

**IMPLEMENTASI ALGORITMA K-MEANS UNTUK *CLUSTERING* DATA
PENYAKIT DI PUSKESMAS KOTAGEDE 2 YOGYAKARTA**

SKRIPSI

**Disusun untuk memenuhi sebagai persyaratan
mencapai derajat Sarjana**



Disusun Oleh:

**MARWAH SILVIA SASMITA
1900018169**

**PROGRAM STUDI INFORMATIKA
FAKULTAS TEKNOLOGI INDUSTRI
UNIVERSITAS AHMAD DAHLAN
YOGYAKARTA
2024**

**IMPLEMENTATION OF K-MEANS ALGORITHM FOR CLUSTERING DISEASE
DATA AT KOTAGEDE 2 HEALTH CENTER YOGYAKARTA**

SKRIPSI

**Prepared to fulfill the requirements
to achieve the Bachelor degree**



Arranged by:

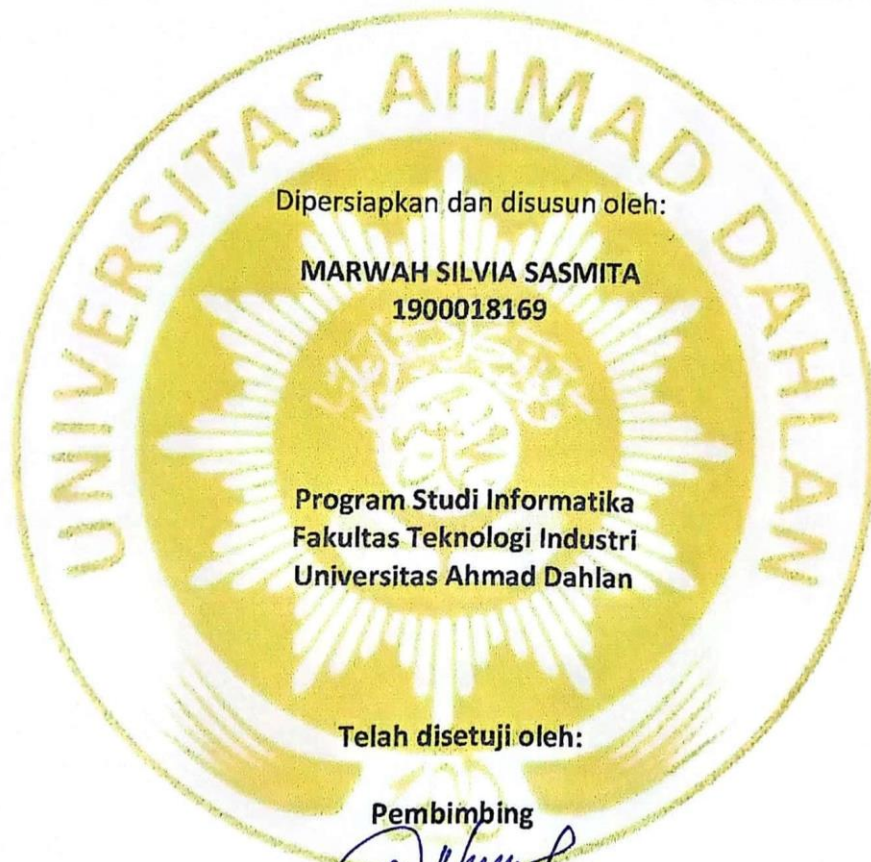
**MARWAH SILVIA SASMITA
1900018169**

**INFORMATICS STUDY PROGRAM
FACULTY OF INDUSTRIAL TECHNOLOGY
AHMAD DAHLAN UNIVERSITY
YOGYAKARTA
2024**

LEMBAR PERSETUJUAN PEMBIMBING

SKRIPSI

**IMPLEMENTASI ALGORITMA K-MEANS UNTUK *CLUSTERING* DATA
PENYAKIT DI PUSKESMAS KOTAGEDE 2 YOGYAKARTA**



Dipersiapkan dan disusun oleh:

MARWAH SILVIA SASMITA

1900018169

Program Studi Informatika
Fakultas Teknologi Industri
Universitas Ahmad Dahlan

Telah disetujui oleh:

Pembimbing

Sri Winiarti, S.T., M.Cs.

