Internasional Journal of Integrative Sciences (IJIS)
Vol.2, No.10, 2023: 1557-1570

Application of Advanced Class Determination System Using K-Means Clustering Method (Case Study: SMK Al-Badar Balaraja)

Perani Rosyani[1]*, Fabian Syawali[2]
Universitas Pamulang
Corresponding Author: Perani Rosyani dosen00837@unpam.ac.id

A R T I C L E I N F O

Keywords: Information System, Application Program, Advanced Class, K-Means Clustering

Received : 10 August
Revised : 18 September
Accepted: 23 October

A B S T R A C T

This research develops an information system and application program to classify advanced students at SMK Al-Badar Balaraja using K-Means Clustering. The main problems addressed are unclear classification and the absence of criteria for advanced classes. The research aims to design the system, develop the app, and determine advanced classes based on report grades, attendance, achievements, extracurricular activities, and student organization involvement. The K-Means Clustering algorithm is used to classify student data into clusters based on final evaluations. Focusing on 180 grade 11 Office Administration students, the system and app were successfully designed and tested, indicating reliable identification of advanced students. This research contributes to advanced student classification and enhances education quality at SMK Al-Badar Balaraja

DOI: https://doi.org/10.55927/ijis.v2i10.6347
ISSN-E: 2985-905
https://journal.formosapublisher.org/index.php/ijis

INTRODUCTION

Advanced classes are intended to provide a wide range of opportunities to students who have outstanding potential compared to their normal peers, enabling them to develop better. Additionally, they offer a special space for them to optimize their abilities and are not hindered by the weaknesses of other students (Susanti, 2019).

SMK Al-Badar is one of the vocational schools in the Balaraja District, Tangerang Regency. After discussions with the school authorities, it was revealed that the educational method that has been applied at SMK Al-Badar, particularly for 10th-grade students majoring in office administration, tends to be generic. This means that the approach used is generally uniform and does not adapt the learning process according to the differences in craftsmanship, skills, interests, and talents of the students. With such a strategy, students' excellence emerges randomly and heavily relies on their motivation to learn and the learning environment. Therefore, it is important to develop the potential of each student at SMK Al-Badar to enable them to improve. As an effort to motivate students to achieve excellence, SMK Al-Badar will implement an advanced class program.

This research will be focused on 10th-grade students who are about to proceed to the 11th grade, majoring in Office Administration at SMK Al-Badar. The number of students in this class is 164, which is part of the total number of students at SMK Al-Badar, totaling 822 students.

The data mining technique to be used in this research is Clustering. Clustering is one of the data mining techniques that functions to group a set of data or objects into clusters so that each cluster will contain data that is as similar as possible and different from objects in other clusters (Lestari, 2019).

K-means Clustering is a data analysis method or Data mining method that conducts a modeling process without supervision (unsupervised) and is one of the methods that group existing data into several clusters (Sembiring, Winata, & Kusnasari, 2022).

LITERATURE REVIEW

By using the K-means algorithm, the existing data can be grouped into several clusters, where each cluster is represented by its center. This data is classified based on the final evaluation, including report grades, attendance, achievements, extracurricular activities, and student organization involvement. With this system in place, the school can assist in identifying suitable choice classes based on predefined criteria. The K-means algorithm functions by grouping data into clusters, where data with similar characteristics are placed in one cluster, while data with different characteristics are placed in another cluster.

The K-means Clustering method has several advantages, including low complexity, fast computation, the capability to handle large datasets, and the ability to adjust cluster membership. However, this method also has some drawbacks, such as the need to determine the number of clusters in advance, sensitivity to outliers, inability to handle varying clusters, sensitivity to data scale, and the potential for different outcomes when different initial centroids are used (Widyadhana, et al, 2021).

In the development of an application for determining students who qualify for the Advanced Class using the K-Means Clustering method, there are several challenges that need to be addressed. The student data collected needs to be processed into more structured and useful information. Additionally, there is currently no clear guidance on establishing criteria and mapping for the Advanced Class. Therefore, the aim of this research is to design a web-based application that utilizes the K-Means Clustering method as a solution to these issues (Sulistiyawati & Supriyanto, 2021).

