Journal of Artificial Intelligence and Engineering Applications

Website: https://ioinformatic.org/

15th October 2024. Vol. 4. No. 1; e-ISSN: 2808-4519

Analysis and Comparison of the Performance of K-Means Algorithm and X-Means Algorithm in Disease Type Clustering in Mitra Medika Hospital

Herdawani Afdilla[1]*, Wilda Rina Hasibuan[2]

[1, 2]*Universitas Muhammadiyah Sumatera Utara*
herdawaniafdilla@gmail.com[1]*, wildarinahasibuan77@gmail.com[2]

Abstract

The system used by the hospital is currently still manual in managing patient data and information. What happened at Mitra Medika Hospital is that it is difficult to provide medical needs related to the patient's illness, considering the many types of illnesses that provide many medical needs. Several inpatients have used BPJS facilities with various illnesses suffered by patients to undergo further examinations in order to recover from the illness they are suffering from. Mitra Medika Hospital only sees medical needs based on the illness suffered by the patient, but seeing the large amount of patient history data makes it very difficult for Mitra Medika Hospital to find out the group of illnesses that patients often experience. This study uses a quantitative approach which starts from a theoretical framework, expert ideas, or researchers' understanding based on their experience, then developed into problems and their solutions that are submitted to obtain justification (verification) or assessment in the form of empirical data support in the field. Here, a data mining pattern is applied where this data mining is a very large data mining (big data). Cluster 0: From 245 Men (Suffering Between Diseases 1-5) Cluster 1: From 255 Women (Suffering Between Diseases 6-10) By using the K-Means Algorithm and the X-Means Algorithm, clustering can be produced. By using the Disease History data, the K-Means Algorithm and the X-Means Algorithm methods can be applied to determine clusters. By using web programming, it can produce an Analysis and Comparison of the Performance of the K-Means Algorithm and the X-Means Algorithm in Clustering Types of Diseases at Mitra Medika Hospital.

Keywords: *Clustering, K-Means Algorithm, X-Means Algorithm, Types of Disease, Comparison.*

1. Introduction

Health is an important asset for every level of society, because health is a measure of the mental, physical and social well-being of each individual. Law Number 24 of 2011 concerning social security is one of the government's efforts in providing health services for all levels of society. The system currently used by the hospital is still manual in managing patient data and information, therefore there is a need for the Mitra Medika Amplas Hospital Service system to provide complete and detailed information to hospital patients. Several inpatients have used BPJS facilities with various diseases suffered by patients to undergo further examinations in order to recover from the disease they are suffering from. The problem that occurs at Mitra Medika Hospital is the difficulty of providing medical needs related to the disease experienced by patients, considering the many types of diseases so that they also provide many medical needs such as medical equipment and medicines. Mitra Medika Hospital only sees medical needs based on the patient's illness, but seeing the large amount of patient history data makes it very difficult for Mitra Medika Hospital to find out the group of illnesses that patients often experience. Therefore, a method is needed that can help Mitra Medika Hospital in grouping types of illnesses. In this study, researchers built a website-based system where the system will later be used by Mitra Medika Hospital as a container for information about patient data that can be accessed easily and effectively by Hospital employees.

1.1. Scope of Problem

The research data taken is data from 2024. The data taken is patient data based on gender and type of disease. The system to be built is website-based and uses HTML, PHP, CSS, Javascript programming languages and the database uses MySQL and XAMPP.

1.2. Analysis

The word analysis or analysis comes from ancient Greek. Namely 'analusis' which means to release. In various fields of science, linguistics, and nature as well as in the field of analysis or analysis of social sciences are widely applied. Everything in this life can be analyzed, but the methods and methods of analysis differ from one part to another. To be able to study a problem known as the scientific method [1].
The word analysis is adapted from the English word "analysis" which etymologically comes from the ancient Greek ἀνάλυσις (pronounced analysis). Analysis is a word consisting of 2 syllables, namely "ana" which means back and "luein" which means to release or break down. When combined, the word means re-decomposition. Analysis is the process of breaking down complex topic and content problems into smaller parts to make them easier to understand, according to its origin [2].