NIPM. 19751216 200103 011 0880702

LEMBAR PERSETUJUAN PENGUJI

SKRIPSI IMPLEMENTASI ALGORITMA K-MEANS UNTUK *CLUSTERING DATA* PENYAKIT DI PUSKEMAS KOTAGEDE 2 YOGYAKARTA

Dipersiapkan dan disusun oleh:

MARWAH SILVIA SASMITA
1900018169

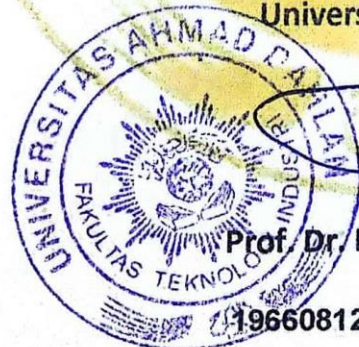
Telah dipertahankan di depan Dewan Penguji
pada tanggal 13 November 2023
dan dinyatakan memenuhi syarat
Susunan Dewan Penguji

Ketua : Sri Winiarti, S.T., M.Cs.  19/03/2024

Penguji 1 : Dr. Murinto S.Si., M.Kom  19/03/2024

Penguji 2 : Miftahurrahma Rosyda, S.Kom., M.Eng.  19/03/2024

Yogyakarta, 20 Maret 2024
Dekan Fakultas Teknologi Industri
Universitas Ahmad Dahlan




Prof. Dr. Ir. Siti Jamilatun, M.T.

19660812 199601 011 0784324

LEMBAR PERNYATAAN KEASLIAN

SURAT PERNYATAAN

Yang bertanda tangan di bawah ini:

Nama : MARWAH SILVIA SASMITA

NIM : 1900018169

Prodi : INFORMATIKA

Judul TA/Skripsi : IMPLEMENTASI ALGORITMA K-MEANS UNTUK *CLUSTERING* DATA PENYAKIT
DI PUSKEMAS KOTEGEDE 2 YOGYAKARTA

Dengan ini saya menyatakan bahwa Laporan Tugas Akhir ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar Ahli Madya/Kesarjanaan di suatu Perguruan Tinggi, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Yogyakarta, 13 November 2023

Mengetahui

Dosen Pembimbing



Sri Winiarti, S.T., M.Cs.

19751216 200103 011 0880702

Yang Menyatakan,



Marwah Silvia Sasmita

1900018169

Pernyataan Tidak Plagiat

Saya yang bertanda tangan di bawah ini:

Nama : Marwah Silvia Sasmita
NIM : 1900018169
Email : marwah1900018169@webmail.uad.ac.id
Program Studi : S1 Informatika
Fakultas : Teknologi Industri
Judul Tesis : Implementasi Algoritma K-means Untuk Clustering Data Penyakit di Puskesmas Kotagede 2 Yogyakarta

Dengan ini menyatakan bahwa:

1. Hasil karya yang saya serahkan ini adalah asli dan belum pernah diajukan untuk mendapatkan gelar keserjanaan baik di Universitas Ahmad Dahlan maupun di institusi pendidikan lainnya.
2. Hasil karya saya ini bukan saduran/terjemahan melainkan merupakan gagasan, rumusan, dan hasil pelaksanaan penelitian/implementasi saya sendiri, tanpa bantuan pihak lain, kecuali arahan pembimbing akademik dan narasumber penelitian.
3. Hasil karya saya ini merupakan hasil revisi terakhir setelah diujikan yang telah diketahui dan disetujui oleh pembimbing.
4. Dalam karya saya ini tidak terdapat karya atau pendapat yang telah ditulis atau dipublikasikan orang lain, kecuali yang digunakan sebagai acuan dalam naskah dengan menyebutkan nama pengarang dan dicantumkan dalam daftar pustaka.

