

# PENGELOMPOKKAN PENYAKIT PASIEN MENGGUNAKAN ALGORITMA K-MEANS

## TUGAS AKHIR

Diajukan Sebagai Salah Satu Syarat  
Untuk Memperoleh Gelar Sarjana Teknik  
Pada Jurusan Teknik Informatika

Oleh

**RAHAYU ANGGRAINI**

**11750125043**



**FAKULTAS SAINS DAN TEKNOLOGI**

**UNIVERSITAS ISLAM NEGERI SULTAN SYARIF KASIM RIAU**

**PEKANBARU**

**2023**



**Hak Cipta Dilindungi Undang-Undang**  
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:  
a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.  
b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.  
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

© Hak cipta milik UIN Suska Riau

State Islamic University of Sultan Syarif Kasim Riau

## LEMBAR PERSETUJUAN

### PENGELOMPOKKAN PENYAKIT PASIEN MENGGUNAKAN ALGORITMA K-MEANS

#### TUGAS AKHIR

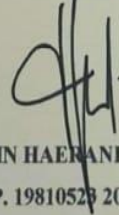
Oleh

RAHAYU ANGGRAINI

NIM. 11750125043

Telah diperiksa dan disetujui sebagai Laporan Tugas Akhir  
di Pekanbaru, pada tanggal 12 Januari 2023

Pembimbing I,



Dr. ELIN HAERANI, S.T., M.Kom.

NIP. 19810523 200710 2 003



- Hak Cipta Dilindungi Undang-Undang**
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
    - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
    - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
  2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

## LEMBAR PENGESAHAN

### PENGELOMPOKAN PENYAKIT PASIEN MENGGUNAKAN ALGORITMA K-MEANS

Oleh

RAHAYU ANGGRAINI

NIM. 11750125043

Telah dipertahankan di depan sidang dewan penguji  
sebagai salah satu syarat untuk memperoleh gelar Sarjana Teknik  
pada Universitas Islam Negeri Sultan Syarif Kasim Riau

Pekanbaru, 12 Januari 2023

Mengesahkan,

Ketua Jurusan,

  
IWAN ISKANDAR, M.T.

NIP. 19821216 201503 1 003



DR. HARTONO, M.PD.

NIP. 19640301 199203 1 003

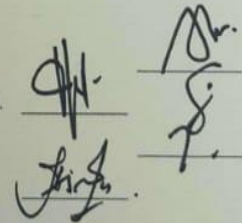
#### DEWAN PENGUJI

Ketua : Dr. Alwis Nazir, M.Kom.

Pembimbing I : Dr. Elin Haerani, S.T., M.Kom.

Penguji I : Jasril, S.Si., M.Sc.

Penguji II : Iis Afrianty, S.T., M.Sc.



#### Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

## SURAT PERNYATAAN

Saya yang bertandatangan di bawah ini :  
Nama : Rahayu Anggraini  
NIM : 11750125043  
Tempat/ Tgl. Lahir : Tanjung Medan, 02 Juli 1999  
Fakultas/Pascasarjana : Sains dan Teknologi/S1  
Prodi : Teknik Informatika  
Judul Skripsi :  
PENGELOMPOKKAN PENYAKIT PASIEN MENGGUNAKAN ALGORITMA K-MEANS

Menyatakan dengan sebenar-benarnya bahwa :

1. Penulisan Skripsi dengan judul sebagaimana tersebut di atas adalah hasil pemikiran dan penelitian saya sendiri.
2. Semua kutipan pada karya tulis saya ini sudah disebutkan sumbernya.
3. Oleh karena itu Skripsi saya ini, saya nyatakan bebas dari plagiat.
4. Apa bila dikemudian hari terbukti terdapat plagiat dalam penulisan Skripsi saya tersebut, maka saya bersedia menerima sanksi sesuai peraturan perundang-undangan.

Demikian Surat Pernyataan ini saya buat dengan penuh kesadaran dan tanpa paksaan dari pihak manapun juga.

Pekanbaru, 19 Januari 2023  
Yang membuat pernyataan



Rahayu Anggraini  
NIM : 11750125043

*\* pilih salah satu sesuai jenis karya tulis*



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

© Hak cipta milik UIN Suska Riau

State Islamic University of Sultan Syarif Kasim Riau



# Pengelompokan Penyakit Pasien Menggunakan Algoritma K-Means

Rahayu Angraini\*, Elin Haerani, Jasril, Iis Afrianty

Fakultas Sains dan Teknologi, Teknik Informatika, Universitas Islam Negeri Sultan Syarif Kasim Riau, Pekanbaru, Indonesia  
Email: <sup>1,\*</sup>11750125043@students.uin-suska.ac.id, <sup>2</sup>elin.haerani@uin-suska.ac.id, <sup>3</sup>jasril@uin-suska.ac.id,  
<sup>4</sup>iis.afrianty@uin-suska.ac.id

Email Penulis Korespondensi : 11750125043@students.uin-suska.ac.id  
Submitted 20-11-2022; Accepted 20-12-2022; Published 30-12-2022

## Abstrak

Kesehatan menjadi salah satu faktor terpenting selain pendidikan dan pendapatan. Setiap orang memiliki hak asasi yang sama untuk mendapatkan pelayanan kesehatan yang baik. Suatu lembaga instansi pemerintahan yang berfungsi untuk melayani seluruh masyarakat yang membutuhkan pelayanan pengobatan di Indonesia yaitu puskesmas. Puskesmas Ujung Batu yang terletak di kecamatan Ujung Batu Kabupaten Rokan Hulu sebagai salah satu instansi pemerintahan. Di puskesmas Ujung Batu menyimpan Data Rekam Medis Pasien hanya menggunakan penyakit saja. Oleh karena itu, data rekam medis tersebut perlu diolah dengan menggunakan Clustering atau Pengelompokan dengan menggunakan metode K-Means Algoritma ini melakukan partisi data ke dalam cluster sehingga data dengan karakteristik yang sama dikelompokkan ke dalam satu cluster yang sama dan data yang mempunyai karakteristik yang berbeda dikelompokkan ke dalam cluster yang lain. Data yang digunakan terdiri dari 3875 record dan 5 atribut yaitu Jenis Kelamin, Jenis Peserta, Diagnosis, Status, Padang, Alamat. Dari pengujian menggunakan Algoritma K-means menunjukkan hasil clustering yaitu cluster 1 sebanyak 710 data sedangkan cluster 2 dengan jumlah 3165 data. Hasil dari penelitian menunjukkan bahwa penggunaan 2 cluster menjadi cluster terbaik dengan nilai Silhouette Coefficient menunjukkan hasil dengan nilai SC yaitu 0,646.

