Terbit online pada laman web jurnal: **http://ejurnal.unim.ac.id/index.php/submit**

SUBMIT
(Jurnal Ilmiah Teknologi Informasi dan Sains)
Vol. 4 No. 1 (2024) 1 - 9 ISSN Media Elektronik: 2798-6861

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

Marwah Silvia Sasmita[*1], Ir. Sri Winiarti, S.T., M.Cs [2]

[1,2] Program Studi Informatika, Universitas Ahmad Dahlan, Yogyakarta, Indonesia
Email: [1] marwah1900018169@webmail.uad.ac.id, [2]sri.winiarti@tif.uad.ac.id

(Naskah masuk: 20 Maret 2024, diterima untuk diterbitkan: 5 April 2024)

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 denga n baik maka dilakukanlah analisis kebutuhan sistem, setelah melakukan analisis kebutuhan sistem selanjutnya merancang sistem. Sistem ini dirancang untuk bisa melalukan 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 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*

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

Abstract

Kotagede 2 Community Health Center serves patients from various regions every day, which generates a lot of medical record data. Currently, the data is only archived without further analysis, causing data hoarding. To overcome this, a system is needed that can cluster medical record data based on

similarities, to provide new insights for medical personnel about diseases, risk factors, and disease patterns. The purpose of this research is to generate two disease clusters and evaluate the results with the Davies Bouldin Index which is useful for knowing the groups of diseases that are common and rare, the average age of sufferers, and the region of origin of sufferers. Data collection was done through an interview process with one of the medical record staff at Kotagede 2 Community Health Center Yogyakarta. Medical record data obtained from the health center as much as 300 data that will be processed into datasets so that they can be processed by the system. In order for the system to run well, a system needs analysis is carried out, after analyzing system needs, then designing the system. This system is designed to be able to perform data clustering operations using the k-means algorithm. In this research the data will be grouped into two clusters. This clustering process is then implemented into the website for easy use. The accuracy test is carried out after the clustering results are found, the accuracy test in this study uses the Davies Bouldin Index method. The result of this study is the formation of two clusters that have been tested using the Davies Bouldin Index with an accuracy value of 0.73. The results of clustering from a total of 300 data obtained the number of cluster 0 members as many as 183 patient data with many diseases suffered, namely asthma, hypertension, headache, gerd, acute nasopharyngitis, fever, rhinitis, nasopharyngitis and chronic pharyngitis, migraine, diabetes mellitus 2, cough, and cataract, for the average age of patients in cluster 0 is 34 years with the patient's area of origin from Kelurahan Rejowinangun and cluster members as many as 117 patient data with rare diseases namely typhoid, anemia, dbd, vertigo, gout, measles, vertigo, constipation, and cholesterol with an average age of 29 years with the patient's area of origin from Kelurahan Banguntapan.

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

1 PENDAHULUAN

Puskesmas merupakan salah satu lembaga yang memberi pelayanan dalam bidang kesehatan. Puskesmas menyediakan pelayanan kesehatan mulai dari rawat jalan, rawat inap, dan gawat darurat(Baharudin, Faza, and Herfiyanti 2021). Puskesmas Kotagede 2 Yogyakarta merupakan salah satu puskesmas di wilayah Kecamatan Kotagede yang terletak di Jalan Ki Penjawi No. 04 Kotagede yang memiliki visi : puskemas dengan pelayanan prima mewujudkan masyarakat Rejowangun yang sehat dan mandiri, dan misi dari puskesmas ini adalah mengembangan layanan promotif dan preventif di bidang kesehatan, memberikan pelayanan kesehatan yang terjangkau dengan berorientasi pada peningkatan mutu dan keselamatan pelayanan, dan meningkatkan kerjasama dan saling bersinergi dengan lintas sektor terkait.

Puskesmas Kotagede 2 setiap harinya melayani pasien dari berbagai wilayah. Seiring dengan bertambahnya jumlah pasien tersebut pasti akan menghasilkan banyak data rekam medis pasien, data ini akan terus bertambah seiring berjalannya aktivitas di puskesmas tersebut. Data rekam medis sediri adalah data atau dokumen yang berisi tentang catatan pemeriksaan, pengobatan, identitas pasien, dan pelayanan kesehatan lainnya (Peraturan Menteri Kesehatan No.55 Tahun 2013). Hingga saat ini penggunaan data rekam medis sebatas hanya sebagai arsip dan tidak dianalisa lebih lanjut. Hal ini hanya menyebabkan terjadinya pemimbunan data baik dalam database maupun data dalam bentuk fisik. Untuk mengolah data rekam medis yang ada maka dibutuhkan sistem yang dapat mengelompokkan data rekam medis berdasarkan kemiripan data yang telah ditentukan.

