

Pemanfaatan Algoritma K-Means untuk Pengelompokan Pasien Penyakit Infeksi Saluran Pernafasan Akut (ISPA)

Friska Selvina Agoestina¹, Heru Satria Tambunan², Rizki Alfadillah Nasution³

^{1,2}Program Studi Sistem Informasi, STIKOM Tunas Bangsa, Pematangsiantar
Email: ¹ friskaselvina1@gmail.com

ABSTRACT

This study discusses Acute Respiratory Infections (ARI) at the Bah Biak Health Center. Bah Biak Health Center is one of the health centers in Marihat in Pematangsiantar city. Every day the number of patients who come and perform medical treatment at this health center is quite a lot. The high number of patient visits at this puskesmas causes the amount of medical record data to be very large as well. So far, data containing information about patients at the puskesmas has not been used properly. This information can actually be used as knowledge for puskesmas, especially for patients who have a history of ARI disease. Therefore, the purpose of this study was to classify patients with ARI at the Puskesmas. The method used is K-Means clustering. The results of this study were able to classify ARI patients into 2 clusters, cluster 1 gave a high recommendation of 72 patients, and cluster 2 gave a low recommendation of 70 patients. The cluster process stops at the 5th iteration of data. Based on the manual calculation process using Ms. Excel and testing using Rapidminer 5.3, yielded the same value. It can be concluded that in this case, the K-Means algorithm can classify ARI patients at the Bah Biak health center well.

Keywords: Data Mining, ARI, Bah Biak, Clustering, Puskesmas

ABSTRAK

Penelitian ini membahas tentang Penyakit Infeksi Saluran Pernafasan Akut (ISPA) pada Puskesmas Bah Biak. Puskesmas Bah Biak merupakan salah satu puskesmas yang ada di Marihat di kota Pematangsiantar. Setiap harinya jumlah pasien yang datang dan melakukan perawatan medis di puskesmas ini cukup banyak. Tingginya jumlah kunjungan pasien pada puskesmas ini menyebabkan jumlah data rekam medis menjadi sangat banyak pula. Selama ini data yang berisi informasi mengenai pasien di puskesmas belum dimanfaatkan dengan baik. Informasi tersebut sebenarnya dapat dijadikan suatu pengetahuan bagi puskesmas, khususnya bagi pasien yang memiliki riwayat penyakit ISPA. Oleh sebab itu tujuan penelitian ini untuk mengelompokkan pasien penyakit ISPA pada Puskesmas tersebut. Metode yang digunakan adalah K-Means klustering. Hasil penelitian ini mampu mengelompokkan pasien penyakit ISPA ke dalam 2 kluster, kluster 1 memberikan rekomendasi tinggi berjumlah 72 Pasien, dan kluster 2 memberikan rekomendasi rendah berjumlah 70 Pasien. Proses kluster berhenti pada data iterasi ke 5. Berdasarkan proses perhitungan manual dengan menggunakan Ms. Excel dan pengujian dengan menggunakan Rapidminer 5.3, menghasilkan nilai yang sama. Dapat disimpulkan bahwa untuk kasus ini, algoritma K-Means dapat mengelompokkan pasien penyakit ISPA yang ada di puskesmas Bah Biak dengan baik.

Kata Kunci: Data Mining, ISPA, Bah Biak, Pengelompokan, Puskesmas

1. Pendahuluan

Puskesmas bah biak pematang marihat merupakan salah satu Puskesmas yang terletak di kecamatan Siantar Marimbun Kota Pematangsiantar. Setiap tahunnya jumlah pasien yang berobat semakin meningkat. Saat ini Jumlah pasien semakin meningkat, peningkatan jumlah pasien berbanding lurus dengan

jenis penyakit pasien. Dari data jumlah pasien terhadap jenis penyakit Infeksi Saluran Pernafasan Akut tersebut, penyakit ini dikelompokkan dan paling banyak di derita pasien. Dengan adanya data penyakit pasien ini, maka pihak puskesmas dalam melakukan penyuluhan kesehatan, menyediakan obat-obatan dan peralatan kesehatan dalam melakukan pelayanan kesehatan yang tepat sasaran. Untuk itu diperlukan

pengelompokan (Kluster) terhadap penyakit ISPA. Metode yang digunakan adalah Klustering.

Klustering merupakan salah satu metode dalam data mining yang digunakan untuk mencari dan mengelompokkan data yang memiliki kesamaan karakteristik dengan lainnya. Data yang diolah dengan teknik data mining akan menghasilkan suatu pengetahuan (knowledge) yang baru bersumber dari informasi data lama [1]. Metode Klustering yang ada pada saat ini sangat banyak, salah satunya adalah K-Means. K-Means mempunyai sifat relatif efektif dan efisien. K-Means merupakan suatu algoritma yang digunakan dalam pengelompokan secara partisi yang memisahkan data ke dalam kelompok yang berbeda-beda. Algoritma ini mampu meminimalkan jarak antara data ke kluster nya [2]. Metode ini mempartisi data ke dalam Kluster/kelompok sehingga data yang memiliki karakteristik yang sama dan data yang mempunyai karakteristik yang berbeda dikelompokkan ke dalam kelompok yang lain [3]. Data tersebut diolah menggunakan Ms.Excel untuk ditentukan nilai centroid dalam 2 Kluster yaitu Kluster tingkat tinggi dan Kluster tingkat rendah. Sehingga diperoleh penilaian berdasarkan indeks penyakit menurut upaya antisipasi pada pasien penyakit infeksi saluran pernapasan akut (ISPA) di Puskesmas Bah Biak.