**Kata Kunci:** Kesehatan; Penyakit; Clustering; K-Means; Silhouette Coefficient.

## Abstract

Health is one of the most important factors besides education and income. Everyone has the same human rights to get good health service. A government agency that functions to serve all people who need medical services in Indonesia, namely the puskesmas. Ujung Batu Health Center which is located in Ujung Batu sub-district, Rokan Hulu Regency as one of the government agencies. The Ujung Batu health center stores patient medical record data, only sorting out the disease. Therefore, the medical record data needs to be processed using clustering or grouping using the K-Means method. This algorithm partitions the data into clusters so that data with the same characteristics are grouped into the same cluster and data with different characteristics are grouped into another cluster. The data used consists of 3875 records and 5 attributes, namely Gender, Participant Type, Diagnosis, Return Status, Address. From the test using the K-means algorithm, the clustering results show that cluster 1 has 710 data while cluster 2 has 3165 data. The results of the study show that the use of 2 clusters is the best cluster with a Silhouette Coefficient value showing results with a SC value of 0.646.

**Keywords:** Health; Public health center; Clustering; K-Means; Silhouette Coefficient.

## 1. PENDAHULUAN

Kesehatan menjadi salah satu faktor terpenting selain pendidikan dan pendapatan. Setiap orang memiliki hak asasi yang sama untuk mendapatkan pelayanan kesehatan yang baik. Kondisi kesehatan itu dipengaruhi oleh beberapa faktor diantaranya yaitu lingkungan dan pelayanan kesehatan. Mutu pelayanan kesehatan dipengaruhi oleh beberapa faktor seperti ketersediaan dan kualitas sarana pelayanan, obat dan perbekalan kesehatan, serta manajemen kesehatan[1]. Suatu lembaga instansi pemerintahan yang berfungsi untuk melayani seluruh masyarakat yang membutuhkan pelayanan pengobatan di Indonesia yaitu puskesmas[2].

Puskesmas Ujung Batu yang terletak di kecamatan Ujung Batu merupakan salah satu puskesmas yang terletak di kabupaten Rokan Hulu. Puskesmas Ujung Batu merupakan salah satu instansi pemerintah di bawah Kementerian Kesehatan yang fungsinya memberikan pelayanan kepada masyarakat. Puskesmas dalam menyimpan data rekam medis pasien banyak mengurutkan penyakit yang paling banyak diderita saja dan tidak memperhatikan alamat, jenis peserta, dan jenis kelaminnya. Oleh karena itu, pihak puskesmas membutuhkan pengolahan data penyakit untuk mengetahui informasi mengenai penyakit yang diderita oleh pasien di beberapa wilayah yang ada di kecamatan Ujung Batu. Dengan informasi tersebut pihak puskesmas dapat mengambil tindakan kebijakan dalam antisipasi pengobatan dan pencegahan penyakit dengan melakukan sosialisasi. Untuk menghasilkan informasi yang tepat maka data rekam medis perlu diolah dengan menggunakan metode data mining.

Data mining merupakan suatu metode untuk pengolahan data yang digunakan untuk menemukan pola-pola atau model baru yang sempurna dan dapat dipahami dalam suatu database yang besar[3]. Data mining adalah proses penambangan data dengan jumlah besar untuk memperoleh pengetahuan dengan menyatukan teknik statistik, kecerdasan buatan, machine learning dan visualisasi informasi[4].

Algoritma yang digunakan dalam penelitian ini adalah algoritma K-Means. Algoritma K-Means merupakan salah satu algoritma yang termasuk dalam penerapan *clustering* data mining. Algoritma ini melakukan partisi data ke dalam cluster sehingga data dengan karakteristik yang sama dikelompokkan ke dalam satu cluster yang sama dan data yang mempunyai karakteristik yang berbeda dikelompokkan ke dalam cluster yang lain[5]. Kemiripan anggota suatu objek terhadap *cluster* diukur dengan kedekatan objek terhadap nilai *mean* pada *cluster* atau dapat disebut sebagai centroid atau pusat *cluster*[6].

Beberapa penelitian sebelumnya telah banyak menggunakan Algoritma K-means, diantaranya Penelitian terdahulu pengelompokan Algoritma K-Means[7]. penelitian mengelompokkan data calon mahasiswa baru ini menghasilkan 4 cluster yaitu pada cluster 0, jurusan ilmu keperawatan sebanyak 3795 pendaftar calon mahasiswa baru dari jumlah 25000 ( 15%), cluster 1 jurusan pendidikan kedokteran, terdapat 8296 pendaftar calon mahasiswa baru (33%), cluster 2 dengan jurusan Kedokteran Gigi sebanyak 7427 pendaftar dari jumlah calon mahasiswa baru (30%). Dan cluster 3 dengan jurusan pendidikan Dokter, terdapat sebanyak 5482 pendaftar dari jumlah mahasiswa baru (22%).

Pada penelitian[8] Pengelompokkan penyebab penyebaran ISPA menghasilkan 2 cluster, yang mana cluster 1 tingkat tinggi berjumlah 10 kabupaten dan cluster 2 tingkat rendah berjumlah 2 kabupaten.

