

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

Herdawani Afdilla

Department of Information System, Universitas Muhammadiyah Sumatera Utara, Indonesia

---

---

### ABSTRACT

The system currently used by the hospital is still manual in managing patient data and information. What happens at Mitra Medika Hospital is that it is difficult to provide medical needs related to the diseases experienced by patients, considering that there are many types of diseases, so they provide many medical needs. Several inpatients have used BPJS facilities for various diseases suffered by the patient to carry out further examinations so that they can recover from the disease they are suffering from. Mitra Medika Hospital only looks at medical needs based on the disease suffered by the patient, however, seeing the large amount of patient history data makes it very difficult for Mitra Medika Hospital to find out the groups of diseases that patients often experience. This research uses a quantitative approach which starts from a theoretical framework, expert ideas, and researchers' understanding based on their experience, then developed into problems and solutions that are proposed to obtain justification (verification) or assessment in the form of empirical data support in the field. Here we apply a data mining pattern where data mining is extracting very large data (big data). Cluster 0: Of 245 Men (Suffering from 1-5 Diseases) Cluster 1: Of 255 Women (Suffering from 6-10 Diseases) Using the K-Means Algorithm and X-Means Algorithm can produce clustering. By using disease history data, you can apply the K-Means Algorithm and X-Means Algorithm methods to determine clusters. By using web programming, we can produce an analysis and comparison of the performance of the K-Means algorithm and the X-Means algorithm in clustering disease types in hospitals. Medika Partners.

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



This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.

---

### Corresponding Author:

Herdawani Afdilla,  
Department of Information System,  
Universitas Muhammadiyah Sumatera Utara,  
Jalan Kapten Muktar Basri No 3 Medan 20238, Indonesia.  
Email: herdawaniafdilla@gmail.com

---

---

## 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. Without well-being in health, every aspect of people's lives will be disrupted. Thus, the government is required to provide optimal health services for all levels of society. 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 used by the hospital is currently still manual in managing patient data and information, therefore there needs to be a 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. With the data and information left behind in the form of documents such as, patient name, age, type of disease.

The problem that occurs at Mitra Medika Hospital is the difficulty of providing medical needs related to the patient's illness, considering the many types of illnesses so that they also provide many medical needs such as medical equipment and medicines. 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. Therefore, a method is needed that can help Mitra Medika Hospital in grouping types of illnesses. One method that can be used is data mining. In data mining, there are several types of algorithms with

different functions. Here the researcher uses two algorithms, namely the K-Means and X-Means Algorithms where the researcher makes a comparison between the two algorithms. The K-Means algorithm is an algorithm that can cluster or group types of classes based on data taken by researchers (Rizki, 2022). While the X-Means Algorithm aims to perfect the K-Means algorithm. So that this clustering can run perfectly.

In this study, researchers built a website-based system where the system will be used by Mitra Medika Hospital as a container for information about patient data that can be accessed easily and effectively by hospital employees. This system is designed using the PHP programming language and its database uses MySQL.

## 2. RESEARCH METHOD

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). The data collection technique used by researchers to collect data in this study is Literature Study, which is 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 Mitra Medika Hospital. In addition, Field Study, which is the collection of data obtained through research by going directly to the research location to find facts related to the research subject. There are 2 data analysis techniques in this study, namely using the K-Means Algorithm and the X-Means Algorithm.

## 3. RESULTS AND DISCUSSION

### A. Results

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:

#### 1. Login Form

Login Form of Analysis and Comparison of Performance of K-Means Algorithm and X-Means Algorithm in Clustering Disease Types in Mitra Medika Hospital can be seen in Figure 1.

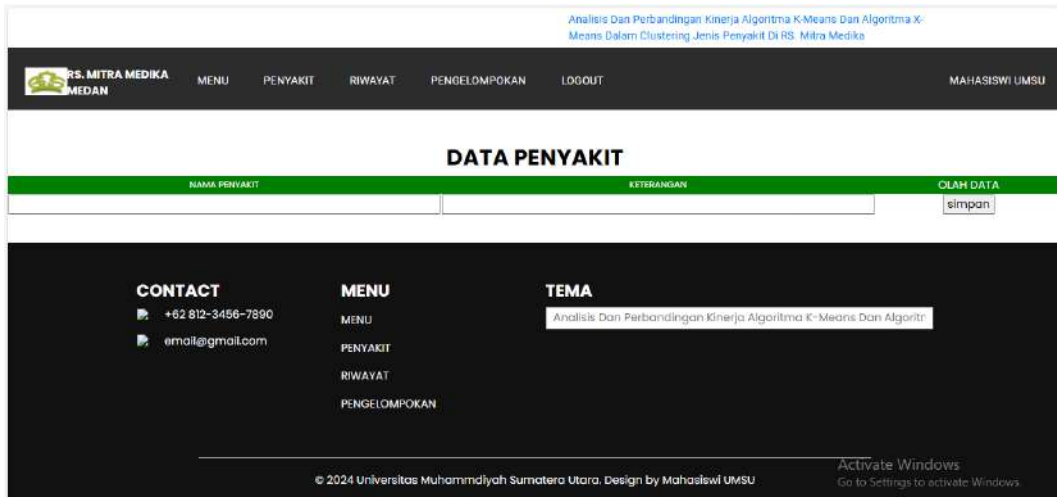
#### 2. Menu Form

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 2.



