analysis [8]. This process involves calculating the distance between the data and the cluster centroid, as well as evaluating the clustering results using the Davies-Bouldin index to assess the quality of the resulting clusters [9]. Next, the clustering results will be analyzed using the R language program to identify the characteristics of each group of applicants. This analysis will include an understanding of the skills most commonly possessed by applicants in each cluster, as well as their proposed job preferences. As such, the results of this analysis can provide useful insights for schools in tailoring the curriculum and training programs offered to students. Finally, the research results will be compiled in the form of a report that includes findings, conclusions, and recommendations. This report is expected to contribute to the development of vocational education in Indonesia, as well as assist SMK graduates in improving their chances of being accepted into the workforce. Through this methodology, the research is expected to provide a clear picture of the dynamics of job applicants of skill-based SMK graduates.

RESULTS OF RESEARCH

Data Collection

The data was obtained through accessing the page https://portalkerja.co.id/ to see some job applicants who are SMK graduates in Indonesia, as well as by distributing questionnaires through https://docs.google.com/forms/d/e/1FAIpQLSekv7ZVdcKRplaSTKD9fUFtXexhzOdq5hP8tSHIq1NiSRFZw/viewform?vc=0&c=0&w=1&flr=0 to collect answers from SMK graduates in several regions in Indonesia. Based on data from these sources, it is concluded that there are three (3) Departments that have the highest number of applicants, namely: Light Vehicle Engineering with 50 people, Network and Computer Engineering with 30 people, and Multimedia with 20 people.

Table 1. Applicant Data Based on Expertise

No.	Name of Applicant	Basic Skills	Soft Skills	ademic Skills
1	BAYU PERMANA	87	81	64
2	TORIQ NUR A.	79	92	77
3	M. SARIFUDIN	81	64	50
4	HIDAYAT N. C.	78	81	78
5	CITRA KIRANA	77	87	64
6	SEPTIAN K.	64	87	87
7	ADI HIDAYAT	98	64	87
8	KHARUNISAH F.	73	81	67
9	BAMBANG S.	64	64	76
10	BUDI S.	81	64	64

11	BRIGITA TALAN	92	80	65
12	ARDIANSYAH T.	98	75	64
13	VERESCA ABIGAIL	87	78	64
14	ZIDAN ZULKIFLI	87	64	82
15	AGUS NURYANTO	64	87	87
16	MARSELINA L.	64	87	87
17	PUTU GEDE	98	87	87
18	ASMIRA P.	64	86	87
19	ABDUR F.	87	64	64
20	STEFAN I. M.	90	80	87

By obtaining this data, the objects are grouped into 3 clusters with attributes TKR, TKJ, MM, as shown in the table below:

Table 2. Division of attributes

No.	**Name of Applicant**	**KD**	**SS**	**KA**	**TKR**			**TKJ**			**MM**		
					64	**78**	**70**	**75**	**64**	**65**	**75**	**98**	**65**
1	BAYU PERMANA	87	81	64									
2	TORIQ NUR A.	79	92	77									
3	M. SARIFUDIN	81	64	50									
4	HIDAYAT N. C.	78	81	78									
5	CITRA KIRANA	77	87	64									
6	SEPTIAN K.	64	87	87									
7	ADI HIDAYAT	98	64	87									
8	KHARUNISAH F.	73	81	67									
9	BAMBANG S.	64	64	76									
10	BUDI S.	81	64	64									
11	BRIGITA TALAN	92	80	65									
12	ARDIANSYAH T.	98	75	64									
13	VERESCA A.	87	78	64									
14	ZIDAN ZULKIFLI	87	64	82									
15	AGUS NURYANTO	64	87	87									
16	MARSELINA L.	64	87	87									
17	PUTU GEDE	98	87	87									
18	ASMIRA P.	64	86	87									
19	ABDUR F.	87	64	64									
20	STEFAN I. M.	90	80	87									

Based on the table above, we can then determine the Centroid Value, to determine the centroid value is determined randomly, as follows: TKR = (64,78,70), TKJ = (75,64,65), MM = (75,98,65). The next step is to calculate the distance based on the Centroid point that has been determined using the following formula:

$$D_e = \sqrt{(x_i - s_i)^2 + (y_i - t_i)^2}$$

where :

- De is the Euclidean
- Distance i is the number of objects,
- (x,y) is the object coordinate and (s,t) is the centroid coordinate[12].

