

Design and Application of K-Means Method to Predict Sales at Arya Elektrik Stores

Michael Setiawan^{1)*}, Rino²⁾

¹⁾²⁾Buddhi Dharma University

Jl. Imam Bonjol No.41-Karawaci, Tangerang, Indonesia

¹⁾michaellsetiawan@gmail.com

²⁾rino@ubd.ac.id

Article history:

Received 26 Sept 2022;
Revised 19 Oct 2022;
Accepted 8 Nov 2022;
Available online 14 Dec 2022

Keywords:

Arya Elektrik Shop
Cluster
K-means
Product
Selling

Abstract

In a shop, the product is a staple that is sold and bought. There are products in the store between products that sell well and products that don't sell. Given this problem, it is necessary to create a system that can classify products that sell, products that sell well, and products that don't sell well, which was carried out at the Arya Elektrik Store and carried out from March to July 2022. The K-Means algorithm is not affected by the order of objects used. used, this is proven when the author tries to randomly determine the starting point of the cluster center of one of the objects at the start of the calculation. The number of cluster memberships generated is the same when using another object as the starting point for the cluster center. However, this only affects the number of iterations performed. The purpose is to create applications and analyze product sales at the Arya Elektrik Store using the K-Means method. With this system, it can provide convenience benefits for analyzing the grouping of product sales at the Arya Elektrik Store, determining and classifying product sales that are selling well, very selling, and less selling. The method used to collect data is observation and interviews. With this application, shop owners can see the results of grouping these products. So, if there are products that don't sell well, shop owners can look for other alternatives so that products that don't sell can be sold.

I. INTRODUCTION

The rapid development of technology and human knowledge by utilizing natural resources is very high. From discoveries from the 18th century, in this world, many electric tools are used to meet human needs in daily activities [1]. now, electric tools are a necessary requirement for humans in doing work, learning, and communicating. even around us today, there are a lot of home appliances, office equipment using electrical tools such as electric plugs or what are often called plugs that can drain electricity for charging cables, television cables, refrigerator cables, and others. Then the light bulb is also one example of electrical equipment that is always used in every building, house, office, or in a shopping center. The business opportunity in the field of electrical equipment is quite guarantee in the future. However, during the pandemic caused by the transmission of COVID-19, it has had a huge impact on everyone's economy, including in the trade sector. The number of visitors must be minimized, and shop closing hours cannot be as normal as before this pandemic period, thus making the time to transact trade feel shorter. Arya Elektrik Store is one of the store businesses in the field of electrical equipment that has been impacted by the COVID-19 pandemic. Arya Elektrik store was founded in 2009 and is located on Jl. Mendut Raya Block CC No. 18, Bencong Indah, Kelapa Dua District, Tangerang, Banten. This shop is a business managed by housewives who take care of their children at home every day while earning income, but over time it continues to grow until now. During this pandemic, many companies, businesses, and even stores are going through tough times. With the lack of income during this pandemic, the Arya Elektrik Store experienced a decline in revenue so the owner of the Arya Elektrik Store decided to reduce the stock items that entered the store. This makes the author want to help shop owners to process sales data so that in the coming year, the owner can provide stock of goods that match the best-selling items and generate profits instead of having to buy stock of items that rarely sell. In everyday life, we need

* Corresponding author

electric tools that suit our needs, with this modern era, there are many tools that have better functions and will produce items that are rarely needed. And to provide a good stock of goods means having to look at and reduce the stock of items that rarely sell. Therefore, the author chose the K-Means method, where this clustering method is suitable for grouping data on goods sold from 2020 where the pandemic has an impact on store businesses, and sales data for 2021, which is the endemic year of COVID-19. Through the K-Means method, data can also be clustered for products that sell well with those that don't [2]. So that in this study, sales items can be produced that can be predicted to be sold based on data from January 2020 to December 2021 which will later be continued by the owner of the Arya Electric Store to decide which items that are not selling well want to be dismissed or paused in determining the addition of stock in the coming year.

II. RELATED WORKS/LITERATURE REVIEW

K-Means

K-Means is one of the most widely used clustering algorithms because it is easy to implement. The K-Means algorithm is a distance-based algorithm that tries to partition data into several clusters [3].

Sale

Selling is the science and art of personal influence carried out by the seller to invite others to be willing to buy the goods and services offered. So, the sale can create a process of exchanging goods and/or services between the seller and the buyer [4].