Penelitian lain mengelompokkan penyakit [9] Dalam Penelitian ini menggunakan 2 cluster yang mana pada cluster 0 mengelompokkan penyakit akut, sedangkan cluster 1 mengelompokkan penyakit tidak akut. Berdasarkan pengujian dengan algoritma k-means menghasilkan sebanyak 241 data pada cluster 0 dan 9 data pada cluster 1 dengan nilai Davies Bouldin untuk algoritma k-means sebesar -0,453.

Penelitian terkait Algoritma K-Means[10] penelitian ini menghasilkan 3 cluster di mana dari 34 jumlah Tamu Asing berdatangan provinsi menghasilkan 1 provinsi termasuk cluster tingkat tinggi, 2 provinsi cluster tingkat sedang dan 3 provinsi termasuk cluster rendah. Hasil nya hampir 90% provinsi Indonesia masih memiliki potensi wisata yang rendah.

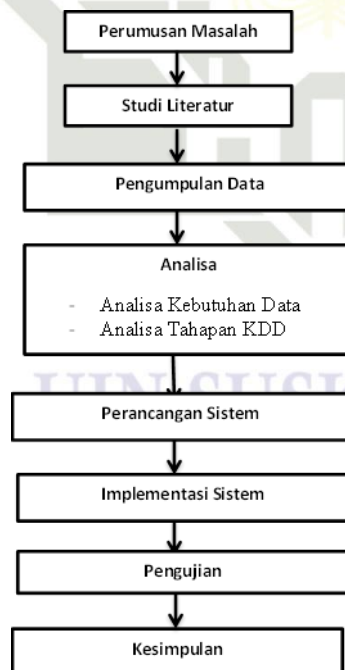
Penelitian terkait Algoritma K\_Means[11] pada penelitian pengelompokkan resiko jantung tersebut mendapatkan hasil 3 cluster, di mana pada cluster 1 kategori usia dengan tingkat resiko penyakit jantung dengan jumlah data sebanyak 355 cukup rendah (low), cluster 2 sebanyak 208 data dengan kategori usia tingkat resiko penyakit jantung sedang (medium) dan pada cluster 3 kategori usia dengan tingkat kategori usia cukup tinggi atau high sebanyak 462 data.

Berdasarkan penjabaran permasalahan diatas, maka akan di lakukan Penelitian “Clustering Pengelompokkan Penyakit Pasien Menggunakan Algoritma K-Means ( Studi Kasus : Puskesmas Ujung Batu )”. Sehingga dapat menghasilkan informasi untuk pengelompokkan penyakit pasien yang terdapat di beberapa wilayah yang dapat membantu pihak puskesmas dalam melakukan pencegahan dan pengobatan dengan sosialisasi.

## 2. METODOLOGI PENELITIAN

### 2.1 Tahapan Penelitian

Metodologi penelitian adalah suatu proses atau tahapan yang disusun secara sistematis dan logis dalam melakukan penelitian yang berguna untuk mencapai tujuan yang diharapkan. Adapun tahapan metodologi penelitian yang digunakan pada penelitian ini dapat dilihat pada gambar berikut :



Gambar 1. Tahapan Penelitian

Keterangan gambar di atas ialah sebagai berikut :

#### a. Perumusan Masalah

Pada tahap ini dilakukan identifikasi permasalahan yang menjadi topik penelitian. Adapun perumusan masalah dalam penelitian ini yaitu Bagaimana Pengelompokkan Penyakit Pasien menggunakan Algoritma K-Means Pada Puskesmas Ujung Batu.

## b. Studi Literatur

Studi literatur adalah proses memahami dan mempelajari teori-teori yang terkait dengan penelitian yang dilakukan yang didapat dari buku referensi tentang Data Mining, jurnal penelitian tentang Data Mining, Clustering, Algoritma K-Means, jurnal penelitian tentang metode *Silhouette Coefficient* ataupun artikel.

## c. Pengumpulan Data

Dalam penelitian ini data yang digunakan dalam pengelompokan penyakit pasien ini merupakan data sekunder yang diperoleh dari penelitian sebelumnya yang dilakukan oleh Ezu Pratama yang merupakan data Rekam Medis Pasien Di Puskesmas Ujung Batu. Data yang diperoleh kemudian akan diolah K-menggunakan metode K-Means.

## d. Analisa

Tahapan selanjutnya ialah tahapan analisa. Berikut tahapan analisa dalam penelitian ini :

### a. Kebutuhan Data

Tahapan analisa kebutuhan data terdiri dari data sekunder dan atribut. Berikut penjelasan dari analisa kebutuhan data.

#### 1. Data Sekunder

Penelitian ini menggunakan data sekunder yang didapatkan pada penelitian sebelumnya yang dilakukan oleh Ezu Pratama yang merupakan data Rekam Medis Pasien Di Puskesmas Ujung Batu. Data sekunder yang diperoleh akan digunakan untuk pengelompokan data pada algoritma K-Means. Data yang digunakan seluruh data sekunder yang terdapat.

#### 2. Atribut Data

Penelitian ini menggunakan atribut data rekam medis pasien yang berjumlah 3.974 record dengan 5 atribut yaitu jenis kelamin, Jenis Peserta, Diagnosa, Status Pulang, Alamat. Adapun sampel dataset dapat dilihat pada Tabel

Tabel 1. Dataset

	Jenis Peserta	Diagnosa	Status Pulang	Alamat
P1	Pegawai Pemerintah Dengan Perjanjian Kerja	Acute upper respiratory infections of multiple and unspecified sites	Berobat Jalan	Kelurahan Ujungbatu
P2	PNS Daerah	Nonsuppurative otitis media	Berobat Jalan	Desa Ngaso
P3	Pensiun PNS	Chronic ischaemic heart disease	Rujuk Vertikal	Desa Ngaso
P4	PNS Daerah	Diarrhoea and gastroenteritis of presumed infectious origin	Berobat Jalan	Desa Ujungbatu Timur
P5	Pensiun PNS	Atopic dermatitis	Berobat Jalan	Desa Pematang Tebih
P6	PNS Daerah	Pure hypercholesterolaemia	Berobat Jalan	Desa Ngaso
.....	.....	.....	.....	.....
P3922	PBI (APBD)	Myalgia	Berobat Jalan	Ujung Batu Timur
P3923	Pekerja Mandiri	Hyperemesis gravidarum	Sembuh	Pematang Tebih
P3924	Pegawai	Supervision of normal pregnancy	Berobat Jalan	Desa Ngaso