Ada Banyak cabang dibidang ilmu komputer yang mampu menyelesaikan masalah yang kompleks, hal ini terbukti dari banyaknya penelitian-penelitian yang sudah pernah dilakukan, seperti bidang sistem pendukung keputusan [4]–[11], bidang jaringan saraf tiruan [11]–[20], maupun penelitian dalam bidang data mining [25]–[32]. Beberapa penelitian terdahulu yang berhubungan dengan topik ini diantaranya: Penelitian untuk mengelompokkan Laju Pertumbuhan PDRB Kota Surabaya menggunakan Metode K-Means. Data dibagi menjadi 3 Kluster: tinggi, sedang dan rendah. Hasil yang diperoleh terdapat 9 kategori/sektor dengan kluster tinggi, 5 kategori/sektor dengan kluster sedang dan 3 kategori/sektor dengan kluster rendah [33]. Selanjutnya penelitian yang dilakukan untuk

mengelompokkan jumlah kasus terkonfirmasi Covid-19 dan jumlah kematian akibat virus ini di Asia Tenggara. Data yang digunakan adalah data statistik negara berdasarkan area kasus COVID-19 yang dikonfirmasi dan tercatat pada laboratorium, serta kematian pada bulan April 2020 dari WHO (Kesehatan Dunia Organisasi). Pada penelitian ini Data dibagi menjadi 3 Kluster: tinggi (C1), sedang (C2) dan rendah (C3). Hasil yang diperoleh terdapat empat negara dengan Kluster level tinggi (C1), satu negara dengan Kluster level sedang (C2), dan 6 negara dengan Kluster level rendah (C3) [34]. Berikutnya adalah penelitian menggunakan metode K-Means untuk mengelompokkan garis kemiskinan menurut kabupaten/kota di Sumatera Utara. Data yang digunakan adalah data garis kemiskinan menurut kabupaten/kota (rupiah/kapita/bulan) di Provinsi Sumatera Utara tahun 2017-2019. Data bersumber dari Badan Pusat Statistik Sumatera Utara. Pengelompokan tersebut dibagi menjadi 3 kluster yaitu garis kemiskinan kategori tinggi, garis kemiskinan kategori sedang, dan garis kemiskinan kategori rendah. Hasil untuk kategori tinggi terdiri dari 5 kabupaten/kota, kategori sedang terdiri dari 18 kabupaten/kota dan kategori rendah terdiri dari 10 kabupaten/kota [35].

Berdasarkan uraian latar belakang permasalahan di atas, maka penulis menggunakan teknik data mining menggunakan Algoritma K-Means Klustering untuk mengangkat sebuah judul yaitu “Pengelompokan Pasien Penyakit Infeksi Saluran Pernapasan Akut (ISPA) Menggunakan Metode Algoritma K-Means Klustering”. Dengan adanya penelitian ini diharapkan dapat mengetahui pengelompokan data pasien penyakit Infeksi Saluran Pernapasan Akut (ISPA) di Puskesmas Bah Biak Pematangsiantar.

2. Metode Penelitian

2.1. Lokasi dan Waktu Penelitian

Dataset penelitian diperoleh langsung dari Puskesmas Bah Biak Kota Pematangsiantar.

Tabel 1. Data Pasien Penyakit ISPA

Informasi Pasien			Gejala				
No	Jenis Kelamin	Usia	Batuk	Pilek	Demam	Sesak	Sakit Kepala
1	L	11	3	3	3	0	0
2	L	48	3	0	3	0	3
3	P	18	4	4	0	4	4
4	L	54	4	4	4	0	4
5	P	60	2	0	2	0	0
6	P	9	0	3	3	0	3
7	P	12	3	3	3	0	0
8	P	30	2	0	0	2	0
9	L	14	0	0	2	0	2
10	P	5	4	4	0	4	4
11	P	64	5	5	5	5	5
12	P	75	4	4	0	4	4
13	L	50	2	0	0	2	0
14	L	42	3	3	3	0	0
15	L	15	3	0	0	3	3

DOI: 10.54259/satesi.v2i1.804

Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)

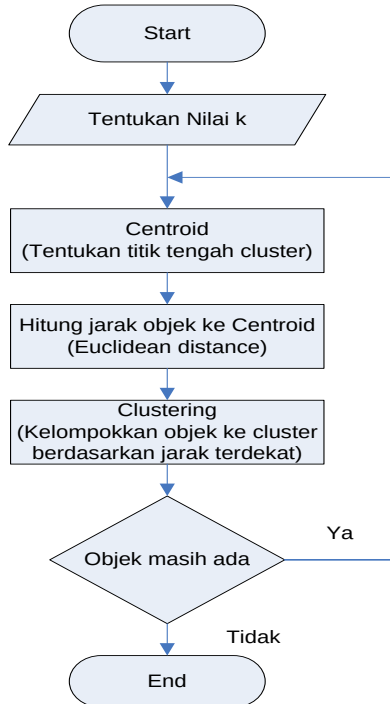
Informasi Pasien			Gejala				
No	Jenis Kelamin	Usia	Batuk	Pilek	Demam	Sesak	Sakit Kepala
16	L	52	2	2	0	0	0
17	L	13	3	3	3	0	0
18	L	35	2	0	2	0	0
19	P	15	3	3	3	0	0
20	P	3	2	0	2	0	0
21	P	51	3	3	3	0	0
22	P	54	0	3	3	0	3
23	L	45	2	0	2	0	0
24	P	7	3	3	0	0	3
25	L	43	2	2	0	0	0
26	L	35	4	4	0	4	4
27	P	63	2	2	0	0	0
28	P	25	2	2	0	0	0
29	P	35	4	4	4	0	4
30	L	54	5	5	5	5	5
31	P	75	4	4	4	4	0
32	P	54	3	3	3	0	0
33	L	42	2	2	0	0	0
34	L	15	4	4	4	4	0
35	L	52	2	2	0	0	0
36	L	13	0	3	3	0	3
37	L	35	3	3	3	0	0
38	L	4	2	2	0	0	0
39	L	14	4	4	0	4	4
40	P	62	3	3	0	3	0
41	P	67	5	5	5	5	5
42	P	77	3	3	3	0	0
43	P	24	2	0	2	0	0
44	L	11	4	4	0	4	4
45	P	65	4	4	4	4	0
46	P	64	3	3	3	0	0
47	L	10	2	2	0	0	0
48	P	29	4	4	0	4	4
49	L	51	3	3	3	0	0
50	P	7	5	5	5	5	5
51	L	80	3	3	0	0	3
52	L	57	4	4	0	4	4
53	P	60	2	2	0	0	0
54	P	23	5	5	5	5	5
55	P	75	2	0	0	2	0
56	P	45	2	0	2	0	0
57	L	32	1	0	0	0	0
58	L	23	3	3	0	3	0
59	L	42	2	0	0	2	0
60	P	27	3	0	3	3	0
61	L	56	5	5	5	5	5
62	P	78	4	4	0	4	4
63	P	65	2	0	0	2	0
64	L	32	3	0	3	3	0
65	L	55	3	3	3	0	0
66	L	12	2	2	0	0	0
67	P	5	4	4	0	4	4
68	L	66	2	2	0	0	0
69	P	22	4	4	0	4	4
70	P	52	4	4	4	4	0
71	P	70	2	0	0	2	0
72	P	34	2	2	0	0	0
73	L	8	4	4	0	4	4
74	P	33	2	0	0	2	0
75	P	50	1	0	0	0	0
76	P	15	3	3	0	3	0
77	L	43	2	0	0	2	0
78	L	45	4	4	0	4	4
79	L	75	2	0	0	2	0
80	L	37	3	0	0	3	3
81	L	20	2	0	0	2	0
82	L	28	4	4	0	4	4
83	P	54	2	0	0	2	0
84	L	34	3	0	0	3	3
85	P	37	1	0	0	0	0