With the existence of this application, it is expected that the school can be more effective in identifying students who meet the requirements for admission to the Advanced Class, ultimately contributing to the improvement of the quality of education at SMK Al-Badar (Salman, Syarifudin, & Kamal, 2021).

RESULTS AND DISCUSSION

Analysis of Proposed System

The analysis of the proposed system in the advanced class grouping system at SMK Al-Badar Balaraja using the K-means Clustering method aims to understand how the system operates and identify the challenges it faces, thereby enabling its development into a new computerized system. In the design phase of the proposed system analysis, the creation of a workflow diagram for the application and the procedures for advanced class grouping using the K-means Clustering method at SMK Al-Badar Balaraja is carried out.

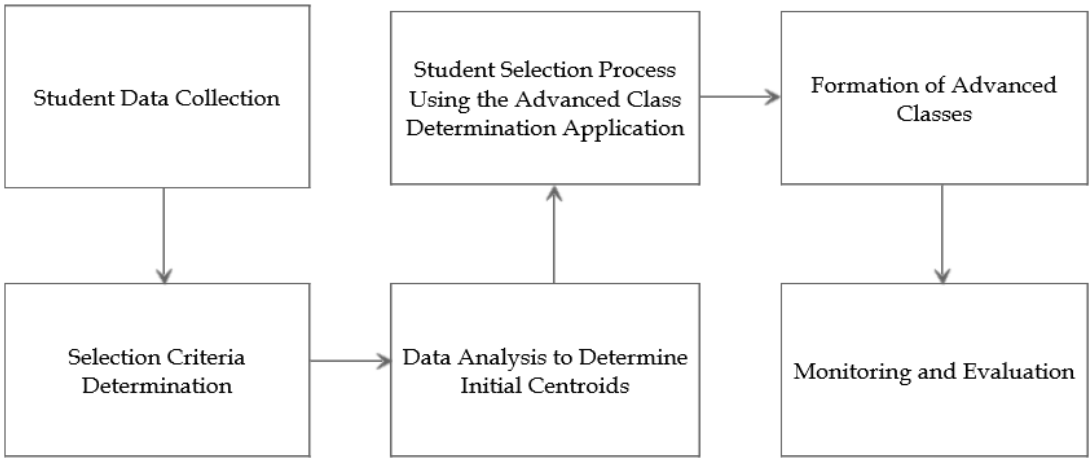

Figure 1. Analysis of Proposed System

METHODOLOGY

Research Methodology

This research, titled 'Application of Advanced Class Determination System Using K-means Clustering Method (Case Study: SMK Al-Badar Balaraja),' involves several stages designed to achieve an appropriate methodology and system design. The following are the stages conducted in the research:

1. Literature Review

   In this stage, a literature review is conducted to search for, read, and study sources such as scientific journals and other scholarly works. This aims to enhance the theoretical understanding of the K-means Clustering Algorithm as the research's theoretical foundation.
2. Data Collection

   The data used in this research are obtained through interviews with one of the teachers at SMK Al-Badar Balaraja. The collected data include criteria for determining advanced class students. Additionally, data on students' scores from various criteria are collected from the class representatives for use in the analysis.
3. Website Design

   The website design stage involves implementing the K-means Clustering Algorithm using various visual aids. Flowcharts are used for algorithmic steps, use cases for actors and functions, activity diagrams for activity visualization, and sequence diagrams for system component interactions.
4. Implementation

   In this stage, the design created in the previous stage is implemented. In brief, the K-means Clustering Algorithm is applied to the dataset following the algorithm's steps. The result is the clustering of advanced class students based on report scores, attendance, achievements, extracurricular activities, and organizational involvement.
5. Testing

   The testing stage involves testing the implementation of the K-means Clustering Algorithm by comparing the calculation results using pre-processed data and without pre-processing. The comparison results are used to evaluate the performance and the class grouping obtained.
6. Documentation

   A research report is created as documentation for the entire research and website development process. This report is prepared incrementally, starting from the introduction and proceeding through the conclusion and the provided recommendations.

Determining the Criteria Values for the Dataset

Table 1 serves as a guide for establishing criteria values in the dataset based on various factors. It explains how to determine these values in a more specific manner.