1.3. Comparison

Comparison is a method of study or investigation by making a comparison between two or more objects of study to increase and deepen knowledge about the objects being studied. So in this comparison there is an object to be compared that is already known previously, but this knowledge is not yet clear and firm. In the perception of legal science, comparison is something different from other sciences. [3] the concept of comparison has no specific definition either in terms of laws, literature or the opinions of scholars, but comparison is only a method, so it can be taken from other social sciences.

1.4. Algorithm K-Means

The K-Means algorithm is one of the partitional algorithms, because the K-Means algorithm is based on determining the initial number of groups by defining the initial centroid value [4]. The K-Means algorithm uses an iterative process to obtain a cluster database. It takes the desired number of initial clusters as input and produces the final centroid point as output. The K-Means algorithm method will select the k pattern as the starting point of the centroid randomly. The number of iterations to reach the cluster centroid will be influenced by the initial candidate cluster centroid randomly. So that a way is obtained in developing the algorithm by determining the cluster centroid which is seen from the high initial data density in order to get higher performance [5].
In its completion, the K-Means Algorithm will produce a centroid point which is used as the target of the K-Means Algorithm. After the iteration of the K-Means Algorithm stops, each object in the dataset becomes a member of a cluster. The cluster value is determined by searching all objects to find the cluster with the closest distance to the object. The K-means algorithm will group data items in a dataset into a cluster based on the closest distance [6].

1.5. Algorithm X-Means

The X-Means algorithm was developed by Dan Pelleg and Andre Moore in 2000. In this algorithm the number of clusters is calculated dynamically using the upper and lower bounds provided by the user [7]. The X-Means algorithm means taking advantage of the Bayesian Information Criterion (BIC) to control the cluster separation process. In other words, if splitting one cluster into two clusters [8]. increasing the BIC score, then having two clusters is more likely than one cluster. In this paper, we suggest using the Minimum Noisy Description Length (MNDL. as the cluster separation criterion, which leads to a more precise prediction of the number of clusters. The X-Means clustering algorithm is used to overcome one of the main drawbacks of the K-Means clustering algorithm, namely the need for prior knowledge of the number of clusters (K) [9].
In this method, the actual value of K is estimated in an unsupervised manner and only based on the data set itself. The X-Means algorithm is a method that improves the drawbacks of the K-Means algorithm [10]. This algorithm was developed in 2000 by Dan Pelleg and Andre Moore. In this method, the number of clusters formed is dynamically calculated using the upper and lower bounds specified by the user. The process of this algorithm consists of two stages that are repeated until completion. The X-Means Clustering algorithm is one of the clustering methods developed from the K-Means Clustering algorithm grouping method. The X-Means Clustering algorithm is carried out by comparing the Davis Buldin Index (DBI) value of the X-Means Algorithm with the DBI value of the K-Means Algorithm on variations in group size. The result of this grouping is a group of data that is grouped based on several input variables to produce groups of less, sufficient, good, and superior [11].

1.6. Website

A website is an information page provided via the internet so that it can be accessed worldwide as long as it is connected to the internet network. A website is a component or collection of components consisting of text, images, animated sounds so that it is more of an interesting information media to visit.

2. Research Framework

2.1. Research Framework

In the research framework, it is done to find something systematically by using scientific methods and applicable sources. With this research framework process, it can provide good and accurate research results. There are several stages of the research framework carried out in solving problems. The stages are as follows:

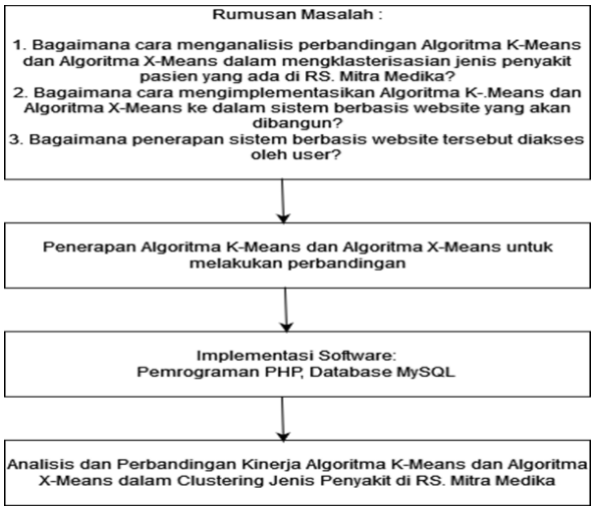

Fig. 1: Research Framework

2.2. Research Supporting Data

Table 1: Medical History Data

No.	Gender	Type of Disease
1	2	3
2	2	10
3	2	3
4	1	10
5	1	2
6	2	2
7	1	6
8	2	10
9	1	9
10	2	1
11	2	9
12	2	8
13	1	7
14	2	1
15	2	8
16	2	3
17	1	5
18	2	8
19	1	10
20	1	8
21	2	8
22	2	5
23	2	6
24	1	3
25	1	9
26	1	1
27	2	8
28	2	3
29	2	1