Pernyataan ini saya buat dengan sesungguhnya. Apabila di kemudian hari terbukti ada penyimpangan dan ketidakbenaran dalam pernyataan ini maka saya bersedia menerima sanksi akademik berupa pencabutan gelar yang telah diperoleh karena karya saya ini, serta sanksi lain yang sesuai dengan ketentuan yang berlaku di Universitas Ahmad Dahlan.

Yogyakarta, 13 November 2023

Menyatakan


Marwah Silvia Sasmita

DA5ALX045174354

Pernyataan Persetujuan Akses

Saya yang bertanda tangan di bawah ini:

Nama : Marwah Silvia Sasmita
NIM : 1900018169
Email : marwaha1900018169@webmail.uad.ac.id
Program Studi : S1 Informatika
Fakultas : Teknologi Industri
Judul Tesis : Implementasi Algoritam K-Means Untuk Clustering Data
Penyakit Di Puskesmas Kotagede 2 Yogyakarta

Dengan ini Saya menyerahkan hak sepenuhnya kepada Perpustakaan Universitas Ahmad Dahlan untuk menyimpan, mengatur akses serta melakukan pengelolaan terhadap karya saya ini dengan mengacu pada ketentuan akses tesis elektronik sebagai berikut (beri tanda pada kotak):



Saya (~~tidak mengizinkan~~ **mengizinkan**)* karya tersebut diunggah ke dalam aplikasi Repository Perpustakaan Universitas Ahmad Dahlan.

Demikian pernyataan ini Saya buat dengan sebenarnya.

Yogyakarta, 13 November 2023

Yang Menyatakan



Marwah Silvia Sasmita

Mengetahui,

Dosen Pembimbing Skripsi


Sri Winiarti, S.T., M.Cs.

Motto

“Jika kamu lelah jangan menyerah tengoklah ke belakang dah lihat seberapa jauh kamu bertahan”

(Kim Namjoon)

“Bukan orang lain, bukan siapa pun, tapi kamu yang harus jadi orang pertama yang mengapresiasi dirimu sendiri”

(Alvi Syahrin)

Persembahan

Dengan mengucapkan puji syukur kepada Allah SWT, yang telah memberikan kekuatan, kesehatan, dan karunia yang engkau berikan sehingga saya dapat menyelesaikan skripsi ini dengan baik. Sebagai tanda terima kasih, saya persembahkan skripsi ini untuk :

1. *Superhero*-ku Bapak Sugeng , beliau memang tidak sempat merasakan pendidikan sampai bangku perkuliahan akan tetapi beliau mampu mendidik penulis, memberi semangat dan dukungan kepada penulis sehingga penulis dapat menyelesaikan studinya.
2. Pintu surgaku Ibunda Suwarni, yang telah melahirkan dan merawat saya dengan sepenuh hati dan selalu mengirimkan do'a sehingga penulis bisa menyelesaikan tugas akhir ini.
3. Adik-adik kecilku Safa dan Ara, yang selalu menghibur penulis.
4. Teman – teman ku, Hilwa adya tsaqofa, Dinda Karmila, Dea Puspita, dan Rosena milenia yang telah mewarnai masa – masa perkuliahan, membantu, memotivasi, dan memberi semangat pada penulis.
5. Terakhir, terima kasih untuk diri sendiri karena telah berusaha dan bertahan sejauh ini. Terima kasih sudah mampu mengendalikan diri dari tekanan di luar keadaan dan tetap memutuskan untuk tidak menyerah sesulit apapun proses penyusunan skripsi ini.

KATA PENGANTAR

Puji Syukur kamu panjatkan kehadiran Allah SWT, atas limpahan rahmat, nikmat dan karunia-Nya kepada penulis sehingga dapat menyelesaikan tugas akhir dengan judul “Implementasi Algoritma K-Means Untuk *Clustering* Data Penyakit Di Puskesmas Kotagede 2 Yogyakarta” dengan baik dan lancar. Skripsi ini disusun untuk memenuhi sebagian persyaratan dalam menyelesaikan derajat Sarjana Komputer di Program Studi Informatika, Fakultas Teknologi Industri, Universitas Ahmad Dahlan Yogyakarta.

