

EVALUASI K-MEANS DAN FUZZY C-MEANS CLUSTERING ANALISIS DATA HIV DI TINGKAT PROVINSI

Gabriel Arnold Budianto

Teknik Informatika, Universitas Tarumanagara, Jakarta

E-mail: Gabriel.535210061@stu.untar.ac.id

ABSTRAK

HIV (Human Immunodeficiency Virus) merupakan penyakit serius yang menyerang sistem kekebalan tubuh dengan menginfeksi darah dan sel-sel di dalam tubuh. Penyakit ini sangat berbahaya karena penyebarannya yang cepat, terutama melalui hubungan seksual berisiko, air liur, aktivitas menyusui oleh ibu yang terinfeksi, hingga penggunaan alat medis yang terkontaminasi seperti pada proses transfusi darah. Kota Surabaya di Provinsi Jawa Timur menjadi salah satu daerah dengan jumlah kasus HIV yang tinggi. Oleh karena itu, diperlukan analisis terhadap penyebaran kasus HIV di wilayah ini serta evaluasi layanan kesehatan yang diberikan berdasarkan standar nasional di berbagai puskesmas. Penelitian ini memanfaatkan data dari situs katalog.data.co.id dan menerapkan metode clustering K-Means serta Fuzzy-C-Means untuk mengevaluasi jumlah penderita HIV serta cakupan layanan kesehatan nasional yang mereka terima. Hasil dari pengelompokan ini dapat digunakan sebagai dasar pemetaan penyebaran HIV guna mencegah penyebaran lebih lanjut melalui tindakan penanganan yang tepat.

Kata kunci

Evaluasi, Clustering, HIV, K-Means, Fuzzy-C-Means

ABSTRACT

HIV, or Human Immunodeficiency Virus, is a life-threatening disease that weakens the immune system by attacking the blood and body cells. Its rapid transmission, which can occur through unprotected sex, contact with infected saliva, breastfeeding, or contaminated medical equipment such as during blood transfusions, poses a serious health threat. Surabaya, located in East Java Province, is one of the regions with a high number of HIV cases. This study aims to analyze the distribution of HIV cases in Surabaya and assess treatment data aligned with Indonesia's national health standards using clustering techniques. Data was sourced from katalog.data.co.id. The research applied K-Means and Fuzzy-C-Means clustering methods to determine the number of infected individuals and the extent to which they receive standardized healthcare services in each health center. The results of this clustering evaluation can help in mapping the spread of HIV and aid efforts to control and prevent its expansion.

Keywords

Evaluation, Clustering, HIV, K-Means, Fuzzy-C-Means

1. PENDAHULUAN

Human Immunodeficiency Virus (HIV) adalah virus yang menyerang sistem kekebalan tubuh, khususnya sel darah putih dan limfosit. Virus ini menyebabkan penurunan daya tahan tubuh, membuat penderitanya lebih rentan terhadap infeksi lain. Menariknya, beberapa individu dapat hidup dengan HIV tanpa menunjukkan gejala yang mencolok dan tidak selalu membutuhkan pengobatan langsung.

Antara tahun 2019 hingga 2023, HIV/AIDS menjadi perhatian publik karena banyak kasus terjadi pada anak-anak di bawah umur, khususnya di Surabaya, yang merupakan provinsi dengan jumlah kasus tertinggi kedua setelah DKI Jakarta dengan total 42.399 penderita. Penyebaran HIV di Indonesia banyak ditemukan di daerah perkotaan dan

dapat terjadi melalui hubungan seksual berisiko, transfusi darah, air liur dari penderita, serta dari ibu menyusui kepada bayinya. Mengingat risiko penyebarannya yang tinggi, perlu ada keterlibatan aktif dari seluruh pihak terkait, termasuk instansi kesehatan dan aparat keamanan nasional, untuk menanggulangi kasus ini secara efektif.

