

## PERNYATAAN

Saya menyatakan bahwa skripsi ini minim plagiasi, dan apabila di kemudian hari terbukti terdapat plagiat dalam skripsi ini, maka saya bersedia menerima sanksi sesuai ketentuan peraturan perundang-undangan.

Bojonegoro, 12 Juni 2024



Putri Ayu Ningtiyas  
NIM. 2120200501



# UNUGIRI

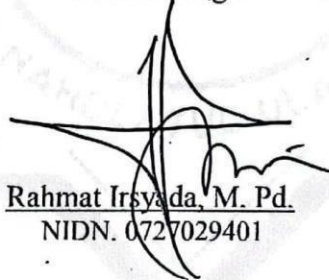
## HALAMAN PERSETUJUAN

Nama : Putri Ayu Ningtiyas  
NIM : 2120200501  
Program Studi/Fakultas : Teknik Informatika/ Fakultas Sains dan Teknologi  
Judul : Perbandingan Metode K-Means dan Fuzzy C-Means  
untuk Pengelompokan Diagnosa Kanker Payudara.

Telah disetujui dan dinyatakan memenuhi syarat untuk diajukan dalam sidang skripsi.

Bojonegoro, 13 Juni 2024

Pembimbing I



Rahmat Irsyada, M. Pd.  
NIDN. 0727029401

Pembimbing II




Aprillia Dwi Ardianti, S. Si., M. Pd.  
NIDN. 0726048902

## HALAMAN PENGESAHAN

Nama : Putri Ayu Ningtiyas  
NIM : 2120200501  
Program Studi/Fakultas : Teknik Informatika/ Fakultas Sains dan Teknologi  
Judul : Perbandingan Metode K-Means dan Fuzzy C-Means  
untuk Pengelompokan Diagnosa Kanker Payudara.


Telah dipertahankan di hadapan penguji pada tanggal 20 Juni 2024

Dewan Penguji  
Penguji I



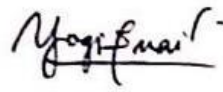
Zaki Alawi, S. Kom., M.M  
NIDN. 0709068906

Tim Pembimbing  
Pembimbing I




Rahmat Irsyada, M. Pd.  
NIDN. 0727029401

Penguji II



Dr. H. Yogi Prana Izza, Lc. M.A.  
NIDN.0731127601

Pembimbing II



Aprillia Dwi Ardianti, S. Si., M. Pd.  
NIDN. 0726048902

Mengetahui,  
Dekan Fakultas Sains dan Teknologi



Muhammad Jahhar Vikri, M.Kom  
NIDN. 0712078803

Mengetahui,  
Ketua Program Studi



Uca Pradema Sanjaya, M. Kom  
NIDN. 07120728903

## **MOTTO**

Lebih baik bergerak dan merasakan gagal, daripada tidak sama sekali.

(Putri Ayu Ningtiyas)

Buat dan ciptakan peluang sendiri, jangan hanya menunggu orang lain.

(Putri Ayu Ningtiyas)

## **PERSEMBAHAN**

Persembahan tugas akhir dan rasa terimakasih saya ucapkan untuk :

1. Tuhan Yang Maha Esa yang telah melancarkan semua proses mulai dari bimbingan, seminar proposal, pengerjaan skripsi sampai seminar hasil ujian skripsi
2. Ayah dan ibu tercinta yang tanpa lelah sudah mendukung semua keputusan dan pilihan dalam hidup saya serta tidak pernah putus mendoakan saya.
3. Bapak Rahmat Irsyada, M. Pd selaku dosen pembimbing 1 saya dan Ibu Aprillia Dwi Ardianti, S. Si., M. Pd yang telah memberikan tenaga, waktu, dan pikiran untuk membimbing dan mengarahkan skripsi saya hingga saya dinyatakan lulus.
4. Sahabat seperjuangan yang sudah berkenan bertukar pikiran, memotivasi sehingga saya bisa menyelesaikan skripsi ini.
5. Terakhir, saya ucapkan terimakasih kepada diri sendiri yang sudah berjuang dan bertahan sampai sejauh ini.

