



Clustering Hierarchical and Non-Hierarchical Algorithm: Clustering of 21st Century Hr Skills In SMKN Students In Indonesia

Wahyu Nurul Faroh¹, Rahadyan Tajuddien², Umi Narimawati³, Azhar Affandi⁴, Sidik Priadana⁵

^{1,2}Universitas Pamulang, Tangerang Selatan, Banten, Indonesia

³Universitas Komputer Indonesia (UNIKOM), Bandung, Jawa Barat, Indonesia

^{4,5}Universitas Pasundan, Bandung, Bandung, Jawa Barat, Indonesia

E-mail: dosen01061@unpam.ac.id

Article Info

Article History:

Received: 10 March 2022

Revised: 23 March 2022

Published: March 2022

e-ISSN: 2623-2324

p-ISSN: 2654-2528

DOI: 10.5281/zenodo.6418031

Abstract:

This study conducted a grouping sample of the data of State Vocational High School students in Bogor Regency by utilizing the data mining process using clustering techniques. The method in this study uses the Cross-School Standard Process for Data Mining (CSISP-DM). While the algorithm used for cluster determination is a hierarchical and non-hierarchical algorithm (K-Means). The hierarchical algorithm is a method that does not specify the number of clusters so that the results or output of a dendrogram with a certain number of clusters are determined by the distance until only 2 clusters are formed, while K-Means is a non-hierarchical clustering data method that can group student data into several clusters based on the similarity of the data, so that student data with the same characteristics are grouped in one cluster and those with different characteristics are grouped in another cluster. The attributes used are life and career skills (life and career skills), learning and innovation skills (learning and innovation skills) and information media and technology skills. The results of this study we can conclude that the percentage of 21st century skills in students with a sample of 10737 students at SMKN Bogor is life and career skills (life and career skills) 38% with a frequency of 4130 students, learning and innovation skills (learning and innovation skills) 46 % with a frequency of 4977 Students and information media and technology skills 35% with a frequency of 3796 Students.

Keywords: Clustering, Algoritma Hierarki, Algoritma Non-Hierarki K-Means, Keterampilan SDM Abad 21, Siswa SMKN.

INTRODUCTION

Since Schwab (2015) introduced the term fourth Industrial Revolution, the transformation of this world has become increasingly rapid. In the past, when the first industrial revolution appeared, production machines were simply driven by the power of water and steam. Furthermore, in the era of the second industrial revolution, the use of electric power for mass production became very massive. Entering the era of the third industrial revolution, the use of electronics and information technology is not undeniable. So that many productions apply machine automation. In the current era, the fourth industrial revolution, namely the integration of technology by blurring the lines between the physical, digital and biological fields. The hallmark of this fourth industrial revolution is the speed, scope and impact of the system. Due to its growing speed, it causes disruption in almost all countries.

Referring to these increasingly massive changes, the education sector certainly needs to move forward with it. Education itself is the basic human foundation to be able to form individuals who are ready to fill and mobilize the industrial/business potential that exists in a country. To accommodate it, Education 4.0 was created. Hussin (2018) explains that Education 4.0 is a form of response to the needs of the fourth Industrial Revolution, which aligns humans with technology to create new possibilities.

The insistence on the fulfillment of skills through Education 4.0 is an urgent matter to be realized. But it is not easy. Moreover, if it is implemented simultaneously in the country because the ICT (Information Communication and Technology) aspect is not evenly distributed. Muawiyah et al. (2018) explained that ICT itself is a skill that must be possessed by students. The implementation of ICT in the learning process will improve students' basic ICT skills, especially for vocational students. Whereas SMK itself, according to Lopez-Mayan and Nicodemo (2012), is a vocational school that is oriented to providing special qualifications to access the labor market.