Informasi Pasien			Gejala				
No	Jenis Kelamin	Usia	Batuk	Pilek	Demam	Sesak	Sakit Kepala
86	L	12	4	4	0	4	4
87	P	65	3	3	3	0	0
88	L	32	3	0	3	3	0
89	L	17	5	5	5	5	5
90	L	52	2	0	0	2	0
91	L	19	3	0	3	3	0
92	P	36	3	0	0	3	3
93	L	43	2	0	0	2	0
94	L	9	4	4	0	4	4
95	P	54	2	0	0	2	0
96	P	27	2	0	0	2	0
97	P	24	4	4	0	4	4
98	L	60	3	0	0	3	3
99	P	40	1	0	0	0	0
100	P	20	4	4	0	4	4
101	L	67	3	3	3	0	0
102	L	50	1	0	0	0	0
103	L	12	2	0	0	2	0
104	L	29	3	3	3	0	0
105	L	25	5	5	5	5	5
106	L	76	3	0	3	3	0
107	L	38	5	5	5	5	5
108	L	60	2	0	0	2	0
109	P	34	3	3	3	0	0
110	P	8	4	4	0	4	4
111	P	54	5	5	5	5	5
112	L	80	2	0	0	2	0
113	P	27	1	0	0	0	0
114	P	15	4	4	4	4	0
115	L	55	3	0	3	3	0
116	P	27	5	5	5	5	5
117	L	69	2	0	0	2	0
118	P	25	4	4	0	4	4
119	L	53	4	4	4	4	0
120	L	47	5	5	5	5	5
121	L	70	3	0	3	3	0
122	L	22	4	4	4	4	0
123	L	30	2	0	2	0	0
124	P	17	3	3	0	3	0
125	L	45	5	5	5	5	5
126	P	60	2	2	0	0	0
127	P	32	5	5	5	5	5
128	P	70	4	4	0	4	4
129	P	68	2	0	0	2	0
130	L	38	5	5	5	5	5
131	P	80	4	4	0	4	4
132	P	67	2	0	0	2	0
133	P	46	5	5	5	5	5
134	L	87	3	3	0	3	0
135	L	56	4	4	0	4	4
136	L	74	3	0	0	3	3
137	L	51	1	0	0	0	0
138	L	25	5	5	5	5	5
139	L	81	2	2	0	0	0
140	P	17	3	3	0	3	0
141	P	41	4	4	0	4	4
142	P	57	4	4	4	4	0

Tabel 1 merupakan Sumber data penelitian yang diperoleh dari Puskesmas Bah Biak Kota Pematangsiantar yang terdiri dari 142 pasien dengan melakukan observasi untuk mendapatkan validasi data yang jelas. Data ini nantinya akan di kluster menjadi 2 bagian yaitu kluster tinggi dan kluster rendah.

2.2. Flowchart Penelitian

Berikut ini akan disajikan Flowchart penelitian menggunakan metode K-Means.



Gambar 1. Diagram alir dari metode Klustering K-Means [36]

Tahapan demi tahapan metode *K-Means* sebagai berikut [37]–[41]:

1. Atur jumlah *cluster* (*k*) pada data set.
2. Atur nilai pusat (*Centroid*).

Pada tahap awal Pengaturan nilai *Centroid* dilakukan secara acak. Pada tahap iterasi digunakan rumus persamaan (1) seperti berikut.

$$V_{ij} = \frac{1}{N_i} \sum_{k=0}^{N_i} X_{kj} \quad \dots\dots\dots(1)$$

Keterangan :

V_{ij} = *Centroid* rata-rata *cluster* ke-I variabel ke-j
 N_i = Jumlah anggota *cluster* ke-i
 i, k = Indeks *cluster*
 j = Indeks variabel
 X_{kj} = nilai ke-k variabel ke-j untuk *cluster*

3. Berdasarkan masing-masing *record*, jarak terdekat dihitung dengan *Centroid*.

Beberapa cara yang biasa digunakan untuk melakukan pengukuran jarak data ke pusat *cluster*, diantaranya *Euclidean*, *Manhattan/City Block*, dan *Minkowsky*. Setiap cara memiliki kelebihan dan

kekurangan masing-masing. Untuk penulisan pada bab ini, jarak *Centroid* yang digunakan adalah *Euclidean Distance*, dengan rumus seperti dibawah ini:

$$De = \sqrt{(x_i - s_i)^2 + (y_i - t_i)^2} \dots\dots\dots(2)$$

Keterangan :

De = *Distance Euclidean*
 i = Jumlah objek ²
 (x, y) = Koordinat objek
 (s, t) = Koordinat *Centroid*

4. Kelompokkan objek berdasarkan jarak ke *Centroid* terdekat
5. Ulangi langkah ke-3 hingga langkah ke-4, lakukan iterasi hingga *Centroid* bernilai optimal

3. Hasil dan Pembahasan

3.1. Penentuan Jumlah Kluster

Menentukan jumlah Kluster dilakukan dengan menggunakan data pengelompokan pasien penyakit ISPA dalam sebanyak 142 Pasien. Berikut ini adalah cara mencari nilai rata-rata dengan menjumlahkan data pasien dengan sesuai umur.

3.2. Penentuan Nilai K

Menetapkan nilai *k* jumlah kluster pasien sebanyak 2 Kluster (*k*-2). Kluster yang dibentuk yaitu Kluster tinggi dan rendah.

3.3. Penentuan Nilai Centroid (Pusat Kluster)

Penentuan pusat Kluster awal ditentukan secara random yang diambil dari data yang ada dalam range. Adapun nilai untuk Kluster tinggi (C1) diambil dari nilai tertinggi yang terdapat pada tabel 1, untuk Kluster rendah (C2) diambil dari nilai rata-rata setiap variabel yang juga terdapat pada tabel 1.