Table 1. Criteria for the Dataset

Criteria	Description	Value
K1 (Report Card Score)	(1200 - 1400)	5
	(1120 - 1199)	4
	(1050 - 1119)	3
	(950 - 1049)	2
	(0 - 950)	1
K2 (Attendance)	Never absent	5
	Absent once	4
	Absent twice	3
	Absent three times	2
	Absent more than three times	1
K3 (Achievements)	Regional/national level	5
	Inter-school level	4
	School level	3
	Class level	2
	None	1
K4 (Extracurricular Activities)	Participated in 2 or more activities	5
	Participated in 1 activity	3
	Did not participate	1
K5 (Organizational Involvement)	Involved in school and class organizations	5
	Involved in school organization	4
	Involved in class organization	3
	Not involved	1

Dataset

Table 2 is a dataset of 164 student data that will be grouped into 5 classes.

Table 2. Student Dataset

No	ID	Student Name	K1	K2	K3	K4	K5
1	001	Adel Nuralita	3	3	3	3	3
2	052	Adi Suryana	3	5	3	1	3
3	045	Aghni Maula Rizkia	3	5	2	1	3
4	012	Ahmad Rifaldi	2	1	1	3	1
5	021	Ai Indriani	2	3	1	3	1
.....							
164	161	Zalfa	2	1	3	3	4

RESULTS

Manual Calculation of K-Means Clustering Method

In this manual calculation process, it will be explained in more detail below:

Determining the Number of Clusters

There are 5 clusters to be formed, namely:

- Cluster 1 (C1) = Class A
- Cluster 2 (C2) = Class B
- Cluster 3 (C3) = Class C
- Cluster 4 (C4) = Class D
- Cluster 5 (C5) = Class E

Determining the Centroid Values

In this step, initial values will be determined for each cluster to be used as the initial center points. The initial centroids are formed by selecting samples from the student dataset to serve as the initial representation of each cluster. This process aims to initiate the initial classification before we recalculate and update the centroids based on the attribute averages within the cluster in the subsequent iterations.

Table 3. Initial Centroid Values

Initial Centroids						
	Taken from the ID Data	K1	K2	K3	K4	K5
C1	129	5	5	4	3	3
C2	107	4	3	3	3	4
C3	001	3	3	3	3	3
C4	156	3	1	2	3	1
C5	020	2	1	1	3	1

Calculating Euclidean Distance

Sample Example:

Table 4. Sample Dataset

No	ID	Student Name	K1	K2	K3	K4	K5
1	001	Adel Nuralita	3	3	3	3	3

The specific formula is used in calculating the distance between data and the cluster centroid for data allocation to the nearest cluster:

$$D_{ik} = \sqrt{\sum_{j}^{m} (Cij - Ckj)^2}$$

Cij: The first data point.
Ckj: The second data point.
Dik: Euclidean distance is a method of calculating the distance between data points x and y using mathematical calculations.

Calculation of Euclidean Distance between Data 1 (Adel Nuralita) and each centroid:

Distance To C1 = √((3 - 5)^2 + (3 - 5)^2 + (3 - 4)^2 + (3 - 3)^2 + (3 - 3)^2)
= 3

Distance To C2 = √((3 - 4)^2 + (3 - 3)^2 + (3 - 3)^2 + (3 - 3)^2 + (3 - 4)^2)
≈ 1.41421

Distance To C3 = √((3 - 3)^2 + (3 - 3)^2 + (3 - 3)^2 + (3 - 3)^2 + (3 - 3)^2)
= 0

Distance To C4 = √((3 - 3)^2 + (3 - 1)^2 + (3 - 2)^2 + (3 - 3)^2 + (3 - 1)^2)
= 3

Distance To C5 = √((3 - 2)^2 + (3 - 1)^2 + (3 - 1)^2 + (3 - 3)^2 + (3 - 1)^2)
≈ 3.60555

Table 5. Clustering Results in the First Iteration

No	ID	Student Name	Distance to Cluster					Result
			C1	C2	C3	C4	C5	
1	001	Adel Nuralita	3	1.4	0	3	3.6	Class C
2	052	Adi Suryana	3	3.1	2.8	5	5.3	Class C
3	045	Aghni Maula Rizkia	3.4	3.3	3	4.8	5.1	Class C
4	012	Ahmad Rifaldi	6.1	4.5	3.6	1.4	0	Class E
5	021	Ai Indriani	5.1	4.1	3	2.4	2	Class E
.....								
164	161	Zalfa	5.1	2.8	2.4	3.3	3.6	Class C