Data as of February 2020 shows that the highest unemployment rate is occupied by vocational school graduates at 8.49% (Central Bureau of Statistics, 2020). Every year, SMK produces around 1.4 million graduates, but not all of them enter college or are absorbed in the world of work. This is an irony because Vocational High Schools (SMK) which were originally intended to produce graduates who are ready to work have actually become the biggest contributor to unemployment. The amount of 11.41% is the largest number of unemployed who come from SMK (CNN Indonesia, 26-2-2018). The Minister of Education and Culture, Muhadjir Effendy, said the same thing, that the quality of Vocational Schools in Indonesia is still lacking. As a result, many SMK graduates are unemployed. He even admitted that there were only around 219 SMKs that already had company standards (Kompas.com, 26-4-2018).

The existence of the Covid-19 pandemic has an impact on the decline in economic growth. Of course, this condition will affect SMK graduates this year. Many graduates who try to find work are predicted to end up unemployed. Vocational graduates should not be disturbed when there is an economic shock, like today. Therefore, Ball et al. (2016) directs schools to complete academic skills and life and career skills for students' success when leaving school. Furthermore, the importance of students' mastery of non-cognitive skills, especially as this field is an emerging priority in education reform and in the global market. These are all the scope of 21st century skills (The 21st century skills).

Based on the explanation above, researchers need to conduct an assessment of the 21st century skills possessed by vocational students. Given that the government has issued a policy of revitalizing vocational schools, there are already many vocational schools that are required to be ready to face the demands of the fourth Industrial Revolution. Then it's just a matter of how the SMK students are able to take advantage of the existing opportunities. From here, this research needs to be done.

LITERATURE REVIEW

Life and Career Skills

Singh & Gera (2015) stated that life skills can be described as a person's ability to behave adaptively and positively which enables him to survive by dealing effectively with the demands and challenges of life in his daily life. With this ability, the individual can make informed decisions, solve problems, think critically and creatively, communicate effectively, build relationships, empathize with others, manage and cope with their lives in a healthy and productive way.

Life skills themselves are generic, depending on how much life experience a person has had. However, in the context of vocational students, who are teenagers, the life experiences (challenges) they experience are not as numerous and complex as adults. For this reason, it needs to be honed and improved sequentially. Including when they choose to have a career. These skills will immediately form a desire in students when completing their formal education. Will he choose to work, continue his studies or become an entrepreneur? Life and career skills are very useful in understanding the psychological aspects and maturity of him. So how can the school provide the experience and skill maturation of their students.

Learning and innovation Skills

As long as humans live life, the learning process will always take place. So that humans will always develop and have a lot of thoughts. This is what makes learning and innovation skills will always emerge. Paiwithayasiritham (2014) explains that learning and innovation skills are divided into creativity and innovation, critical thinking and problem solving and communication and collaboration. For this reason, learning and innovation skills are included as skills in the 21st century. Often in schools, these learning and innovation skills are supported by qualified school facilities. But what about schools with minimal facilities. This is what needs to be studied through this research.

Information Communication and Technology (ICT)

Technological developments that occur today require changes in the way of teaching, teaching methods to students' perspectives. Information communication and technology will certainly be implemented and run by many schools. In fact, ICT facilities are always highlighted as the school's excellence. But not necessarily these advantages can be enjoyed and understood by all students.

On the other hand, it is a must to be part of 21st century skills. Nowadays, technology cannot be separated from daily life. So it can be said that technology is a part of a student's life and is a must to be able to master it. In many cases, ICT is able to facilitate students in learning. But as far as ICT is mastered by vocational students. For this reason, it is necessary to explore the meaning of ICT in these students.

In this study, ICT can be described as learning that applies the concepts of computer and multimedia learning. The application of the concept will be more interesting, fun and believed to improve learning outcomes (Nurdyansyah & Riananda, 2016). On this basis, it is necessary for a vocational student to master ICT.

Data Mining

Data mining is a process that uses statistical, mathematical, artificial intelligence, and machine learning techniques to extract and identify useful information and related knowledge from large databases.

1. Grouping Data Mining

According to Larose, data mining is divided into several groups based on the tasks that can be done, namely::

- a. Description Sometimes researchers and analysts simply want to try to find ways to describe the patterns and trends that exist in the data.