### b. Analisa Tahapan Metode

Tahapan ini melakukan analisa terhadap data Rekam Medis Pasien Puskesmas Ujung Batu dimulai dari tahap data *selection*, *pre-processing*, *transformation* dan data mining. Adapun analisa proses KDD yaitu sebagai berikut:

#### 1. Data Selection

Pada Tahap Data *Selection* menggunakan data rekam medis pasien yang mana atribut nya adalah jenis kelamin, jenis peserta, diagnosa, status pulang, alamat. Dan dari data yang didapatkan digunakan semua atribut sehingga tidak ada proses seleksi data.

#### 2. Pre-Processing

Pada tahap ini digunakan untuk melengkapi data, penghapusan data dan menghasilkan noise. Dalam penelitian ini data yang kosong atau missing value akan dihapus. Jumlah data kosong adalah 99 yang telah dihapus sehingga data berubah menjadi 3875 data. data yang telah di cleaning atau pembersihan data dapat dilihat pada Tabel 2.

#### 3. Transformation

Proses Transformasi Data adalah proses perubahan data yang dapat diolah menjadi bentuk yang bisa diolah pada proses data mining. Pada penelitian ini inialisasi yang dapat dilihat sebagai berikut :



**Tabel 2.** Inisialisasi Jenis Kelamin

Jenis Kelamin	Inisialisasi
Laki-Laki	1
Perempuan	2

**Tabel 3.** Inisialisasi Jenis Peserta

jenis peserta	Inisialisasi
© Pegawai Pemerintah Dengan perjanjian kerja	1
PNS daerah	2
Pensiun PNS	3
PBI ( APBN )	4
Pekerja Mandiri	5
PBI ( APBD )	6
.....	.....
Pegawai Bumn	13
Dewan Perwakilan Rakyat Daerah	14

**Tabel 4.** Inisialisasi Status Pulang

Status Pulang	Inisialisasi
Berobat Jalan	1
Rujuk Vertikal	2
Sembuh	3

**Tabel 5.** Inisialisasi Alamat

Alamat	Inisialisasi
Kelurahan Ujung Batu	1
Desa Ngaso	2
Desa Ujungbatu Timur	3
Desa Pematang Tebih	4
Desa Sukadamai	5

**Tabel 6.** inisialisasi diagnosa

Diagnosa	Inisialisasi
Acute upper respiratory infections of multiple and unspecified sites	1
Non-suppurative otitis media	2
Chronic ischaemic heart disease	3
Diarrhoea and gastroenteritis of presumed infectious origin	4
Atopic dermatitis	5
Pure hypercholesterolaemia	6
.....	.....
follow-up examination after treatment for malignant neoplasm	426
other specified chronic obstructive pulmonary disease	427

4. Data Mining dengan Algoritma K-Means

Algoritma K-Means adalah salah satu algoritma clustering untuk mengelompokkan data menjadi beberapa cluster atau kelompok. Algoritma K-Means juga merupakan metode pengelompokkan data non-hierarchical yaitu partition clustering yang berbasis titik pusat (centroid)[12]. Metode K-means digunakan sebagai metode cluster untuk data dengan ukuran yang besar. Data yang digunakan dalam Algoritma K-Means hanya dengan atribut data numeric[13]. Data-data dipilih menjadi beberapa *cluster* atau kelompok dengan berdasarkan kriteria yang sudah ditentukan dan kemudian dikumpulkan menjadi satu cluster. Di mana setiap *cluster* memiliki titik pusat yang disebut dengan centroid [14]. Algoritma K-Means akan memilih *k* sebagai titik awal centroid secara acak dan random. Dibutuhkan jumlah *cluster* awal yang diinginkan untuk sebagai masukan dan menghasilkan titik centroid akhir sebagai output[15]. Berikut ini persamaan untuk menghitung K-Means [16].

- 1) Menentukan nilai *k* sebagai sebagai jumlah cluster yang ingin dibentuk.
- 2) Inisialisasi *k* sebagai *centroid* awal cluster yang dapat dilakukan secara *Random*.
- 3) Hitung jarak setiap data ke masing-masing centroid menggunakan persamaan *Euclidean* sebagai berikut :

$$d(x,y)=\sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (1)$$



- 4) Kelompokkan setiap data berdasarkan jarak terdekat antara data *centroidnya*.
- 5) Tentukan posisi *centroid* baru (*k*).
- 6) Kembali kelangkah 3 jika posisi *centroid* baru dengan *centroid* lama tidak sama.

Dimana:

$x_1$  : Titik data Pertama

$x_2$  : Titik data kedua

$n$  : Jumlah karakteristik ( atribut) dalam data mining

$d(x,y)$  :Euclidean distance yaitu jarak data pada titik  $x$  dan  $y$  menggunakan kalkulasi matematika.

### e. Perancangan Sistem

Perancangan adalah tahapan perencanaan tentang merancang sistem yang akan dibangun berdasarkan analisa yang telah dilakukan. Adapun perancangan system yaitu menggunakan perancangan UML (*Unified Modelling Language*)

### f. Implementasi Sistem

Implementasi adalah tahapan dimana hasil analisis dan desain diterapkan pada sistem. Pada penelitian ini implementasi sistem menggunakan Bahasa Pemrograman PHP.

### g. Pengujian

Pengujian merupakan suatu tahapan untuk menguji hasil clustering dari sistem yang telah dibangun. Pada penelitian ini hasil clustering akan diuji dengan menggunakan metode *Silhouette Coefficient* (SC).