30	2	2
31	1	3
32	2	1
33	2	2
34	1	5
35	2	8
36	1	1
37	1	3
38	1	3
39	1	9
40	1	1
41	1	4
42	2	7
43	1	8
44	1	6
45	2	7
46	2	7
47	2	2
48	2	9
49	2	1
50	1	5
….	…..	…..
500	1	1

3. Research Methodology

This study uses a quantitative approach which starts from a theoretical framework, expert ideas, or researcher understanding based on their experience, then developed into problems and their solutions that are proposed to obtain justification (verification) or assessment in the form of empirical data support in the field. Here, applying a data mining pattern where data mining is a very large data excavation (big data). After that, the researcher conducted a literature study of journals related to the problems studied. Then the researcher conducted a system design starting with designing a database and designing an interface display and comparing the two algorithms, namely the K-Means Algorithm and the X-Means Algorithm and obtaining the results and conclusions of this study.

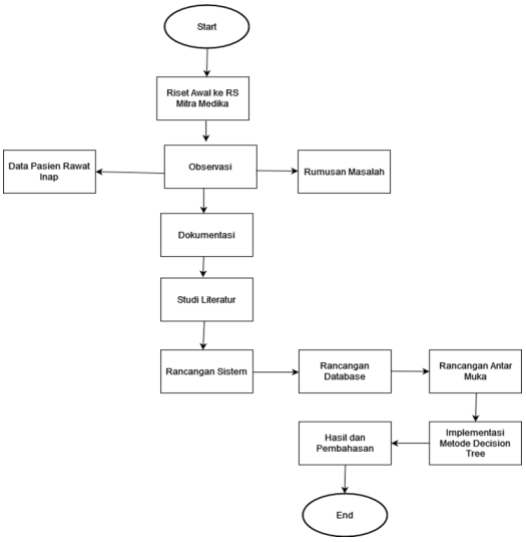

Fig. 2: Research Stages

3.1. Data Collection Techniques

The data collection techniques used by researchers to collect data in this study are as follows:

1. Literature Study, namely a data collection technique used by collecting data and information related to the problem being studied by studying and reviewing scientific books, written works related to the problem being studied and other literature references related to data mining and the K-Means Algorithm and the X-Means Algorithm to compare the two algorithms in classifying patient data at RS. Mitra Medika.
2. Field Study, namely data collection obtained through research by going directly to the research location to find facts related to the research subject, namely; observation, namely collecting data on inpatient data at RS. Mitra Medika, documentation, conducting documentation at RS. Mitra Medika.

3.2. Data Analysis Techniques

There are 2 Data Analysis Techniques in this study, namely using the K-Means Algorithm and the X-Means Algorithm. The following are the steps of the two Algorithms:

1. K-Means Algorithm; determine k as the number of clusters formed; To determine the number of clusters k is done with several considerations such as theoretical and conceptual considerations that may be proposed to determine how many clusters.
2. Means Algorithm; Increase-Params, in this step apply the K-Means Algorithm initially for K - clusters until convergenc. Where k is equal to the lower limit provided by the user, Improve Structure, this structure improvement step begins by breaking each cluster center into two children in opposite directions along a randomly selected vector. After that run the K-Means Algorithm locally in each cluster for two clusters. The decision of each cluster center itself by comparing the BIC values, If K >= kmax (upper bound) stop and report to the model the best scoring found during the dance, otherwise go to step 1.

4. Results and Discussion

4.1. Result

The results of the Analysis and Comparison of the Performance of the K-Means Algorithm and the X-Means Algorithm in Clustering Disease Types at Mitra Medika Hospital can be seen as follows:

a. Login Form
The Login Form of the Analysis and Comparison of the Performance of the K-Means Algorithm and the X-Means Algorithm in Clustering Disease Types at Mitra Medika Hospital can be seen in Figure 3.