Pemanfaatan teknologi seperti machine learning dapat menjadi solusi dalam menganalisis data penyebaran HIV. Sebuah studi sebelumnya oleh Rut Indra Lita Sinaga dan rekan-rekannya menggunakan metode K-Means untuk mengelompokkan jumlah kasus AIDS berdasarkan provinsi dari tahun 2016 hingga 2018. (Lesmana, A.A., et al., 2019) Metode ini menghasilkan dua kelompok: cluster rendah dan tinggi. Selain itu, metode Fuzzy-C-Means dapat digunakan untuk mengevaluasi seberapa besar pengaruh data terhadap setiap cluster melalui sistem keanggotaan.

Penelitian ini bertujuan untuk mengevaluasi persebaran HIV di Surabaya melalui data dari puskesmas di berbagai kecamatan dan kelurahan, mengelompokkan tingkat risiko tinggi dan rendah, serta mengevaluasi penanganan kesehatan berdasarkan standar nasional.

2. METODE PENELITIAN

Penelitian ini menggunakan metode K-Means adalah metode klasterisasi yang mengelompokkan data ke dalam klaster secara tegas, di mana setiap data hanya masuk ke satu klaster. Metode ini cepat dan sederhana, cocok untuk data yang terpisah jelas. Sementara itu, Fuzzy C-Means memungkinkan satu data menjadi bagian dari beberapa klaster sekaligus dengan tingkat keanggotaan tertentu. FCM lebih akurat untuk data yang tumpang tindih, meski lebih kompleks dan lambat dibandingkan K-Means.

2.1 Sumber Data

Data yang digunakan dalam penelitian ini diperoleh dari Provinsi Jawa Timur dan diakses melalui situs katalog.data.co.id pada bulan September 2023. Dataset tersebut memuat informasi jumlah kasus HIV berdasarkan jenis kelamin (laki-laki dan perempuan) serta jumlah pasien yang telah menerima penanganan sesuai dengan standar kesehatan nasional Indonesia.

2.1.2 Dataset

Dataset terdiri dari 304 entri yang mencakup fitur seperti kecamatan, kelurahan, jenis kelamin, bentuk penanganan, dan jumlah penderita HIV. Format data yang digunakan ditampilkan dalam Gambar 1.