Pengelompokkan data ini akan membagi data rekam medis pasien menjadi dua kelompok yaitu penyakit yang banyak di derita dan penyakit yang jarang di derita beserta rentang usia penderitanya dan juga wilayah penderita berasal. Melalui pengelompokan data penyakit ini diharapkan dapat memberikan pengetahuan baru bagi pihak manajemen puskemas dalam mengambil keputusan dan mempertimbangkan untuk agenda penyuluhan terkait penyakit dan pertimbangan untuk peningkatan fasilitas kesehatan. Selain itu pengelompokkan data rekam medis ini diharapkan dapat membantu tenaga medis untuk memahami penyebab, faktor risiko, dan pola penyakit sehingga memungkinkan untuk pengembangan strategi pencegahan dan pengobatan yang lebih efek

Dengan adanya ilmu *data mining* beserta dengan berbagai metodenya bisa dimanfaatkan untuk penggalian data dan pengekstrakan tumpukan berskala besar agar bisa menghasilkan informasi dan pengetahuan yang bermanfaat. Beberapa contoh fungsi yang ada pada datang mining antara lain prediksi, klasifikasi, *clustering*, asosiasi, dan estimasi. Untuk bisa menjalankan fungsi tersebut dierlukan metode seperti apriori untuk asosiasi, Support Vector Machine unntuk prediksi, regresi untuk estimasi, k-means untuk *clusterisasi,* dan C4.5 untuk klasifikasi.

Jeri Wandana dan kawan-kawan (Wandana, Defit, and Sumijan 2020) melakukan penelitian terkait penerapan algoritma K-Means dalam melakukan klaterisasi data rekam medis pasien pengguna layanan BPJS Kesehatan. Penelitian

tersebut melibatkan data rekam medis di rumah sakit Prof. Dr. Tabrani periode Oktober 2019 hingga Desember 2019. Hasil penelitian tersebut ditemukan pasien BPJS Kesehatan dapat dibedakan menjadi 3 *cluster* dengan kriteria yang berbeda, yaitu *cluster* 0 (diare/disentri), *cluster* 1 (penyakit beragam), dan *cluster* 2 (dyspepsia). Penelitian serupa juga dilakukan oleh Fitria Kurnia dan kawan- kawan (Kurnia et al. 2019) dalam melakukan pengelompokkan diagnosa penyakit mata berdasarkan rentang usia menggunakan metode K-Means. Penelitian tersebut berhasil dilakukan dengan menggunakan 6689 record data dan didapatkan informasi bahwa yang rentan terhadap penyakit myopia adalah pada usia balita, anak-anak, remaja, dan dewasa. Sedangkan untuk penyakit katarak sering terjadi pada usia tua.

Berdasarkan keberhasilan yang telah dicapai pada penelitian terdahulu, penerapan metode K-Means sangat dimungkinkan dalam melakukan pengelompokkan data penyakit yang ada di puskesmas. Pada penelitian ini, *clustering* dilakukan untuk mengelompokan data penyakit menjadi dua *cluster* yang bertujuan untuk mengelompokan data penyakit yang memiliki karakteristik yang sama menggunakan metode K-Means. K-Means *clustering* adalah algoritma non-hirarki yang digunakan untuk pengelompokan data kedalam satu atau beberapa *cluster* berdasarkan karakteristik data itu sendiri, sehingga hasil akhir data dalam satu *cluster* akan memiliki tingkat variasi yang kecil. Penggunaan metode K-Means didasarkan atas pertimbangan sifat algoritmanya yang mampu untuk peminimalan jarak antara data ke *cluster*nya.

2 Landasan Teori

2.1 Data mining

Data Mining adalah adalah suatu proses menggunakan satu atau lebih teknik kecerdasan buatan (artificial intelligence) dan pembelajaran mesin (machine learning) dengan tujuan menganalisis dan mengekstraksi informasi atau ilmu pengetahuan [9]. Definisi lainnya adalah suatu tahapan untuk menambang atau menggali informasi tersembunyi dari suatu database atau sebuah pusat data besar. Apabila disederhanakan, data mining berarti sebuah proses penambangan data menggunakan suatu teknik yang nantinya dianalisis sehingga seuatu informasi atau pengetahuan dapat didapatkan.

Beberapa tahapan dalam *data mining*, yaitu(Saripurna and Ristamaya 2021):

1. *Data Cleaning*

Data Cleaning adalah proses yang bertujuan meniadakan data tidak relevan atau sesuai, data noise, dan data-data lain yang tidak berkaitan dengan tujuan akhir proses *data mining*.

2. *Data Integration*

Data Integration atau integrasi data adalah proses dengan tujuan untuk penggabungan data dari berbagai database ke suatu database yang baru.

3. *Data Selection*

Seleksi data adalah suatu proses pemilihan data yang cocok dengan keperluan analisa, data yang tidak cocok akan dikembalikan ke dalam database.