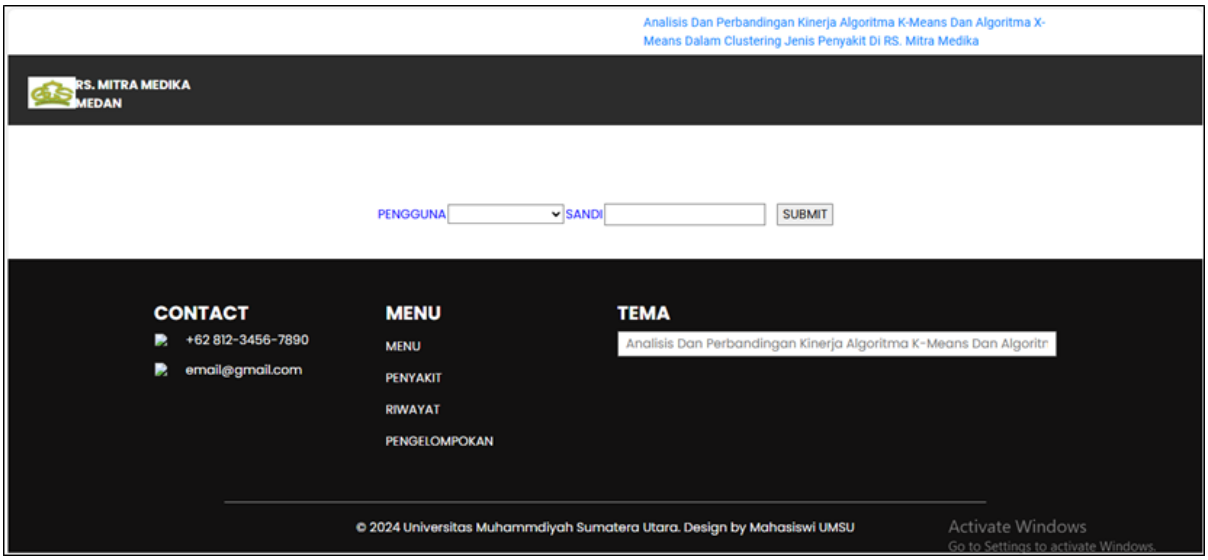

Fig. 3: Login Form

b. Menu Form
The Menu Form of Analysis and Comparison of K-Means Algorithm Performance and X-Means Algorithm in Clustering Disease Types at Mitra Medika Hospital can be seen in Figure 4.

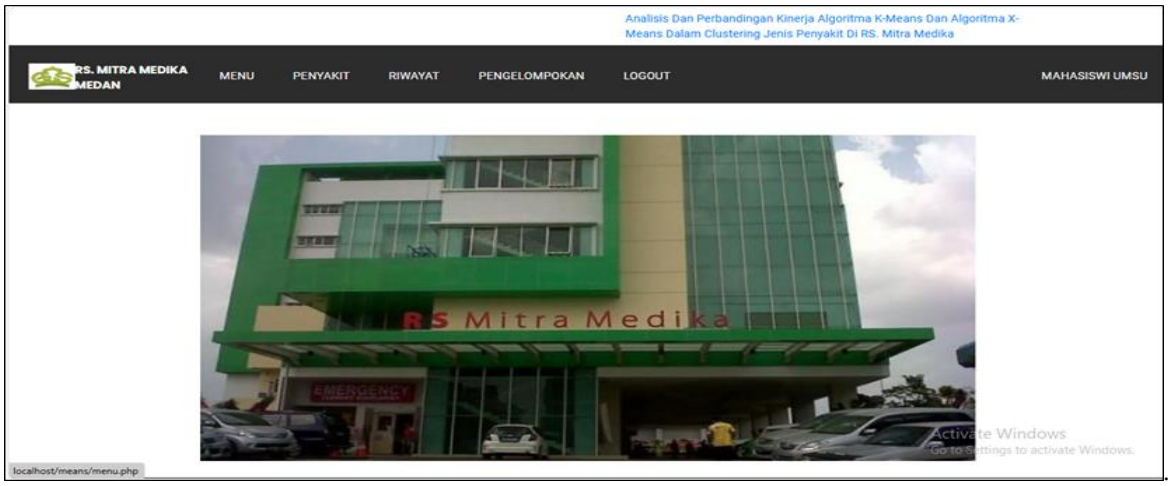

Fig. 4: Menu Form

c. Disease Form

Disease Form from Analysis and Comparison of K-Means Algorithm Performance and X-Means Algorithm in Clustering Disease Types at Mitra Medika Hospital can be seen in Figure 5.