## KATA PENGANTAR

Syukur Alhamdulillah penulis panjatkan kehadiran Allah SWT atas rahmat dan hidayah-Nya penyusunan skripsi yang berjudul “Perbandingan Metode *K-Means* dan *Fuzzy C-Means* untuk Pengelompokan Diagnosa Kanker Payudara” dapat di selesaikan dengan baik. Skripsi ini merupakan syarat untuk memperoleh gelar sarjana (S.Kom) pada Program Studi Teknik Informatika Fakultas Sains dan Teknologi Universitas Nahdlatul Ulama Sunan Giri Bojonegoro

Dalam penulisan skripsi ini, penulis mendapatkan banyak motivasi, inspirasi, dukungan, dan bimbingan dari berbagai pihak, sehingga penulis dapat menyelesaikan skripsi ini tepat waktu. Penulis ingin menggunakan kesempatan ini untuk mengucapkan rasa terimakasih kepada:

1. Bapak K. H. Jauharul Ma'arif, M. Pd. I. selaku Rektor Universitas Nahdlatul Ulama Sunan Giri Bojonegoro.
2. Bapak Muhammad Jauhar Vikri, M. Kom. selaku Dekan Fakultas Sains dan Teknologi.
3. Bapak Ucta Pradema Sanjaya, M. Kom. selaku Ketua Program Studi Teknik Informatika Universitas Nahdlatul Ulama Sunan Giri Bojonegoro.
4. Bapak Rahmat Irsyada, M. Pd selaku Dosen Pembimbing I yang telah memberikan bimbingan serta saran yang membangun demi terselesaikannya proposal skripsi ini.
5. Ibu Aprillia Dwi Ardianti, S. Si., M.Pd selaku Dosen Pembimbing II yang telah memberikan bimbingan serta saran yang membangun demi terselesaikannya proposal skripsi ini.
6. Bapak Sunu Wahyudhi, M.Pd selaku Dosen Wali yang telah memberikan bimbingan serta nasehat selama perkuliahan.
7. Kedua orang tua dan keluarga besar yang selalu mendukung, mendoakan dan memotivasi penulis dalam menyelesaikan skripsi.
8. Sahabat dan teman-teman yang selalu mendukung penulis dalam menyelesaikan skripsi ini.
9. Semua pihak yang telah membantu dan tidak dapat penulis sebutkan satu per satu.

Penulis menyadari bahwa Skripsi ini masih kurang sempurna sehingga kepada pembaca, kiranya dapat memberikan saran yang sifatnya membangun agar kekurangan-

kekurangan yang ada dapat diperbaiki. Penulis juga berharap semoga skripsi ini dapat berguna pada diri pribadi penulis, almamater, khususnya dalam rangka meningkatkan kualitas pendidikan di masa yang akan datang. Amin.

Bojonegoro, 12 Juni 2024

Putri Ayu Ningtias



## ABSTRACT

Ningtiyas, Putri Ayu. 2024. "Comparison of k-means and fuzzy c-means methods for classifying breast cancer diagnoses". Thesis. Informatics Engineering Study Program, Faculty of Science and Technology, Nahdatul Ulama Sunan Giri University. Main Supervisor Rahmat Irsyada, M.Pd and Assistant Supervisor Aprillia Dwi Ardianti, S. Si., M.Pd. Breast cancer is one of the causes of death in women. Breast cancer is said to be a very malignant disease. However, breast cancer is not a contagious disease. Thus, breast cancer requires special treatment, so that in this case early breast cancer prevention can be carried out. If breast cancer is not prevented early or even gets more optimal treatment, breast cancer cases and death rates will continue to increase. In the medical environment, breast cancer examination or detection uses mammography, mammography is a screening method that is able to identify cancer based on the variable physical symptoms of the disease that appears. Therefore, to identify cancer diagnoses, it is necessary to add additional tools that can help identify symptoms of breast cancer, whether the results of the patient's mammography are classified as malignant or benign cancer. However, this research focuses on the clustering method that will be used to group breast cancer diagnoses based on the patient's physical symptom variables. This research aims to find out which algorithm is more optimal to use for grouping breast cancer data. The algorithms used in this research are the k-means and fuzzy c-means algorithms. In this research, determining the best cluster for each algorithm uses the Dunn Index and Davies Bouldin Index using R software. From comparative research on the k-means and fuzzy c-means methods, it is found that the k-means algorithm is a more optimal algorithm used in grouping data. breast cancer compared to the fuzzy c-means algorithm, with a Davies Bouldin Index k-means value of 1.177676, while fuzzy c-means is 1.323947, the closer the value is to zero, the more optimal the results of the Davies Bouldin Index value.

