

Decision Making System for Visitor Satisfaction of Labuhanbatu Library using K-Means Method

Martiwi Sitorus^{*}, Masrizal, & Rahma Muti'ah

Information Systems Study Program, Universitas Labuhanbatu, Kab. Labuhanbatu, Indonesia

Abstract

The K-Means research technique is used in this study. By using Satisfaction Variables and visits to college libraries, this research was conducted at the Labuhanbatu library. Researchers used a sample of 35 visitors for this study, which were tested using the Rapidminer data processing application. Where the results of this study showed the results of 18 people were satisfied and 17 people were neutral. Where it is stated that most visitors are satisfied with library services.

Keywords: K-means; Satisfaction; Library; Labuhanbatu; ULB.

1. Introduction

Labuhabatu University was founded in 1998, namely by having several faculties and with various types of facilities and infrastructure intended to support the smooth running of lecture activities. Library. Wrong one important facility at the university of harbors, the library as media provision of information is required to meet user needs such as services, information, completeness of library materials, comfort of space and internet access. Currently, the library has developed into a digital facility that can be seen without having to visit the actual library. Students can easily search for various books and journals online, especially on campuses. without having to go to the library again. Even so, the library facilities are still used because of several aspects of the advantages that exist. At this time, the need for information is easy to obtain, both from print media and non-print media such as the internet. The library as a medium for campus facilities and infrastructure certainly has advantages and disadvantages where the weakness of the manual library is the difficulty of getting the books you want, the limited number of the latest books, while the advantages of the manual library are the convenience of reading books directly, easy to get books at critical times, and the convenience of renting books.

Campus library services must pay attention to visitor satisfaction because it will encourage more student visits to the library, which results in crowded spaces. Not only satisfaction but the form of service must also be considered by the campus, Good service will give a good impression to these visitors, and they will always return to using existing library services, such as borrowing books and getting book layouts according to their type. Student activities in the library are often to add knowledge as well as working on final assignments or theses, not a few also use the village library facilities to support the GPA, where the GPA greatly influences students who use scholarships and students who want to work in large companies and for the sake of to get good study results. The library according to Saputra & Nugroho, (2017), reveals that the library is a work unit of a certain agency or institution that manages library materials, both in the form of books and not in the form of books (non-book material), which are arranged systematically according to certain rules so that it can be used as a source of information by each user. Meanwhile, according to Lu et al., (2015) a library is a room that contains a collection of books, the collection of books is classified with the aim that visitors can easily find the book they want. The library is also not infrequently visited by students to get additional knowledge about thesis where the thesis can help students in preparing their thesis, the library also provides scripts from year to year which are arranged according to the year's order which makes it easier for students to find the thesis they need, but unfortunately scripts in the library are not to be loaned because there are many considerations that the thesis will be lost which makes the librarian not give permission to borrow the thesis (Kenedi et al., 2017). Students really need a library as an

^{*} Corresponding author.

E-mail address: martiwisitorus07@gmail.com

information service center provided by the library to fulfill students' rights in searching and carrying out tasks that require help with book references, so that the library is considered the heart of the college (Marisca, 2014). The campus library is generally a service unit which is often interpreted as an activity in providing services to its users or students. The library is a source of information that has an important role in the management and dissemination of information. One of the libraries that offers both electronic and printed information is the library at Labuhanbatu University. Books, journals, and theses of various types make up printed information.

This research is also in line with the IAN 2020 study which states that visitor service and satisfaction are important aspects that must be considered to encourage students to use the campus library as a university resource (Isnaini Fathoni & Siyamto, 2022). Based on this description, researchers use the Labuhanbatu University Library as a research vessel because researchers want to see whether it is true that satisfaction and good service can attract students to visit the library as has been done by many other researchers who use satisfaction and service variables as research variables, researchers also Wanting to see whether suggestions from various students can help the development of Labuhanbatu University, the researcher is interested in conducting research on "DECISION MAKING SYSTEM (SPK) VISITOR SATISFACTION AT LABUHANBATU UNIVERSITY LIBRARY USING K-MEANS METHOD"

2. Literature Review

2.1. Visitor Satisfaction

Satisfaction is a post-visit evaluation, where the alternatives visited at least equal or exceed visitor expectations (Saidani et al., 2019). Satisfaction in the context of library visits is where librarians or visitors are satisfied with the library visited both in terms of arrangement, completeness of library materials and comfort of the room. This is also in line with Erpurini et al., (2022), which states that visitor satisfaction is defined as individual satisfaction based on meeting expectations (what is expected) with actual reality.