id	nama_kecamatan_kode_kecamatan_namanya_kecamatan_pis_kecamatan	jumlah_penduduk_kecamatan	jumlah_penderita_hiv_kecamatan	jumlah_penderita_hiv_kecamatan_persentase	jumlah_penderita_hiv_kecamatan_persentase_standar
1	Puskesmas Kel. Tam. 1	3	0	0	0
2	Puskesmas Kel. Tam. 2	3	0	0	0
3	Puskesmas Kel. Tam. 3	3	0	0	0
4	Puskesmas Kel. Tam. 4	3	0	0	0
5	Puskesmas Kel. Kran.	3	0	0	0
6	Puskesmas Kel. Laki	3	1	0	0
7	Puskesmas Kel. Laki 2	3	1	0	0
8	Puskesmas Kel. Laki 3	3	1	0	0
9	Puskesmas Kel. Laki 4	3	1	0	0
10	Puskesmas Kel. Laki 5	3	1	0	0
11	Puskesmas Kel. Laki 6	3	1	0	0
12	Puskesmas Kel. Laki 7	3	1	0	0
13	Puskesmas Kel. Laki 8	3	1	0	0
14	Puskesmas Kel. Laki 9	3	1	0	0
15	Puskesmas Kel. Laki 10	3	1	0	0
16	Puskesmas Kel. Laki 11	3	1	0	0
17	Puskesmas Kel. Laki 12	3	1	0	0
18	Puskesmas Kel. Laki 13	3	1	0	0
19	Puskesmas Kel. Laki 14	3	1	0	0
20	Puskesmas Kel. Laki 15	3	1	0	0
21	Puskesmas Kel. Laki 16	3	1	0	0
22	Puskesmas Kel. Laki 17	3	1	0	0
23	Puskesmas Kel. Laki 18	3	1	0	0
24	Puskesmas Kel. Laki 19	3	1	0	0
25	Puskesmas Kel. Laki 20	3	1	0	0
26	Puskesmas Kel. Laki 21	3	1	0	0
27	Puskesmas Kel. Laki 22	3	1	0	0
28	Puskesmas Kel. Laki 23	3	1	0	0
29	Puskesmas Kel. Laki 24	3	1	0	0
30	Puskesmas Kel. Laki 25	3	1	0	0
31	Puskesmas Kel. Laki 26	3	1	0	0
32	Puskesmas Kel. Laki 27	3	1	0	0
33	Puskesmas Kel. Laki 28	3	1	0	0
34	Puskesmas Kel. Laki 29	3	1	0	0
35	Puskesmas Kel. Laki 30	3	1	0	0
36	Puskesmas Kel. Laki 31	3	1	0	0
37	Puskesmas Kel. Laki 32	3	1	0	0
38	Puskesmas Kel. Laki 33	3	1	0	0
39	Puskesmas Kel. Laki 34	3	1	0	0
40	Puskesmas Kel. Laki 35	3	1	0	0
41	Puskesmas Kel. Laki 36	3	1	0	0
42	Puskesmas Kel. Laki 37	3	1	0	0
43	Puskesmas Kel. Laki 38	3	1	0	0
44	Puskesmas Kel. Laki 39	3	1	0	0
45	Puskesmas Kel. Laki 40	3	1	0	0
46	Puskesmas Kel. Laki 41	3	1	0	0
47	Puskesmas Kel. Laki 42	3	1	0	0
48	Puskesmas Kel. Laki 43	3	1	0	0
49	Puskesmas Kel. Laki 44	3	1	0	0
50	Puskesmas Kel. Laki 45	3	1	0	0
51	Puskesmas Kel. Laki 46	3	1	0	0
52	Puskesmas Kel. Laki 47	3	1	0	0
53	Puskesmas Kel. Laki 48	3	1	0	0
54	Puskesmas Kel. Laki 49	3	1	0	0
55	Puskesmas Kel. Laki 50	3	1	0	0
56	Puskesmas Kel. Laki 51	3	1	0	0
57	Puskesmas Kel. Laki 52	3	1	0	0
58	Puskesmas Kel. Laki 53	3	1	0	0
59	Puskesmas Kel. Laki 54	3	1	0	0
60	Puskesmas Kel. Laki 55	3	1	0	0
61	Puskesmas Kel. Laki 56	3	1	0	0
62	Puskesmas Kel. Laki 57	3	1	0	0