4. *Data Transformation*

Transformasi data atau *data transformation* adalah suatu proses yang bertujuan untuk mengubah suatu data ke bentuk lain sehingga mudah disesuaikan untuk proses *data mining*.

5. *Data Mining*

Data Mining atau penambangan data adalah suatu proses untuk menetapkan metode untuk mendapatkan pengetahuan tersembunyi dari dalam data.

6. *Pattern Evolution*

Pattern Evolution atau Evaluasi pola untuk mengidentifikasi sejumlah pola data berdasarkan perhitungan tertentu yang dianggap merepresentasikan suatu pengetahuan.

7. *Knowledge Presentation*

Knowledge presentation atau presentasi pengetahuan adalah suatu proses menyajikan dan memvisualisasikan pengetahuan yang telah didapatkan kepada user.

2.2 Clustering

Clustering atau klasterisasi merupakan suatu proses untuk mengelompokkan dan membagi pola data ke dalam beberapa kelompok data yang nantinya akan terbentuk pola dengan kemiripan maksimum yang terklasifikasi pada suatu *cluster* yang sama dan terpisah dari pola berbeda (kemiripan minimum) ke *cluster* yang berbeda. *Cluster* sendiri didefinisikan sebagai kelompok data yang memiliki kesamaan karakterisitik antar satu dengan yang lainnya dan tidak memiliki kesamaan karakteristik terhadap data yang berbeda.

Dalam melakukan *clustering* terdapat beberapa pendekatan, dua pendekatan yang utamanya antara lain pendekatan hirarki dan pendekatan partisi. *Clustering* dengan pendekatan *hierarchical clustering* yaitu pendekatan yang mengelompokan data dengan membuat suatu tingkatan berupa dendrogram dimana data yang mirip akan ditempatkan pada tingkatan yang berdekatan. Pendekatan *clustering* partisi atau partition based *clustering* yaitu pendekatan yang mengelompokan data dengan membagi data kedalam *cluster* yang ada (Siahaan 2019).

2.3 Algoritma K-Means

Algoritma K-Means merupakan salah satu algoritma pembelajaran unsupervised learning sederhana yang sering digunakan dalam masalah

pengelompokan. Algoritma ini merupakan salah satu cara mengelompokan data nonhierarki yang membagi data ke dalam satu atau lebih kelompok (Hutagalung and Sonata 2021). Keunggulan algorimta K-Means ini adalah mudah diimplementasikan, sederhana, mudah disesuaikan, cepat, dan juga algoritma ini paling sering digunakan dalam proses data mining. Langkah – langkah dari K-Means Clustering (Hutagalung and Sonata 2021) :

a. Penentuan berapa banyak k klaster yang akan dibentuk.
b. Memilih centroid awal secara random sebanyak k *cluster.*
c. Perhitungan jarak antar setiap data input pada masing-masing centroid menggunakan rumus jarak Eucledian sampai ditemukan jarak terdekat setiap data dengan centroid. Rumus Euclidien Distance :

$$Z(x_i, y_i) = \sqrt{\Sigma_{i=1}^{n}(x_i - y_i)^2} \quad (2.1)$$

Keterangan :
$Z(x_i, y_i)$ = Jarak objek antar nilai data dan centroid
x_i = Data ke I pada atribut data ke k
y_i = Nilai centroid
n = Banyaknya atribut data

d. Melakukan perhitungan nilai centroid baru. Nilai centroid baru bisa diperoleh dari rata-rata dari semua data atau objek dalam klaster tertentu.
e. Melakukan perhitungan jarak pada setiap objek dengan menggunakan pusat centroid yang baru sama seperti pada langkah c.
f. Ulangi langkah c dan d sampai tidak adanya perubahan anggota *cluster*

2.4 Davies Bouldin Index

Metode Davies Bouldin Index merupakan salah satu cara yang dapat digunakan untuk pengukuran validitas suatu cluster untuk pemaksimalan jarak inter-klaster diantara cluster dan peminimalan jarak antara titik dalam sebuah cluster(Septiani, Fauzan, and Huda 2022). Davies Bouldin Index memiliki skema evaluasi dari internal cluster, yang baik atau tidaknya suatu hasil cluster dapat dilihat dari kuantitas dan kedekatan antar data hasil kluster. Tahapan dalam perhitungan Davies Bouldin Index sebagai berikut (Az-zahra et al. 2021):

a. Hitung nilai *Sum of Square Within-Cluster* (SSW) untuk mengetahui kohesi pada sebuah *cluster* ke – i.

$$SSW_i = \frac{1}{mi} \Sigma_{j=i}^{mi} d(x_j, c_i) \quad (2.2)$$

Keterangan :
mi = jumlah data dalam *cluster* ke-i
c_i = centroid *cluster* ke-i
$d(x_j, c_i)$ = Jarak setiap data ke centroid I yang dihitung menggunakan jarak eulidean.