Determining the New Cluster Centroids

Re-establish the new Cluster centroids based on the averages. The new Clusters are obtained using the following formula:

$$Average = \frac{Sum\ of\ attribute\ values\ in\ the\ Cluster}{Number\ of\ data\ points\ in\ the\ Cluster}$$

In this step, there are several Clusters (Cluster-1, namely Class A, will be used as a sample for calculating the average), and each Cluster contains a number of data points that have been classified into it. Each data point has several attributes (Report Score, Attendance, Achievement, Extracurricular, and

Organizational Activity). The goal is to calculate the averages of these attributes for each Cluster and use these averages as the new Cluster centroids.

Table 6. Attributes in Cluster 1 Iteration 1

No	NIS	Nama Siswa	K1	K2	K3	K4	K5	Hasil
1	097	Ainun Nisa	5	5	2	3	3	Class A
2	099	Dian Nafisa	4	5	2	3	3	Class A
3	095	Eva Nadiatul	4	5	2	3	3	Class A
4	056	Fiantika Nur Apriliani	4	5	2	3	1	Class A
5	058	Iin Febriani	4	5	2	3	1	Class A
.....								
20	150	Usfuriah Nazibah	5	5	3	3	3	Class A

Determine the new centroids
Cluster 1, K1(Report Card Score) =
(5+4+4+4+4+4+5+4+4+4+4+5+4+4+5+5+5+4+5+5)/20 = 4.4000
Cluster 1, K2(Attendance) =
(5+5+5+5+5+5+5+5+5+5+5+5+5+5+5+5+5+5+4+5)/20 = 4.9500
Cluster 1, K3(Achievement) =
(2+2+2+2+2+2+1+2+2+2+2+2+2+2+3+4+4+2+4+3)/20 = 2.3500
Cluster 1, K4(Extracurricular Activities) =
(3+3+3+3+3+3+3+3+3+3+3+3+3+3+3+3+3+5+3+3+3)/20 = 3.1000
Cluster 1, K5(Organizational Involvement) =
(3+3+3+1+1+3+1+1+3+3+3+3+3+3+3+3+5+3+3+3)/20 = 2.7000
Perform these calculations for clusters 2, 3, 4, and 5 to obtain the new cluster values.

Table 7. The New Centroid Value

The Second Centroid					
	K1	K2	K3	K4	K5
C1	4.4000	4.9500	2.3500	3.1000	2.7000
C2	3.8966	3.5862	2.1379	2.5172	3.8276
C3	2.6667	3.8750	2.0000	2.5833	2.3750
C4	2.8667	1.9333	1.7333	2.7333	1.0000
C5	1.7568	1.9459	1.0000	2.4595	1.3243

If the clustering results change in the subsequent iteration, new centroid values will be sought. If the centroids (central points) of each cluster do not change, the calculation is stopped. In determining the advanced class in this

study, the final calculation is performed with 9 iterations, or 9 rounds of distance calculations for each cluster.

Table 8. Clustering Results

Cluster	Class
CLASS A	34
CLASS B	28
CLASS C	35
CLASS D	27
CLASS E	40
TOTAL	164

DISCUSSION

Clustering Results

Thus, the clustering results provide an overview of how many students are included in each class or cluster that has been formed. For instance, "Class A" has 34 students, "Class B" has 28 students, "Class C" has 35 students, "Class D" has 27 students, and "Class E" has 40 students. The total number of students in the dataset is 164 students.

These results can be used to better understand the distribution of students based on the characteristics measured in the K-means analysis

CONCLUSIONS AND RECOMMENDATIONS

Conclusion

Based on the research conducted on the Application System for Determining Advanced Classes Using the K-means Clustering Method at SMK Al-Badar Balaraja, several conclusions have been drawn:

1. This research has successfully designed an effective application program for grouping advanced class students at SMK Al-Badar Balaraja using the K-means Clustering method. The application utilizes factors such as report card scores, attendance, achievements, extracurricular activities, and student organizational involvement in the grouping process.
2. The testing results of the application indicate that the method and application have achieved their objectives effectively. This application can be relied upon in the process of determining advanced classes at SMK Al-Badar Balaraja.