Dalam penyusunan skripsi ini tentunya penulis mendapatkan arahan dan bimbingan dari beberapa pihak. Penulis mengucapkan terimakasih kepada pihak-pihak yang terlibat sehingga skripsi ini dapat terselesaikan dengan baik. Oleh karena itu, penulis mengucapkan terimakasih kepada :

1. Prof. Dr. Muchlas, M.T. selaku Rektor Universitas Ahmad Dahlan Yogyakarta.
2. Ir. Sunardi, S.T., M.T., Ph.D. selaku Dekan Fakultas Teknologi Industri Universitas Ahmad Dahlan Yogyakarta.
3. Dr. Murinto, S. Si, M.Kom. selaku Kaprodi Informatika, Fakultas Teknologi Industri Universitas Ahmad Dahlan Yogyakarta.
4. Arfiani Nur Khusna, S.T., M.Kom. selaku Dosen Pembimbing Akademik yang telah mendampingi dan memberi arahan selama masa studi.
5. Ir. Sri Winiarti, S.T., M.Cs. selaku Dosen Pembimbing dalam penyusunan skripsi yang telah memberikan bimbingan, dukungan, dan arahan kepada penulis dalam menyelesaikan skripsi ini. Semoga Ibu senantiasa diberikan kesehatan dan kebahagiaan selalu.
6. Bapak dan Ibu Dosen Informatika Universitas Ahmad Dahlan Yogyakarta yang telah banyak memberikan ilmunya selama masa studi.
7. Puskesmas Kotagede 2 Yogyakarta yang telah memberikan informasi yang dibutuhkan dalam penelitian ini.
8. Kedua orang tua penulis Bapak Sugeng dan Ibu Suwarni tercinta yang selalu memberikan dukungan, semangat dan doanya sehingga penulis bisa menyelesaikan pendidikan dengan baik.
9. Teman – teman kuliah penulis Hilwa, Dinda, Rosena, dan Dea yang telah membuat hari-hari penulis dalam menjalani perkuliahan menjadi menyenangkan.

Penulis menyadari bahwa penelitian yang telah ditulis ini masih jauh dari kata sempurna, sehingga kritik dan saran yang membangun dari pembaca sangat penulis harapkan. Akhir kata, semoga penelitian ini dapat bermanfaat bagi semua pihak.

Yogyakarta, 20 Desember 2023

Marwah Silvia Sasmita

DAFTAR ISI

IMPLEMENTASI ALGORITMA K-MEANS UNTUK <i>CLUSTERING</i> DATA PENYAKIT DI PUSKESMAS KOTAGEDE 2 YOGYAKARTA	i
LEMBAR PERSETUJUAN PEMBIMBING.....	ii
LEMBAR PERSETUJUAN PENGUJI	iii
LEMBAR PERNYATAAN KEASLIAN.....	iv
Pernyataan Tidak Plagiat.....	v
Pernyataan Persetujuan Akses.....	vi
Motto	vii
Persembahan	viii
KATA PENGANTAR.....	ix
DAFTAR ISI	x
DAFTAR GAMBAR	xiii
DAFTAR TABLE.....	xiv
DAFTAR KODE PROGRAM.....	xv
DAFTAR LAMPIRAN	xvi
ABSTRAK.....	xvii
BAB I	1
PENDAHULUAN	1
1.2 Rumusan Masalah	4
1.3 Batasan Penelitian.....	4
1.4 Tujuan Penelitian.....	4
1.5 Manfaat Penelitian	5
BAB II	5
TINJAUAN PUSTAKA	5
2.1 Kajian Penelitian Terdahulu.....	5
2.2 Landasan Teori	12
2.2.1 <i>Data Mining</i>	12
2.2.2 Clustering.....	14
2.2.3 Algoritma K-Means.....	16
2.2.4 Metode Davies Bouldin Index	17