b. Hitung nilai *Sum of Square Between-Cluster* (SSB) yang berfungsi untuk mengetahui jarak antar *cluster.*

$$SSBij = d(xi, xj) \quad (2.3)$$

Keterangan :
$d(xi, xj)$ = Jarak antara data ke-I dan data ke-j di *cluster* lain

c. Hitung ratio ini berfungsi untuk mengetahui nilai perbandingan antar *cluster* ke – I dan *cluster* ke – j untuk dapat mengetahui nilai rasio yang dimiliki oleh masing- masing *cluster*. Variabel i dan j merupakan representasi dari jumlah *cluster,* untuk menghitung nilai rasio menggunakan rumus berikut :

$$R_{ij} \ldots, n = \frac{SSW_i + SSW_j + \cdots + SSW_n}{SSB_{i,j} + \cdots + SSB_{ni,nj}} \quad (2.4)$$

Keterangan :
SSW_I = *Sum Of Square Within Cluster* pada centroid i
$SSB_{i,j}$ = *Sum of Square Between Cluster* data ke i dengan j pada *cluster* yang berbeda

Pada perhitungan ration ini *n* akan berlanjut sebanyak jumlah *cluster. Cluster* yang dipilih dengan syarat n_i tidak sama dengan n_j.

d. Davies Bouldin Index (DBI)

Hasil ratio yang sudah didapatkan dari rumus 2.4 digunakan untuk mencari hasil DBI nya menggunakan rumus:

$$DBI = \frac{1}{k} \Sigma_{i=1}^{k} max_{i \neq j}(R_{i,j}, \ldots k) \quad (2,5)$$

Yang dimana *Ri,j* adalah ratio dari nilai SSW dan SSB. Variable *k* adalah jumlah *cluster* yang digunakan. Syarat penentuan baik atau tidaknya suatu *cluster* adalah apabila nilai DBI yang diperoleh semakin kecil (non negative >= 0) maka kelompok tersebut semakin baik (Deny Jollyta, Muhammad Siddik, Herman Mawekang 2021).

3 Metodologi Penelitian

3.1 Tahapan Penelitian

Tahapan penelitian diawali dengan kajian penelitian, pengumpulan data, pengolahann data, analisis kebutuhan sistem, perancangan sistem, implementasi, uji akurasi, dan penarikan

kesimpulan. Untuk lebih jelasnya dapat dilihat pada Gambar 3.1.

Gambar 3. 1 Tahapan penelitian

3.1.1 Kajian penelitian

Pada tahapan kajian penelitian ini dilakukan kajian terhadap penelitian terdahulu, dan memahami konsep penelitian yang akan dilakukan.

3.1.2 Pengumpulan data

Data yang diperoleh dari hasil wawancara dengan salah satu staff bagian rekam medis yang ada di Puskesmas Kotagede 2 Yogyakarta. Data yang digunakan adalah data rekam medis variable usia, diagnosa, wilayah mulai dari bulan November 2022 – November 2023.

Tabel 3. 1 Data Pasien

No	Usia	Diagnosa	Wilayah	Jenis Kelamin
1	12	Asma	Rejowangun	L
2	25	Tipus	Banguntapan	P
3	19	Anemia	Prenggan	P
4	47	Hipertensi	Tamanan	L
5	8	DBD	Banguntapan	P
6	15	Sakit kepala	Giwangan	P
7	18	Tipus	Purbayan	L
8	29	Vertigo	Purbayan	L
9	27	Gerd	Purbayan	P
10	34	Nasopharingits Akut	Purbayan	L
11	26	Demam	Prenggan	P
12	13	Tipus	Banguntapan	L
13	17	Anemia	Purbayan	P
14	36	Rinitis, nasofaringitis dan faringitis kronik	Rejowangun	P
15	40	Migren	Prenggan	L

3.1.3 Pengolahan Data

Data penyakit yang ada kemudian diseleksi dan diolah sehingga menjadi dataset yang dapat diproses oleh sistem. Pengolahan data ini dimulai dari seleksi data yang awalnya ada empat atribut yaitu, usia, diagnosis, wilayah dan jenis kelamin yang kemudian di seleksi menjadi 3 atribut yang akan digunakan yaitu usia, diagnosis, dan wilayah. Pada pengolahan data ini juga dilakukan inisialisasi agar program dapat memproses data. Data awal ini bisa dilihat pada Tabel 3.2

Tabel 3. 2 Data awal

No	Usia	Diagnosa	Wilayah	Jenis Kelamin
1	12	Asma	Rejowangun	L
2	25	Tipus	Banguntapan	P
3	19	Anemia	Prenggan	P
4	47	Hipertensi	Tamanan	L
5	8	DBD	Banguntapan	P
6	15	Sakit kepala	Giwangan	P
7	18	Tipus	Purbayan	L
8	29	Vertigo	Purbayan	L
9	27	Gerd	Purbayan	P
10	34	Nasopharingits Akut	Purbayan	L
11	26	Demam	Prenggan	P
12	13	Tipus	Banguntapan	L
13	17	Anemia	Purbayan	P
14	36	Rinitis, nasofaringitis dan faringitis kronik	Rejowangun	P
15	40	Migren	Prenggan	L