Predict

Prediction is the thought of a quantity, for example the demand for one or more products in the future period [5].

Website

Website is a collection of pages that display text data information, still or motion image data, animation data, sound, video and or a combination of all of them, both static and dynamic which form one a series of interrelated buildings, each of which is linked by a network of pages (hyperlinks) [6].

III. METHODS

According to Chailles "Data mining is the core of Knowledge Discovery Databases (KDD) where algorithms explore data, build models, and find unknown patterns. KDD is a problem solving by analyzing the existing data in the database. KDD has the following process stages" [7].

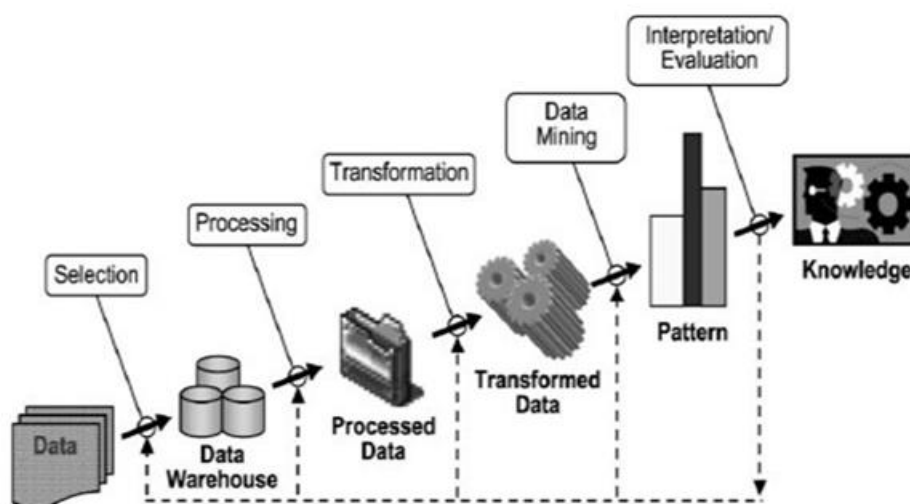


Fig. 1 Knowledge Discovery in Database

The KDD Stages process consists of:

1. Data Selection

In data selection, data is selected from a set of data in the operational database for the purpose of digging up information before the KDD process begins. The data that has been selected will be used for the data mining process which will then be stored in a separate file from the operational database.

2. Pre-Processing and Cleaning Data

In the pre-processing and data cleaning process, duplication of data is removed, checks for data that has noise and is inconsistent, corrects errors in data such as errors in writing (typography), and can enrich data with appropriate external data.

3. Transformation

After completing the data cleaning process, then in the data transformation process, it will be processed by combining it into a special format before it can be applied. For example, association and clustering analysis methods that can only accept input in the form of categorical data. Therefore, numerical data needs to be divided into several intervals, such as age data attributes can be transformed into age ranges.

4. Data Mining

In this process, the search for interesting patterns or information in the selected data is carried out using certain algorithmic techniques or methods that are in accordance with the goals and processes of KDD.

5. Interpretation / Evaluation

The evaluation process is a process of translating patterns resulting from data mining that is displayed in a form that is easily understood and understood by interested parties. At this evaluation stage, it includes the stage of testing the pattern or information to ascertain whether the pattern or information found contradicts the facts or previous hypotheses.

Research

1. Observation

Observation is a data collection technique that is carried out through an observation, accompanied by notes on the situation or the behavior of the target object [8].

2. Interview

Interview defined as an interaction in which there is exchange or sharing of rules, responsibilities, feelings, beliefs, motives, and information [9].

3. Literature review

Literature review is research that examines or reviews critically knowledge, ideas, or findings contained in the body of literature academically-oriented literature, as well as formulating his theoretical and methodological contributions to a particular topic [10].

IV. RESULTS

The results and discussion will explain the stages in the application of the K-Means algorithm from sales history data so that it can form clusters that can categorize items from the data. The first step in performing calculations is to determine the initial centroid value and the number of clusters. The number of clusters is determined based on sales data, namely desirable, moderately desirable and less desirable. The initial data will be processed into the calculation of the K-Means algorithm which has 62 types of goods according to the category of each item which can be seen in Table 1 and the early 2020 centroid (k) which can be taken randomly in Table 2 and the early 2021 centroid is also taken randomly in Table 3 as follows.