Table 3. Calculation of Centroid Points and Distance Points

1 and 2 Determining the Centroid and Calculating the Distance												
				TKR			**TKJ**			MM		
Nama	**KD**	**SS**	**KA**	C1(X1,Y1,Z1)			C2(X2,Y2,Z2)			**C3(X3,Y3,Z3)**		
				64	78	70	75	64	65	75	98	**65**
BAYU PERMANA	87	81	64	29,09			20,00			**39,45**		
TORIQ NUR A.	79	92	77	26,65			30,23			**34,44**		
M. SARIFUDIN	81	64	50	33,30			15,26			**48,38**		
HIDAYAT N. C.	78	81	78	17,80			21,47			**25,79**		
CITRA KIRANA	77	87	64	25,57			22,09			**40,55**		
SEPTIAN K.	64	87	87	19,24			33,67			**26,94**		
ADI HIDAYAT	98	64	87	35,68			32,54			**25,51**		
KHARUNISAH F.	73	81	67	17,97			16,40			**34,94**		
BAMBANG S.	64	64	76	6,32			16,31			**24,62**		
BUDI S.	81	64	64	22,83			6,08			**34,54**		
BRIGITA TALAN	92	80	65	32,45			22,69			**40,04**		
ARDIANSYAH T.	98	75	64	37,11			25,08			**42,25**		
VERESCA A.	87	78	64	28,09			17,69			**38,33**		
ZIDAN ZULKIFLI	87	64	82	24,10			21,66			**20,02**		
AGUS NURYANTO	64	87	87	19,24			33,67			**26,94**		
MARSELINA L.	64	87	87	19,24			33,67			**26,94**		
PUTU GEDE	98	87	87	39,06			39,27			**33,67**		
ASMIRA P.	64	86	87	18,36			33,03			**26,13**		
ABDUR F.	87	64	64	27,59			12,04			**36,07**		
STEFAN I. M.	**90**	**80**	**87**	**29,27**			**31,29**			**23,90**		

After calculating the distance matrix by entering the formula:

C1 = ((64+87)/2,(78+81)/2,(70+64)/2)
= (29,09)

C2 = ((75+87)/2,(64+81)/2,(65+64)/2)
= (20,00)

C3 = ((75+87)/2,(98+81)/2,(65+64)/2)
= (39,45)

This formula is repeated or iterated until the 20th / last data and all points are filled, then determining cluster members according to the minimum distance from the centroid. By referring to the distance matrix. This can be seen in the acquisition of values in table 3 above which are colored successively in Yellow for Cluster 1, Green for Cluster 2, and Blue for Cluster 3, then each Cluster is merged or grouped into 1 cluster and calculated the 2nd Centorid Point and its Distance Point.

Table 4. Calculation of Centorid Points and Distances of Stage 2

Nama	**KD**	**SS**	**KA**	**TKR**			**TKJ**			**MM**		
				C1(X1,Y1,Z1)			C2(X2,Y2,Z2)			C3(X3,Y3,Z3)		
				68	83	83	85	75	63	93	74	86
TORIQ NUR A.	79	92	77	11.91			14.68			16.17		
HIDAYAT N. C.	78	81	78	11.04			16.38			16.89		
SEPTIAN K.	64	87	87	5.65			31.64			29.05		
BAMBANG S.	64	64	76	9.04			24.79			30.99		
AGUS NURYANTO	64	87	87	5.65			31.64			29.05		
MARSELINA L.	64	87	87	5.65			31.64			29.05		
ASMIRA P.	64	86	87	5.74			31.65			29.07		

BAYU PERMANA	87	81	64	26.61	0.25	22.47
M. SARIFUDIN	81	64	50	35.43	13.83	37.92
CITRA KIRANA	77	87	64	20.62	7.04	26.90
KHARUNISAH F.	73	81	67	16.52	12.23	27.47
BUDI S.	81	64	64	23.13	5.14	25.16
BRIGITA TALAN	92	80	65	29.77	7.18	20.64
ARDIANSYAH T.	98	75	64	35.36	13.26	22.23
VERESCA A.	87	78	64	26.67	1.75	22.54
ABDUR F.	87	64	64	26.93	4.13	22.84
ADI HIDAYAT	98	64	87	30.48	27.70	5.82
ZIDAN ZULKIFLI	87	64	82	19.38	19.52	7.93
PUTU GEDE	98	87	87	30.10	27.28	3.30
STEFAN I. M.	90	80	87	22.35	24.57	2.42