2.2. Library Service

According Pratiwi, (2018), library services are meetings between users and librarians directly with the aim of users wanting to obtain the information they need. If the service provided is bad, it will also make the image of the library worse so that users are reluctant to come back to the library. User service is something that needs to be considered because service greatly impacts visitor visits.

By utilizing the library materials provided, library services are designed to help users meet their information needs. If library services can meet the needs of users and meet their expectations, they will be considered satisfactory (Putera & Wahyono, 2018).

2.3. Libraries

According to Ayuni et al., (2019), states that librarians and users are both elements that form a library ecosystem that forms a relationship (relationship) that cannot be separated and needs and benefits one another. Librarians manage information resources (knowledge) in a professional manner so that they can be accessed and distributed optimally. There are various approaches needed to maximize access to library resources. In a perfect world, implementation and improvement of library management systems related to access to reading resources should be carried out on an ongoing basis in accordance with the progress of the needs of the knowledge society.

Librarians can be interpreted as assessments or responses given by users of what they get in the library which can give an impression to the users themselves, Librarians can be in the form of individuals, groups or institutions that use the library, whether they are members of the library or not (Akbar, 2017).

2.4. K-Means Clustering

The K-Means algorithm is a clustering algorithm that belongs to the Unsupervised learning group which is used to group data into several parts with a partition system. (Filki, 2022). When compared to other methods, the K-Means clustering algorithm method is a fairly good classification method because it is very accurate.

Ran et al., (2021), states that the K-means clustering algorithm is an iterative clustering analysis algorithm that has been considered an effective way to solve planning problems. This algorithm has been widely used in the fields of document classification, subscriber classification, trip data analysis, criminal network analysis, call log detailed analysis, and so on. It is very difficult to determine the number of clusters and sensitive to initialize the cluster center. (Aggarwal & Sharma, 2019). Clustering is one method of data mining and has become a valid instrument for solving complex problems in computer science and statistics. Clustering is the process of grouping data points into two or more groups so that data points belonging to the same group are more similar to each other than those in different groups, based solely on the information available with the data points.

2.5. The K-Means function

K-Means Clustering is used to classify data into groups. To group data, the method is to look at the numbers in the K-Means variable. This algorithm works iteratively from each point until a K-Means group is formed. According to the convenience feature, general data will be maintained. The probability of belonging to the same group increases with the similarity of the data.

Data Clustering using the K-Means Clustering Method is generally carried out with the following basic algorithm (Aggarwal & Sharma, 2019):

- a. Define.Amount.Clusters
- b. Allocation.data.to.in.clusters.in a manner.random
- c. Calculate the centroid or average of the data in each cluster
- d. Return to step 3, if there is still data moving clusters

2.6. Data Mining

The term Data Mining is used. to check the knowledge discovery database, Dogan & Birant, (2021) Says that, Data mining is the process of extracting and finding useful knowledge and information using certain algorithms or methods and techniques in accordance with the knowledge or information sought. Output on data mining can be used as an alternative decision making. The process of automatically searching big data for information is called useful data mining. In order to discover new and useful patterns, data mining methods are applied to large databases. However, not all jobs that involve searching for information can be described as data mining (Rehman et al., 2022).

2.7. Decision Support System

A decision support system or DSS (Decision Support System) is an interactive information system that provides information, modeling, and data manipulation (Tinambunan et al., 2021). A Decision Support System (Decision Support System) is a software product that developed specifically to assist management in the decision-making process. The purpose of using this system is as a "second opinion" or "information source" which can be used as material for consideration before a manager decides on a particular policy. Sutton et al., (2020) System for making decisions (DSS) or decision support systems. To find a solution to a particular problem, decision making involves selecting the best alternative based on a set of criteria.

3. Methods

This study used primary data by giving questionnaires to visitors to the Labuhanbatu University library. Data as many as 35 visitors became the source of this research. This study uses the K-means method to determine visitor satisfaction clustering. Whole and partial groupings are two types of groupings that fall under the category of groupings. In the sense of partition, it can be said that all data is compressed into one group if it can be combined into one. On the other hand, a data set is said to exhibit aberrant behavior—also referred to as outliers, noise, or "unattractive background"—if there is data that doesn't fit the majority group. DBSCAN and K-Means are two methods that can find these outliers (with some additional processing steps).