*Silhouette Coefficient* adalah metode evaluasi *clustering* yang menggabungkan cohesian dan separation. Metode ini digunakan untuk mengukur kekuatan dan melihat kualitas hasil clustering pada proses clustering[17]. *Silhouette Coefficient* adalah metode pengujian kualitas *cluster* dengan menghitung jarak rata-rata elemen dalam cluster yang sama dengan jarak rata-rata dengan cluster yang lain[18]. Apabila nilai *Silhouette Coefficient* mendekati = 1 maka struktur dari cluster berada pada kategori baik dan jika bernilai = 0 maka struktur dari cluster termasuk pada kategori tidak jelas[19]. Berikut ini adalah rentang nilai *Silhouette Coefficient* dan interpretasinya dapat dilihat pada Tabel 9 :

Tabel 7..Rentang Nilai dan Interpretasi dari Silhouette Coefficient

RentangNilai	Interpretasi
0,71 – 1,0	Struktur yang kuat
0,51 – 0,7	Struktur yang beralasan
0,26 – 0,5	Struktur yang lemah
0,25	Tidak ditemukan struktur yang substansial

### h. Kesimpulan

Tahap akhir adalah kesimpulan berdasarkan hasil perhitungan metode K-Means dan implemenatsi sistem.

## 3. HASIL DAN PEMBAHASAN

### 3.1 Penerapan Algoritma K-Means

Setelah dilakukan transformasi maka selanjutnya masuk ke tahap perhitungan K-Means dengan tujuan hasil data akhir akan menghasilkan kelompok data. Adapun langkah-langkah perhitungan algoritma K-Means ialah sebagai berikut :

- a. Tentukan jumlah *cluster k* yang ingin dibentuk. Dalam penelitian ini dibentuk kelompok atau *cluster* sebanyak 2 *cluster*. Sampel data yang adakan digunakan dapat dilihat pada tabel 8 berikut :

Tabel 8. Dataset

	X1	X2	X3	X4	X5
P1	2	1	1	1	1
P2	2	2	2	1	2
P3	2	3	3	2	2
P4	2	2	4	1	3
P5	1	3	5	1	4
P6	2	2	6	1	2
P7	2	4	7	1	3
P8	2	4	1	1	5
P9	1	4	8	1	2
P10	1	4	9	1	3
.....	.....	.....	.....	.....	.....
P3874	2	6	111	1	3
P3875	2	5	335	3	4

- b. Inialisasi *k* pusat *cluster* ( centroid) awal dengan cara mengambil data dari sumber secara random acak. Adapun

centroid awal sebagai berikut, dapat dilihat pada Tabel 9 di bawah ini

**Tabel 9.** Centroid Awal

Centroid Awal	Objek	X1	X2	X3	X4	X5
C1	P929	1	5	13	1	4
C2	P2666	2	4	1	1	5

c. Hitung jarak setiap data ke pusat *cluster* (centroid) menggunakan rumus *Euclidian Distance*. Berikut perhitungan

*Euclidean Distance* yang terdapat pada **Rumus 1** :  
Jarak Data ke pusat cluster

$$C1 = \sqrt{(1-2)^2 + (5-1)^2 + (13-1)^2 + (1-1)^2 + (4-1)^2}$$

$$C1 = 17,068$$

$$C2 = \sqrt{(2-2)^2 + (4-1)^2 + (1-1)^2 + (1-1)^2 + (5-1)^2}$$

$$C2 = 4,472$$

Jarak Data ke pusat cluster

$$C1 = \sqrt{(2-2)^2 + (4-2)^2 + (13-2)^2 + (1-1)^2 + (4-2)^2}$$

$$C1 = 12,727$$

$$C2 = \sqrt{(2-2)^2 + (4-2)^2 + (1-2)^2 + (1-1)^2 + (5-2)^2}$$

$$C2 = 4,9$$

Lakukan Perhitungan jarak untuk data ke-3..... N terhadap awal *cluster* hingga didapatkan matrik jarak.

Selanjutnya hasil dapat dilihat pada Tabel 10 di bawah ini :

**Tabel 10.** Hasil Perhitungan jarak setiap Data pada iterasi ke -1

	X1	X2	X3	X4	X5	Jarak C1	Jarak C2
P1	2	1	1	1	1	5	13.038
P2	2	2	2	1	2	3.742	11.619
P3	2	3	3	2	2	3.873	10.488
P4	2	2	4	1	3	4.123	9.592
P5	1	3	5	1	4	4.359	8.246
P6	2	2	6	1	2	6.164	7.937
P7	2	4	7	1	3	6.325	6.245
P8	2	4	1	1	5	0	12.124
P9	1	4	8	1	2	7.681	5.477
P10	1	4	9	1	3	8.307	4.243
.....	.....	.....	.....	.....	.....	.....	.....
P3874	2	6	111	1	3	110.036	98.015
P3875	2	5	335	3	4	334.009	322.008

d. Melakukan pengelompokkan data kedalam *cluster* dengan jarak yang terpendek. Menempatkan data kedalam *cluster* dengan jarak yang terpendek. Adapun hasil penempatan cluster setiap data pada iterasi ke-1 dapat dilihat pada Tabel 11 sebagai berikut

**Tabel 11.** Hasil Pengelompokkan cluster iterasi 1

	X1	X2	X3	X4	X5	Jarak C1	Jarak C2	Hasil
P1	2	1	1	1	1	5	13.038	1
P2	2	2	2	1	2	3.742	11.619	1
P3	2	3	3	2	2	3.873	10.488	1
P4	2	2	4	1	3	4.123	9.592	1
P5	1	3	5	1	4	4.359	8.246	1
P6	2	2	6	1	2	6.164	7.937	1
P7	2	4	7	1	3	6.325	6.245	1
P8	2	4	1	1	5	0	12.124	1
P9	1	4	8	1	2	7.681	5.477	2
P10	1	4	9	1	3	8.307	4.243	2
.....	.....	.....	.....	.....	.....	.....	.....	.....
P3874	2	4	1	1	3	110.036	98.015	2
P3875	1	4	1	1	2	334.009	322.008	2

e. Setelah didapatkan anggota dari setiap *cluster* kemudian Hitung pusat *cluster* atau *centroid* yang baru, yang mana akan digunakan pada iterasi selanjutnya. Untuk hasil perhitungan *centroid* yang baru pada iterasi ke-2 dapat dilihat pada Tabel 12 sebagai berikut :