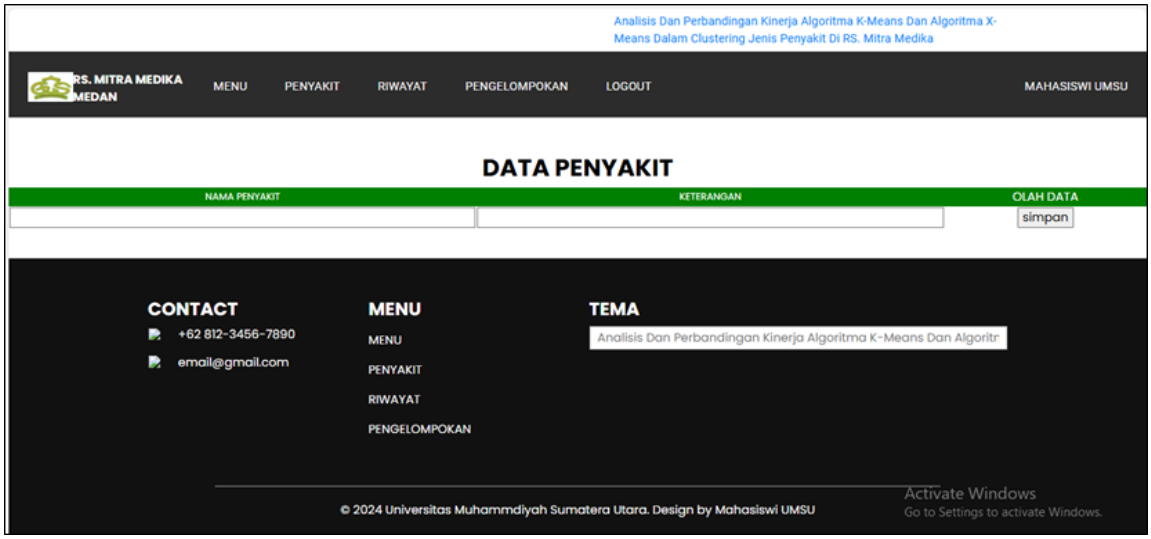

Fig. 5: Disease Form

d. Cluster Form

Cluster Form from Analysis and Comparison of K-Means Algorithm Performance and X-Means Algorithm in Clustering Disease Types at Mitra Medika Hospital can be seen in Figure 6.