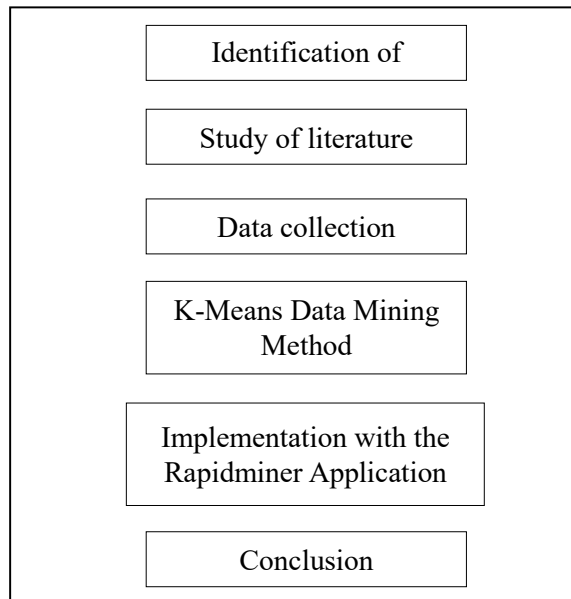


Figure 1: Research Framework

The method used to collect research data is called data collection techniques. That is, when writing or producing scientific work, the writer must choose the right data collection method. K-Means Clustering is a technique for dividing a group of objects into several groups (amount of positive integers) based on the same characteristics or attributes. K-means is a popular cluster data mining analysis. The Rapidminer application is a freeware application in which there are various kinds of data processing methods that are ready to be used easily. This application is used to process research data to make it more accurate. Rapidminer is also known as a software application used to analyze data.

4. Result and Discussion

4.1. Analysis of the K-Means Method

This manual calculation is done to simplify and understand the stages of using the K-Means clustering algorithm. In this example the researcher took a sample of 10 library visitors. The number of clusters was determined to be 2 clusters where cluster 1 visitors were satisfied with library services and library facilities, cluster 2 visitors were dissatisfied with library services and facilities in the library. The following is a sample data table that will be used for manual calculation examples of the K-Means algorithm.

Table 1. Visitor sample data

Visitor Name	Service	Modern Facilities
P1	5	3
P2	4	4
P3	4	3
P4	5	3
P5	4	1
P6	5	4
Q7	4	3
Q8	5	3
Q9	4	3
P10	5	3

The table above shows library visitor data where data is obtained with a Likert scale giving categories in terms of library services and modern facilities. The following shows the Likert scale schematic data

Table 2. Likert scale

Description	Number
Very Satisfied Visitors	5
Satisfied Visitors	4
Quite Satisfied Visitors	3
Dissatisfied visitor	2
Very Dissatisfied Visitors	1

The table above shows library visitor data where data is obtained with a Likert scale giving categories in terms of library services and modern facilities. The following shows the Likert scale schematic data

Determination of the initial center of the cluster is obtained randomly or randomly, for the initial center of the cluster or centroid obtained:

Clusters1 (4, 4) data 2nd,
 Clusters2 (4,3) 3rd data
 Cluster center distance calculation

To measure the distance between the data and the center of the cluster, a formula is used Euclidean distance as follows

$$d = \sqrt{(x1 - y1)^2 + (x2 - y2)^2}$$

Where:

d = distance
 x= Data
 y = Cluster Center

The data distance to the center of cluster 1 is:

$$d1 = \sqrt{(5 - 4)^2 + (3 - 4)^2} = 1,414$$

$$d2 = \sqrt{(4 - 4)^2 + (4 - 4)^2} = 0$$

$$d3 = \sqrt{(4 - 4)^2 + (3 - 4)^2} = 1$$

$$d4 = \sqrt{(5 - 4)^2 + (3 - 4)^2} = 1,414$$

$$d5 = \sqrt{(4 - 4)^2 + (1 - 4)^2} = 2$$

$$d6 = \sqrt{(5 - 4)^2 + (4 - 4)^2} = 1$$

$$d7 = \sqrt{(5 - 4)^2 + (4 - 4)^2} = 1$$

$$d8 = \sqrt{(5 - 4)^2 + (3 - 4)^2} = 1,414$$

$$d9 = \sqrt{(4 - 4)^2 + (3 - 4)^2} = 1$$

$$d10 = \sqrt{(5 - 4)^2 + (3 - 4)^2} = 1,414$$

The distance between the data and the center of cluster 2 is:

$$d1 = \sqrt{(5 - 4)^2 + (3 - 3)^2} = 1$$

$$d2 = \sqrt{(4 - 4)^2 + (4 - 3)^2} = 1$$

$$d3 = \sqrt{(4 - 4)^2 + (4 - 3)^2} = 0$$

$$d4 = \sqrt{(5 - 4)^2 + (3 - 3)^2} = 1$$

$$d5 = \sqrt{(4 - 4)^2 + (1 - 3)^2} = 2$$

$$d6 = \sqrt{(5 - 4)^2 + (4 - 3)^2} = 1,414$$

$$d7 = \sqrt{(5 - 4)^2 + (4 - 3)^2} = 0$$

$$d8 = \sqrt{(5 - 4)^2 + (3 - 3)^2} = 1$$

$$d9 = \sqrt{(4 - 4)^2 + (3 - 3)^2} = 1$$

$$d10 = \sqrt{(5 - 4)^2 + (3 - 3)^2} = 1$$

The distance calculated from the Euclidian distance will be compared and the closest distance between the data and the cluster center will be selected. This distance indicates that the data is in the same group as the nearest cluster center

Table 3.Iteration Results 1

Data	C1	C2	C1	C2
P1	1.414	1		C2
P2	0	1	C1	
P3	1	0		C2
P4	1.414	1		C2
P5	3	2		C2
P6	1	1.414	C1	
Q7	1	0		C2
Q8	1.414	1		C2
Q9	1	0		C2
P10	1.414	1		C2

Determine the position of the new centroid by calculating the average value of the data in the same centroid. And repeat steps 2 and 3, then if the group results in the 2nd iteration table are the same as in 1st position, then the process is stopped but the iteration table positions are different then the iteration process is continued. Update a central point.centroid can be done with the following formula.

$$\mu_k = \frac{1}{N_k} \sum_{q=1}^{N_k} x_q$$

Where:

- μk = centroid point of the K-th cluster
- Nk = the amount of data in the K-cluster
- Xq = qth data in Kth cluster

A new cluster center is obtained, namely:

C1 (4,5, 4), C2 (4,5, 2,75). After getting the new cluster center, calculate the distance between the data and the new cluster center using the Euclidean distance formula as in the previous step.

In the 2nd iteration table, the cluster position does not change or is the same as the cluster position in the 1st iteration, so the iteration process is stopped. The result of the clustering is that 2 members are in cluster 1, visitors who are satisfied with library services and modern facilities owned by the library, Cluster 1 has 8 members who are not satisfied with the services and facilities provided by the library.

Table 4.Results Table Iteration 2

Data	C1	C2	C1	C2
P1	1.118	0.55		C2
P2	0.5	1.34	C1	
P3	1.118	0.55		C2
P4	1.118	0.55		C2
P5	3.04	1.82		C2
P6	0.5	1.34	C1	
Q7	1.118	0.55		C2
Q8	1.118	0.55		C2
Q9	1.118	0.55		C2
P10	1.118	0.55		C2

Implementation of K-Means Using Rapidminer

After explaining the theoretical manual calculation of K-Means, this time we will explain the process of the K-Means method in an applicative manner where the K-Means process will be carried out using the Rapidminer software. Rapidminer is free (open source) software. To help users make the best decisions, Rapidminer uses a variety of descriptive and predictive methods. The Rapidminer application used is version 5 because this version provides a user friendly interface to make it easier to use the k means data mining clustering technique.

In mathematics, Euclidean distance is known as the usual straight line distance between two points in Euclidean space. By knowing this distance, the Euclidean space becomes a metric space with an associated norm, which is usually called the Euclidean norm. This clustering also determines the performance in each cluster with the average value of the performance vector. Following are the results of the data at the performance level of each cluster.



Figure 3. PerformanceVector

Figure 5 displays performance vector data where the average centroid value is 26,761 with 2 clusters. The average cluster data in each cluster consists of cluster 1 of 28,244 and cluster 2 of 25,190.

To determine $k = 2$ data is the best cluster data, Elbow method data testing is carried out. Elbow method is the information used to find out how many clusters are in the data set. The annotated variation is plotted as a function of the number of clusters, and the angular curve is chosen as the number of clusters. is the information used to find out how many clusters are in the data set. The annotated variation is plotted as a function of the number of clusters, and the angular curve is chosen as the number of clusters.