63	Puskesmas Kel. Laki 58	3	1	0	0
64	Puskesmas Kel. Laki 59	3	1	0	0
65	Puskesmas Kel. Laki 60	3	1	0	0
66	Puskesmas Kel. Laki 61	3	1	0	0
67	Puskesmas Kel. Laki 62	3	1	0	0
68	Puskesmas Kel. Laki 63	3	1	0	0
69	Puskesmas Kel. Laki 64	3	1	0	0
70	Puskesmas Kel. Laki 65	3	1	0	0
71	Puskesmas Kel. Laki 66	3	1	0	0
72	Puskesmas Kel. Laki 67	3	1	0	0
73	Puskesmas Kel. Laki 68	3	1	0	0
74	Puskesmas Kel. Laki 69	3	1	0	0
75	Puskesmas Kel. Laki 70	3	1	0	0
76	Puskesmas Kel. Laki 71	3	1	0	0
77	Puskesmas Kel. Laki 72	3	1	0	0
78	Puskesmas Kel. Laki 73	3	1	0	0
79	Puskesmas Kel. Laki 74	3	1	0	0
80	Puskesmas Kel. Laki 75	3	1	0	0
81	Puskesmas Kel. Laki 76	3	1	0	0
82	Puskesmas Kel. Laki 77	3	1	0	0
83	Puskesmas Kel. Laki 78	3	1	0	0
84	Puskesmas Kel. Laki 79	3	1	0	0
85	Puskesmas Kel. Laki 80	3	1	0	0
86	Puskesmas Kel. Laki 81	3	1	0	0
87	Puskesmas Kel. Laki 82	3	1	0	0
88	Puskesmas Kel. Laki 83	3	1	0	0
89	Puskesmas Kel. Laki 84	3	1	0	0
90	Puskesmas Kel. Laki 85	3	1	0	0
91	Puskesmas Kel. Laki 86	3	1	0	0
92	Puskesmas Kel. Laki 87	3	1	0	0
93	Puskesmas Kel. Laki 88	3	1	0	0
94	Puskesmas Kel. Laki 89	3	1	0	0
95	Puskesmas Kel. Laki 90	3	1	0	0
96	Puskesmas Kel. Laki 91	3	1	0	0
97	Puskesmas Kel. Laki 92	3	1	0	0
98	Puskesmas Kel. Laki 93	3	1	0	0
99	Puskesmas Kel. Laki 94	3	1	0	0
100	Puskesmas Kel. Laki 95	3	1	0	0
101	Puskesmas Kel. Laki 96	3	1	0	0
102	Puskesmas Kel. Laki 97	3	1	0	0
103	Puskesmas Kel. Laki 98	3	1	0	0
104	Puskesmas Kel. Laki 99	3	1	0	0
105	Puskesmas Kel. Laki 100	3	1	0	0
106	Puskesmas Kel. Laki 101	3	1	0	0
107	Puskesmas Kel. Laki 102	3	1	0	0
108	Puskesmas Kel. Laki 103	3	1	0	0
109	Puskesmas Kel. Laki 104	3	1	0	0
110	Puskesmas Kel. Laki 105	3	1	0	0
111	Puskesmas Kel. Laki 106	3	1	0	0
112	Puskesmas Kel. Laki 107	3	1	0	0
113	Puskesmas Kel. Laki 108	3	1	0	0
114	Puskesmas Kel. Laki 109	3	1	0	0
115	Puskesmas Kel. Laki 110	3	1	0	0
116	Puskesmas Kel. Laki 111	3	1	0	0
117	Puskesmas Kel. Laki 112	3	1	0	0
118	Puskesmas Kel. Laki 113	3	1	0	0
119	Puskesmas Kel. Laki 114	3	1	0	0
120	Puskesmas Kel. Laki 115	3	1	0	0
121	Puskesmas Kel. Laki 116	3	1	0	0
122	Puskesmas Kel. Laki 117	3	1	0	0
123	Puskesmas Kel. Laki 118	3	1	0	0
124	Puskesmas Kel. Laki 119	3	1	0	0
125	Puskesmas Kel. Laki 120	3	1	0	0
126	Puskesmas Kel. Laki 121	3	1	0	0
127	Puskesmas Kel. Laki 122	3	1	0	0
128	Puskesmas Kel. Laki 123	3	1	0	0