BAB III	20
DAFTAR ISI	
METODOLOGI PENELITIAN	20
3.3.1 Kajian Penelitian	21
3.3.2 Pengumpulan Data	21
3.3.3 Alat dan Bahan	22
3.3.4 Pengolahan data.....	22
3.3.5 Analisis Kebutuhan Sistem	23
3.3.6 Perancangan sistem	23
3.3.7 Implementasi.....	24
3.3.8 Uji Akurasi.....	24
3.3.9 Kesimpulan	24
BAB IV	26
HASIL DAN PEMBAHASAN	26
4.1 Hasil Pengumpulan Data	26
4.2 Pengolahan data.....	27
4.3 Analisis kebutuhan sistem	29
4.4 Perancangan Sistem	30
4.5 Implementasi.....	39
4.5.1 Halaman Beranda	40
4.5.2 Halaman Tampil Data dan Seleksi data	44
4.5.3 Tampilan menentukan jumlah <i>cluster</i>	45
4.5.4 Tampilan hasil <i>cluster</i>	47
4.4.5 Tampilan hasil uji akurasi	51
4.6 Uji Akurasi.....	52
4.7 Pembahasan	57
BAB V	59
KESIMPULAN DAN SARAN	59
5.1 Kesimpulan	59
5.2 Saran.....	60
DAFTAR PUSTAKA.....	61
LAMPIRAN	64
1. Tempat pengambilan data (Puskesmas Kotagede 2 Yogyakarta)	64
2. Data penyakit yang ada di puskesmas.....	64

DAFTAR GAMBAR

Gambar 2. 1 Ilustrasi tahapan algoritma K-Means (Juniar ,2020)	17
Gambar 3. 1 Tahapan penelitian	20
Gambar 3. 2 Alur sistem	23
Gambar 4. 1 Alur sistem python.....	31
Gambar 4. 2 Halaman Dahsboard	32
Gambar 4. 3 Halaman tampil data dan seleksi data.....	33
Gambar 4. 4 Halaman penentuan <i>cluster</i>	33
Gambar 4. 5 Halaman hasil <i>clustering</i>	34
Gambar 4. 6 Hasil standarisasi data	37
Gambar 4. 7 Hasil uji akurasi	38
Gambar 4. 8 Anggota <i>cluster 0</i>	38
Gambar 4. 9 Anggota <i>cluster 1</i>	39
Gambar 4. 10 Alur sistem pada web	40
Gambar 4. 11 Halaman beranda	41
Gambar 4. 12 Halaman tampil data dan seleksi data.....	44
Gambar 4. 13 Menentukan jumlah <i>cluster</i>	46
Gambar 4. 14 Jumlah anggota setiap <i>cluster</i>	47
Gambar 4. 15 Anggota <i>cluster 0</i>	48
Gambar 4. 16 Anggota <i>cluster 1</i>	48
Gambar 4. 17 Diagram pie hasil <i>clustering</i>	50
Gambar 4. 18 Hasil uji akurasi DBI	51

DAFTAR TABLE

Tabel 2. 1 Kajian Pustaka	9
Tabel 4. 1 Data penyakit.....	26
Tabel 4. 2 Data awal	27
Tabel 4. 3 Tabel inialisasi umur	27
Tabel 4. 4 Tabel inialisasi diagnosa	28
Tabel 4. 5 Tabel inialisasi wilayah	28
Tabel 4. 6 Data yang sudah diolah.....	29
Tabel 4. 7 Hasil load dataset.....	35
Tabel 4. 8 Hasil seleksi data.....	36
Tabel 4. 21 Percobaan uji akurasi.....	52
Tabel 4. 9 Tabel data pasien.....	53
Tabel 4. 10 Pemilihan centroid awal	53
Tabel 4. 11 Hasil Iterasi 1.....	53
Tabel 4. 12 Penentuan centroid baru.....	54
Tabel 4. 13 Hasil iterasi 2.....	54
Tabel 4. 14 Perhitungan DBI nilai SSW	55
Tabel 4. 15 Mencari nilai SSB.....	55
Tabel 4. 16 Kode program perhitungan <i>clustering</i>	56
Tabel 4. 17 Anggota <i>cluster 0</i>	56
Tabel 4. 18 Anggota <i>cluster C1</i>	57
Tabel 4. 19 Source code akurasi DBI.....	57