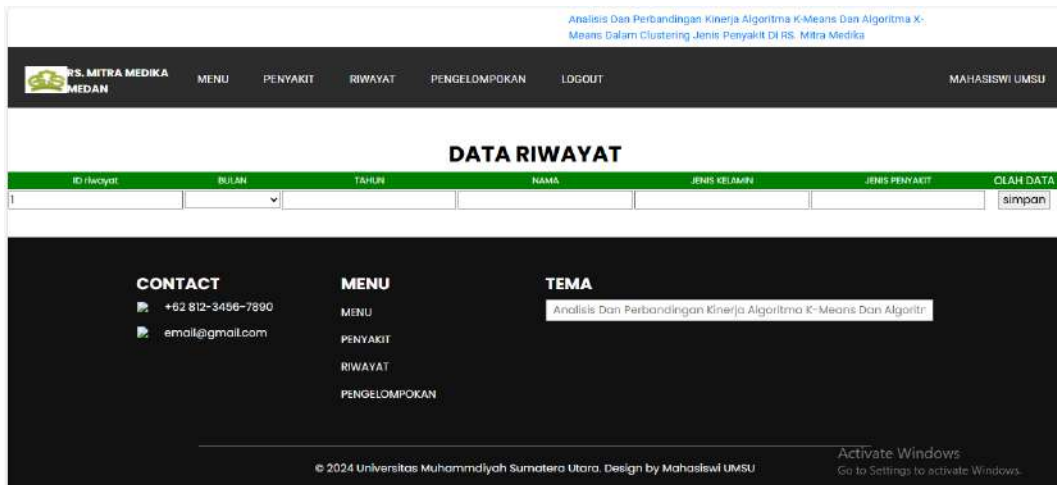
3. Disease Form

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 3.



4. History Form

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



## 5. 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 5.

## 6. 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 6.

## B. Discussion

The discussion includes device requirements, results used and testing in this study.

### 1. Device Requirements

The hardware and software requirements for creating the application are as follows:

- a. One laptop unit with the following specifications:
  - 1) Minimum Core 2 Duo Processor
  - 2) Minimum 1 GB RAM
  - 3) Minimum 80 GB Hardisk
- b. Software with the following specifications:
  - 1) Windows Operating System
  - 2) Notepad++
  - 3) Appserv

### 2. History Data from Mitra Medika Hospital

### 3. K-Means and X-Means Methods

This study uses the K-Means and X-Means Methods which are used as the process of finding results. The following are the stages of the K-Means and X-Means Methods:

- a. K-Means

The following is the calculation of the K-Means method:

- 1) Setting the initial value of the midpoint

Determine the cluster center randomly on the initial data in table III.2. The first cluster used in the calculation is the minimum and maximum values for gender, namely C0(1, 2) and the minimum and maximum values for disease type C1(1, 10).

Cluster	C0	C1
0	1	1
1	2	10

- 2) Calculate the distance of an object to the centroid using the Euclidean distance formula  
The calculation of the distance between the first student and the center of the first cluster is:

$$d_{\text{jeniskelamin}} = \sqrt{(\text{Jenis\_Kelamin} - \text{Cluster 0 C0})^2 + (\text{Jenis\_Penyakit} - \text{Cluster 0 C1})^2}$$

$$d_{\text{jenispenyakit}} = \sqrt{(\text{Jenis\_Kelamin} - \text{Cluster 1 C0})^2 + (\text{Jenis\_Penyakit} - \text{Cluster 1 C1})^2}$$

- 3) Perform object clustering by inserting each object into a cluster (group) based on its minimum distance.