129	Puskesmas Kel. Laki 124	3	1	0	0
130	Puskesmas Kel. Laki 125	3	1	0	0
131	Puskesmas Kel. Laki 126	3	1	0	0
132	Puskesmas Kel. Laki 127	3	1	0	0
133	Puskesmas Kel. Laki 128	3	1	0	0
134	Puskesmas Kel. Laki 129	3	1	0	0
135	Puskesmas Kel. Laki 130	3	1	0	0
136	Puskesmas Kel. Laki 131	3	1	0	0
137	Puskesmas Kel. Laki 132	3	1	0	0
138	Puskesmas Kel. Laki 133	3	1	0	0
139	Puskesmas Kel. Laki 134	3	1	0	0
140	Puskesmas Kel. Laki 135	3	1	0	0
141	Puskesmas Kel. Laki 136	3	1	0	0
142	Puskesmas Kel. Laki 137	3	1	0	0
143	Puskesmas Kel. Laki 138	3	1	0	0
144	Puskesmas Kel. Laki 139	3	1	0	0
145	Puskesmas Kel. Laki 140	3	1	0	0
146	Puskesmas Kel. Laki 141	3	1	0	0
147	Puskesmas Kel. Laki 142	3	1	0	0
148	Puskesmas Kel. Laki 143	3	1	0	0
149	Puskesmas Kel. Laki 144	3	1	0	0
150	Puskesmas Kel. Laki 145	3	1	0	0
151	Puskesmas Kel. Laki 146	3	1	0	0
152	Puskesmas Kel. Laki 147	3	1	0	0
153	Puskesmas Kel. Laki 148	3	1	0	0
154	Puskesmas Kel. Laki 149	3	1	0	0
155	Puskesmas Kel. Laki 150	3	1	0	0
156	Puskesmas Kel. Laki 151	3	1	0	0
157	Puskesmas Kel. Laki 152	3	1	0	0
158	Puskesmas Kel. Laki 153	3	1	0	0
159	Puskesmas Kel. Laki 154	3	1	0	0
160	Puskesmas Kel. Laki 155	3	1	0	0
161	Puskesmas Kel. Laki 156	3	1	0	0
162	Puskesmas Kel. Laki 157	3	1	0	0
163	Puskesmas Kel. Laki 158	3	1	0	0
164	Puskesmas Kel. Laki 159	3	1	0	0
165	Puskesmas Kel. Laki 160	3	1	0	0
166	Puskesmas Kel. Laki 161	3	1	0	0
167	Puskesmas Kel. Laki 162	3	1	0	0
168	Puskesmas Kel. Laki 163	3	1	0	0
169	Puskesmas Kel. Laki 164	3	1	0	0
170	Puskesmas Kel. Laki 165	3	1	0	0
171	Puskesmas Kel. Laki 166	3	1	0	0
172	Puskesmas Kel. Laki 167	3	1	0	0
173	Puskesmas Kel. Laki 168	3	1	0	0
174	Puskesmas Kel. Laki 169	3	1	0	0
175	Puskesmas Kel. Laki 170	3	1	0	0
176	Puskesmas Kel. Laki 171	3	1	0	0
177	Puskesmas Kel. Laki 172	3	1	0	0
178	Puskesmas Kel. Laki 173	3	1	0	0
179	Puskesmas Kel. Laki 174	3	1	0	0
180	Puskesmas Kel. Laki 175	3	1	0	0
181	Puskesmas Kel. Laki 176	3	1	0	0
182	Puskesmas Kel. Laki 177	3	1	0	0
183	Puskesmas Kel. Laki 178	3	1	0	0
184	Puskesmas Kel. Laki 179	3	1	0	0
185	Puskesmas Kel. Laki 180	3	1	0	0
186	Puskesmas Kel. Laki 181	3	1	0	0
187	Puskesmas Kel. Laki 182	3	1	0	0
188	Puskesmas Kel. Laki 183	3	1	0	0
189	Puskesmas Kel. Laki 184	3	1	0	0
190	Puskesmas Kel. Laki 185	3	1	0	0
191	Puskesmas Kel. Laki 186	3	1	0	0
192	Puskesmas Kel. Laki 187				