*Keywords : Clustering, Davies Bouldin Index, Dunn Index, K-Means, Fuzzy C-Means.*

## ABSTRAK

Ningtiyas, Putri Ayu. 2024. *Perbandingan metode k-means dan fuzzy c-means untuk mengelompokkan diagnosa kanker payudara*. Skripsi. Program Studi Teknik Informatika Fakultas Sains dan Teknologi Universitas Nahdatul Ulama Sunan Giri. Pembimbing Utama Rahmat Irsyada, M.Pd dan Pembimbing Pendamping Aprillia Dwi Ardianti, S. Si., M.Pd. Kanker payudara merupakan salah satu penyebab kematian pada Wanita. kanker payudara disebut sebagai penyakit yang sangat ganas. Namun, kanker payudara bukanlah penyakit yang menular. Dengan demikian, kanker payudara memerlukan penanganan secara khusus, sehingga dalam hal tersebut dapat dilakukan pencegahan kanker payudara secara dini. Jika tidak dilakukan pencegahan kanker payudara secara dini atau bahkan tidak mendapatkan penanganan yang lebih optimal, maka kasus penyakit kanker payudara dan tingkat kematian akan terus meningkat. Pada lingkungan medis pemeriksaan atau deteksi kanker payudara menggunakan mamografi, mamografi adalah metode skrining yang mampu mengidentifikasi kanker berdasarkan variabel gejala-gejala fisik dari penyakit yang muncul. Oleh karena itu, untuk mengidentifikasi diagnosa kanker perlu ditambahkan perangkat tambahan yang mampu membantu mengidentifikasi gejala kanker payudara, apakah hasil dari mamografi tersebut pasien tergolong kanker ganas atau kanker jinak. Namun, pada penelitian ini difokuskan kepada metode *clustering* yang akan digunakan untuk mengelompokkan diagnosa kanker payudara tersebut berdasarkan variabel gejala fisik pasien. Penelitian ini mempunyai tujuan untuk mengetahui algoritma mana yang lebih optimal digunakan untuk pengelompokkan data kanker payudara, algoritma yang digunakan dalam penelitian ini adalah algoritma k-means dan fuzzy c-means. Pada penelitian ini penentuan cluster terbaik dari masing-masing algoritma menggunakan Dunn Index dan Davies Bouldin Index dengan memanfaatkan software R. Dari penelitian perbandingan metode k-means dan fuzzy c-means diperoleh bahwasannya algoritma k-means merupakan algoritma yang lebih optimal digunakan dalam pengelompokkan data kanker payudara disbanding algoritma fuzzy c-means, dengan nilai Davies Bouldin Index k-means sebesar 1.177676, sementara fuzzy c-means sebesar 1.323947, semakin mendekati nilai nol maka semakin optimal hasil dari nilai Davies Bouldin Indexnya.

Kata kunci : *Clustering*, Davies Bouldin Index, Dunn Index, K-Means, Fuzzy C-Means.