Tabel 2. Centroid Data Awal Iterasi 1

Kluster	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆
C1	87	5	5	5	5	5
C2	3	0	0	0	0	0

Keterangan : X₁ = Usia, X₂ = Batuk, X₃ = Pilek, X₄ = Demam, X₅ = Sesak, X₆ = Sakit Kepala

3.4. Menghitung Jarak Setiap Data Terhadap Centroid

Setelah data nilai pusat Kluster awal ditentukan, maka langkah selanjutnya adalah menghitung jarak masing-masing data terhadap pusat Kluster dengan *Euclidean Distance*. Untuk menghitung jarak setiap data Pasien Penyakit Infeksi Saluran Pernapasan Akut terhadap pusat Kluster dapat menggunakan rumus yang perhitungannya dapat kita lihat sebagai berikut :

Dilakukan perhitungan jarak terhadap data Pasien Penyakit ISPA dengan titik pusat (*centroid*) pada Kluster pertama.

$$D_{(1,1)} = \sqrt{(11 - 87)^2 + (3 - 5)^2 + (3 - 5)^2 + (3 - 5)^2 + (0 - 5)^2 + (0 - 5)^2}$$

= 76.406806

Begitu seterusnya sampai dengan D_(1.142)

Perhitungan jarak data Pasien Penyakit ISPA serta data Pasien dengan centroid Kluster kedua.

$$D_{(2.1)} = \sqrt{(11-3)^2 + (3-0)^2 + (3-0)^2 + (3-0)^2 + (0-0)^2 + (0-0)^2} = 9.539392014$$

Begitu seterusnya sampai dengan D_(2.142)

Hasil perhitungan jarak data dengan titik pusat Kluster pada iterasi 1 menggunakan *Euclidean Distance* dapat dilihat pada tabel 3.

Tabel 3. Jarak Centroid Iterasi 1

No	C1	C2	Jarak Terpendek
1	76.406806	9.539392014	9.539392014
2	39.7869325	45.29900661	39.78693253
3	69.2098259	17	17
4	33.436507	51.623638	33.43650699
5	28.6705424	57.07013229	28.67054237
...	...	...	...
138	62	24.67792536	24.67792536
139	11.3578167	78.0512652	11.35781669
140	70.4414651	14.93318452	14.93318452
141	46.3141447	38.83297568	38.83297568
142	30.4795013	54.58937626	30.47950131

3.5. Menentukan Posisi Kluster atau Pengelompokkan

Dalam menentukan posisi Kluster masing-masing data pasien serta data berdasarkan jarak minimum data terhadap pusat Kluster. Data yang memiliki jarak terkecil dengan centroid akan menjadi anggota pada kelompok tersebut. Berikut ini merupakan posisi data dengan tiap Kluster pada iterasi 1 dengan bantuan angka 1 sebagai tanda yang menyatakan data menjadi anggota pada Kluster.

Tabel 4. Pengelompokkan Posisi Kluster 1

No	C1	C2
1		1
2	1	
3		1
4	1	
5	1	
...	...	...
138		1
139	1	
140		1
141		1
142	1	

Tabel 5. Hasil Dari Kluster Iterasi 1

C1	63
C2	79

Selanjutnya dalam metode K-Means, perhitungan berhenti apabila Kluster pada iterasi yang dihasilkan sama pada iterasi sebelumnya. Maka selanjutnya mencari Kluster pada iterasi berikutnya sampai nilai iterasinya sama.

3.6. Menghitung Centroid Baru Menggunakan Hasil Dari Setiap Anggota Pada Masing-Masing Kluster

Untuk mencari nilai centroid selanjutnya dengan menggunakan centroid baru pada iterasi 1 dengan menjumlahkan nilai sesuai yang tertera pada Kluster yang terdapat pada tabel 4, adapun centroid baru untuk mencari Kluster selanjutnya adalah dengan menjumlahkan nilai yang terpilih pada Kluster tersebut kemudian membagikannya sebanyak jumlah nilai.

Contoh perhitungan titik pusat baru pada Kluster X₁, X₂, X₃, X₄, X₅ dan X₆ adalah sebagai berikut.

$$C_1(X_1) = \frac{48+54+60+64+75+50+52+51+54+63+54+75+54+52+62+67+77+65+64+51+80+57+60+75+56+78+65+55+66+52+70+50+45+75+54+65+52+54+60+67+50+76+60+54+80+55+69+53+47+70+45+60+70+68+80+67+46+87+56+74+51+81+57}{63} = 61.65079365$$

Begitu seterusnya hingga C₂(X₆).

Tabel 6. Centroid Data Baru Iterasi 1

Kluster	Nilai					
	X1	X2	X3	X4	X5	X6
C1	61.650794	2.984127	2.253968	1.666667	2.142857	1.380952
C2	24.164557	3.012658	2.379747	1.481013	2.240506	1.772152

3.7. Dilakukan kembali langkah ke 4 sampai 6

Jika nilai centroid hasil iterasi dengan nilai centroid sebelumnya bernilai sama atau nilai centroid sudah optimal serta posisi cluster data pasien serta tidak mengalami perubahan lagi maka proses iterasi berhenti. Namun jika nilai centroid tidak sama atau belum optimal serta posisi data masih berubah maka proses iterasi berlanjut pada iterasi berikutnya.

Tabel 7. Centroid Data Iterasi 2

Cluster	Nilai					
	X1	X2	X3	X4	X5	X6
C1	61.65079	2.98413	2.25397	1.66667	2.14286	1.38095
C2	24.16456	3.01266	2.37975	1.48101	2.24050	1.77215

Setelah data nilai pusat Kluster iterasi 2 didapat, maka langkah selanjutnya adalah melakukan perhitungan jarak terhadap data pasien serta dengan titik pusat (centroid) pada Kluster pertama.

$$D_{(1.1)} = \sqrt{((11 - 61.650794)^2 + (3 - 2.984126984)^2 + (3 - 2.253968)^2 + (3 - 1.666667)^2 + (0 - 2.142857)^2 + (0 - 1.380952)^2)} = 50.7379183$$

Begitu seterusnya hingga D_(1.142)

Perhitungan jarak data Indeks Pembangunan Manusia serta data pasien dengan centroid Kluster kedua, seperti berikut.

$$D_{(2.1)} = \sqrt{((11 - 24.16455696)^2 + (3 - 3.012658228)^2 + (3 - 2.379747)^2 + (3 - 1.481013)^2 + (0 - 2.240506)^2 + (0 - 1.772152)^2)} = 13.57048812$$

Begitu seterusnya hingga D_(2.142)

Hasil perhitungan jarak data dengan titik pusat Kluster pada iterasi 3 menggunakan Eulidean Distance dapat dilihat pada tabel berikut.