TABLE 1
Sales Data 2020 and 2021

No	Product Name	Quantity	
		2020	2021
1	Lampu Bohlam Blub 3 w	66	64
2	Lampu Bohlam Blub 5 w	63	72
3	Lampu Bohlam Blub 7 w	83	70
4	Lampu Bohlam Blub 9 w	98	84
5	Lampu Bohlam Blub 11 w	91	112

...
62	Black tape	21	35
	Total	3390	4396

TABLE 2
Early centroid 2020

Centroid	C1	C2	C3
Jan	5	6	21
Feb	3	9	11
Mar	7	11	7
Apr	4	2	6
Mei	2	3	5
Jun	3	2	4
Jul	3	2	6
Agus	6	4	6
Sep	2	1	4
Okt	4	4	5
Nov	3	1	8
Des	2	5	9

TABLE 3
Early centroid 2021

Centroid	C1	C2	C3
Jan	1	4	8
Feb	1	2	11
Mar	2	5	9
Apr	3	3	7
Mei	1	3	13
Jun	2	5	7
Jul	3	5	12
Agus	2	6	7
Sep	1	5	9
Okt	1	11	11
Nov	2	7	15
Des	1	7	11

Furthermore, from the three centroid values that have been randomly determined in Table 2 and Table 3, the distance between the initial data and the centroid is calculated using the Euclidean Distance calculation and grouped based on their proximity so that three clusters will be formed according to the number of centroids formed. In iteration 1 will form the data as in Table 4 and Table 5 as follows.

TABLE 4

First iteration of data 2020

No	Product Name	C1	C2	C3
1	Lampu Bohlam Blub 3 w	10,488	8,2462	13,9284
2	Lampu Bohlam Blub 5 w	10,247	6,8557	15,0665
3	Lampu Bohlam Blub 7 w	17,635	14,177	10,4403
4	Lampu Bohlam Blub 9 w	22,978	18,33	14,4914
5	Lampu Bohlam Blub 11 w	20,469	16,155	20,025
6	Lampu Bohlam Blub 15 w	23,728	20,224	15,3297
7	Lampu Bohlam Blub 18 w	20,543	19,079	0
8	Lampu Bohlam Blub 23 w	25,357	21	15,7162
9	Lampu Kapsul Blub 20 w	14,933	12,369	12,3693
10	Lampu Kapsul Blub 30 w	15,811	11,747	14,8997
11	Lampu Kapsul Blub 40 w	7,874	10,392	17,3781
...
62	Black tape	7,4162	11,446	25

TABLE 5
First iteration of data 2021

No	Product Name	C1	C2	C3
1	Lampu Bohlam Blub 3 w	15,297	7,280	18,111
2	Lampu Bohlam Blub 5 w	16,492	9,539	16,673
3	Lampu Bohlam Blub 7 w	15,427	7,141	17,321
4	Lampu Bohlam Blub 9 w	19,647	12,288	15,556
5	Lampu Bohlam Blub 11 w	29,086	17,916	11,314
6	Lampu Bohlam Blub 15 w	31,843	20,664	16,062
7	Lampu Bohlam Blub 18 w	29,052	17,349	11,136
8	Lampu Bohlam Blub 23 w	30,364	19,468	0,000
9	Lampu Kapsul Blub 20 w	26,325	14,629	10,247
10	Lampu Kapsul Blub 30 w	28,723	17,607	8,426
11	Lampu Kapsul Blub 40 w	26,211	16,613	9,644
...
62	Black tape	6,245	10,954	26,019

The iteration continues until the average value of the centroid is calculated and the results of the centroid value do not change as in the previous iteration. And the last iteration can be seen in Table 6 for sales data in 2020 and Table 7 for sales data in 2021.