# DAFTAR ISI

|  | Halaman     |
|--|-------------|
| <b>HALAMAN SAMPUL LUAR</b> .....                     | <b>i</b>    |
| <b>HALAMAN SAMPUL DALAM</b> .....                    | <b>ii</b>   |
| <b>HALAMAN PERNYATAAN KEASLIAN TULISAN</b> .....     | <b>iii</b>  |
| <b>HALAMAN PERSETUJUAN</b> .....                     | <b>iv</b>   |
| <b>HALAMAN PENGESAHAN</b> .....                      | <b>v</b>    |
| <b>HALAMAN MOTTO DAN PERSEMBAHAN</b> .....           | <b>vi</b>   |
| <b>KATA PENGANTAR</b> .....                          | <b>vii</b>  |
| <b>ABSTRACT</b> .....                                | <b>ix</b>   |
| <b>ABSTRAK</b> .....                                 | <b>x</b>    |
| <b>DAFTAR ISI</b> .....                              | <b>xi</b>   |
| <b>DAFTAR TABEL</b> .....                            | <b>xv</b>   |
| <b>DAFTAR GAMBAR</b> .....                           | <b>xvi</b>  |
| <b>DAFTAR LAMPIRAN</b> .....                         | <b>xvii</b> |
| <b>BAB I PENDAHULUAN</b> .....                       | <b>1</b>    |
| 1.1 Latar Belakang.....                              | 1           |
| 1.2 Rumusan Masalah .....                            | 3           |
| 1.3 Batasan Masalah.....                             | 3           |
| 1.4 Tujuan Penelitian.....                           | 3           |
| 1.5 Manfaat Penelitian.....                          | 3           |
| 1. Manfaat Praktis.....                              | 3           |
| 2. Manfaat Teoritis .....                            | 3           |
| <b>BAB II TINJAUAN PUSTAKA DAN DASAR TEORI</b> ..... | <b>4</b>    |
| 2.1 Tinjauan Pustaka .....                           | 4           |
| 2.2 Dasar Teori .....                                | 8           |
| 2.2.1 Kanker Payudara .....                          | 8           |
| 2.2.2 Data Mining.....                               | 9           |
| 2.2.3 Tahapan Data Mining .....                      | 10          |
| 2.2.4 Clustering .....                               | 11          |
| 2.2.5 Algoritma K-Means.....                         | 11          |
| 2.2.6 Algoritme Fuzzy C-Means .....                  | 12          |

|  |           |
|--|-----------|
| 2.2.7 System Development Life Cycle (SDLC) ..... | 13        |
| 2.2.7.1 Analisis .....                           | 14        |
| 2.2.7.2 Desain .....                             | 14        |
| 2.1.7.2.1 Skenario Pengujian R .....             | 14        |
| 2.2.7.3 Implementasi .....                       | 15        |
| 2.1.7.3.1 PHP .....                              | 15        |
| 2.1.7.3.2 MySQL .....                            | 15        |
| 2.2.7.4 Testing .....                            | 15        |
| 2.1.7.4.1 Skenario Pengujian Sistem .....        | 15        |
| 2.1.7.5 Maintenance .....                        | 15        |
| <b>BAB III METODE PENELITIAN.....</b>            | <b>16</b> |
| 3.1 Obyek Penelitian.....                        | 16        |
| 3.2 Identifikasi Msalah .....                    | 16        |
| 3.3 Metode Penelitan .....                       | 16        |
| 3.4 Data Mining.....                             | 18        |
| 3.4.1 Tahap Perhitungan Manual .....             | 18        |
| 3.4.2 Model Pengembangan Sistem.....             | 25        |
| 3.4.2.1 Analisis.....                            | 26        |
| 3.4.2.1.1 Analisis Data .....                    | 26        |
| 3.4.2.1.2 Analisis Kebutuhan .....               | 26        |
| 3.4.2.1.3 Analisis Pengguna.....                 | 28        |
| 3.4.2.1.4 Analisis Sistem.....                   | 28        |
| 3.4.2.2 Desain.....                              | 29        |
| 3.4.2.2.1 Alur Sistem .....                      | 29        |
| 3.4.2.2.2 Use Case.....                          | 30        |
| 3.4.2.2.3 Rancangan Desain Aplikasi .....        | 31        |
| 3.4.2.3 Implementasi .....                       | 34        |
| 3.4.2.4 Testing.....                             | 34        |
| 3.4.2.4.1 Skenario Pengujian Sistem.....         | 35        |
| 3.4.2.5 Black box testing.....                   | 35        |