To provide the best variation of performance clusters the researcher used four tests which consisted of $k=2$, $k=3$, $k=4$, and $k=5$. The following shows the average centroid distance test data.

Table 5. Average Value of Mass Center Distance

K	Avg, within centroid distance
2	26,761
3	12,447
4	7,536
5	5.180

From the test using the Rapidminer application, there is an average value of the center of mass distance where cluster 2 has a center of mass distance of 26,761, in the test with clusters of 3 clusters has a center of mass distance of 12,447, the third test with clusters of 4 clusters obtains data of the center of mass distance of 7,536 , the fourth test by clustering as many as 5 clusters obtained the result of the distance to the center of mass of 5,180. when the data that has been obtained is sufficient for testing the best performance level in choosing the right grouping, the Elbow Method test is carried out.

The data in Figure 6 displays the Elbow method by showing an elbow-shaped line where in the figure the results shown show that cluster 2 is the best performance.

Services Against Library Visits

In this study, library services were said to be good and received positive views, which became one of the factors for repeat visits by visitors or students both reading in the library and borrowing library books, the services provided became a benchmark for visitors' interest in coming to make repeat visits. Therefore the library must continue to pay attention and improve the quality of its services.

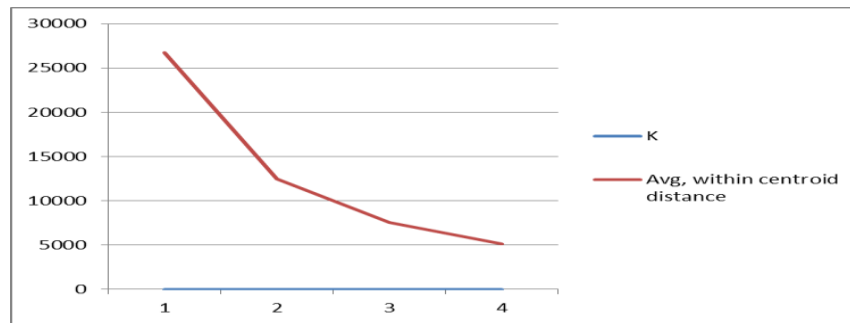


Figure 6. Elbow Method

Satisfaction with Library Visits

The satisfaction response to library visits obtained is good, which indicates that visitors make repeat visits because visitors feel satisfied with the library, therefore visitor satisfaction has an influence so that visitors want to make repeat visits and visitor satisfaction is one of the factors that must be considered. noticed by the library.

5. Conclusion

Based on the results of the study it can be concluded that the best cluster value consists of 2 clusters which have been determined through the technical elbow method. Grouping with the application of data mining was carried out using the k-means clustering method which obtained the results of cluster 1 as many as 18 visitors who were satisfied and 17 obtained dissatisfied or neutral results.

The results of this study also stated that satisfaction and service greatly influenced visits by library visitors, therefore the university and the library had to pay more attention to and increase the level of satisfaction and service development.

References

- Aggarwal, D., & Sharma, D. (2019). Application of clustering for student result analysis. *International Journal of Recent Technology and Engineering*, 7(6), 50–53.
- Akbar, G. H. (2017). Analisa kepuasan pengunjung terhadap kualitas jasa perpustakaan kota x dengan metode service quality. *Jurnal Media Teknologi*, 04(01), 13–24. <https://jurnal.unigal.ac.id/index.php/mediateknologi/article/view/2375%0Ahttps://jurnal.unigal.ac.id/index.php/mediateknologi/article/download/2375/2072>
- Ayuni, N. W. D., Utthavi, W. H., & Wahyuni, A. I. R. (2019). Analisis Kepuasan Pengunjung Terhadap Layanan Perpustakaan Politeknik Negeri Bali. *Epigram*, 16(2), 111–120. <https://doi.org/10.32722/epi.v16i2.1972>
- Dogan, A., & Birant, D. (2021). Machine learning and data mining in manufacturing. *Expert Systems with Applications*, 166, 114060. <https://doi.org/https://doi.org/10.1016/j.eswa.2020.114060>
- Erpurini, W., Alamsyah, N., Kencana, R., Kunci, K., Produk, K., Pelayanan, K., Konsumen, K., & Konsumen, K. (2022). Pengaruh Kualitas Produk dan Kualitas Pelayanan Terhadap Kepuasan Konsumen dan Dampaknya Pada Kepercayaan Konsumen Lazada. *Investasi Dan Syariah (EKUITAS)*, 3(4), 767. <https://doi.org/10.47065/ekuitas.v3i4.1524>
- Filki, Y. (2022). Algoritma K-Means Clustering dalam Memprediksi Penerima Bantuan Langsung Tunai (BLT) Dana Desa. *Jurnal Informatika Ekonomi Bisnis*, 4, 166–171. <https://doi.org/10.37034/infv4i4.166>
- Isnaini Fathoni, M., & Siyamto, Y. (2022). Pengaruh Kualitas Pelayanan Terhadap Kepuasan Pengunjung Perpustakaan Di Dinas Perpustakaan Dan Kearsipan. *Jurnal Ilmiah Keuangan Akuntansi Bisnis*, 1(2), 89–97. <https://doi.org/10.53088/jikab.v1i2.16>
- Kanedi, I., Utami, F. H., & Zulita, L. N. (2017). SISTEM PELAYANAN UNTUK PENINGKATAN KEPUASAN PENGUNJUNG PADA PERPUSTAKAAN ARSIP DAN DOKUMENTASI KOTA BENGKULU.