Gambar 1: Data Penderita HIV (Sumber:katalog.co.id)

2. 2 Klasterisasi Data

Pada Penelitian ini menerapkan pendekatan *unsupervised learning* dengan metode klasterisasi untuk mengevaluasi dataset yang diambil dalam format CSV dari katalog.data.co.id. Dua algoritma utama digunakan dalam proses ini, yaitu K-Means dan Fuzzy C-Means, dengan evaluasi menggunakan indeks silhouette dan *Davies-Bouldin Index*. Klasterisasi bertujuan untuk mengelompokkan data mentah agar lebih mudah dianalisis dan dipahami.

2. 3 Metode K-Means

Algoritma K-Means bekerja dengan menentukan jumlah klaster (nilai K) yang diinginkan, lalu membagi data ke dalam klaster secara acak sebelum menghitung rata-rata untuk menentukan pusat klaster. K-Means merupakan metode sederhana yang digunakan untuk membagi entitas menjadi kelompok yang disebut klaster berdasarkan kesamaan karakteristik.(Nugroho, D.Y., Wibowo, A.Muhammadiyah, M., 2019). Algoritma ini menggunakan jarak Euclidean untuk mengukur kesamaan antar data, dengan rumus:

$$d(b_i, a_t) = \sqrt{\sum_{j=1}^l (b_{ij} - a_{tj})^2}$$

Keterangan :

- a. d adalah jarak antara data dengan pusat klaster
- b. Bi adalah nilai data
- c. at adalah pusat klaster

2. 4 Metode Fuzy C-Means

metode Fuzzy C-Means (FCM) menggunakan konsep tingkat keanggotaan (membership degree) untuk setiap data terhadap setiap klaster. Berbeda dari K-Means yang hanya mengelompokkan data secara tegas, FCM memungkinkan suatu data memiliki keanggotaan di lebih dari satu klaster, sehingga lebih fleksibel dalam menangani data yang menyebar secara tidak teratur.

FCM sangat berguna ketika objek data memiliki karakteristik yang tumpang tindih antara beberapa klaster. Formula matematis dari metode ini melibatkan pusat klaster (V_{kj}), derajat keanggotaan (μ_{ik}), dan nilai data (X_{ij}) dengan rumus :

$$V_{kj} = \frac{\sum_{i=1}^n ((\mu_{ik})^w X_{ij})}{\sum_{i=1}^n (\mu_{ik})^w}$$

2. 5 Evaluasi Clustering Silhoutte Dan Davies-Bouldin Index

Evaluasi tambahan dilakukan menggunakan Clustering Silhoutte dan Davies-Bouldin Index untuk mengukur performa K-Means dan Fuzzy C-Means. Indeks ini mengevaluasi sejauh mana klaster saling terpisah dan seberapa kompak klaster tersebut. Semakin rendah nilai indeks ini, maka semakin baik hasil klasterisasi.

Evaluasi Clustering: Silhouette

Nilai Silhouette	Interpretasi
0.71 - 1.00	Struktur yang dihasilkan kuat
0.51 - 0.70	Struktur yang dihasilkan baik
0.26 - 0.50	Struktur yang dihasilkan lemah
≤ 0.25	Tidak terstruktur

e

Experiment	Davies Bouldin Index Value					
	Simple K-Means			K-Means with centroid based min SSE		
	k=2	k=3	k=4	k=2	k=3	k=4
1	0.4296	0.0524	0.0406	0.1366	0.0493	0.0372
2	0.4296	0.0491	0.0358	0.2392	0.0540	0.0409
3	0.4296	0.0524	0.0322	0.2392	0.0505	0.0373
4	0.4296	0.0661	0.0367	0.2392	0.0493	0.0409
5	0.2845	0.0493	0.0391	0.2392	0.0493	0.0372
6	0.2845	0.0679	0.0515	0.2392	0.0493	0.0319
7	0.2845	0.0493	0.0320	0.2392	0.0493	0.0373
8	0.4296	0.0576	0.0407	0.2392	0.0491	0.0409
9	0.4296	0.0493	0.0365	0.1366	0.0493	0.0365
10	0.4296	0.0661	0.0407	0.2392	0.0493	0.0406
Average	0.38607	0.05595	0.03858	0.21868	0.04987	0.03807

Gambar 2. Evaluasi Davies Bouldin Index Value

3. HASIL DAN PEMBAHASAN

Pengelompokan data menggunakan algoritma K-Means menunjukkan bahwa seluruh data berhasil dikelompokkan berdasarkan jenis dan jumlah penderita HIV. Visualisasi hasil pengelompokan dan kurva *elbow* dari metode ini menggambarkan sebaran serta jumlah optimal klaster.

