

Penerapan Algoritma *Density Based Spastial Clustering Algorithm With Noise* Untuk Pengelompokkan Penyakit Pasien

Dia Ayu Puspitasari
Universitas Buana Perjuangan Karawang
Karawang, Indonesia
if18.diapuspitarsari@mhs.ubpkarawang.ac.id

Yana Cahyana
Universitas Buana Perjuangan Karawang
Karawang, Indonesia
yana.cahyana@ubpkarawang.ac.id

Santi Arum Puspita Lestari
Universitas Buana Perjuangan Karawang
Karawang, Indonesia
santi.arum@ubpkarawang.ac.id

Abstract— Setiap tahun kuantiti pasien yang menjalankan pengobatan di rumah sakit As-sofwan semakin meningkat dengan beragam jenis penyakit yang berbeda-beda. Dari jumlah data pasien dengan beragam jenis penyakit yang berbeda, belum terlihat adanya pendataan secara teratur terhadap tingkat keparahan penyakit yang diderita oleh pasien sehingga menyulitkan pihak rumah sakit dalam melakukan penyediaan obat-obatan dan peralatan medis. Maka penelitian ini bertujuan untuk mengelompokkan penyakit pasien kedalam 2 *cluster* yaitu banyak dan sedikit. Pengelompokan data pasien dilakukan dengan tahapan data mining *clustering*. Dilakukan perhitungan menggunakan jupyter dengan bahasa pemrograman python.

Kata kunci — *Clustering, Data Mining, DBSCAN, Penyakit, Python*

I. PENDAHULUAN

Penyakit merupakan suatu keadaan dimana pikiran dan tubuh mengalami semacam rasa nyeri dan ketidaknyamanan bagi yang mengalaminya [1] Setiap tahun kuantiti pasien yang menjalankan pengobatan di rumah sakit As-sofwan semakin meningkat, berdasarkan sumber data yang diambil dari rumah sakit As-sofwan pada tahun 2019 rumah sakit As-sofwan melayani pasien dengan jumlah sekitar 593 orang per-tahun. Dari jumlah data pasien dengan berbagai jenis penyakit yang berbeda, belum terlihat adanya pendataan secara teratur terhadap tingkat keparahan penyakit yang di alami oleh pasien pada rumah sakit As-sofwan sehingga menyulitkan pihak rumah sakit dalam melakukan penyediaan obat-obatan dan peralatan kesehatan.

Untuk melancarkan proses pengolahan data yang banyak tersebut, maka suatu institusi memerlukan sebuah sistem dalam mengambil langkah untuk mempermudah mendapati penyakit yang terbanyak dan sedikit yang dialami oleh pasien pada rumah sakit As-sofwan. Salah satu cara dalam mengolah data yang banyak dapat memanfaatkan Teknik Data Mining. Data mining merupakan metode mencari pengetahuan dari data yang berskala tinggi. Data diproses menggunakan teknik data mining yang selanjutnya akan menghasilkan suatu pengetahuan yang baru, kemudian hasil dari proses tersebut dapat dimanfaatkan dalam menentukan sebuah keputusan. Dalam teknik pengolahan data mining terdapat berbagai macam cara yang dapat digunakan seperti, *classification, grouping, estimation, forecast* dan *asosiation* [2].

Berdasarkan permasalahan tersebut, maka dalam penelitian ini metode yang digunakan dalam pengelompokkan penyakit pasien yaitu *clustering*. Salah satu teknik *clustering* yang sering digunakan dalam data mining yaitu algoritma *Density Based Spatial Clustering Of Application With Noise* pada dasarnya DBSCAN mampu menemukan beberapa *cluster* yang bersifat bebas dan acak yang sanggup menemukan *cluster* dengan makin sederhana jika menemukan *noise* pada salah satu *cluster* tersebut [3].