Thus, this research has successfully designed an information system, developed an application program, and determined advanced classes at SMK Al-Badar Balaraja using the K-means Clustering method. This research is expected to contribute to student grouping in advanced classes and improve the quality of education at SMK Al-Badar Balaraja.

Implementation of the Program

The program implementation involves the application of the previously prepared system analysis and design results. The developed application will be tailored to the established requirements. The purpose of creating this application is to ensure that users can easily operate the Advanced Class Determination System Application using the K-Means Clustering Method at SMK Al-Badar Balaraja.

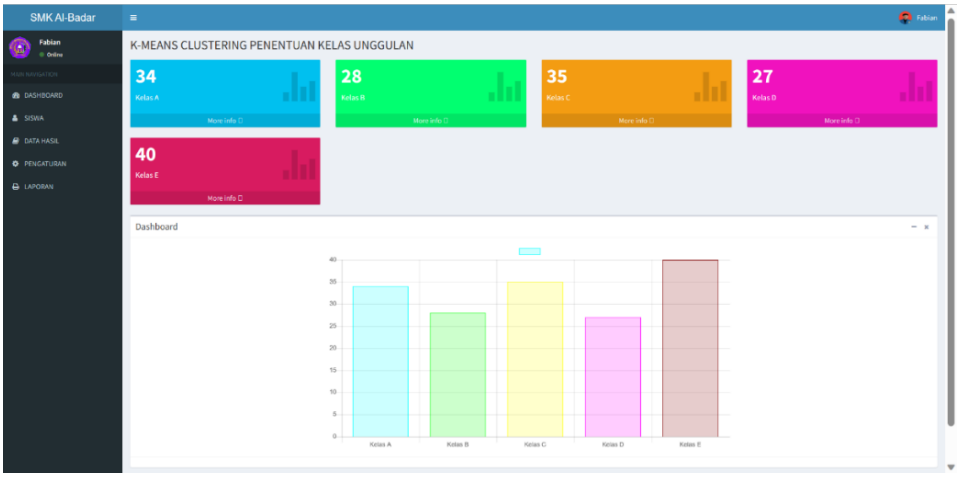

Figure 2. Main Page Display

The Main Page (Dashboard) visually displays the number of students in each class through a bar chart, providing clear information about the distribution of students in each class. This facilitates users in quickly viewing and understanding the comparison of the number of students in each class in the advanced class determination application.

Recommendations

Based on the conclusions above, this research offers several recommendations for further development:

1. There is a need for further development of the student clustering information system using the K-means Clustering method. In this development, the collected student data should be processed into more structured and useful information for the school. Additionally, clear mapping of advanced class levels and objective criteria for selecting eligible students should be established.
2. In the application development, attention should be paid to system security and user access aspects. The system must have adequate security levels and appropriate user access rights, such as admin and student access.
3. Furthermore, a more in-depth evaluation of the use of the K-means Clustering method in student clustering is required. Sensitivity to

initializations and outliers in the dataset should be considered to obtain more accurate clustering results.
4. This research can serve as a reference for future studies related to determining advanced classes and the application of the K-means Clustering method in grouping objects.

FURTHER STUDY

Research Limitations

To maintain focus and relevance in this research, the following limitations will be applied:

1. There are five clusters used to determine advanced classes at SMK Al-Badar, namely Class A, Class B, Class C, Class D, and Class E.
2. This application will only be implemented for 10th-grade students who are transitioning to 11th grade and are enrolled in the office administration program at SMK Al-Badar Balaraja.
3. The system will prioritize security by providing limited access rights to administrators and students.
4. The data used in this system will include report card scores, attendance, achievements, extracurricular activities, and organizational involvement of SMK Al-Badar Balaraja students.
5. The development of this system will focus on using the K-means Clustering algorithm.

ACKNOWLEDGMENT

The researcher would like to express gratitude to SMK Al-Badar Balaraja for granting permission to conduct the research at their institution. Special thanks also go to the research supervisors for their valuable guidance throughout the research process.

REFERENCES

Afif, H., & Arif, M. (2022). Penerapan Data Mining Untuk Pengelompokan Kelas Siswa Unggulan Berdasarkan Nilai Raport Menggunakan Metode K-Means Clustering. *Jurnal Teknik Informatika dan Desain Komunikasi Visual, 1*(2).