Tabel 8. Jarak Centroid Iterasi 2

No	C1	C2	Jarak Terpendek
1	50.7379183	13.57048812	13.57048812
2	14.156826	24.13765735	14.15682598
3	43.8469905	7.200996477	7.200996477
4	8.91689183	30.16756865	8.916891832
5	3.92224117	36.04576797	3.922241172
...	...	...	...
142	6.04242997	33.08093138	6.04242997

Data yang memiliki jarak terkecil dengan centroid akan menjadi anggota pada kelompok tersebut. Posisi data dengan tiap Kluster pada iterasi 2 dengan menggunakan bantuan angka 1 sebagai tanda yang mengatakan data menjadi anggota pada Kluster dapat dilihat pada tabel berikut.

Tabel 9. Jarak Centroid Iterasi 2

No	C1	C2
1		1
2	1	
3		1
4	1	
5	1	
...	...	...
142	1	

Tabel 10. Hasil Dari Kluster Iterasi 2

C1	68
C2	74

Untuk mencari nilai centroid selanjutnya dengan menggunakan centroid baru pada iterasi 2 dengan menjumlahkan nilai sesuai yang tertera pada Kluster yang terdapat pada tabel 9. Adapun centroid baru untuk mencari Kluster selanjutnya adalah dengan menjumlahkan nilai yang terpilih pada Kluster tersebut kemudian membagikannya sebanyak jumlah nilai. Contoh perhitungan titik pusat baru adalah sebagai berikut.

$$C_1(X_1) = \frac{48+54+60+64+75+50+52+51+54+45+43+63+54+75+54+52+62+67+77+65+64+51+80+57+60+75+45+56+78+65+55+66+52+70+50+43+45+75+54+65+52+43+54+60+67+50+76+60+54+80+55+69+53+47+70+45+60+70+68+80+67+46+87+56+74+51+81+57}{68} = 60.33823529$$

Begitu seterusnya hingga $C_2(X_6)$

Tabel 11. Centroid Data Baru Iterasi 2

Cluster	Nilai					
	X1	X2	X3	X4	X5	X6
C1	60.338235	2.911765	2.117647	1.602941	2.044118	1.279412
C2	22.837838	3.081081	2.513514	1.527027	2.337838	1.891892

Setelah data nilai pusat Kluster iterasi 2 didapat, maka langkah selanjutnya adalah melakukan perhitungan jarak terhadap data pasien serta pada Kluster pertama.

$$D_{(1,1)} = \sqrt{((11 - 60.338235)^2 + (3 - 2.911765)^2 + (3 - 2.117647)^2 + (3 - 1.602941)^2 + (0 - 2.044118)^2 + (0 - 1.279412)^2)} = 49.4248407$$

Begitu seterusnya hingga $D_{(2,142)}$

Hasil perhitungan jarak data dengan titik pusat Kluster pada iterasi 3 menggunakan Eulidean Distance dapat dilihat pada tabel berikut.

Tabel 12. Jarak Centroid Iterasi 3

No	C1	C2	Jarak Terpendek
1	49.4248407	12.31227184	12.31227184
2	12.8769126	25.46216345	12.87691263
3	42.5564352	5.999786945	5.999786945
4	7.88838663	31.46680164	7.888386628
5	3.37684974	37.38691251	3.376849743
...	...	...	...
142	5.20380484	34.38844299	5.203804836

Data yang memiliki jarak terkecil dengan centroid akan menjadi anggota pada kelompok tersebut. Posisi data dengan tiap Kluster pada iterasi 2 dengan menggunakan bantuan angka 1 sebagai tanda yang mengatakan data menjadi anggota pada Kluster dapat dilihat pada tabel 13 berikut.

Tabel 13. Pengelompokan Posisi Kluster Iterasi 3

No	C1	C2
1		1
2	1	
3		1
4	1	
5	1	
...	...	...
142	1	

Tabel 14. Hasil Dari Kluster Iterasi 3

C1	71
C2	71

Untuk mencari nilai centroid selanjutnya dengan menggunakan centroid baru pada iterasi 3 dengan menjumlahkan nilai sesuai yang tertera pada Kluster yang terdapat pada tabel 13. Adapun centroid baru untuk mencari Kluster selanjutnya adalah dengan menjumlahkan nilai yang terpilih pada Kluster tersebut kemudian membagikannya sebanyak jumlah nilai. Contoh perhitungan titik pusat baru adalah sebagai berikut.

$$C_1(X_1) = \frac{48+54+60+64+75+50+42+52+51+54+45+43+63+54+75+54+42+52+62+67+77+65+64+51+80+57+60+75+45+42+56+78+65+55+66+52+70+50+43+45+75+54+65+52+43+54+60+67+50+76+60+54+80+55+69+53+47+70+45+60+70+68+80+67+46+87+56+74+51+81+57}{71} = 59.56338$$

Begitu seterusnya hingga $C_2(X_6)$. Maka diperoleh centroid data baru iterasi 3 sebagai berikut.

Tabel 15. Centroid Data Iterasi 3

Cluster	Nilai					
	X1	X2	X3	X4	X5	X6
C1	59.56338	2.887324	2.098592	1.577465	1.985915	1.225352
C2	22.02817	3.112676	2.549296	1.549296	2.408451	1.971831

Setelah data nilai pusat Kluster iterasi 3 didapat, maka langkah selanjutnya adalah melakukan perhitungan jarak terhadap data pasien pada Kluster pertama.

$$D_{(1,1)} = \sqrt{((11 - 59.56338)^2 + (3 - 2.887324)^2 + (3 - 2.098592)^2 + (3 - 1.577465)^2 + (0 - 1.985915)^2 + (0 - 1.225352)^2)}$$

Begitu seterusnya hingga D_(2,142)

Tabel 16. Jarak Centroid Iterasi 4

No	C1	C2	Jarak Terpendek
1	48.6487008	11.55982859	11.55982859
2	12.1344711	26.26803987	12.13447106
3	41.7924824	5.307056458	5.307056458
4	7.30176762	32.26468532	7.301767619
5	3.31752185	38.2032563	3.317521846
...	...	...	...
142	4.78027422	35.19016833	4.78027422