II. METODE PENELITIAN

A. Data Mining

Data Mining atau disebut juga dengan *pattern recognition* adalah sebutan yang sering dipakai dalam mendapatkan sebuah pengetahuan yang meliputi sekumpulan data yang berukuran sangat tinggi dengan objek utamanya yaitu untuk mendapatkan, menelusuri, atau menggali suatu informasi dari data yang dimiliki [4]

B. Pasien

Penyakit merupakan dimana kondisi tubuh mengalami rasa nyeri yang berlebihan sehingga mengakibatkan keterhambatannya kegiatan sehari-hari. Menurut surat keputusan Menteri Kesehatan RI no.269/KEMENKES/PER/III/2008 tentang rekam klinis, bahwa pasien merupakan “setiap orang yang melakukan konsultasi kesehatannya untuk memperoleh pelayanan Kesehatan yang di perlukan baik secara langsung kepada dokter maupun dokter gigi”. Menurut [4] pasien merupakan seorang yang mendapatkan perawatan medis atau terkena penyakit, cedera yang memerlukan pertolongan dokter untuk menyembuhkannya.

C. Algoritma DBSCAN

Algoritma DBSCAN membentuk wilayah yang memiliki kerekatan tinggi menjadi beberapa kelompok dengan dua model yang diperlukan yaitu radius ketegangan di simbolkan dengan ϵ , dan minimum objek yang menjadi batas kerapatan (*density threshold*) untuk memastikan kerapatan dari wilayah tersebut, yang disimbolkan $MinObj$ [5] keunggulan dari metode DBSCAN dengan metode *clustering* yang lain, adalah keahlian dalam menemukan *outlier/noise*. Karena konsep *density-based* yang dimanfaatkan, yaitu objek yang tidak memiliki fungsi kedekatan jarak dengan objek yang lainnya akan dikenali sebagai *outlier*. Selain itu, dalam DBSCAN tidak perlu mengetahui jumlah *cluster* (k) yang akan terbentuk, sebagaimana yang ada pada k -Means dan k -Medoids. Perhitungan algoritma DBSCAN menggunakan fungsi *Euclidean Distance*.

$$d_{ij} = \sqrt{\sum_{a=1}^p (x_{ia} - x_{ja})^2, i=1, \dots, n; j=1, \dots, n.} \quad (1)$$

Dimana:

D_{ij} = Nilai Euclidean Distance

X_{ia} = Variabel ke- a dari obyek i ($i=1, \dots, n; a=1, \dots, p$)

Langkah-langkah dalam algoritma DBSCAN sebagai berikut:

1. Tentukan titik *point* p awal secara acak
2. Ambil semua *point* yang *density reachable* terhadap titik p
3. Jika p adalah *core point* maka *cluster* terbentuk
4. Jika p adalah *border point*, tidak ada yang merupakan hubungan *density-reachable* dari p dan DBSCAN akan menguji *point* selanjutnya.
5. Lakukan proses sampai semua *point* telah selesai diproses

D. Evaluasi

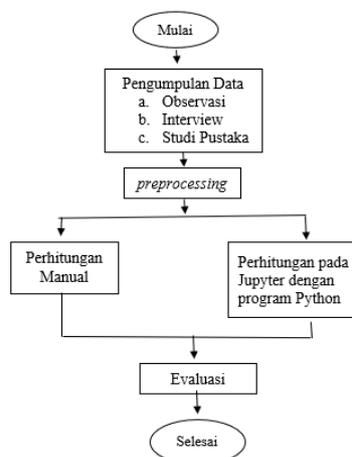
Evaluasi dilakukan untuk memastikan apakah model dan algoritma yang telah dibangun sesuai dengan tujuan atau tidak. Menurut [6] *Davies Bouldin Index* (DBI) Merupakan salah satu cara untuk melengkapi kinerja dari sebuah metode atau algoritma yang digunakan untuk memperoleh jumlah *cluster* yang terbaik. *Cluster* yang mempunyai nilai kohesi terkecil dan separasi yang terbesar adalah *cluster* yang baik (Zulfa Nabila dkk, 2021):

$$DBI = \frac{1}{k} \sqrt{\sum_{i=1}^k \max_{i \neq j} (R_{ij})} \quad (2)$$

K adalah kuantitas *cluster* yang digunakan. Semakin kecil nilai DBI yang didapatkan, maka semakin baik *cluster* yang dihasilkan.

E. Prosedur Penelitian

Prosedur penelitian dimulai dengan pengumpulan data dengan rincian sebagai berikut:



Gambar 1. Prosedur Penelitian

III. HASIL DAN PEMBAHASAN

A. Pengumpulan Data

Data yang dihasilkan dari rumah Sakit As sofwan Bekasi merupakan data rekam medis dengan parameter yang digunakan yaitu Nama, Jenis Kelamin, Umur dan Diagnosis tahun 2019/2020 yang berjumlah 593 data yang dapat dilihat pada Tabel 1.