TABLE 6
Last iteration of data 2020

No	Product Name	C1	C2	C3
1	Lampu Bohlam Blub 3 w	10,667	5,901	9,841
2	Lampu Bohlam Blub 5 w	9,249	4,189	10,297
3	Lampu Bohlam Blub 7 w	16,826	9,347	4,975

4	Lampu Bohlam Blub 9 w	22,282	15,184	7,847
5	Lampu Bohlam Blub 11 w	21,557	16,031	11,071
6	Lampu Bohlam Blub 15 w	23,901	17,075	9,255
7	Lampu Bohlam Blub 18 w	19,703	13,955	10,257
8	Lampu Bohlam Blub 23 w	24,128	17,370	11,128
9	Lampu Kapsul Blub 20 w	14,523	8,837	7,689
10	Lampu Kapsul Blub 30 w	15,978	10,406	7,232
11	Lampu Kapsul Blub 40 w	8,846	8,096	14,350
...
62	Black tape	5,9031	12,951	23,375

TABLE 7
Last iteration of data 2021

No	Product Name	C1	C2	C3
1	Lampu Bohlam Blub 3 w	12,941	4,748	14,492
2	Lampu Bohlam Blub 5 w	14,041	5,323	12,740
3	Lampu Bohlam Blub 7 w	12,941	5,547	13,559
4	Lampu Bohlam Blub 9 w	17,587	8,568	11,846
5	Lampu Bohlam Blub 11 w	27,160	15,707	6,862
6	Lampu Bohlam Blub 15 w	29,869	18,413	10,642
7	Lampu Bohlam Blub 18 w	27,070	15,003	6,470
8	Lampu Bohlam Blub 23 w	28,059	16,392	7,806
9	Lampu Kapsul Blub 20 w	23,888	11,888	6,811
10	Lampu Kapsul Blub 30 w	26,413	14,487	5,878
11	Lampu Kapsul Blub 40 w	24,114	13,424	7,510
...
62	Black tape	4,450	11,103	22,684

From the calculation of the initial iteration to the end of each annual data, the results for the 2020 data are in Table 8 and the 2021 data are in Table 9.

TABLE 8
Calculation results of K-means data 2020

Less Selling	Black tape	Selling	Colokan listrik 2 lubang 2 m	Very Selling	Kabel Listrik NYM
	Colokan listrik 3 lubang 2 m		Colokan listrik 5 lubang 2 m		Lampu Bohlam Blub 11 w
	Colokan listrik 4 lubang 2 m		Fitting Gantung		Lampu Bohlam Blub 15 w
	Colokan listrik 6 lubang 2 m		isi Cutter		Lampu Bohlam Blub 18 w
	Cutter		Kabel Listrik NYA		Lampu Bohlam Blub 23 w
	Double Tape		Kabel Listrik NYY		Lampu Bohlam Blub 7 w
	Fitting colok saklar on/off		Kabel roll 3 lubang 12 m		Lampu Bohlam Blub 9 w
	Fitting Plafon		Kabel Ties 2.5		Lampu Kapsul Blub 20 w
	Gunting		Kabel Ties 3.6		Lampu Kapsul Blub 30 w
	Kabel HDMI		Kabel Ties 4.8		Lampu Stick Blub 15 w

	Kabel Klip		Lampu Bohlam Blub 3 w		Lampu TL 40 w
	Kabel LAN		Lampu Bohlam Blub 5 w		
	Kabel roll 3 lubang 5 m		Lampu Kapsul Blub 40 w		
	Kabel roll 3 lubang 9 m		Lampu Stick Blub 12 w		
	Kabel roll 4 lubang 12 m		Lampu Stick Blub 7 w		
	Kabel roll 4 lubang 5 m		Lampu Stick Blub 9 w		
	Kabel roll 4 lubang 9 m		Lampu TL 18 w		
	Kunci L		Lampu TL 38 w		
	Lampu Kapsul Blub 50 w		Obeng min		
	Lampu Stick Blub 5 w		Saklar 2 lampu		
	Obeng kembang		Steker kotak 3 lubang		
	Saklar 1 lampu		Steker kotak 4 lubang		
	Solatip				
	Steker kotak 1 lubang				
	Steker kotak 2 lubang				
	testpen				
	Tongkat Lampu				
	TV Antenna AIO -235				
	TV Antenna TYS				

Based on Table 8, the results of the K-means calculation on the sales data in 2020 are divided into 3 categories of goods, namely 29 items categorized as "less selling", 22 items categorized as "selling", and 11 items categorized as "very selling".