Based on table 4 above, the results of the calculation of centroid points and distance points in stage 2 have been obtained until finally the final cluster is obtained according to the predetermined color. The details are that Yellow represents Cluster 1 which is the field of expertise of the Light Vehicle Engineering (TKR) department, Green represents Cluster 2 which is the field of expertise of the Network and Computer Engineering (TKJ) department, Cluster 3 represents Blue which is the field of expertise of the Multimedia department, as in table 5 below:

Table 5. Final Clustering

Nama	KD	SS	KA	TKR	TKJ	MM	Clustering
				C1(X1,Y1,Z1)	C2(X2,Y2,Z2)	C3(X3,Y3,Z3)	
				68 83 83	85 75 63	93 74 86	
TORIQ NUR A.	79	92	77	11.91	14.68	16.17	**Cluster 1 (TKR)**
HIDAYAT N. C.	78	81	78	11.04	16.38	16.89	
SEPTIAN K.	64	87	87	5.65	31.64	29.05	
BAMBANG S.	64	64	76	9.04	24.79	30.99	
AGUS NURYANTO	64	87	87	5.65	31.64	29.05	
MARSELINA L.	64	87	87	5.65	31.64	29.05	
ASMIRA P.	64	86	87	5.74	31.65	29.07	
BAYU PERMANA	87	81	64	26.61	0.25	22.47	**Cluster 2 (TKJ)**
M. SARIFUDIN	81	64	50	35.43	13.83	37.92	
CITRA KIRANA	77	87	64	20.62	7.04	26.90	
KHARUNISAH F.	73	81	67	16.52	12.23	27.47	
BUDI S.	81	64	64	23.13	5.14	25.16	
BRIGITA TALAN	92	80	65	29.77	7.18	20.64	
ARDIANSYAH T.	98	75	64	35.36	13.26	22.23	
VERESCA A.	87	78	64	26.67	1.75	22.54	
ABDUR F.	87	64	64	26.93	4.13	22.84	
ADI HIDAYAT	98	64	87	30.48	27.70	5.82	**Cluster 3 (MM)**
ZIDAN ZULKIFLI	87	64	82	19.38	19.52	7.93	
PUTU GEDE	98	87	87	30.10	27.28	3.30	
STEFAN I. M.	90	80	87	22.35	24.57	2.42	

The results of this study show that the application of K-Means Clustering on job applicant data of SMK graduates can identify several groups with different characteristics. After conducting the analysis, it was found that there are three (3) main clusters formed based on

the applicants' skills and job preferences. The first cluster consists of applicants who have high technical expertise, such as Light Vehicle Engineering, while the second cluster is dominated by applicants who have expertise in Computer, namely Network and Computer Engineering majors. The third cluster includes applicants who have an educational background in Multimedia, this result shows a significant variation in skills and preferences in the industry, which can help the industry in recruiting SMK graduates as needed.

Implementation in R Language Program

After obtaining the Cluster data using K-Means, the next is coding in the R Program Language to see the multivariate statistical picture, the following is an example of the script used:

```
data_siswa <- data.frame(
  Kelompok1 = c(79,78,64,64,64,64,64,87,81,77,73,81,92,98,87,87,98,87,98,90),
  Kelompok2 = c(92,81,87,64,87,87,86,81,64,87,81,64,80,75,78,64,64,64,87,80),
  Kelompok3 = c(77,78,87,76,87,87,87,64,50,64,67,64,65,64,64,64,87,82,87,87)
)
print(data_siswa)
data_scaled <- scale(data_siswa)
set.seed(123)

kmeans_result <- kmeans(data_scaled, centers = 3, nstart = 25)
data_siswa$Kelulusan <- factor(kmeans_result$cluster, labels = c("TKR", "TKJ", "MM"))
# Melihat hasil klasterisasi
print(data_siswa)
#Visualisasi hasil klasterisasi menggunakan plot 2D
ggplot(data_siswa, aes(x = TKR, y = TKJ, color = MM)) + geom_point(size = 3) +
  labs(title="Hasil Klasterisasi K-Means untuk Kategori Kelulusan", x = "TKR",
       y = "TKJ") + theme_minimal()

# Menambahkan kolom untuk nomor klaster dan label klaster ke data asli
data_siswa$Nomor_Klaster <- kmeans_result$cluster
data_siswa$Label_Klaster <- factor(data_siswa$Nomor_Klaster, levels = c(1, 2, 3)
# Menampilkan data dengan nomor klaster dan label klaster
print(data_siswa[, c("TKR", "TKJ", "MM", "Nomor_Klaster", "Label_Klaster")])
```

The above script snippet is used to initiate data based on the results of normalization using K-Means Clustering using Microsoft Excel, Data Scaling/standardization: scale(data_siswa) is used to perform standard scaling, so that the variables have an average of 0 and a standard deviation of 1, which is important in K-Means so that the results are more accurate. K-Means Clustering: The kmeans function is used with centers = 3 because we want to group the data into three categories. The parameter nstart = 25 specifies that the algorithm will try 25 times to minimize the variance between clusters. Cluster Labeling: A new column Graduation was added to the data to store the clustering results. Visualization: Using ggplot2, the clusterization results are visualized in the form of a scatter plot [13].