Table 1 Rekam Medis

JANUARI			
NAMA	JK	UMUR	DIAGNOSIS
SARNIH BINTI SANIM	PEREMPUAN	50	TYPHOID FEVER
SALEH	LAKI LAKI	63	ACUTE ABDOMEN
MUHAMAD JASKASIH NUR	LAKI LAKI	4	ISPA
AMANDA AGUSTIAN	PEREMPUAN	1	KDK
NURHAKIKI	LAKI LAKI	5	FEVER
NURHAYATI	PEREMPUAN	25	G2PIA
HILDA NURBAYANI	PEREMPUAN	24	G2PIA
ADINDA PRIATNA	PEREMPUAN	4	ISPA
ALYA ADRIANA	PEREMPUAN	1	ISPA
IDA HAMIDAH	PEREMPUAN	42	GERD
JUMRIK	LAKI LAKI	44	HEPATITIS

B. Preprocessing Data

1. Cleaning Data

Menyiapkan informasi yang akan ditangani dengan menghapus atau mengubah informasi yang salah, terfragmentasi, tidak penting, disalin, atau dalam konfigurasi yang tidak sesuai kebutuhan. Data yang dihapus yaitu Nama, Jenis Kelamin, Tanggal Lahir, Nama Dokter, Nomor Peserta, Nomor Rekam Medis, dan Biaya sehingga menghasilkan dataset yang sesuai dengan yang dibutuhkan.

Table 2 Rekam Medis Cleaning

DIAGNOSIS	ID Umur	JAN	FEB	MAR	APR	MEI	JUN	JUL	AGS	SEP
DYPEPSIA	1	18	2	1	11	5	11	8	12	5
ACUTE ABDOMEN	1	8	3	3	4	1	1	5	1	11
DADRS	1	11	12	5	7	1	5	7	3	4
KDK	3	4	1	2	1	2	2	3	1	2
TBC	3	1	2	1	3	2	2	2	2	3

2. Transformation Data

Merubah data kedalam bentuk yang dibutuhkan untuk menentukan jumlah analisis penyakit setiap bulan Januari hingga September. Bagian penentuan diubah menjadi numerik untuk estimasi yang lebih sederhana sehingga dataset layak untuk ditangani untuk mendapatkan perhitungan yang lebih tepat.

Table 3 Rekam Medis Transformasi

ID Umur	JAN	FEB	MAR	APR	MEI	JUN	JUL	AGS	SEP
1	18	2	1	11	5	11	8	12	5
1	8	3	3	4	1	1	5	1	11
1	11	12	5	7	1	5	7	3	4
3	4	1	2	1	2	2	3	1	2
3	1	2	1	3	2	2	2	2	3

C. Perhitungan

1. Perhitungan Manual

a. Algoritma DBSCAN

Perhitungan manual DBSCAN menggunakan *Microsoft Excel 2019* menghasilkan perhitungan sebanyak 3 iterasi dengan nilai Epsilon dan Minpts 4 sehingga *core point* yang dihasilkan yaitu 4,38 dan 45.

2. Program Python

Perhitungan sistem ini menggunakan aplikasi Jupyter dengan memanfaatkan metode DBSCAN. Metode DBSCAN membutuhkan model input, yaitu Epsilon dan Minpts. Nilai yang digunakan untuk Epsilon dan Minpts yaitu bernilai 4. Hasil *clustering* dan nilai *noise* dapat dilihat pada Gambar 2 dan 3.

```
np.unique(labels)
print(labels)

[-1 -1 -1  0  0 -1  0 -1 -1  0 -1 -1 -1 -1  0  0 -1 -1  0  0  0 -1  0  0
  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0]
```

Gambar 2 Hasil Perhitungan Program Python Algoritma DBSCAN

```
n_noise = list(dbscan_cluster.labels_).count(-1)
print('Estimated no. of noise points: %d' % n_noise)
```