Tabel 3. 3 Inisialisasi umur

Kode	Usia	Keterangan
1	0 - 5	Balita
2	6 sampai 9	Anak-anak
3	10 sampai 18	Remaja
4	19 sampai 59	Dewasa
5	60+	Lansia

Tabel 3. 4 Inisialisasi diagnosa

Kode	Jenis Penyakit
1	Hipertensi Esensial
2	Nasopharingits Akut
3	Diabetes Melitus 2
4	Rinitis, nasofaringitis dan faringitis kronik
5	Batuk
6	Demam
7	Asma
8	Sakit kepala
9	Katarak
10	Myalgia
11	Migren
12	Pneumonia
13	Campak
14	Sembelit
15	Anemia
16	Tipus
17	Vertigo
18	DBD
19	Asam urat
20	Kolestrol

Tabel 3. 5 Inisialisasi wilayah

Kode	Wilayah
1	Rejowinagun
2	Prenggan
3	Purbayan
4	Pandeyan
5	Sorosutan
6	Giwangan
7	Warungbroto
8	Banguntapan
9	Tamanan
10	Mujamuju
11	Singosaren
12	Jagalan

Setelah dilakukan seleksi dan inisiasi data, maka tampilan data yang sudah diolah seperti pada Tabel 3.6 dibawah ini.

Tabel 3. 6 Data yang sudah diolah

No	Usia	Diagnosa	Wilayah
1	3	7	1
2	4	16	8
3	4	15	2
4	4	1	9
5	2	18	8
6	3	8	6
7	3	16	3
8	4	17	3
9	4	10	3
10	4	2	3
11	4	6	2
12	3	16	8
13	3	15	3
14	4	4	1
15	4	11	2

3.1.4 Analisis Kebutuhan Sistem

3.1.4.1 Kebutuhan Fungsional

Kebutuhan fungsional ini mencakup semua proses maupun fitur yang dapat di lakukan oleh sistem sehingga sistem ini bisa mencapai tujuan yang diinginkan dan juga kebutuhan fungsional ini berisi informasi yang harus dihasilkan oleh sistem. Di bawah ini adalah kebutuhan fungsional yang dapat di lakukan oleh sistem.

1. Sistem dapat mengunggah file dalam format excel.
2. Sistem mampu melakukan proses perhitungan klastering.
3. Sistem mampu melakukan proses pengujian hasil klaster menggunakan metode davies bouldin index.
4. Sistem dapat menampilkan hasil clustering data.
5. Sistem dapat menampilkan visualisasi data berdasarkan hasil klaster.

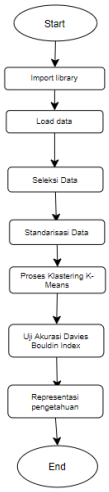
3.1.4.2 Kebutuhan Non Fungsional

Kebutuhan non fungsional merupakan kebutuhan yang digunakan sebagai batasan pada fungsi yang ada pada sistem, Kebutuhan non fungsional pada sistem ini antara lain:

1. Sistem dapat dijalankan di berbagai web browser.
2. Sistem memiliki desain antar muka yang mudah dipahami oleh user.
3. Jenis dokumen yang dapat di input adalah file.xlsl.

3.1.5 Perancangan Sistem

Sistem ini dirancang untuk bisa melakukan operasi *clustering* data menggunakan metode K-Means *clustering*. Alur dari sistem ini digambarkan pada Gambar 3.2



Gambar 3. 2 Alur Sistem

3.1.5.1 Import library

Tahap awal dalam pembuatan progam ini adalah import library yang akan digunakan.

```
import pandas as pd
import numpy as np
from sklearn import datasets
from sklearn.cluster import KMeans
import matplotlib.pyplot as plt
from sklearn.preprocessing import StandardScaler
from sklearn.metrics import davies_bouldin_score
```

Kode program 3. 1 Import library

3.1.5.2 Load data

Tahap selanjutnya merupakan tahap load dataset yang dimana data yang akan diimport ke dalam program yang selanjutnya dataset ini akan menampilkan data berupa no, usia, diagnosa, wilayah, dan jenis kelamin dalam periode Nov 2022 – Nov 2023.

```
df = pd.read_excel('medis2.xlsx', index_col=0,
names=['no', 'usia', 'diagnosa', 'wilayah',
'jenis_kelamin'])
df
```