Darsono, V., & Andrianti, A. (2022). Penerapan Data Mining Algoritma K-Means Untuk Rekomendasi Pemilihan Bidang Studi Perguruan Tinggi Pada Siswa SMKN 1 Kota Jambi. *JAKAKOM, 2*(2), 161-171.

Hafiz, M. F., Faisal, I., & Lubis, I. (2022). Perancangan Klasifikasi Pemetaan Kelas Siswa Unggulan Menggunakan Metode K-Means Clustering Pada SMA Negeri 3 Tanjung Balai Berbasis Web. *Jurnal Ilmu Komputer dan Informatika, 6*(2).

Hutagalung, J., Syahputra, Y. H., & Tanjung, Z. P. (2022). Pemetaan Siswa Kelas Unggulan Menggunakan Algoritma K-Means Clustering. *Jurnal Teknik Informatika dan Sistem Informasi, 9*(1), 606-620.

Lestari, W. (2019). Clustering Data Mahasiswa Menggunakan Algoritma K-Means Untuk Menunjang Strategi Promosi (Studi Kasus : STMIK Bina Bangsa Kendari). *Jurnal Sistem Informasi dan Sistem Komputer (SIMKOM), Vol. 4, No. 2*, pp. 35-48. doi:10.51717/simkom.v4i2.37

Primanda, R. P., Alwi, A., & Mustikasari, D. (2021). 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. 89-101. doi:10.24269/jkt.v5i1.686

Salman, Syarifudin, E., & Kamal, Y. (2021). Peran Organisasi Dan Kepemimpinan Pendidikan Dalam Upaya Peningkatan Mutu Pendidikan Di Yayasan Al-Badar Balaraja Tangerang. *Jurnal Online Universitas Islam Negeri Sultan Maulana Hasanuddin Banten, 6*(1), 122-145. doi:10.32678/annidhom.v6i1.4863

Santoso, H., Magdalena, H., & Wardhana, H. (2022). Aplikasi Dynamic Cluster Pada K-Means Berbasis Web Untuk Klasifikasi Data Industri Rumahan. *Jurnal Manajemen, Teknik Informatika, dan Rekayasa Komputer, 21*(3), 541-554.

Satria, C., & Anggrawan, A. (2021). Aplikasi K-Measn Berbasis Web Untuk Klasifikasi Kelas Unggulan. *Jurnal Manajemen, Teknik Informatika, dan Rekayasa Komputer, 21*(1), 111-124.

Sembiring, S. N., Winata, H., & Kusnasari, S. (2022). Pengelompokan Prestasi Siswa Menggunakan Algoritma K-Means. *Jurnal Sistem Informasi TGD, Vol. 1, No. 1*, pp. 31-40. doi:10.53513/jursi.v1i1.4784

Sulistiyawati, A., & Supriyanto, E. (2021). Implementasi Algoritma K-means Clustring dalam Penetuan. *TEKNOKOMPAK, Vol. 15, No. 2*, pp. 25-36. doi:10.33365/jtk.v15i2.1162

Susanti, W. (2019). Penerapan Manajemen Kurikulum pada Kelas Unggulan. *JPPI (Jurnal Pendidikan Islam Pendekatan Interdisipliner), Vol. 3, No. 1*, pp. 42-62. doi:10.36915/jpi.v3i1.49

Syahputri, N. (2020). Penerapan Data Mining Asosiasi pada Pola Transaksi dengan Metode Apriori. *J-SAKTI (Jurnal Sains Komputer & Informatika), Vol. 4, No.2*, pp. 728-736. doi:10.30645/j-sakti.v4i2.345

Wati, A. R. (2020). Strategi Pengelolaan Kelas Unggulan Dalam Meningkatkan Prestasi Belajar Siswa. *JDMP (Jurnal Dinamika Manajemen Pendidikan), Vol. 5, No. 1*, pp. 46-57. doi:10.26740/jdmp.v5n1

Widyadhana, D., Hastuti, R. B., Kharisudin, I., & Fauzi, F. (2021). Perbandingan Analisis Klaster K-Means dan Average Linkage. *PRISMA, Prosiding Seminar Nasional Matematika, Vol. 4, No. 1*, pp. 584-594. Retrieved from https://journal.unnes.ac.id/sju/index.php/prisma/article/view/45032