Estimated no. of noise points: 13

Gambar 3 Hasil Noise Algoritma DBSCAN

D. Evaluasi

Evaluasi *cluster* menggunakan metode *Davies Bouldien Index* (DBI) pada dataset rekam medis yang menghasilkan 2 *cluster* dengan nilai Epsilon 4 dan Minpts 4 menghasilkan nilai sebesar 1.45.

```
from sklearn.metrics import davies_bouldin_score

dbscan = DBSCAN (eps=4, min_samples=4)
labels= dbscan.fit_predict(df)
db_index=davies_bouldin_score(df, labels)
print(db_index)

1.422086806557305
```

Gambar 4 Evaluasi

IV. KESIMPULAN DAN SARAN

Berdasarkan pada hasil penelitian, dapat disimpulkan bahwa metode *clustering* memanfaatkan metode *Density Based Spatial Clustering Algorithm With Noise* dapat diimplementasikan dengan menghasilkan nilai *Davies Bouldien Index* sebesar 1.422086806557305 pada dataset rekam medis. Dan penerapan teknik *clustering Density Based Spatial Clustering Algorithm With Noise* pada pengelompokan penyakit pasien menghasilkan 2 *cluster* dengan menggunakan aplikasi jupyter pada bahasa pemrograman python sedangkan dari hasil perhitungan manual didapatkan iterasi sebanyak 3 kali yaitu *core point* yang dihasilkan adalah 4,38,45.

Diharapkan dalam penelitian selanjutnya peneliti dapat menambah atribut yang lain dalam melakukan perhitungan analisis data, dapat menambahkan dataset yang lebih banyak untuk dilakukan clustering agar hasil akurasi lebih optimal dan dapat dilakukan perbandingan antara metode yang telah digunakan dengan yang lainnya.

PENGAKUAN

Naskah ilmiah ini adalah sebagai dari penelitian Tugas Akhir milik Dia Ayu Puspitasari, dengan judul Penerapan Algoritma *Density Based Spastial Clustering Algorithm With Noise* Untuk Pengelompokkan Penyakit Pasien yang dibimbing oleh Yana Cahyana dan Santi Arum Puspitalestari.

DAFTAR PUSTAKA

- [1] D. U. Iswavigra, S. Defit, and G. W. Nurcahyo, "Data Mining dalam Pengelompokan Penyakit Pasien dengan Metode K-Medoids," *J. Inf. dan Teknol.*, vol. 3, pp. 181–189, 2021, doi: 10.37034/jidt.v3i4.150.
- [2] C. A. Sugianto, A. H. Rahayu, and A. Gusman, "Algoritma K-Means untuk Pengelompokkan Penyakit Pasien pada Puskesmas Cigugur Tengah," *J. Inf. Technol.*, vol. 2, no. 2, pp. 39–44, 2020, doi: 10.47292/joint.v2i2.30.
- [3] R. Adha, N. Nurhaliza, U. Soleha, P. Studi, S. Informasi, and F. Sains, "Perbandingan Algoritma DBSCAN dan K-Means Clustering untuk Pengelompokan Kasus Covid-19 di Dunia," vol. 18, no. 2, pp. 206–211, 2021.
- [4] S. S. Sundari and N. Ariani, "Penerapan Data Mining Untuk Pengelompokan Penyakit Dengan Algoritma Fuzzy C-Means (Studi Kasus : UPT Puskesmas Salawu)," *J. VOI (Voice Informatics)*, vol. 8, no. 2, pp. 63–76, 2019.
- [5] M. Windy Rohalidyawati, Rita Rahmawati, "JURNAL GAUSSIAN, Volume 9, Nomor 2, Tahun 2020, Halaman 162-169 Online di: <https://ejournal3.undip.ac.id/index.php/gaussian/> SEGMENTASI," vol. 9, pp. 162–169, 2020.
- [6] D. Jollyta, S. Efendi, M. Zarlis, and H. Mawengkang, "Optimasi Cluster Pada Data Stunting: Teknik Evaluasi Cluster Sum of Square Error dan Davies Bouldin Index," *Pros. Semin. Nas. Ris. Inf. Sci.*, vol. 1, no. September, p. 918, 2019, doi: 10.30645/senaris.v1i0.100.
- [7] W. Alim, S. Dedy, and K. Noraini, "Analisis Kesesuaian Penggunaan Lahan Aktual Terhadap Rencana Tata Ruang Wilayah (rtrw) (Studi Kasus : Kecamatan Serengan dan Kecamatan Pasar Kliwon , Surakarta , Jawa Tengah)," p. 8, 2017.