	KD	SS	KA	TKJ	Nomor_Klaster	Label_Klaster
1	87	81	64	81	2	TKJ
2	79	92	77	92	3	MM
3	81	64	50	64	1	TKR
4	78	81	78	81	2	TKJ
5	77	87	64	87	2	TKJ
6	64	87	87	87	3	MM
7	98	64	87	64	1	TKR
8	73	81	67	81	2	TKJ
9	64	64	76	64	1	TKR
10	81	64	64	64	1	TKR
11	92	80	65	80	2	TKJ
12	98	75	64	75	2	TKJ
13	87	78	64	78	2	TKJ
14	87	64	82	64	1	TKR
15	64	87	87	87	3	MM
16	64	87	87	87	3	MM
17	98	87	87	87	2	TKJ
18	64	86	87	86	3	MM
19	87	64	64	64	1	TKR
20	90	80	87	80	2	TKJ

Figure 2. Results of Running the R Language Program

K-Means Clustering Based on Expertise on Applicant Data at Vocational High School Job Fairs Using the R Language Program

The image above is the result of running the R language program after data initiation as in the script snippet added some instructions to produce clusters according to the K-Means algorithm [14]. To display the K-Means Cluster visualization, the following script is added:

```
data_siswa$Nomor_Klaster <- kmeans_result$cluster
data_siswa$Label_Klaster <- factor(data_siswa$Nomor_Klaster,
                                   levels = c(1, 2, 3),
                                   labels = c("TKR", "TKJ", "MM"))
print(data_siswa[, c("TKR", "TKJ", "MM", "Nomor_Klaster", "Label_Klaster")])
#Menampilkan Visualisasi Klaster
fviz_cluster(list(data = data_scaled, cluster = clusters),
             geom = "point",
             ellipse.type = "convex",
             ggtheme = theme_minimal(),
             main = "Visualisasi Klasterisasi Hirarki")
```