Tabel 12. Pusat Cluster Baru

	X1	X2	X3	X4	X5
C1	1.636	4.432	85.063	1.322	2.935
C2	1.596	4.203	2.584	1.018	2.92

f. Selanjutnya dilakukan pengecekan hasil *cluster* dan pusat *cluster*. Jika hasil *cluster* dan pusat *cluster* sama, maka iterasi berhenti. Jika *cluster* dan pusat *cluster* berbeda, maka iterasi dilanjutkan mulai dari langkah ke-3 menggunakan *centroid* yang baru. Pada penelitian ini iterasi di lakukan sampai iterasi ke-9. Dikarenakan posisi *cluster* tidak berubah maka penelitian di berhentikan. Adapun hasil akhir posisi *cluster* dari *Centroid* iterasi ke-3 dapat dilihat pada Tabel 13 sebagai berikut :

Tabel 13. Pengelompokkan cluster pada iterasi ke-9

	X1	X2	X3	X4	X5	Jarak C1	Jarak C2	Hasil
P1	2	1	1	1	1	238.005	39.72	2
P2	2	2	2	1	2	236.986	38.613	2
P3	2	3	3	2	2	235.978	37.571	2
P4	2	2	4	1	3	234.984	36.607	2
P5	1	3	5	1	4	233.978	35.576	2
P6	2	2	6	1	2	232.986	34.623	2
P7	2	4	7	1	3	231.971	33.53	2
P8	2	4	1	1	5	237.979	39.585	2
P9	1	4	8	1	2	230.975	32.547	2
P10	1	4	9	1	3	229.972	31.535	2
.....	.....	.....	.....	.....	.....	.....	.....	.....
P3874	2	4	1	1	3	127.982	70.495	2
P3875	1	4	1	1	2	96.049	294.483	1

### 3.2 Interpretation

*Interpretation* merupakan sebuah tahapan setelah melakukan proses data mining yang mana proses tersebut menghasilkan pola informasi yang dapat mudah dipahami. Di bawah ini akan dijelaskan hasil dari pengelompokkan menggunakan data 3875 dengan 5 atribut.

Dari tabel tiap tiap atribut maka dapat disimpulkan C1-C2 adalah sebagai berikut:

Tabel 14. Kesimpulan C1-C2

Cluster	Kesimpulan
C1	<p>Jumlah data 710</p> <ul style="list-style-type: none"> <li>• Diagnosa Pasien penderita penyakit menular sebanyak 28 data yang di dominasi oleh penyakit <i>Acute upper respiratory infection, unspecified</i>.</li> <li>• Pasien dengan jenis peserta berobat terbanyak adalah dengan kategori Pekerja Mandiri</li> <li>• Pasien yang berobat banyak yang dilakukan dengan berobat jalan</li> <li>• Pasien di Dominasi oleh masyarakat desa Sukadamai</li> <li>• Jumlah Pasien Laki-laki 247 dan Pasien Perempuan 463</li> </ul>
C2	<p>Jumlah data 165</p> <ul style="list-style-type: none"> <li>• Diagnosa Pasien penderita penyakit menular sebanyak 218 data yang di dominasi penyakit <i>Acute upper respiratory infections of multiple and unspecified sites</i>.</li> <li>• Pasien jenis Peserta berobat terbanyak adalah kategori Pekerja Mandiri</li> <li>• Pasien berobat banyak dilakukan tindakan berobat jalan</li> <li>• Pasien di dominasi oleh masyarakat desa Ngaso</li> <li>• Jumlah pasien laki-laki 1181 dan pasien perempuan 1984</li> </ul>

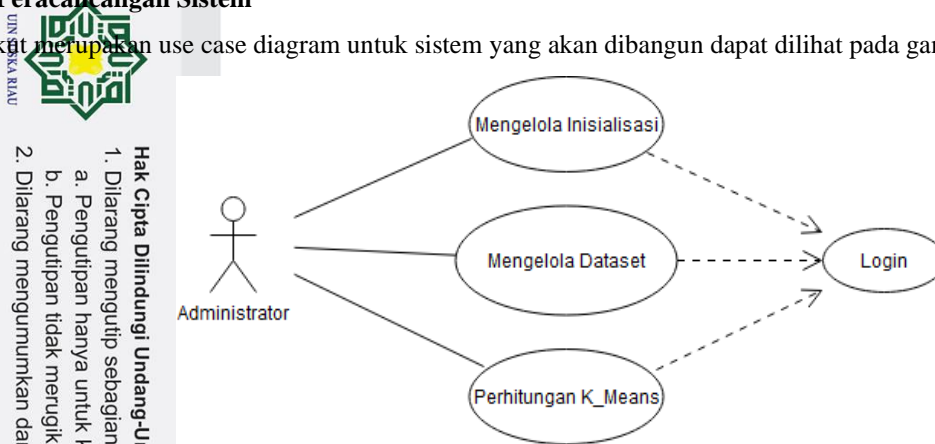
Berdasarkan tabel di atas maka dapat disimpulkan bahwa pada penelitian Pengelompokkan Penyakit Pasien Menggunakan Algoritma K-Means menunjukkan bahwa penggunaan 2 *cluster* menjadi *cluster* terbaik. Dari pengujian yang telah didapatkan hasil clustering yaitu *cluster* 1 sebanyak 710 dengan pasien penderita penyakit menular sebanyak 28 data yang didominasi penyakit *Acute upper respiratory infection, unspecified* dengan jumlah pasien Laki-laki 247 dan Pasien Perempuan 463 yang di dominasi oleh masyarakat desa Sukadamai. Sedangkan *cluster* 2 jumlah 3165 data dengan pasien penderita penyakit menular sebanyak 218 data yang didominasi penyakit *Acute upper respiratory infections of*



multiple and unspecified sites dengan jumlah pasien laki-laki 1181 dan pasien perempuan 1984 yang di dominasi oleh masyarakat Desa Ngaso.