- 4) 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:

$$\begin{aligned} \text{C1(new)} &= \text{First Total/First Non-Empty Total Value} \\ &= 379/255 = 1.486275 \end{aligned}$$

$$\begin{aligned} \text{C2(new)} &= \text{Second Total/Total First Blank Value} \\ &= 737/245 = 2.890196 \end{aligned}$$

$$\begin{aligned} \text{C3(new)} &= \text{Third Total/Total Value Not Blank Second} \\ &= 385/245 = 1.571429 \end{aligned}$$

$$\begin{aligned} \text{C4(new)} &= \text{Fourth Total/Total Second Blank Value} \\ &= 1968/255 = 8.032653 \end{aligned}$$

So that:

Cluster	C0	C1
0	1.486275	2.890196
1	1.571429	8.032653

- 5) Repeat steps 1 to 4 until the cluster value does not change, so that it becomes as follows:

**Cluster 0: Out of 245 Men (Suffering Between Diseases 1-5)**

**Cluster 1: Of 255 Women (Suffering Between 6-10 Diseases)**

b. X-Means

The following is the calculation of the X-Means method:

- 1) Setting the initial value of the midpoint

Determine the cluster center randomly on the initial data in table III.2. The first cluster used in the calculation is the total gender/total data and type of disease/total data:

Cluster	C0	C1
0	236/500=0.472	Rata-rata=5.41
1	264/500=0.528	Maksimal=10

- 2) Calculate the distance of an object to the centroid using the Euclidean distance formula  
The calculation of the distance between the first student and the center of the first cluster is:

$$d_{\text{jeniskelamin}} = \sqrt{(\text{Jenis\_Kelamin} - \text{Cluster 0 C0})^2 + (\text{Jenis\_Penyakit} - \text{Cluster 0 C1})^2}$$

$$d_{\text{jenispenyakit}} = \sqrt{(\text{Jenis\_Kelamin} - \text{Cluster 1 C0})^2 + (\text{Jenis\_Penyakit} - \text{Cluster 1 C1})^2}$$

- 3) Perform object clustering by inserting each object into a cluster (group) based on its minimum distance
- 4) 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:  
 $C1(\text{new}) = \text{First Total} / \text{First Non-Empty Total Value}$   
 $= 237 / 346 = 0.684971$   
 $C2(\text{new}) = \text{Second Total} / \text{Total First Blank Value}$   
 $= 1329 / 154 = 3.84104$   
 $C3(\text{new}) = \text{Third Total} / \text{Total Value Not Blank Second}$   
 $= 385 / 154 = 3.422078$   
 $C4(\text{new}) = \text{Fourth Total} / \text{Total Second Blank Value}$   
 $= 1968 / 346 = 8.935065$

So that:

Cluster	C0	C1
0	0.684971098	3.841040462
1	3.422077922	8.935064935

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

### C. Program Trial

The system test aims to ensure that the system is ready to use. The instrument used to conduct this test is by using Blackbox Testing:

#### Login Form Blackbox Testing

No	Form Login	Keterangan	Validitas
1.	Klik Tombol Submit	Aplikasi menampilkan form menu	Valid

#### Menu Form Blackbox Testing

No	Form Menu	Keterangan	Validitas
1.	Klik Tombol Menu	Aplikasi menampilkan form Menu	Valid
2.	Klik Tombol Penyakit	Aplikasi menampilkan form Penyakit	Valid
3.	Klik Tombol Riwayat	Aplikasi menampilkan form Riwayat	Valid
4.	Klik Tombol Cluster	Aplikasi menampilkan form Cluster	Valid
5.	Klik Tombol Hasil	Aplikasi menampilkan form hasil	Valid

#### Disease Form Blackbox Testing

No	Form Penyakit	Keterangan	Validitas
1.	Klik Tombol Simpan	Aplikasi menyimpan seluruh data di textbox ke dalam table database	Valid
2.	Klik Tombol Ubah	Aplikasi mengubah isi di table database sesuai data yang diubah	Valid
3.	Klik Tombol Hapus	Aplikasi menghapus isi data di database	

#### History Form Blackbox Testing

No	Form Riwayat	Keterangan	Validitas
1.	Klik Tombol Simpan	Aplikasi menyimpan seluruh data di textbox ke dalam table database	Valid
2.	Klik Tombol Ubah	Aplikasi mengubah isi di table database sesuai data yang diubah	Valid
3.	Klik Tombol Hapus	Aplikasi menghapus isi data di database	

### Cluster Form Blackbox Testing

No	Form Cluster	Keterangan	Validitas
1.	Klik Tombol Simpan	Aplikasi menyimpan seluruh data di textbox ke dalam table database	Valid
2.	Klik Tombol Ubah	Aplikasi mengubah isi di table database sesuai data yang diubah	Valid
3.	Klik Tombol Hapus	Aplikasi menghapus isi data di database	

### Result Form Blackbox Testing

No	Form Hasil	Keterangan	Validitas
1.	Klik Tombol Simpan	Aplikasi menyimpan seluruh data di textbox ke dalam table database	Valid
2.	Klik Tombol Ubah	Aplikasi mengubah isi di table database sesuai data yang diubah	Valid
3.	Klik Tombol Hapus	Aplikasi menghapus isi data	Valid