Kode program 3. 2 Load data

3.1.5.3 Seleksi data

Seleksi data ini berfungsi untuk mengambil data yang akan digunakan. Pada data penyakit ini terdapat empat atribut yaitu usia, diagnose, wilayah, dan jenis kelamin. Atribut ini kemudian diseleksi sehingga hanya tersisa atribut usia, diagnosa,wilayah . Atribut inilah yang akan melalui proses clustering.

```
data =newdf.loc[:, ['usia', 'diagnosa', 'wilayah']]

data
```

Kode program 3. 3 Seleksi data

3.1.5.4 Standarisasi data

Pada tahapan ini data diubah sesuai dengan rata-rata dan deviasi standar yang ada sehingga dapat di proses ke tahapan k-means *clustering.*

```
scaler = StandardScaler()

scaler.fit(data)

df_scaled = scaler.transform(data)

df_scaled
```

Kode program 3. 4 Standarisasi data

3.1.5.5 Proses K-Means Clustering

Tahapan selanjutnya yaitu proses k-means clustering. Pada penilitian ini akan menghasilkan dua cluster sehingga nilai n_clusters=2. Setelah melalui proses clustering ini maka dari 300 penyakit akan terbagi menjadi 2 kelompok

.

```
km = KMeans(n_clusters=2)

y_predicted = km.fit_predict(df_scaled)

newdf['cluster'] = y_predicted
```

```
newdf
```

Kode program 3. 5 Proses K-Means

3.1.5.6 Uji Akurasi

Tahap uji akurasi ini berfungsi untuk mengetahui seberapa baik data dalam suatu klaster saling berdekatan. Dataset yang digunakan sebanyak 300 yang kemudian di kelompokan menjadi dua. Setelah pengelompokan ini kemudian dilakukan uji akurasi menggunakan metode Davies Bouldin Index yang mengukur validitas suatu *cluster* untuk pemaksimalan jarak inter-klaster diantara *cluster* dan peminimalan jarak antara titik dalam sebuh *cluster.*

```
akurasi = davies_bouldin_score(data, y_predicted)
akurasi
```

Kode program 3. 6 Uji akurasi DBI

3.1.6 Implentasi

Pada tahapan implementasi, proses *clustering* dan pengujian akurasinya serta desain sistem di implementasikan ke dalam bentuk website agar mudah dalam penggunaannya.

3.1.7 Uji Akurasi

Uji akurasi pada penelitian ini menggunakan metode *Davies bouldin index.* Setelah mendapatkan hasil *cluster* kemudian dihitung nilai SSW, SSB,nilai ratio, dan nilai DBI nya. Syarat baik atau tidaknya suatu *cluster* pada metode ini adalah semakin kecil (non negative >=0) maka *cluster* tersebut semakin baik.

4 Hasil dan Pembahasan

Pada perancangan sistem yang sudah di paparkan sebelumya, program *clustering* yang menggunakan Bahasa *python* kemudian di implementasikan ke dalam bentuk website agar mudah untuk digunakan.

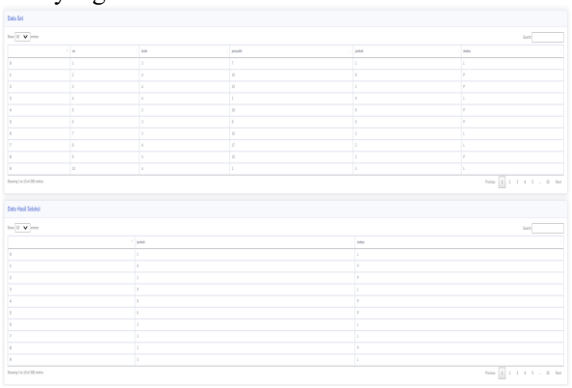
4.1 Halaman Beranda

Halaman beranda ini merupakan halaman awal yang berisi tentang tata cara penggunaan website dan juga upload data.

Gambar 4. 1 Halaman beranda

4.2 Halaman Tampil dan Seleksi Data

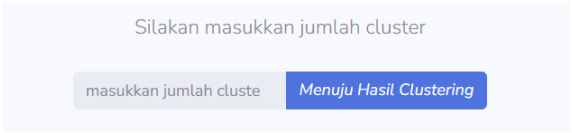
Halaman tampil data dan seleksi data ini merupakan halaman yang berfungsi menampilkan data yang sudah di upload dan juga menampilkan data yang sudah diseleksi.