Tabel 17. Pengelompokan Posisi Kluster Iterasi 4

No	C1	C2
1		1
2	1	
3		1
4	1	
5	1	
...	...	...
142	1	

Tabel 18. Hasil Dari Kluster Iterasi 4

C1	72
C2	70

Untuk mencari nilai centroid selanjutnya dengan menggunakan centroid baru pada iterasi 4 dengan menjumlahkan nilai sesuai yang tertera pada Kluster yang terdapat pada tabel 17. Adapun centroid baru untuk mencari Kluster selanjutnya adalah dengan menjumlahkan nilai yang terpilih pada Kluster tersebut kemudian membagikannya sebanyak jumlah nilai. Contoh perhitungan titik pusat baru adalah sebagai berikut.

$$C_1(X_1) = \frac{48+54+60+64+75+50+42+52+51+54+45+43+63+54+75+54+42+52+62+67+77+65+64+51+80+57+60+75+45+42+56+78+65+55+66+52+70+50+43+45+75+54+65+52+43+54+60+67+50+76+60+54+80+55+69+53+47+70+45+60+70+68+80+67+46+87+56+74+51+81+41+57}{72} = 59.30556$$

Begitu seterusnya hingga C₂(X₆)

Tabel 19. Centroid Data Iterasi 4

Cluster	Nilai					
	X1	X2	X3	X4	X5	X6
C1	59.30556	2.902778	2.125	1.555556	2.013889	1.263889
C2	21.75714	3.1	2.528571	1.571429	2.385714	1.942857

Setelah data nilai pusat Kluster iterasi 4 didapat, maka langkah selanjutnya adalah melakukan perhitungan jarak terhadap data pasien serta data pada Kluster pertama.

$$D_{(1,1)} = \sqrt{((11 - 59.30556)^2 + (3 - 2.902778)^2 + (3 - 2.125)^2 + (3 - 1.555556)^2 + (0 - 2.013889)^2 + (0 - 1.263889)^2)} = 48.6487008$$

Begitu seterusnya hingga D_(2,142)

Hasil perhitungan jarak data dengan titik pusat Kluster pada iterasi 5 menggunakan Euclidean Distance dapat dilihat pada tabel 20 berikut.

Tabel 20. Jarak Centroid Iterasi 5

No	C1	C2	Jarak Terpendek
1	48.3936086	11.28961958	11.28961958
2	11.8952475	26.53189071	11.89524752
3	41.529738	5.137934152	5.137934152
4	7.09829574	32.53303502	7.098295744
5	3.41519842	38.46777607	3.41519842
...	...	...	...
142	4.64246369	35.45860478	4.642463692

Data yang memiliki jarak terkecil dengan centroid akan menjadi anggota pada kelompok tersebut. Posisi data dengan tiap Kluster pada iterasi 2 dengan menggunakan bantuan angka 1 sebagai tanda yang mengatakan data menjadi anggota pada Kluster dapat dilihat pada Tabel 21.

Tabel 21. Pengelompokan Posisi Kluster Iterasi 5

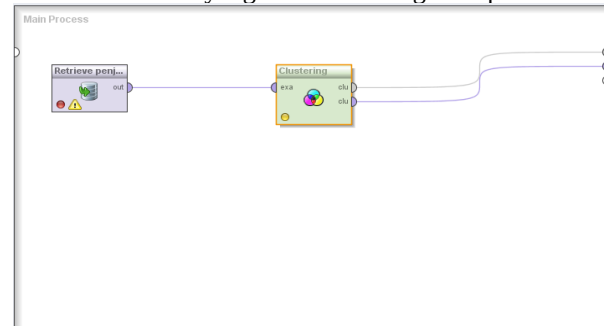
No	C1	C2
1		1
2	1	
3		1
4	1	
5	1	
...	...	...
142	1	

Tabel 22. Hasil Dari Kluster Iterasi 5

C1	72
C2	70

3.4. Implementasi dengan Rapidminer

Berikut adalah proses pengelompokan dan hasil dari metode K-Means yang dilakukan dengan Rapidminer..



Gambar 2. Proses K-Means dengan Rapidminer (Nilai K = 2)

Gambar 2 menjelaskan proses pengelompokan atau pengklusteran metode K-means dengan menggunakan Rapidminer yang diawali dengan mengimport data excel Pasien yang memiliki riwayat penyakit ISPA, kemudian dilanjutkan dengan pemilihan operator metode K-means untuk pengklusterannya. Nilai k = 2, measure types yang digunakan adalah MixedMeasures. Setelah itu dihubungkan ke operator Apply Model untuk menerapkan model yang sudah dipelajari atau dilatih. Tujuannya adalah untuk mendapatkan prediksi pada unlabeled data (data testing) yang belum memiliki label. Tahap berikutnya adalah menghubungkan ke operator Performance untuk mengevaluasi kinerja model yang memberikan daftar

nilai kriteria kinerja secara otomatis sesuai dengan tugas yang diberikan. Hasil nya dapat dilihat pada gambar 3, gambar 4 dan gambar 5.

Cluster Model

```
Cluster 0: 70 items
Cluster 1: 72 items
Total number of items: 142
```

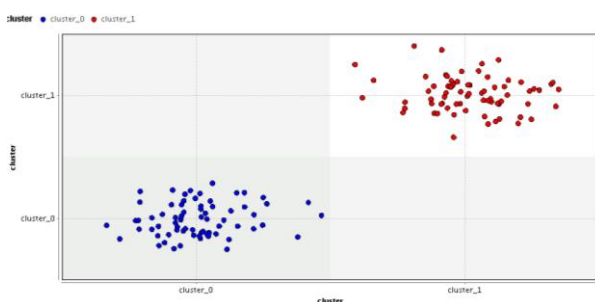
Gambar 3. Hasil Kluster

Berdasarkan gambar 3 dapat dijelaskan bahwa dari dua Kluster yang dihasilkan terdapat 70 items untuk Kluster_0 (C1), dan 72 items untuk Kluster_1 (C2). Untuk hasil akhir Centroid table dapat dilihat pada gambar 4. Sedangkan hasil *plot view* Kluster pasien penyakit ISPA disajikan pada gambar 5.

Gambar 5 berikut menjelaskan bahwa titik warna biru merupakan kelompok Kluster_0 (rendah) yang terdiri dari 70 items dan titik warna merah merupakan kelompok Kluster_1 (tinggi) yang terdiri dari 72 item.