Pada metode Fuzzy C-Means (FCM), digunakan parameter *random_state* yang mengontrol nilai acak dalam proses klasterisasi. Dengan menetapkan nilai awal acak, metode ini berhasil membentuk klaster sesuai penyebaran data, sebagaimana ditunjukkan dalam hasil visualisasi clustering.

Evaluasi performa kedua metode dilakukan menggunakan dua metrik, yaitu Silhouette Score dan Davies-Bouldin Index. Nilai *silhouette* menilai kualitas klaster berdasarkan keserupaan dalam klaster yang sama, sementara Davies-Bouldin mengukur seberapa baik klaster terbentuk berdasarkan jarak dan penyebaran antar klaster.

Hasil perbandingan menunjukkan bahwa secara umum, metode Fuzzy C-Means memberikan hasil yang lebih akurat dan representatif dibandingkan K-Means dalam mengelompokkan data penyebaran HIV di Surabaya. Evaluasi dari grafik dan indeks yang dihasilkan mengindikasikan bahwa FCM lebih sesuai untuk mendeskripsikan sebaran penderita HIV berdasarkan penanganan kesehatan yang sesuai standar nasional.

Pengelompokan data dengan metode K-Means terjadi dimana semua jumlah data per kategori sudah dikelompokan berdasarkan jenis dan jumlah total data yang ada pada Gambar 4. Kemudian statistik data elbow yang muncul dapat dilihat pada Gambar 5.

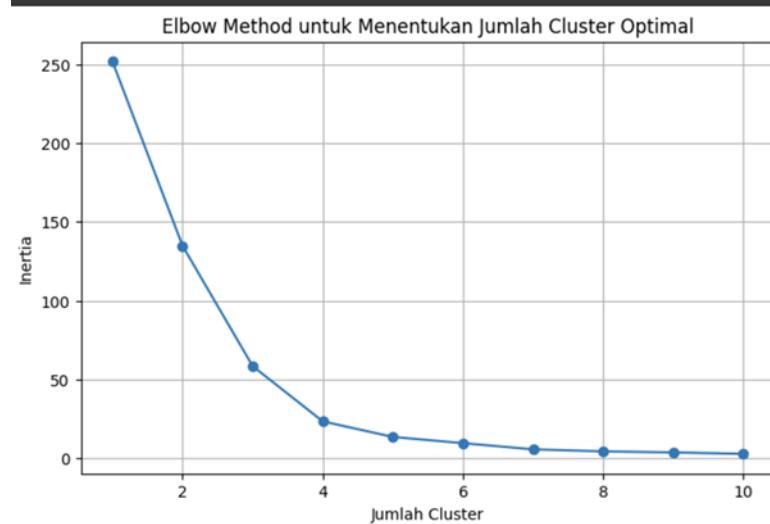
```
[ ] Karakteristik rata-rata setiap cluster:
  ↪ kode_sarana_kesehatan_lama_penanganan \
    Cluster
    0.0      1.033304e+06      3.0
    1.0      1.033316e+06      3.0
    2.0      1.033326e+06      3.0

    jumlah_penduduk_terjangkit_laki2 \
    Cluster
    0.0          1.0
    1.0          0.0
    2.0          0.0

    jumlah_penduduk_terjangkit_perempuan  total_penduduk_terjangkit \
    Cluster
    0.0          0.0          1.0
    1.0          0.0          0.0
    2.0          1.0          1.0

    jumlah_orang_terduga_menderita_hiv_yang_mendapatkan_pelayanan_kesehatan_sesuai_standar \
    Cluster
    0.0          83.400000
    1.0          56.631579
    2.0          1.000000

    Cluster
    Cluster
    0.0      0.0
    1.0      1.0
    2.0      2.0
```