Gambar 4. 2 Halaman tampil dan seleksi data

4.3 Tampilan menentukan jumlah *cluster*

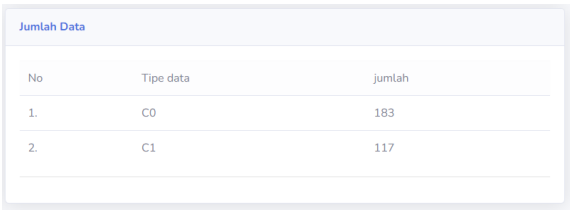
Tampilan ini merupakan salah satu fitur yang berfungsi untuk menentukan berapa jumlah *cluster* yang akan di gunakan



Gambar 4. 3 Menentukan jumlah cluster

4.4 Tampilan hasil cluster

4.4.1 Tampilan jumlah anggota setiap cluster



Gambar 4. 4 Tampilan jumlah anggota tiap cluster

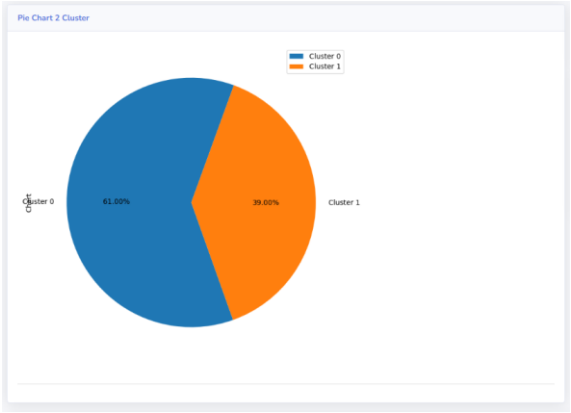
4.4.2 Tampilan data anggota setiap cluster

Gambar 4. 5 Tampilan data nggota cluster 0

Gambar 4. 6 Tampilan data anggota cluster 1

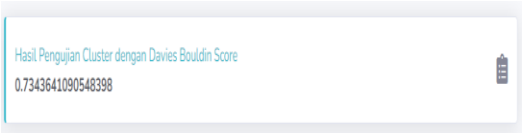
4.4.3 Tampilan Diagram Pie

Pada Gambar 4.7 terdapat diagram pie hasil dari clustering yang telah dilakukan. Warna biru dengan nilai 83.08% untuk cluster 0 yang merupakan penyakit yang banyak diderita dan warna oranye 16.92% untuk cluster 1 yang merupakan penyakit yang jarang diderita..



Gambar 4. 7 Diagram pie hasil clustering

4.5 Tampilan Hasil Uji Akurasi



Gambar 4. 8 Hasil Uji Akurasi DBI

Uji akurasi pada penelitian ini menggunakan metode Davies Bouldin Index yang dimana syarat penentuan baik atau tidaknya suatu *cluster* adalah apabila nilai DBI yang diperoleh semakin kecil (non negative >= 0) maka kelompok tersebut semakin baik. Dalam uji akurasi ini peneliti telah melakukan beberapa kali percobaan untuk mendapatkan hasil uji akurasi yang baik.

Tabel 4. 1 Percobaan uji akurasi

Jumlah data	Jumlah *cluster*	Hasil uji akurasi
10	2	2.15
150	2	0.82
300	2	0.73
300	3	1.95

Dari beberapa percobaan tersebut hasil uji akurasi yang paling baik adalah di nilai 0.73 dengan 300 data dan menggunakan 2 *cluster* .

4.6 Pembahasan

Dari hasil clustering data pasien sebanyak 300 data dari bulan Nov 2022 – Nov 2023 yang dibagi menjadi dua cluster, cluster 0 berjumlah 183 data pasien dengan kasus penyakit yang ada di cluster ini antara lain asma, hipertensi, sakit kepala, gerd, nasopharingitis akut, demam, Rinitis, nasofaringitis dan faringitis kronik, migren, diabetes mellitus 2, batuk, katarak, untuk usia rata – rata penderita pada cluster 0 ini adalah 34 tahun dengan wilayah penderita yang paling dominan berasal dari Kelurahan Rejowinangun.

Untuk cluster 1 berjumlah 117 data pasien yang kasus penyakitnya antara lain tipus, anemia, dbd, vertigo, asam urat, campak, vertigo, sembelit, dan kolestrol dengan rata – rata usia penderita 29 tahun dengan wilayah asal penderita berasal dari Kelurahan Banguntapan.. Hasil cluster ini kemudian di uji menggunakan metode davies bouldin index dengan hasil uji akurasi sebesar 0.734 yang berarti hasil cluster ini merupakan cluster terbaik karena hasil uji akurasi dengan menggunakan dua cluster dan jumlah data sebanyak 300 adalah hasil uji akurasi yang paling mendekati 0 dibanding dengan beberapa percobaan lainnya.

5 Kesimpulan dan saran

5.1 Kesimpulan

Berdasarkan dari hasil dan pembahasan mengenai pengelompokan data penyakit di Puskesmas Kotagede 2 Yogyakarta, maka dapat disimpukan:

1. Hasil implementasi algoritma K-Means untuk clustering pada 300 data penyakit terbagi menjadi 2 cluster. Pada cluster 0 terdapat 183 anggota dan untuk cluster 1 terdapat 117 anggota. Cluster 0 merupakan kasus yang banyak terjadi dan cluster 1 merupakan kasus penyakir yang jarang terjadi.
2. Cluster 0 berjumlah 183 anggota dengan kasus penyakit yang ada di cluster ini antara lain 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 penderita yang paling dominan berasal dari Kelurahan Rejowinangun.
3. Untuk cluster 1 berjumlah 117 anggota yang kasus penyakitnya antara lain tipus, anemia, dbd, vertigo, asam urat, campak, vertigo, sembelit, dan kolestrol dengan rata – rata usia penderita 29 tahun dengan wilayah asal penderita berasal dari Kelurahan Banguntapan.