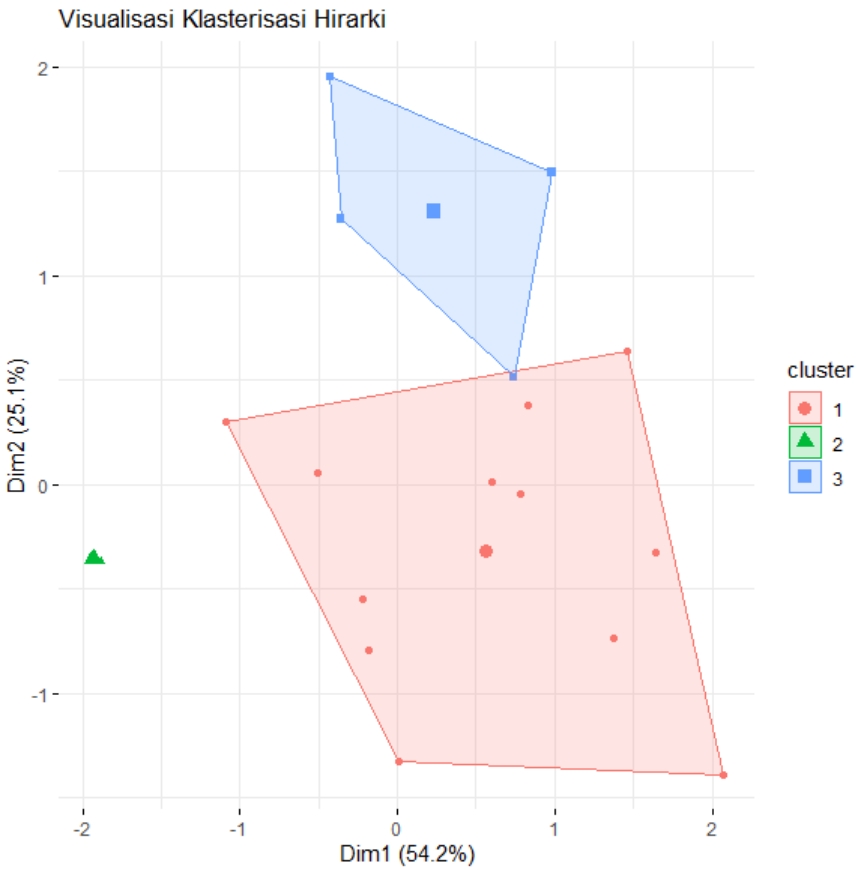

Figure 3. Visualization of K-Means Clustering

Fviz_cluster() is used to display the visualization as shown above. The orange colored graph represents cluster 1 with a circle image pattern, the green color with a triangle image pattern represents cluster 2 while the blue color with a square image pattern represents cluster 3. By using K-Means Clustering, this study successfully provides a deeper insight into the characteristics of job applicants who graduated from SMK. These results are not only beneficial for the school in designing better training programs, but also for the applicants themselves in understanding their strengths and weaknesses. Thus, this research can contribute to improving the quality of vocational education in Indonesia and help SMK graduates in improving their employment opportunities.

CONCLUSIONS

This research shows that K-Means Clustering is an effective method in analyzing job applicant data of skill-based SMK graduates. The analysis identified several clusters with different characteristics, which can provide valuable insights for schools and industries. By understanding the skills and preferences of applicants, schools can adjust their curriculum and training programs to meet the needs of the job market.

Through the application of the K-Means algorithm, the study also found that applicants with technical expertise have a better chance of being accepted into the workforce. In contrast, applicants in the service field experience greater challenges in finding a job. Therefore, it is important for educational institutions to pay attention to industry trends and adjust the skill programs offered to students.
Thus, this research is expected to make a positive contribution to the development of vocational education in Indonesia. In addition, the results of this study can serve as a reference for further research in the field of educational data analysis and curriculum development. As a recommendation, schools and related institutions are advised to continue to monitor industry developments and conduct periodic evaluations of existing education programs, so that SMK graduates can be better prepared to face challenges in the world of work.

REFERENCES

1) Law of the Republic of Indonesia Number 20 Year 2003. "National education system". Jakarta: Directorate of General Secondary Education.
2) Chen Yu, "K-Means Clustering", Indiana University
3) Wiharto, W. and Suryani, E., 2020. The comparison of clustering algorithms k-means and fuzzy c- means for segmentation of retinal blood vessels. Acta Informatica Medica, 28(1), p.42
4) F. Yunita, "Application of Data Mining Using the K-Means Clustering Algorithm in New Student Admissions (Case Study: Indragiri Islamic University)," Journal of Systematics., vol. 7 No. 03, pp. 238-249, 2018, ISSN: 2302-8149.
5) Agustina, N. and Prihandoko, P., 2018. Comparison of K-Means Algorithm with Fuzzy C-Means for Clustering the Discipline Level of Employee Performance. Journal of RESTI (Systems Engineering and Information Technology), 2(3), pp.621-626.
6) Fadhli, Muhammad. (2017). Education quality improvement management. Tadbir: Journal of Education Management Studies, 1(2), 215-240.
7) Firza, Firza, & Sarjono, Sarjono. (2020). Application of K-Means Algorithm in Clustering Method for Major Specialization for Private Students of Pelita Raya Jambi City. Journal of Information Systems Management, 5(3), 371-382
8) Kassambara, A. (2017). Practical Guide to Cluster Analysis in R (First). STHDA (http://www.sthda.com).
9) Gie, W. and Jollyta, D., 2020, July. Comparison of Euclidean and Manhattan for Cluster Optimization Using Davies Bouldin Index: Covid-19 Status of Riau Region. In Proceedings of the National Seminar on Information Science Research (SENARIS) (Vol. 2, pp. 187-191).
10) Hair Jr, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2009). Multivariate Data Analysis (Seventh Ed). Pearson.
11) Härdle, W. K., & Simar, L. (2012). Applied multivariate statistical analysis. In Applied Multivariate Statistical Analysis. https://doi.org/10.1007/978-3-642-17229-8
12) Omar, T., Alzahrani, A., & Zohdy, M. (2020). Clustering Approach for Analyzing the Student's Efficiency and Performance Based on Data. Journal of Data Analysis and Information Processing, 8, 171-182.
13) S. Bhardwaj, "Data Mining Clustering Techniques -A Review," IJCSMC., vol. 6 No. 05, pp. 183-186, 2017, ISSN: 2320-088X.
14) Everitt, B., & Hothorn, T. (2011). An introduction to applied multivariate analysis with R. Springer Science & Business Media.