|   | Halaman   |
|---|-----------|
| 3.4.2.5.1 Angket Uji Kelayakan .....  | 36        |
| 3.4.2.6 Maintenance .....   | 39        |
| 3.5 Jadwal Kegiatan .....   | 39        |
| <b>BAB IV HASIL DAN PEMBAHASAN</b> .....  | <b>40</b> |
| 4.1 Hasil Penelitian.....   | 40        |
| 4.1.1 Implementasi perbandingan algoritma k-means dan fuzzy c-means pada data<br>kanker payudara dengan uji software R..... | 40        |
| 4.1.1.1 <i>Import Dataset</i> .....   | 40        |
| 4.1.1.2 Algoritma k-means .....   | 41        |
| 4.1.1.3 Algoritma fuzzy c-means.....  | 43        |
| 4.1.1.4 Uji validitas cluster.....  | 48        |
| 4.1.1.5 Implementasi algoritma clustering .....   | 49        |
| 4.1.2 Implementasi system.....  | 50        |
| 4.1.2.1 Implementasi halaman <i>login</i> .....   | 50        |
| 4.1.2.2 Implementasi halaman dataset.....   | 51        |
| 4.1.2.3 Implementasi halaman hasil <i>cluster</i> k-means.....  | 52        |
| 4.1.2.4 Implementasi halaman hasil cluster fuzzy c-means .....  | 53        |
| 4.1.2.5 Implementasi halaman hasil akurasi perbandingan .....   | 53        |
| 4.2 Pengujian .....   | 54        |
| 4.2.1 Pengujian system.....   | 54        |
| 4.2.1.1 <i>Testing</i> .....  | 54        |
| 4.2.1.1.1 Hasil pengujian black box.....  | 54        |
| 4.2.1.1.2 Hasil uji kelayakan.....  | 55        |
| 4.3 Pembahasan .....  | 57        |
| <b>BAB V</b> .....  | <b>59</b> |
| <b>PENUTUP</b> .....  | <b>59</b> |
| 5.1 Kesimpulan.....   | 59        |
| 5.2 Saran.....  | 60        |
| <b>DAFTAR PUSTAKA</b> .....   | <b>61</b> |
| <b>LAMPIRAN</b> .....   | <b>68</b> |

## DAFTAR TABEL

|   | Halaman |
|---|---------|
| Tabel 2.1 Penelitian Terdahulu .....  | 4       |
| Tabel 3.1 Centroid awal data .....  | 19      |
| Tabel 3.2 Iterasi awal .....  | 19      |
| Tabel 3.3 Jarak antar centroid .....  | 19      |
| Tabel 3.4 Centroid baru .....   | 20      |
| Tabel 3.5 Hasil iterasi kedua .....   | 20      |
| Tabel 3.6 Centroid baru iterasi ke-5 .....                                  | 20      |
| Tabel 3.7 Hasil iterasi kedua .....   | 21      |
| Tabel 3.8 Derajat keanggotaan Fuzzy C-Means .....                           | 21      |
| Tabel 3.9 Cluster pertama iterasi 1 .....                                   | 22      |
| Tabel 3.10 Cluster kedua iterasi 1 .....                                    | 22      |
| Tabel 3.11 P total iterasi 1 .....  | 23      |
| Tabel 3.12 Cluster pertama iterasi 1 .....                                  | 23      |
| Tabel 3.13 Cluster kedua iterasi 1 .....                                    | 24      |
| Tabel 3.14 P total iterasi 10 .....   | 24      |
| Tabel 3.15 Error fuzzy c-means .....  | 25      |
| Tabel 3.16 Dataset <i>Breast Cancer</i> .....                               | 26      |
| Tabel 3.17 Analisa kebutuhan software (Perangkat Lunak) .....               | 26      |
| Tabel 3.18 Pengujian <i>Black Box</i> .....                                 | 35      |
| Tabel 3.19 Rancangan hasil pengujian .....                                  | 36      |
| Tabel 3.20 Kriteria skala penelitian .....                                  | 36      |
| Tabel 3.21 Instrumen validasi .....   | 37      |
| Tabel 3.22 Rencana penelitian .....   | 39      |
| Tabel 4.1 Hasil uji matriks Euclidean Distance .....                        | 42      |
| Tabel 4.2 Hasil uji matriks Euclidean Distance dan Manhattan Distance ..... | 45      |
| Tabel 4.3 Nilai Dunn Index .....  | 46      |
| Tabel 4.4 Nilai akurasi perbandingan .....                                  | 49      |
| Tabel 4.5 Hasil pengujian sistem clustering kanker payudara .....           | 54      |
| Tabel 4.6 Hasil uji .....   | 55      |