4. Pada penelitian ini dilakukan beberapa percobaan agar mendapatkan hasil uji akurasi yang baik. Percobaan pertama dilakukan dengan menggunakan 10 data penyakit yang dibagi menjadi 2 cluster menghasilkan akurasi sebesar 2.15. Percobaan kedua menggunakan 150 data penyakit yang dibagi menjadi 2 cluster menghasilkan akurasi sebesar 0.82. Percobaan ketiga menggunakan 300 data penyakit yang dibagi menjadi 2 cluster menghasilkan nilai akurasi sebesar 0.73. dan percobaan terakhir menggunakan 300 data yang dibagi menjadi 3 cluster menghasilkan nilai akurasi 1.95. Dari ke-empat percobaan tersebut, nilai akurasi yang memiliki nilai yang paling rendah adalah 0.73 yang dihasilkan oleh cluster 300 data yang dibagi menjadi 2 cluster.

5.2 Saran

Untuk penelitian yang akan dilakukan kedepannya, disarankan untuk menggunakan data yang lebih banyak dan mempertimbangkan untuk menggunakan berbagai metode agar dapat menghasilkan cluster yang lebih baik. Selain itu, hasil cluster pada penelitian ini bisa diberikan label pada penelitian kedepannya. Pada bagian sistem juga diperlukan pengembangan dari segi UI/UX maupun fitur-fitur yang ada.

DAFTAR PUSTAKA

Az-zahra, Alyeska Astri et al. 2021. "Penerapan Algoritma K-Modes Clustering Dengan Validasi Davies Bouldin Index Pada Pengelompokkan Tingkat Minat Belanja Online Di Provinsi Daerah Istimewa Yogyakarta." *Jurnal MSA (Matematika dan Statistika serta Aplikasinya)* 9(1): 24.

Baharudin, Dapis, Riky Faza, and Leni Herfiyanti. 2021. "Perancangan Sistem Informasi Berkas Keluar Rekam Medis Di Puskesmas Baleenedah." *Jurnal Teknologi Informasi* 5(2): 1–7.

Deny Jollyta, Muhammad Siddik, Herman Mawekang, Syahril Efendi. 2021. *Teknik Evaluasi Cluster Solusi Menggunakan Python Dan RapidMiner*. Sleman, Yogyakarta: Deepublish. https://edeposit.perpusnas.go.id/collection/teknik-evaluasi-cluster-solusi-menggunakan-python-dan-rapidminer-sumber-elektronis/47071#.

Hutagalung, Juniar, and Fifin Sonata. 2021. "Penerapan Metode K-Means Untuk Menganalisis Minat Nasabah." *Jurnal Media Informatika Budidarma* 5(3): 1187.

Kurnia, Fitra et al. 2019. "PENERAPAN ALGORITMA K-MEANS UNTUK PENGELOMPOKAN DIAGNOSA PENYAKIT MATA BERDASARKAN RENTANG USIA." *Jurnal SPEKTRO* 2(1).

Saripurna, D, and W Ristamaya. 2021. "Data Mining Untuk Pengelompokan Data Penjualan Cake Dengan Menggunakan Algoritma K-Means Clustering Pada Jofie Bakery." *Jurnal Cyber Tech* (x). https://ojs.trigunadharma.ac.id/index.php/jct/article/view/373.

Septiani, Ike Wahyu, Abd. Charis Fauzan, and Muhamat Maariful Huda. 2022. "Implementasi Algoritma K-Medoids Dengan Evaluasi Davies-Bouldin-Index Untuk Klasterisasi Harapan Hidup Pasca Operasi Pada Pasien Penderita Kanker Paru-Paru." *Jurnal Sistem Komputer dan Informatika (JSON)* 3(4): 556.

Siahaan, Herdianto. 2019. "Implementasi Metode Clustering Partitional Menentukan Item Slow Moving Dan Fast Moving Pada Persediaan Barang (Studi Kasus PT. SAT)." *Jurikom)* 6(2): 171–77. http://ejurnal.stmik-budidarma.ac.id/index.php/jurikom%7CPage%7C171.

Wandana, Jeri, Sarjon Defit, and S Sumijan. 2020. "Klasterisasi Data Rekam Medis Pasien Pengguna Layanan BPJS Kesehatan Menggunakan Metode K-Means." *Jurnal Informasi dan Teknologi*.