- b. Estimation Estimation is almost the same as classification, except that the estimation target variable is more numerical than categorical.
- c. Prediction Prediction is almost the same as classification and estimation, except that in the prediction the value of the result will be in the future.
- d. Classification In classification, there is a categorical variable target.
- e. Clustering Clustering is a method for finding and grouping data that have similar characteristics (similarity) between one data and another. Clustering is a data mining method that is unsupervised.
- f. Association The task of association in data mining is to find attributes that appear at a time. In the business world it is more commonly called shopping cart analysis.

2. Clustering

Clustering is a data mining technique that functions to group data based on the similarity of data in one group and minimize similarities in other groups. 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. There are two clustering methods that we are familiar with, namely hierarchical clustering and non-hierarchical clustering.

3. Hierarchical Clustering

Hierarchical clustering has two types of methods, namely agglomerative and divisive. Agglomerative clustering or also called the "bottomup" method because each object is considered as a single cluster and then the clusters are combined so that only one cluster remains. Divisional clustering is also known as the "top-down" method. In the divisive method, initially all objects are considered to be one unit of the same cluster, then the cluster splitting process is carried out into two clusters and so on until each object is considered a single cluster.

The process of cluster formation in hierarchical clustering is depicted through a two-dimensional diagram called a dendrogram. The dendrogram image represents the formation of clusters either by agglomerative or divisive procedures in the form of a dendrogram. The procedure that is often used in the hierarchical clustering method is the agglomerative procedure. Initially there are n members/observations which are considered n clusters or single groups and in the end produce one cluster or one group containing n members. The most commonly used basic joining operations are single linkage, complete linkage, and average linkage. The three methods have differences in determining the distance (similarity) between objects and groups of objects or groups of objects with groups of objects. Single linkage is a merging operation that uses the minimum distance or nearest neighbor. In contrast, complete linkage uses the farthest neighbor. While the average linkage uses the average distance to determine the distance between objects. Other merging operation methods include Ward's minimal variance, centroid method, median method, and average linkage weighted..

4. Clustering Non-Hierarchical

K-means clustering is one of the non-hierarchical data clustering methods which group data into one or more clusters/groups. Data that has the same characteristics are grouped into one cluster/group and data that has different characteristics are grouped with other clusters/groups so that data that is in one cluster/group has a small degree of variation.

The steps for clustering with the K-Means method are as follows :

1. Select the number of clusters k
2. Initialization of k cluster centers can be done in various ways. However, what is most often done is in a random way. Cluster centers are assigned initial values with random numbers.
3. Allocate all data/objects to the nearest cluster. The proximity of two objects is determined based on the distance between the two objects. Likewise, the proximity of a data to a particular cluster is determined by the distance between the data and the center of the cluster. In this stage, it is necessary to calculate the distance of each data to each cluster

center. The distance between one data and a certain cluster will determine which data belongs to which cluster. To calculate the distance of all data to each cluster center point, Euclidean distance theory can be used.

- Recalculate the cluster center with the current cluster membership. Cluster center is the average of all data/objects in a particular cluster. If desired, you can also use the median (middle value) of the cluster. So the average (mean) is not the only measure that can be used.

Reassign each object using the new cluster center. If the cluster center does not change again then the clustering process is complete. Or, go back to step number 3 until the cluster center does not change anymore.

RESEARCH METHODS

This research uses quantitative data. The data used are real numbers that describe the attributes in this study. Meanwhile, the analytical method used is Modified Hierarchical Clustering and Non-Hierarchical Clustering, where researchers create clusters. This is then made into the specified cluster. While the data used is in the form of data mining, which must be divided into groups based on clustering, a method for finding and classifying data. Larose (2005) next describes the phases in data mining known as CRISP-DM (Cross School Standard Processor Data Mining), namely:

- School Understanding Phase (School Understanding Phase)
- Data Understanding Phase
- Data processing phase (Data Preparation Phase)
- Modeling Phase
- Evaluation Phase
- Deployment Phase The following is the CSISP-DM flowchart above.

RESEARCH RESULTS AND DISCUSSION

Nominal type data such as Skill Names and School Names must be initialized in the data initialization process into numerical/numerical form. To do the initialization can be done with:

- In the Skill Name, the division is done first. Skills -Skills are divided into Skills sections.