Row No.	JK	cluster	X1	X2	X3	X4	X5	X6
1	L	cluster_0	11	2	2	2	0	0
2	L	cluster_1	48	3	0	3	0	3
3	P	cluster_0	18	4	4	0	4	4
4	L	cluster_1	54	4	4	4	0	4
5	P	cluster_1	60	2	0	2	0	0
6	P	cluster_0	9	0	3	3	0	3
7	P	cluster_0	12	3	3	3	0	0
8	P	cluster_0	30	2	0	0	2	0
9	L	cluster_0	14	0	0	2	0	2
10	P	cluster_0	5	4	4	0	4	4
11	P	cluster_1	64	5	5	5	5	5
12	P	cluster_1	75	4	4	0	4	4
13	L	cluster_1	50	2	0	0	2	0
14	L	cluster_1	42	3	3	3	0	0
15	L	cluster_0	15	3	0	0	3	3
16	L	cluster_1	52	2	2	0	0	0
17	L	cluster_0	13	3	3	3	0	0
18	L	cluster_0	35	2	0	2	0	0
19	P	cluster_0	15	3	3	3	0	0

Gambar 4. Tabel Centroid



Gambar 5. Plot View Pasien Penyakit ISPA

4. Kesimpulan

Penelitian ini mampu mengelompokkan 142 pasien penyakit ISPA ke dalam 2 kluster, kluster 1 memberikan rekomendasi tinggi berjumlah 72 Pasien, dan kluster 2 memberikan rekomendasi rendah berjumlah 70 Pasien. Proses kluster berhenti pada data iterasi ke 5. Berdasarkan proses perhitungan manual dengan menggunakan Ms. Excel dan pengujian dengan menggunakan Rapidminer 5.3, menghasilkan nilai

yang sama. Dapat disimpulkan bahwa untuk kasus ini, algoritma K-Means dapat mengelompokkan pasien penyakit ISPA yang ada di puskesmas Bah Biak dengan baik.

Referensi

- [1] L. Suriani, "Pengelompokan Data Kriminal Pada Poldasu Menentukan Pola Daerah Rawan Tindak Kriminal Menggunakan Data Mining Algoritma K-Means Clustering," *Jurnal Sistem Komputer dan Informatika (JSON)*, vol. 1, no. 2, p. 151, 2020.
- [2] I. N. M. Adiputra, "Clustering Penyakit Dbd Pada Rumah Sakit Dharma Kerti Menggunakan Algoritma K-Means," *INSERT: Information System and Emerging Technology Journal*, vol. 2, no. 2, pp. 99–105, 2022.
- [3] E. Rouza, Basorudin, and Efrida, "Identifikasi dan Klasifikasi UMKM di Kabupaten Rokan Hulu Menggunakan Metode K-Means," *Riau Journal of Computer Science*, vol. 7, no. 1, pp. 32–40, 2021.
- [4] P. P. A. N. . F. I. R.H Zer, Masitha, A. P. Windarto, and A. Wanto, "Analysis of the ELECTRE Method on the Selection of Student Creativity Program Proposals," *Journal of Physics: Conference Series*, vol. 1255, no. 1, pp. 1–7, 2019.
- [5] A. P. W. Budiharjo and A. Muhammad, "Comparison of Weighted Sum Model and Multi Attribute Decision Making Weighted Product Methods in Selecting the Best Elementary School in Indonesia," *International Journal of Software Engineering and Its Applications*, vol. 11, no. 4, pp. 69–90, 2017.
- [6] D. R. Sari, N. Rofiqo, D. Hartama, A. P. Windarto, and A. Wanto, "Analysis of the Factors Causing Lazy Students to Study Using the ELECTRE II Algorithm," *Journal of Physics: Conference Series*, vol. 1255, no. 1, pp. 1–6, 2019.
- [7] K. Fatmawati et al., "Analysis of Promotee II Method in the Selection of the Best Formula for Infants Under Three Years," *Journal of Physics: Conference Series*, vol. 1255, no. 1, pp. 1–6, Aug. 2019.
- [8] P. Alkhairi, L. P. Purba, A. Eryzha, A. P. Windarto, and A. Wanto, "The Analysis of the ELECTREE II Algorithm in Determining the Doubts of the Community Doing Business Online," *Journal of Physics: Conference Series*, vol. 1255, no. 1, pp. 1–7, 2019.
- [9] S. Sundari, Karmila, M. N. Fadli, D. Hartama, A. P. Windarto, and A. Wanto, "Decision Support System on Selection of Lecturer Research Grant Proposals using Preferences Selection Index," *Journal of Physics: Conference Series*, vol. 1255, no. 1, pp. 1–7, 2019.
- [10] S. R. Ningsih, R. Wulansari, D. Hartama, A. P. Windarto, and A. Wanto, "Analysis of PROMETHEE II Method on Selection of Lecturer Community Service Grant Proposals," *Journal of Physics: Conference Series*, vol. 1255, no. 1, pp. 1–7, 2019.
- [11] T. Imandasari, M. G. Sadewo, A. P. Windarto, A. Wanto, H. O. Lingga Wijaya, and R. Kurniawan, "Analysis of the Selection Factor of Online Transportation in the VIKOR Method in Pematangsiantar City," *Journal of Physics: Conference Series*, vol. 1255, no. 1, pp. 1–7, 2019.
- [12] E. Siregar, H. Mawengkang, E. B. Nababan, and A. Wanto, "Analysis of Backpropagation Method with Sigmoid Bipolar and Linear Function in Prediction of Population Growth," *Journal of Physics: Conference Series*, vol. 1255, no. 1, pp. 1–6, 2019.
- [13] G. W. Bhawika et al., "Implementation of ANN for Predicting the Percentage of Illiteracy in Indonesia by Age Group," *Journal of Physics: Conference Series*, vol. 1255, no. 1, pp. 1–6, 2019.
- [14] W. Saputra, J. T. Hardinata, and A. Wanto, "Resilient method in determining the best architectural model for predicting open unemployment in Indonesia," *IOP Conference Series*:

- Materials Science and Engineering*, vol. 725, no. 1, pp. 1–7, 2020.
- [15] A. Wanto and J. T. Hardinata, “Estimations of Indonesian poor people as poverty reduction efforts facing industrial revolution 4.0,” *IOP Conference Series: Materials Science and Engineering*, vol. 725, no. 1, pp. 1–8, 2020.
- [16] A. Wanto *et al.*, “Model of Artificial Neural Networks in Predictions of Corn Productivity in an Effort to Overcome Imports in Indonesia,” *Journal of Physics: Conference Series*, vol. 1339, no. 1, pp. 1–6, 2019.
- [17] A. Wanto *et al.*, “Analysis of the Accuracy Batch Training Method in Viewing Indonesian Fisheries Cultivation Company Development,” *Journal of Physics: Conference Series*, vol. 1255, no. 1, pp. 1–6, 2019.
- [18] I. S. Purba *et al.*, “Accuracy Level of Backpropagation Algorithm to Predict Livestock Population of Simalungun Regency in Indonesia,” *Journal of Physics: Conference Series*, vol. 1255, no. 1, pp. 1–6, 2019.
- [19] P. Parulian *et al.*, “Analysis of Sequential Order Incremental Methods in Predicting the Number of Victims Affected by Disasters,” *Journal of Physics: Conference Series*, vol. 1255, no. 1, pp. 1–6, 2019.
- [20] A. Wanto *et al.*, “Analysis of the Backpropagation Algorithm in Viewing Import Value Development Levels Based on Main Country of Origin,” *Journal of Physics: Conference Series*, vol. 1255, no. 1, pp. 1–6, 2019.
- [21] S. Setti, A. Wanto, M. Syafiq, A. Andriano, and B. K. Sihotang, “Analysis of Backpropagation Algorithms in Predicting World Internet Users,” *Journal of Physics: Conference Series*, vol. 1255, no. 1, pp. 1–6, 2019.
- [22] T. Afriliansyah *et al.*, “Implementation of Bayesian Regulation Algorithm for Estimation of Production Index Level Micro and Small Industry,” *Journal of Physics: Conference Series*, vol. 1255, no. 1, pp. 1–6, 2019.
- [23] A. Wanto *et al.*, “Forecasting the Export and Import Volume of Crude Oil, Oil Products and Gas Using ANN,” *Journal of Physics: Conference Series*, vol. 1255, no. 1, pp. 1–6, 2019.
- [24] M. K. Z. Sormin, P. Sihombing, A. Amalia, A. Wanto, D. Hartama, and D. M. Chan, “Predictions of World Population Life Expectancy Using Cyclical Order Weight / Bias,” *Journal of Physics: Conference Series*, vol. 1255, no. 1, pp. 1–6, 2019.
- [25] R. Rahim *et al.*, “C4.5 Classification Data Mining for Inventory Control,” *International Journal of Engineering & Technology*, vol. 7, pp. 68–72, 2018.
- [26] I. Parlina *et al.*, “Naive Bayes Algorithm Analysis to Determine the Percentage Level of visitors the Most Dominant Zoo Visit by Age Category,” *Journal of Physics: Conference Series*, vol. 1255, no. 1, pp. 1–5, 2019.
- [27] I. S. Damanik, A. P. Windarto, A. Wanto, Poningsih, S. R. Andani, and W. Saputra, “Decision Tree Optimization in C4.5 Algorithm Using Genetic Algorithm,” *Journal of Physics: Conference Series*, vol. 1255, no. 1, pp. 1–7, 2019.
- [28] H. Siahaan, H. Mawengkang, S. Efendi, A. Wanto, and A. Perdana Windarto, “Application of Classification Method C4.5 on Selection of Exemplary Teachers,” in *Journal of Physics: Conference Series*, 2019, vol. 1235, no. 1, pp. 1–7.
- [29] D. Hartama, A. Perdana Windarto, and A. Wanto, “The Application of Data Mining in Determining Patterns of Interest of High School Graduates,” *Journal of Physics: Conference Series*, vol. 1339, no. 1, pp. 1–6, 2019.
- [30] M. Widyastuti, A. G. Fepdiani Simanjuntak, D. Hartama, A. P. Windarto, and A. Wanto, “Classification Model C.45 on Determining the Quality of Customer Service in Bank BTN Pematangsiantar Branch,” *Journal of Physics: Conference Series*, vol. 1255, no. 012002, pp. 1–6, 2019.
- [31] W. Katrina, H. J. Damanik, F. Parhusip, D. Hartama, A. P. Windarto, and A. Wanto, “C.45 Classification Rules Model for Determining Students Level of Understanding of the Subject,” *Journal of Physics: Conference Series*, vol. 1255, no. 1, pp. 1–7, 2019.
- [32] S. Sudirman, A. P. Windarto, and A. Wanto, “Data Mining Tools | RapidMiner : K-Means Method on Clustering of Rice Crops by Province as Efforts to Stabilize Food Crops In Indonesia,” *IOP Conference Series: Materials Science and Engineering*, vol. 420, no. 012089, pp. 1–8, 2018.
- [33] N. A. Febriyati, A. D. GS, and A. Wanto, “GRDP Growth Rate Clustering in Surabaya City uses the K- Means Algorithm,” *International Journal of Information System & Technology*, vol. 3, no. 2, pp. 276–283, 2020.
- [34] J. Hutagalung, N. L. W. S. R. Ginantra, G. W. Bhawika, W. G. S. Parwita, A. Wanto, and P. D. Panjaitan, “COVID-19 Cases and Deaths in Southeast Asia Clustering using K-Means Algorithm,” *Journal of Physics: Conference Series*, vol. 1783, no. 1, p. 012027, 2021.
- [35] M. A. Hanafiah, A. Wanto, and P. B. Indonesia, “Implementation of Data Mining Algorithms for Grouping Poverty Lines by District/City in North Sumatra,” *International Journal of Information System & Technology*, vol. 3, no. 2, pp. 315–322, 2020.
- [36] Z. S. Younus *et al.*, “Content-based image retrieval using PSO and k-means clustering algorithm,” *Arabian Journal of Geosciences*, vol. 8, no. 8, pp. 6211–6224, 2015.
- [37] E. Prasetyo, *Data Mining: Konsep dan Aplikasi menggunakan Matlab*. Yogyakarta: Andi Offset, 2012.
- [38] D. T. Larose, *Discovering Knowledge in Data: An Introduction to Data Mining: Second Edition*. New Jersey: John Wiley & Sons, 2005.
- [39] R. Primartha, *Belajar Machine Learning Teori dan Praktik*. Bandung: Informatika Bandung, 2018.
- [40] T. Khotimah, “Pengelompokan Surat dalam Al Qur’an menggunakan Algoritma K-Means,” *Jurnal Simetris*, vol. 5, no. 1, pp. 83–88, 2014.
- [41] I. Parlina, A. P. Windarto, A. Wanto, and M. R. Lubis, “Memanfaatkan Algoritma K-Means dalam Menentukan Pegawai yang Layak Mengikuti Assessment Center untuk Clustering Program SDP,” *CESS (Journal of Computer Engineering System and Science)*, vol. 3, no. 1, pp. 87–93, 2018.