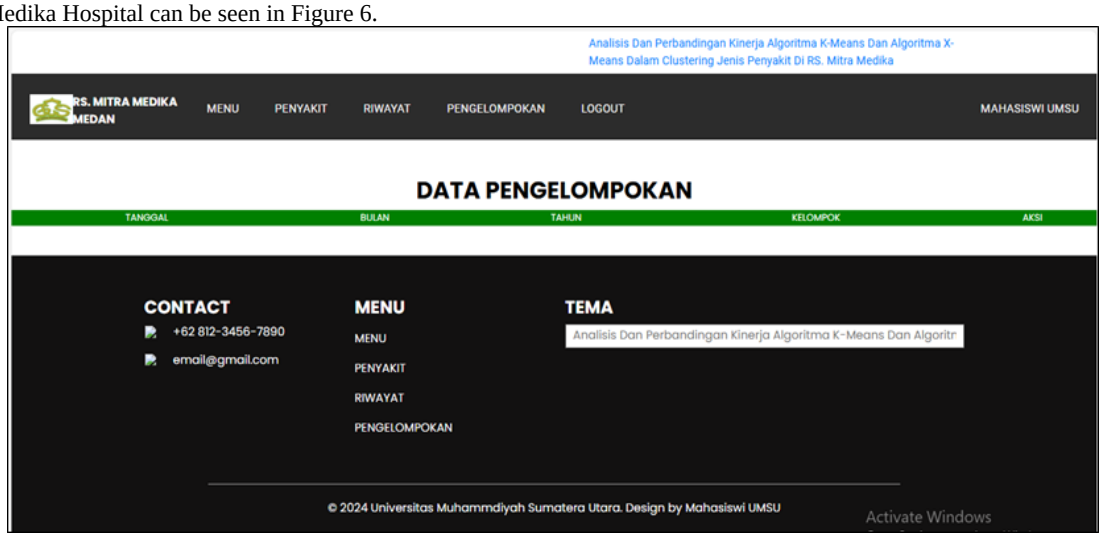

Fig. 6: Cluster Form

e. Result Form

The Result Form of the Analysis and Comparison of the Performance of the K-Means Algorithm and the X-Means Algorithm in Clustering Disease Types at Mitra Medika Hospital can be seen in Figure 7.

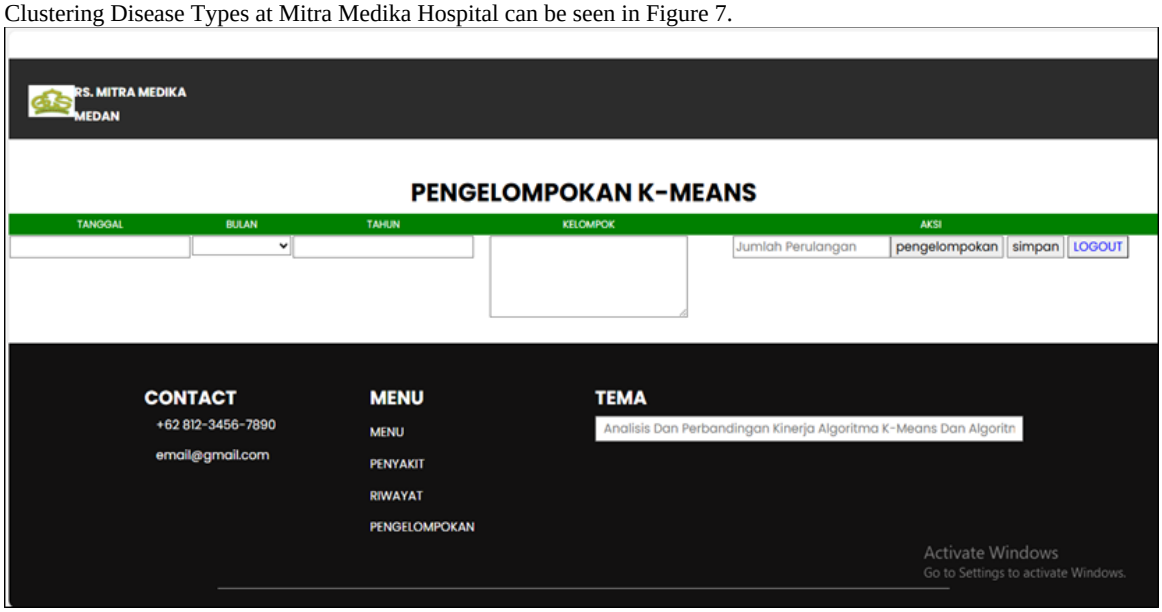

Fig. 7: Result Form

4.2. Discussion

Calculating new cluster centers. The criteria have 2 members, namely gender and type of disease. The latest cluster center for the next cluster is calculated based on the average coordinates to the members are:
C1(new) = First Total/First Total Non-Empty Values
= 237/346 = 0.684971
C2(new) = Second Total/First Total Empty Values
= 1329/154 = 3.84104
C3(new) = Third Total/Second Total Non-Empty Values
= 385/154 = 3.422078
C4(new) = Fourth Total/Second Total Empty Values
= 1968/346 = 8.935065
So:

Cluster	C0	C1
0	0.684971098	3.841040462
1	3.422077922	8.935064935

Repeat steps 1 to 4 until the cluster value does not change, so that it becomes as follows:
Cluster 0: From 245 Men (Suffering Between Diseases 1-5)
Cluster 1: From 255 Women (Suffering Between Diseases 6-10)

4.3. Trial Result

After conducting a trial of the application, it can be concluded that the results obtained are; The design interface is in accordance with the resulting interface, The Decision Tree method has been applied to the application created, The application interface is user friendly so that users can use it easily, The application that has been created runs well, The application that has been created does not have any logical errors.
The disadvantages of the application in this study include; The application that has been created uses two algorithms, making it difficult to search for results; The application that has been created requires History data in Cluster searches; The execution process takes a long time.