Table 1. Initialize Skill Name

Skill Name	Student sample	Frekuensi	Percentage	Initiation
life and career skills (LNCS)	10737	4130	38%	1
learning and innovation skills (LNIS)	10737	4977	46%	2
information media and technology skills (IMNTS)	10737	3796	35%	3

From the data above, we can conclude that the highest percentage of skills mastered by SMKN students is learning and innovation skills (LNIS) with a total data of 46%. While the lowest percentage of skills is in information media and technology skills (IMNTS) with a total of 35%.

- Then the School Names are divided into several parts.

Table 2. School Name Initialization

Nama Sekolah	Jumlah Siswa	Frekuensi	Presentase	Inisialisasi
SMKN 1 Cileungsi	1117	394	35%	1
SMKN 1 Puncak Cisarua	602	145	24%	2

SMKN 1 Gunung Sindur	888	298	34%	3
SMKN 1 Cariu	893	273	31%	4
SMKN 1 Gunungputri	1345	645	48%	5
SMKN 1 Ciomas	928	402	43%	6
SMKN 1 Bojong Gede	1014	488	48%	7
SMKN 1 Kemang	307	85	28%	8
SMKN 1 Leuwiliang	861	260	30%	9
SMKN 1 Cibinong	2022	810	40%	10
SMKN 2 Cibinong	760	330	43%	11

From the data above, it shows that the highest percentage of students' skills seen is SMKN 1 Gunungputri 48% and SMKN 1 Bojong Gede 48%. Meanwhile, the lowest percentage of student skills data is at SMKN 1 Puncak Cisarua 24%.

c. Creating variables and types of School Name and Skill Name.

Table 3. 21st Century Skilled Students

School Name	21st Century Skilled Students			
	Life And Career Skills (LNCS)	Learning And Innovation Skills (LNIS)	Information Media And Technology Skills (IMNTS)	Mean
SMKN 1 Cileungsi	363	524	297	394
SMKN 1 Puncak Cisarua	145	138	152	145
SMKN 1 Gunung Sindur	240	361	293	298
SMKN 1 Cariu	248	320	250	273
SMKN 1 Gunungputri	609	757	570	645
SMKN 1 Ciomas	489	506	402	466
SMKN 1 Bojong Gede	669	625	488	594
SMKN 1 Kemang	81	102	73	85
SMKN 1 Leuwiliang	340	220	220	260
SMKN 1 Cibinong	686	1028	717	810
SMKN 2 Cibinong	260	397	333	330

The data above shows the number of attributes and schooling 21st century skills. The data shows the average number of students with the most students having 21st century skills is SMKN 1 Cibinong with 810 students. Furthermore, in the second place there is SMKN 1 Gunungputri with a total of 645 students. Meanwhile, the school that has the number of students with 21st century skills is SMKN 1 Kemang with a total of 85 students.

d. Statistical descriptive output to show the minimum, maximum and mean data. The data shows the Std Deviation value away from the mean, which means that the data distribution is varied.

Table 4. Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
LNCS	11	81	686	375.45	209.569
LNIS	11	102	1028	452.55	276.853

IMNTS	11	73	717	345.00	188.080
-------	----	----	-----	--------	---------

From the data above, we can see the minimum and maximum frequency of each skill from all SMKN in Bogor Regency. We can see that the minimum number is Information Media And Technology Skills with a total of 73 students.

The results of cluster analysis with the single linkage method. Where, the more clusters selected, the smaller the distance. From the dendogram above with a distance of more than 25, then 1 cluster is obtained (no grouping occurs), whereas if the distance is 25 then 2 clusters are obtained, namely cluster 1 consisting of SMKN 1 Cibinong while other schools enter cluster 2. These results are similar to the previous output table. (cluster membership table) if 2 clusters are selected.

Similarly, if we take the distance is 20, then there are 3 clusters, where cluster 1 consists of SMKN 1 Cileungsi, SMKN 1 Puncak Cisarua, SMKN 1 Gunung Sindur, SMKN 1 Cariu, cluster 2 consists of SMKN 1 Gunungputri and SMKN 1 Bojong Gede while the district is the rest goes to cluster 3. Likewise for other distances which results in the number of clusters that may differ from one another. The more clusters formed, the smaller the distance. That is, the more similar it is within the cluster and the dissimilarity between clusters.