DAFTAR KODE PROGRAM

Kode program 4. 1 Import library.....	35
Kode program 4. 2 Load dataset	35
Kode program 4. 3 Seleksi data	36
Kode program 4. 4 Standarisasi data.....	36
Kode program 4. 5 Proses K-Means	37
Kode program 4. 6 Uji akurasi DBI.....	38
Kode Program 4. 7 Halaman beranda	43
Kode Program 4. 8 Halaman tampil data dan seleksi data.....	45
Kode Program 4. 9 Menentukan jumlah <i>cluster</i>	46
Kode Program 4. 10 Jumlah anggota setiap <i>cluster</i>	48
Kode Program 4. 11 Anggota setiap <i>cluster</i>	49
Kode Program 4. 12 Diagram pie hasil <i>clustering</i>	51
Kode Program 4. 13 Hasil uji akurasi DBI	52

DAFTAR LAMPIRAN

1. Tempat pengambilan data (Puskesmas Kotagede 2 Yogyakarta)64
2. Data penyakit yang ada di puskesmas.....64

Implementasi Algoritma K-Means untuk *Clustering* Data Penyakit di Puskesmas Kotagede 2 Yogyakarta

Marwah Silvia Sasmita
1900018169

ABSTRAK

Puskesmas Kotagede 2 melayani pasien dari berbagai wilayah setiap hari, yang menghasilkan banyak data rekam medis. Saat ini, data tersebut hanya diarsipkan tanpa analisis lebih lanjut, menyebabkan pemimbunan data. Untuk mengatasi hal ini, diperlukan sistem yang dapat mengelompokkan data rekam medis berdasarkan kemiripan, untuk memberikan wawasan baru bagi tenaga medis tentang penyakit, faktor risiko, dan pola penyakit. Tujuan penelitian ini adalah menghasilkan dua cluster penyakit dan mengevaluasi hasilnya dengan Davies Bouldin Index yang bermanfaat untuk mengetahui kelompok penyakit yang banyak dan jarang, rata-rata usia penderitanya, dan wilayah asal penderita.

Pengumpulan data dilakukan melalui proses wawancara dengan salah satu staff bagian rekam medis di Puskesmas Kotagede 2 Yogyakarta. Data rekam medis yang diperoleh dari puskesmas sebanyak 300 data yang akan diolah menjadi dataset sehingga dapat di proses oleh sistem. Agar sistem dapat berjalan dengan baik maka dilakukanlah analisis kebutuhan sistem, setelah melakukan analisis kebutuhan sistem selanjutnya merancang sistem. Sistem ini dirancang untuk bisa melakukan operasi *clustering* data menggunakan algoritma k-means. Pada penelitian ini data akan di kelompokkan menjadi dua *cluster*. Proses *clustering* ini kemudian di implementasikan ke dalam website agar mudah digunakan. Uji akurasi dilakukan setelah hasil *clustering* di temukan, uji akurasi pada penelitian ini menggunakan metode Davies Bouldin Index.

Hasil dari penelitian ini adalah terbentuknya dua *cluster* yang telah di uji menggunakan Davies Bouldin Index dengan nilai akurasi sebesar 0.73. Hasil dari *clustering* dari jumlah 300 data didapatkan jumlah anggota *cluster* 0 sebanyak 183 data pasien dengan penyakit yang banyak diderita yaitu asma, hipertensi, sakit kepala, gerd, nasopharingitis akut, demam, Rinitis, nasofaringitis dan faringitis kronik, migren, diabetes mellitus 2, batuk, dan katarak, untuk usia rata – rata penderita pada *cluster* 0 ini adalah 34 tahun dengan wilayah pasien berasal dari Kelurahan Rejowinangun dan anggota *cluster* sebanyak 117 data pasien dengan penyakit yang jarang diderita yaitu tipus, anemia, dbd, vertigo, asam urat, campak, vertigo, sembelit, dan kolestrol dengan rata – rata usia penderita 29 tahun dengan wilayah asal pasien berasal dari Kelurahan Banguntapan.

Kata kunci: *Clustering; Data Mining; Davies Bouldin Index; K-Means*