TABLE 9
Calculation results of K-means data 2021

Less Selling	Black tape	Selling	Colokan listrik 2 lubang 2 m	Very Selling	Kabel Listrik NYM
	Cutter		Colokan listrik 3 lubang 2 m		Lampu Bohlam Blub 11 w
	Gunting		Colokan listrik 4 lubang 2 m		Lampu Bohlam Blub 15 w
	Kabel HDMI		Colokan listrik 5 lubang 2 m		Lampu Bohlam Blub 18 w
	Kabel LAN		Colokan listrik 6 lubang 2 m		Lampu Bohlam Blub 23 w
	Kunci L		Double Tape		Lampu Kapsul Blub 20 w
	Obeng kembang		Fitting colok saklar on/off		Lampu Kapsul Blub 30 w
	Obeng min		Fitting Gantung		Lampu Kapsul Blub 40 w
	testpen		Fitting Plafon		Lampu Stick Blub 12 w
	Tongkat Lampu		isi Cutter		Lampu Stick Blub 15 w
	TV Antenna AIO -235		Kabel Klip		Lampu Stick Blub 7 w
	TV Antenna TYS		Kabel Listrik NYA		Lampu Stick Blub 9 w
			Kabel Listrik NYN		Lampu TL 40 w

Kabel roll 4 lubang 5 m
Kabel roll 4 lubang 9 m
Kabel Ties 2.5
Kabel Ties 3.6
Kabel Ties 4.8
Lampu Bohlam Blub 3 w
Lampu Bohlam Blub 5 w
Lampu Bohlam Blub 7 w
Lampu Bohlam Blub 9 w
Lampu Kapsul Blub 50 w
Lampu Stick Blub 5 w
Lampu TL 18 w
Lampu TL 38 w
Saklar 1 lampu
Saklar 2 lampu
Solatip
Steker kotak 1 lubang
Steker kotak 2 lubang
Steker kotak 3 lubang
Steker kotak 4 lubang

As for Table 9, the results of the calculation of K-means on sales data in 2021 are divided into 3 categories of goods, namely 12 items categorized as "less selling", 37 items categorized as "selling", and 13 items categorized as "very selling".

App View



Fig. 2 Main View

In Figure 2, there are 2 menus in this application, namely Home and Results. Home contains data to be processed and for Results contains calculations that have been carried out by K-means clustering that have been entered into the application.

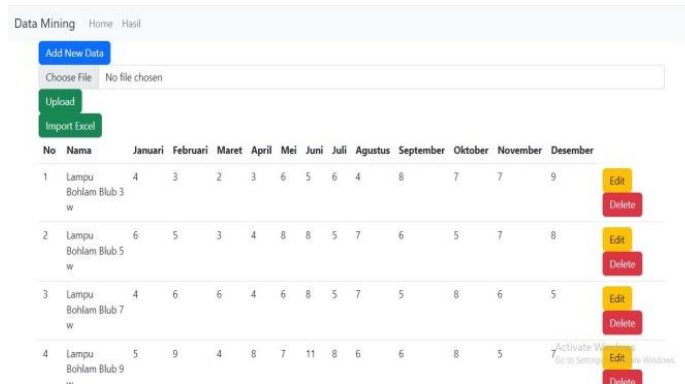


Fig. 3 Home View

In the Home form in Figure 3, it is data processing that you want to do the calculation of K-means clustering and the data can be input manually on the "Add New Data" button. In addition, the input data can also be in the form of excel (.xls). And the data that has appeared can be edited and deleted if an error occurs in inputting data.

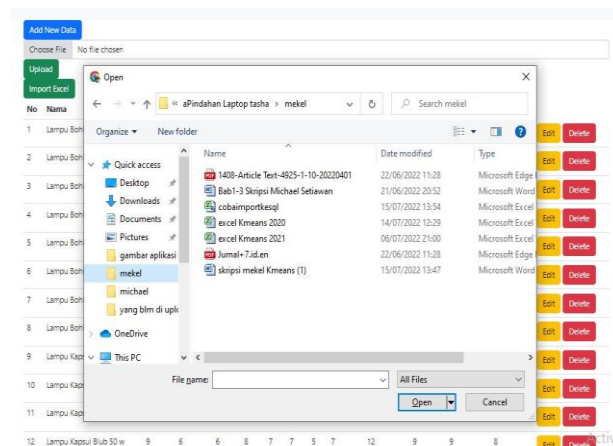


Fig. 4 Excel Input Display

In Figure 4 you can select excel data that has been saved on the device used.