### 3.3 Perancangan Sistem

Berikut merupakan use case diagram untuk sistem yang akan dibangun dapat dilihat pada gambar berikut ini :

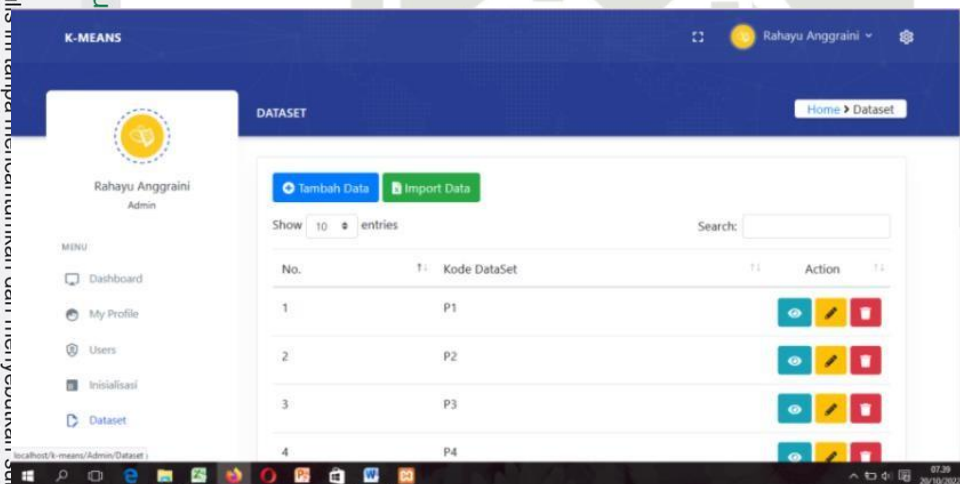


Gambar 2. Use Case Diagram

### 3.4 Implementasi Sistem

#### a. Halaman Dataset

Pada Halaman Dataset terdapat Menu Import Data untuk mengimport Dataset dan terdapat menu Tambah Data serta Action detail, edit, dan hapus data.



Gambar 3. Halaman Dataset

#### b. Halaman Perhitungan

Pada halaman Perhitungan pengguna dapat memasukkan jumlah cluster yang diinginkan serta terdapat tabel yang menunjukkan proses KDD.

No.	Kode	Jarak C1	Jarak C2	Kedekatan
1	P1	238.005	39.72	2
2	P2	236.986	38.613	2
3	P3	235.978	37.571	2
4	P4	234.984	36.607	2
5	P5	233.978	35.576	2
6	P6	232.986	34.623	2
7	P7	231.971	33.53	2
8	P8	237.979	39.585	2
9	P9	230.975	32.547	2
10	P10	230.073	31.535	2

Gambar 4. Halaman Perhitungan

### 3.5 Pengujian

#### 3.5.1 Silhouette Coefficient

Pengujian ini dilakukan untuk melihat kualitas dan kekuatan dari *cluster* yang dibentuk. Hasilnya berupa angka yang apabila semakin mendekati 1, maka semakin baik *cluster* tersebut. Hasil pengelompokan data Penyakit Pasien Menggunakan Algoritma K-Means dengan pengujian menggunakan metode *Silhouette Coefficient*. Dapat dilihat pada Tabel 15 sebagai berikut:

Tabel 15. Hasil pengujian *Silhouette Coefficient*

NO	Jumlah Cluster	Silhouette Coefficient
1	2	0,646
2	3	0,601
3	4	0,549
4	5	0,581

Kemudian yang dapat diambil dari hasil pengujian *Silhouette Coefficient* (SC) yaitu kualitas *cluster* sama dengan 2 menunjukkan kualitas yang paling baik karena nilai *Silhouette Coefficient* dengan  $k = 2$  merupakan nilai yang paling mendekati 1 yaitu sebesar 0,646, dimana nilai SC yang semakin mendekati 1 adalah yang paling optimal karena memiliki kekuatan yang baik dalam *cluster*.

## 4. KESIMPULAN

Berdasarkan hasil penelitian mengenai pengelompokan Penyakit Pasien Menggunakan Algoritma K-Means maka didapatkan kesimpulan yaitu, pengelompokan Penyakit Pasien dengan atribut Jenis Kelamin, jenis peserta, Diagnosa, Status Penderita, dan Alamat menggunakan algoritma K-Means dengan jumlah 3875 data menunjukkan bahwa jumlah *cluster* yang dibentuk sebanyak 2 *cluster*, banyaknya data pada masing-masing *cluster* yaitu *Cluster 1* terdiri dari 710 data dengan 28 penderita Penyakit Menular sebanyak 28 data yang didominasi penyakit *Acute upper respiratory infection*, *Unspecified* didominasi oleh desa Sukadamai dengan Jumlah pasien laki-laki sebanyak 247 data dan Pasien Perempuan 66 data. Pada *Cluster 2* terdiri dari 3165 data dengan Pasien Penyakit Menular sebanyak 218 data yang didominasi penyakit *Acute upper respiratory infections of multiple and unspecified sites* didominasi oleh desa Ngaso dengan Jumlah pasien laki-laki sebanyak 1181 data dan Pasien Perempuan 1984 data. Berdasarkan hasil pengujian nilai evaluasi *Silhouette Coefficient* pada Clustering Penyakit Pasien menggunakan Algoritma K-Means menunjukkan bahwa pembentukan 2 *cluster* dengan nilai terbaik yaitu 0,646. Dimana nilai *Silhouette Coefficient* (SC) yang semakin mendekati 1 adalah yang paling optimal karena memiliki kekuatan yang baik dalam *cluster*. Dari hasil *cluster* di dapat informasi tentang penyakit terbanyak yang ada pada Desa Sukadamai dan Desa Ngaso sehingga dapat membantu pihak puskesmas untuk mengambil tindakan kebijakan dalam antisipasi pencegahan dan pengobatan penyakit dengan sosialisasi.