Pseudocode, 4(1 SE-Articles), 37–46. <https://doi.org/10.33369/pseudocode.4.1.37-46>

- Lu, J., Wu, D., Mao, M., Wang, W., & Zhang, G. (2015). Recommender system application developments: A survey. *Decision Support Systems*, 74, 12–32. <https://doi.org/https://doi.org/10.1016/j.dss.2015.03.008>
- Marisca, M. (2014). Analisis Faktor Konfirmatori untuk Tingkat Kepuasan Pengunjung Perpustakaan Universitas Negeri Yogyakarta. *Skripsi*, 1–131.
- Pratiwi, M. (2018). *Pengaruh kualitas pelayanan, fasilitas dan harga, terhadap kepuasan konsumen di penginapan lebar daun Palembang*. UPT Perpustakaan UIN Raden Fatah Palembang.
- Putera, A. K., & Wahyono. (2018). *PENGARUH KUALITAS PELAYANAN, CITRA MEREK, DAN KUALITAS PRODUK TERHADAP LOYALITAS KONSUMEN MELALUI KEPUASAN KONSUMEN*. 7(1).
- Ran, X., Zhou, X., Mou, L., Tepsan, W., & Deng, W. (2021). A Novel K-Means Clustering Algorithm with a Noise Algorithm for Capturing Urban Hotspots. *Applied Sciences*, 11, 11202. <https://doi.org/10.3390/app112311202>
- Rehman, A. U., Saleem, R. M., Shafi, Z., Imran, M., Pradhan, M., & Alzoubi, H. M. (2022). Analysis of Income on the Basis of Occupation using Data Mining. *2022 International Conference on Business Analytics for Technology and Security (ICBATS)*, 1–4. <https://doi.org/10.1109/ICBATS54253.2022.9759040>
- Saidani, B., Lusiana, L. M., & Aditya, S. (2019). Analisis Pengaruh Kualitas Website dan Kepercayaan Terhadap Kepuasan Pelanggan dalam Membentuk Minat Pembelian Ulang pada Pelanggan Shopee. *Jurnal Riset Manajemen Sains Indonesia*, 10(2), 425–444.
- Saputra, P. A., & Nugroho, A. (2017). Perancangan Dan Implementasi Survei Kepuasan Pengunjung Berbasis Web Di Perpustakaan Daerah Kota Salatiga. *JUTI: Jurnal Ilmiah Teknologi Informasi*, 15(1), 63. <https://doi.org/10.12962/j24068535.v15i1.a636>
- Sutton, R. T., Pincock, D., Baumgart, D. C., Sadowski, D. C., Fedorak, R. N., & Kroeker, K. I. (2020). An overview of clinical decision support systems: benefits, risks, and strategies for success. *Npj Digital Medicine*, 3(1), 17. <https://doi.org/10.1038/s41746-020-0221-y>
- Tinambunan, K., ... J. M.-J. of D. S., & 2021, undefined. (2021). The Application of the SMART Method as a Decision Support System in Determining Majors at Budi Mulia Private High School Tumbajae. *E-Jurnal.Pustakatimur.Org*, 1(1), 2809–171. <http://e-jurnal.pustakatimur.org/index.php/distance/article/view/72>