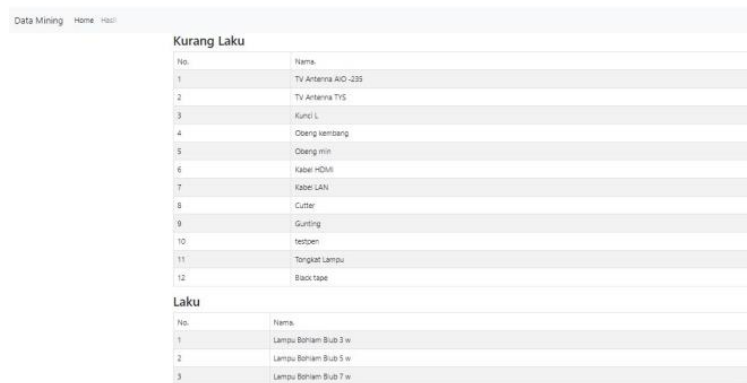


Fig. 5 Result Display

In Figure 5 to generate K-means calculations on existing data in Home.

V. CONCLUSIONS

In this study, modeling was carried out using the K-Means algorithm using data that was processed based on the Knowledge Discovery in Database (KDD) stage. The K-Means model formed can be used to determine the category of sales data. There has been an increase from the data from 2020 to 2021. The categories formed are 3 clusters, namely the "Less Selling", "Selling" and "Very Selling" clusters. Sales data for 2020 with the results of 29 items in the "Less Selling" category, 22 items in the "Selling" category, and 11 items in the "Very Selling" category. Meanwhile, sales data for 2021 resulted in 12 items being categorized as "less selling", 37 items being categorized as "selling", and 13 items being categorized as "very selling".

REFERENCES

- [1] Prasetyo, H., & Sutopo, W. (2018). Industri 4.0: Telaah Klasifikasi Aspek Dan Arah Perkembangan Riset. *J@ti Undip : Jurnal Teknik Industri*, 13(1), 17. <https://doi.org/10.14710/jati.13.1.17-26>.
- [2] Farissa, R. A., Mayasari, R., & Umaidah, Y. (2021). Perbandingan Algoritma K-Means dan K-Medoids Untuk Pengelompokan Data Obat dengan Silhouette Coefficient di Puskesmas Karangsambung. *Journal of Applied Informatics and Computing*, 5(2), 109–116. <https://doi.org/10.30871/jaic.v5i1.3237>.
- [3] Asroni, A., Fitri, H., & Prasetyo, E. (2018). Penerapan Metode Clustering dengan Algoritma K-Means pada Pengelompokan Data Calon Mahasiswa Baru di Universitas Muhammadiyah Yogyakarta (Studi Kasus: Fakultas Kedokteran dan Ilmu Kesehatan, dan Fakultas Ilmu Sosial dan Ilmu Politik). *Semesta Teknik*, 21(1), 60–64. <https://doi.org/10.18196/st.211211>.
- [4] Pramesti, R. W. (2022). Penerapan Metode Peramalan (FORECAST) Penjualan Pada DZIKRAYAT Business Cent. *Repository Muhammadiyah University of Ponorogo*, er Ponorogo, 10.
- [5] Gumanti, A. (2020). Prediksi Pasokan Pengharum Pakaian dengan Metode Trend Moment pada CV. Tunas Kr'unia Abadi Kisaran. *Stmik Royal Sekolah Tinggi Manajemen Informatika Dan Komputer Royal*, 8.
- [6] Yuniva, I. (2019). Perancangan Website Penjualan Pada Toko Automotive Filter Mobil. *Journal CERITA*, 5(2), 146–155. <https://doi.org/10.33050/cerita.v5i2.408>.
- [7] Chailes, A., Hermawan, A., & Kurnaedi, D. (2020). Penerapan Metode Data Mining Untuk Menentukan Pola Pembelian Dengan Menggunakan Algoritma. *Jurnal Algor*, 1(2), 1–8.
- [8] Ananta Gautama, W., (2017). Faktor-Faktor Penyebab Anak Putus Sekolah Dari Mi Mathla'ul Anwar Kota Jawa Kecamatan Way Khilau Kabupaten Pesawaran (Doctoral Dissertation, Uin Raden Intan Lampung).
- [9] Taraporevala, S., & Sahin. (2017). Metode Penelitian. *Physics Education*, 23(4), 8.
- [10] Kurnia, L., (2020). Studi Literatur Penggunaan Local Exhaust Ventilation (Lev) Dalam Meminimalisir Debu Pada Industri Furnitur. 8.