5. Conclusion

The conclusion of the Analysis and Comparison of the Performance of the K-Means Algorithm and the X-Means Algorithm in Clustering Disease Types at Mitra Medika Hospital can be seen as follows:

a. By using the K-Means Algorithm and the X-Means Algorithm, clustering can be produced.
b. By using the Disease History data, the K-Means Algorithm and the X-Means Algorithm methods can be applied to determine clusters.

c. By using web programming, it can produce an Analysis and Comparison of the Performance of the K-Means Algorithm and the X-Means Algorithm in Clustering Disease Types at Mitra Medika Hospital.

6. Suggestions

Suggestions from the Analysis and Comparison of the Performance of the K-Means Algorithm and the X-Means Algorithm in Clustering Disease Types at Mitra Medika Hospital can be seen as follows:

a. It is better if the application that has been created uses only one algorithm.
b. It is better if the application that has been created has instructions for use.
c. It is better to apply it on a desktop basis.

References

[1] Adil, A., Darma, I. M. Y., Heroe Santoso, & Lalu Sofiyandi Prayatna. (2023). Penerapan Algoritma K-Means Berbasis Spasial Untuk Pengelompokan Potensi Virus Covid-19 Di Kabupaten Dompu. *SATIN - Sains dan Teknologi Informasi*, *9*(1), 64–73.\

[2] Al Amin, M., & Juniati, D. (2021). Math Unesa. *Jurnal Ilmiah Matematika*, *9*(2), 437–446. https://media.neliti.com/media/publications/249234-model-infeksi-hiv-dengan-pengaruh-percob-b7e3cd43.pdf

[3] Aldiyatna, K., Rahaningsih, N., & Dana, R. D. (2024). *Penerapan Data Mining UntuK - clustering Penyakit Diare Menggunakan Algoritma K-Means (Studi Kasus : Puskesmas Beber)*. *8*(3), 3124–3131.

[4] [Bayu Prasetyo, R., Agus Pranoto, Y., & Primaswara Prasetya, R. (2023). Implementasi Data Mining Menggunakan Algoritma *K-Means Cluste*ring Penyakit Pasien Rawat Jalan Pada Klinik Dr. Atirah Desa Sioyong, Sulteng. *JATI (Jurnal Mahasiswa Teknik Informatika)*, *7*(4), 2144–2151.

[5] Cahyo Nugroho, A. (2019). Rancang Bangun Sistem Informasi Manajemen Surat Tugas Berbasis Web Menggunakan Waterfall Model. *Jurnal Informatika: Jurnal Pengembangan IT*, *4*(2), 146–151.

[6] Fitri, E. (2020). Sentiment Analysis of the Ruangguru Application Using Naive Bayes, Random Forest and Support Vector Machine Algorithms. *Jurnal Transformatika*, *18*(1), 71.

[7] Hermawan, Y. (2023). Comparison of K-Means Clustering Algorithm with Naive Bayes Clussifier Algorithm to Know Students Reading Interest Using Library Data Mining. *Journal of Computer Science and Big Data*, *1*(September), 20–25.

[8] Mahatmi, M. W., & Sebatubun, M. M. (2022). Strategi Komunikasi Stakeholders Dengan Pendekatan Data Alumni Menggunakan Data Mining. *PRecious: Public Relations Journal*, *2*(1), 26–42.

[9] Nurhakim, B., Septiani, I., Anam, K., & Pratama, D. (2024). Penarapan Algoritma *K-Means Cluste*ring Dalam Menganalisis Resiko Penyakit Stroke. *JATI (Jurnal Mahasiswa Teknik Informatika)*, *8*(1), 318–322.

[10] Noor, H., Dharmawati, A., & Qur'ana, T. W. (2021). Penerapan Algoritma *K- Means Cluste*ring *Analysis* Pada Kasus Penderita Hiv/Aids (Studi Kasus Kabupaten Banjar). *Technologia: Jurnal Ilmiah*, *12*(2), 72.

[11] Parlika, R., & Pratama, A. (2019). Aplikasi Mesin Penjawab Pesan Berbasis Bot Telegram, Php, Dan Mysql. *SCAN - Jurnal Teknologi Informasi dan Komunikasi*, *14*(3), 1–9.