In processing cluster analysis using SPSS, it can be noted that the results of the dendogram in the hierarchical method are not determined by the number of clusters. The result or output of a dendogram with a certain number of clusters is determined by the distance until only 2 clusters are formed. In the dendogram above, the maximum distance is 25 (that is, until only 2 clusters are formed). In contrast to non-hierarchical methods such as K-Means, the number of clusters has been determined beforehand.

The result of K-Means is Final Cluster Centers, which contains three clusters to divide 11 SMKN based on life and career skills (life and career skills) shortened to LNCS, learning and innovation skills (learning and innovation skills) shortened to LNIS and media skills information and technology (information media and technology skills) shortened to IMNTS in 2020.

CONCLUSION

The results of this study we can conclude that the percentage of 21st century skills in students with a sample of 10737 students at SMKN Bogor is life and career skills (life and career skills) 38% with a frequency of 4130 students, learning and innovation skills (learning and innovation skills) 46 % with a frequency of 4977 students and information media and technology skills 35% with a frequency of 3796 students. While the 21st century skills in students at every SMK in Bogor Regency are SMKN 1 Cileungsi 35%, SMKN 1 Puncak Cisarua 24%, SMKN 1 Gunung Sindur 34%, SMKN 1 Cariu 31%, SMKN 1 Gunungputri 48%, SMKN 1 Ciomas 43% , SMKN 1 Bojong Gede 48%, SMKN 1 Kemang 28%, SMKN 1 Leuwiliang 30%, SMKN 1 Cibinong 40% and SMKN 2 Cibinong 43%. The results of the Non-Hierarchical Clustering are in Cluster 1 there are SMKN 1 Gunungputri, SMKN 1 Bojong Gede and SMKN 1 Cibinong. Meanwhile, Cluster 2 is SMKN 1 Puncak Cisarua and SMKN 1 Kemang. Cluster 3 SMKN 1 Cileungsi, SMKN 1 Gunung Sindur, SMKN 1 Cariu, SMKN 1 Ciomas, SMKN 1 Leuwiliang and SMKN 2 Cibinong. With this research, it is hoped that it can be used as a basis for making decisions to determine the SMK development strategy.

BIBLIOGRAPHY

- Augustine. 2011. Marketing Management. Bandung : Alfabeta
- Akbar, Purnomo Setiady and Husaini Usman. 2012. Introduction to Statistics. Jakarta: Earth Literacy
- Ball, A., Joyce, H., dan Anderson-Butcher, D. (2016). Exploring 21st Century Skills and Learning Environments for Middle School Youth. International Journal of School

- CNN Indonesia. (2018). BPS: Lulusan SMK Banyak Menganggur Sepanjang 2017. Kontributor: Elise Dwi Ratnasari 26/02/2018 (diakses pada 7 Februari 2019)
- Hussin, A. (2018). Education 4.0 Made Simple: Ideas for Teaching. *International Journal of Education & Literacy Studies* 6 (3)
- Kompas.com. (2018). Ini Alasan Lulusan SMK Banyak Menganggur. Kontributor: Yohanes Enggar Harususilo 26/04/2018 (diakses pada 7 Februari 2019)
- Larose, D. (2006). *Data Mining Methods and Models*. New Jersey: John Wiley & Sons, Inc.
- Lopez-Mayan, C. dan Nicodemo, C. (2012). Vocational High School or Vocational College? Comparing the Transitions from School to Work. IZA Discussion Paper No 6309
- Muawiyah, D., Yamtinah, S. dan Indriyanti N. (2018). Higher Education 4.0: Assessment on Environmental Chemistry Course in Blended Learning Design. *IOP Conference Series: Journal of Physics: Conf. Series* 1097
- Nurdyansyah dan Riananda, L. (2016). Developing ICT-Based Learning Model to Improve Learning Outcomes IPA of SD Fish Market in Sidoarjo. *Proceedings of International Research Clinic & Scientific Publications of Educational Technology*.