## DAFTAR GAMBAR

|   | Halaman |
|---|---------|
| Gambar 2.1 Bidang Ilmu Data Mining<br>(Bidang Komputer Sains dan Pendidikan Informatika et al., (2017)..... | 9       |
| Gambar 2.2 Proses KDD Data Mining (Benri et al., 2015).....   | 10      |
| Gambar 2.3 Model Waterfall (Praja et al., 2022) .....   | 14      |
| Gambar 3.1 Skenario pengujian data mining pada software R.....  | 17      |
| Gambar 3.2 Model Waterfall (Praja et al.,2022) .....  | 25      |
| Gambar 3.3 Alur Sistem .....  | 29      |
| Gambar 3.4 Use Case Diagram.....  | 30      |
| Gambar 3.5 Tampilan Login Website.....  | 31      |
| Gambar 3.6 Tampilan Halaman <i>Dataset</i> .....  | 32      |
| Gambar 3.7 Tampilan halaman View Cluster K-means .....  | 32      |
| Gambar 3.8 Tampilan halaman View Cluster Fuzzy C-means .....  | 33      |
| Gambar 3.9 Tampilan Halaman View Akurasi .....  | 34      |
| Gambar 3.10 Skenario Pengujian Sistem .....   | 35      |
| Gambar 4.1 Flowchart Olah data.....   | 40      |
| Gambar 4.2 Flowchart Algoritma K-means .....  | 41      |
| Gambar 4.3 Hasil Clustering K-means .....   | 43      |
| Gambar 4.4 Alur Algoritma Fuzzy C-Means .....   | 43      |
| Gambar 4.5 Hasil Cluster Fuzzy C-Means .....  | 47      |
| Gambar 4.6 Hasil Cluster Fuzzy C-Means. ....  | 48      |
| Gambar 4.7 Hasil Cluster Fuzzy C-Means. ....  | 49      |
| Gambar 4.8 Akurasi clustering k-means.....  | 50      |
| Gambar 4.9 Implementasi halaman login. ....   | 51      |
| Gambar 4.10 Implementasi halaman dataset. ....  | 51      |
| Gambar 4.11 Implementasi halaman cluster k-means. ....  | 52      |
| Gambar 4.12 Implementasi halaman cluster fuzzy c-means. ....  | 53      |
| Gambar 4.13 Implementasi halaman hasil akurasi perbandingan .....   | 53      |

## DAFTAR BAGAN

|  | Halaman |
|--|---------|
| Bagan 4.1 Cluster optimum metode elbow ..... | 42      |
| Bagan 4.2 Manhattan Distance .....           | 45      |
| Bagan 4.3 Euclidean Distance .....           | 46      |



**UNUGIRI**

## DAFTAR LAMPIRAN

|  | Halaman |
|--|---------|
| Lampiran 1. Surat Pernyataan Black Box .....                 | 68      |
| Lampiran 2. Angket Software Testing Oleh Test Engineer ..... | 74      |
| Lampiran 3. Cek plagiasi.....                                | 83      |



**UNUGIRI**



**UNUGIRI**