## REFERENCES

- [1] S.S. Suddari and N. Ariani, "Penerapan Data Mining Untuk Pengelompokan Penyakit Dengan Algoritma Fuzzy C-Means (Studi Kasus : JET Puskesmas Salawu)," *J. VOI (Voice Informatics)*, vol. 8, no. 2, pp. 63–76, 2019.
- [2] Dan permata Sari, "PENGELOMPOKKAN PENYAKIT BERDASARKAN LINGKUNGAN DENGAN ALGORITMA K-MEANS PADA PUSKESMAS SUNGAI TARAB 2," vol. 5, no. 2, pp. 75–81, 2021.
- [3] A. Warjo *et al.*, *Data Mining : Algoritma dan Implementasi*. Medan, Indonesia: Yayasan Kita Menulis, 2020.
- [4] N. Hanum Harani and F. A. Nugraha, *SIGMENTASI PELANGGAN MENGGUNAKAN PYTHON*. Bandung: Kreatif Industri Nusantara, 2020.
- [5] Ajindya Khrisna Wardhani, "Implementasi Algoritma K-Means Untuk Pengelompokan Penyakit Pasien Pada Puskesmas Kajen Pakalongan," *J. Transform.*, vol. 14, pp. 30–37, 2016.
- [6] S. Suryadi, "Penerapan Metode Clustering K-Means Untuk Pengelompokan Kelulusan Mahasiswa Berbasis Kompetensi," *J. Inform.*, vol. 6, no. 1, pp. 52–72, 2019, doi: 10.36987/informatika.v6i1.738.
- [7] A. Asriani, H. Fitri, and E. Prasetyo, "Penerapan Metode Clustering dengan Algoritma K-Means pada Pengelompokan Data Calon Mahasiswa Baru di Universitas Muhammadiyah Yogyakarta (Studi Kasus: Fakultas Kedokteran dan Ilmu Kesehatan, dan Fakultas Ilmu Sosial dan Ilmu Politik)," *Semesta Tek.*, vol. 21, no. 1, pp. 60–64, 2018, doi: 10.18196/st.211211.
- [8] N. Purba and H. S. Tambunan, "Penerapan Algoritma K-Means Clustering Pada Penyebaran Penyakit Infeksi Saluran Pernapasan Akut (ISPA) di Provinsi Riau," vol. 2, no. 3, pp. 220–226, 2021.
- [9] T. L. Budi Susno, "Dengan Menggunakan Bahasa Pemrograman," pp. 399–406, 2015, [Online]. Available: <https://www.neliti.com/id/publications/173303/perancangan-sistem-informasi-distribusi-obat-pasien-rawat-inap>.
- [10] R. W. Sari and D. Hartama, "Data Mining : Algoritma K-Means Pada Pengelompokan Wisata Asing ke Indonesia Menurut Provinsi," *Semin. Ngs. Sains Teknol. Inf.*, pp. 322–326, 2018.
- [11] D. Haryadi and D. M. U. Atmaja, "Penerapan Algoritma K-Means Clustering Untuk Pengelompokan Tingkat Risiko Penyakit Jantung," *J. Informatics Commun. Technol.*, vol. 3, no. 2, pp. 51–66, 2021, doi: 10.52661/j\_ict.v3i2.85.
- [12] B. Serasi Ginting and M. Simanjuntak, "Pengelompokan Penyakit Pada Pasien Berdasarkan Usia Dengan Metode K-Means Clustering (Studi Kasus : Puskesmas Bahorok)," *Algoritma. J. Ilmu Komput. dan Inform.*, vol. 6341, no. November, p. 2, 2021.

- [13] M. Wahyudi, Masitha, R. Saragih, and Solikhun, *Data Mining : Penerapan Algoritma K-means Clustering dan K-Medoids Clustering*. Medan, Indonesia: Yayasan Kita Menulis, 2020.
- [14] N. Dwitri, J. A. Tampubolon, S. Prayoga, F. I. R.H Zer, and D. Hartama, “Penerapan Algoritma K-Means Dalam Menentukan Tingkat Penyebaran Pandemi Covid-19 Di Indonesia,” *J. Teknol. Inf.*, vol. 4, no. 1, pp. 128–132, 2020, doi: 10.30605/jti.v4i1.1266.
- [15] S. Santoso and S. Bunyamin, *PEMBUATAN APLIKASI CLUSTERING GANGGUAN JARINGAN MENGGUNAKAN METODE K-MEANS CLUSTERING*. Bandung: Kreatif Industri Nusantara, 2020.
- [16] D. Jollya, W. Ramdhan, and M. Zarlis, *KONSEP DATA MINING DAN PENERAPAN*. Yogyakarta: CV Budi Utama, 2020.
- [17] S. Paembonan and H. Abduh, “Penerapan Metode Silhouette Coefficient untuk Evaluasi Clustering Obat,” *PENA Tek. J. Ilm. Ilmu Ilmu Tek.* vol. 6, no. 2, p. 48, 2021, doi: 10.51557/pt\_jiit.v6i2.659.
- [18] H. Saefudin, A. Azis, and Zohrahayaty, *MACHINE LEARNING & REASONING FUZZY LOGIC ACORDATA MANUAL, MATLAB & RAPID MINER*. Yogyakarta: CV Budi Utama, 2020.
- [19] R. R. Mufidah, “Pengelompokan daerah rawan kekerasan terhadap perempuan dan anak di Jawa Barat menggunakan algoritma k-means,” vol. 07, no. September, pp. 850–857, 2022.

Hak Cipta Dilindungi Undang-Undang  
Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:  
Dilarang menyalin, mendistribusikan, dan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.  
Dilarang mengutip dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

State Islamic University of Sultan Syarif Kasim Riau



UIN SUSKA RIAU