### Cluster Form Blackbox Testing

No	Form Cluster	Keterangan	Validitas
1.	Klik Tombol Simpan	Aplikasi menyimpan seluruh data di textbox ke dalam table database	Valid
2.	Klik Tombol Cluster	Aplikasi menampilkan Cluster dari hasil KNN	Valid

#### 1) Trial Results

After conducting a trial of the application, the following results can be concluded:

- a) The design interface is in accordance with the resulting interface.
- b) The Decision Tree method has been applied to the application created.
- c) The application interface is user friendly so that users can use it easily.
- d) The application that has been created runs well.
- e) The application that has been created does not have any logical errors

#### 2) Application Disadvantages

The shortcomings of the application in this study include:

- a) The application that has been created uses two algorithms, making it difficult to search for results.
- b) The application that has been created requires History data in Cluster search.
- c) The execution process takes a long time.

## 4. 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 is 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 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

## 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-percoba-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 Clustering 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 Classifier 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 Clustering 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 Clustering Analysis Pada Kasus Penderita Hiv/Aids (Studi Kasus Kabupaten Banjar). *Technologia: Jurnal Ilmiah*, 12(2), 72.
- [11] Parlita, R., & Pratama, A. (2019). Aplikasi Mesin Penjawab Pesan Berbasis Bot Telegram, Php, Dan Mysql. *SCAN - Jurnal Teknologi Informasi dan Komunikasi*, 14(3), 1–9.
- [12] Rahmadayanti, F., Anggraini, I., & Susanti, T. (2023). Pengklasterisasian Data Penyakit Hipertensi dengan Menggunakan Metode Algoritma K-Means. *Journal of Information System Research (JOSH)*, 4(2), 737–741.
- [13] Ramadhan, W., & Putra, S. H. (2022). Aplikasi Absensi Mahasiswa dan Dosen Politeknik Ganesha Medan Berbasis Web Menggunakan PHP dan MySql. *Remik*, 6(3), 526–533.
- [14] Rizki, M. I., Taqqiyuddin, T. A., & Cerelia, J. J. (2021). K-Medoids Clustering dengan Jarak Dynamic Time Warping dalam Mengelompokkan Provinsi di Indonesia Berdasarkan Kasus Aktif COVID-19. *Prosiding Seminar Nasional Matematika*, 4, 685–692.
- [15] Saputra Sy, Y. (2022). Klasterisasi Pasien Rawat Inap Peserta BPJS Berdasarkan Jenis Penyakit Menggunakan Algoritma K-Means. *Jurnal Sistim Informasi Dan Teknologi*, 5, 33–37.
- [16] Sugianto, C. A., Rahayu, A. H., & Gusman, A. (2020). Algoritma K-Means untuk Pengelompokkan Penyakit Pasien pada Puskesmas Cigugur Tengah. *Journal of Information Technology*, 2(2), 39–44.
- [17] Surbakti, N. K. (2021). Data Mining Pengelompokan Pasien Rawat Inap Peserta BPJS Menggunakan Metode Clustering (Studi Kasus : RS.U.Bangkalan). *Journal of Information and Technology*, 1(2), 47–53.
- [18] Ucha Putri, S., Irawan, E., Rizky, F., Tunas Bangsa, S., -Indonesia Jln Sudirman Blok No, P. A., & Utara, S. (2021). Implementasi Data Mining Untuk Prediksi Penyakit Diabetes Dengan Algoritma C4.5. *Januari*, 2(1), 39–46.
- [19] Wahyudi, I., Sulthan, M. B., & Suhartini, L. (2021). Analisa Penentuan Cluster Terbaik Pada Metode K-Means Menggunakan Elbow Terhadap Sentra Industri Produksi Di Pamekasan. *Jurnal Aplikasi Teknologi Informasi dan Manajemen (JATIM)*, 2(2), 72–81.
- [20] Wandana, J., Defit, S., & Sumijan, S. (2020). Klasterisasi Data Rekam Medis Pasien Pengguna Layanan BPJS Kesehatan Menggunakan Metode Algoritma K-Means. *Jurnal Informasi Dan Teknologi*, 2, 4–9.
- [21] Zufria, I., & Damaiani Iskandar, I. (2024). Clustering Pasien Rawat Inap Di RS. USU Menggunakan Algoritma K-Means Clustering of Inpatients at USU Hospital Using the Algoritma K-Means Algorithm. *JOURNAL OF COMPUTER SCIENCE AND INFORMATICS ENGINEERING (CoSIE)*, 03(2), 54–63.