Gambar 3. Pengelompokan K-Means**Gambar 4. Elbow K-Means**

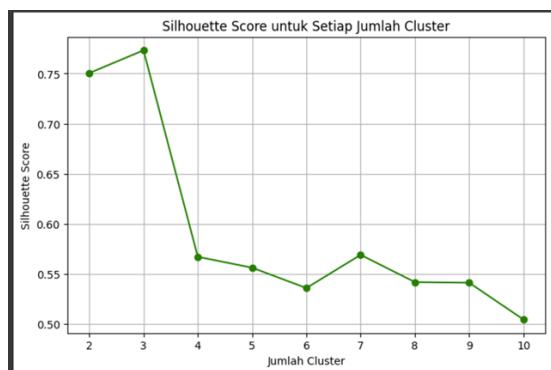
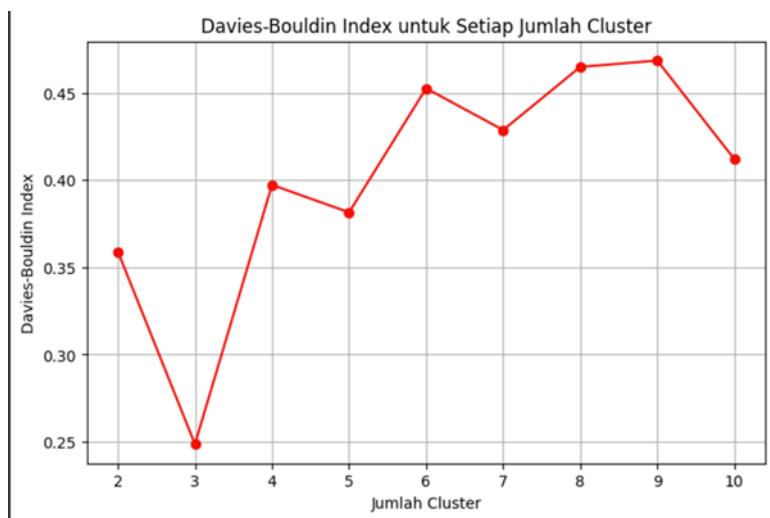
Selanjutnya peforma yang digunakan dalam metode ini yaitu Fuzzy C-Means hasil dapat dilihat pada Gambar 5.

```
 ↪ Distribusi jumlah data per cluster:
  Cluster
    1.0      57
    0.0       5
    2.0       1
  Name: count, dtype: int64
```

Gambar 5. Hasil Clustering Fuzzy C-Means

3.1 Evaluasi Clustering Silhoutte Dan Davies-Bouldin Index

Masing-masing metode ini melampirkan hasil elbow yang ada kemudian setelah di clustering data akan muncul Kembali dengan data yang sebenarnya dapat dilihat pada Gambar 6 metode silhouette dan Gambar 7 davies boildin

**Gambar 6. Silhouette Score****Gambar 7. Davies Boildin Index**

3.2 Perbandingan Peforma Algoritma

Metode clustering yang dilakukan dibandingkan akurasinya dan evaluasi agar dapat terlihat metode mana yang memiliki performa terbaik. Perbandingan evaluasi clustering ini dapat dilihat pada Gambar 8.

→
 Silhouette Score untuk 3 cluster: 0.773
 Davies-Bouldin Index untuk 3 cluster: 0.249

Gambar 8. Perbandingan Akurasi Clustering

4. KESIMPULAN

Berdasarkan hasil evaluasi yang telah dilakukan, algoritma K-Means menghasilkan nilai *silhouette* sebesar 0,773 dengan pembentukan tiga klaster, sedangkan algoritma Fuzzy C-Means memperoleh nilai Davies-Bouldin Index sebesar 0,249 untuk jumlah klaster yang sama. Hasil tersebut mengindikasikan bahwa metode Fuzzy C-Means memiliki performa yang lebih baik dibandingkan K-Means dalam melakukan proses

klasterisasi data penyebaran HIV, khususnya yang berkaitan dengan standar penanganan kesehatan nasional di Provinsi Jawa Timur, Kota Surabaya. Meskipun demikian, diperlukan pengujian lanjutan dengan cakupan data yang lebih besar guna memperoleh hasil evaluasi yang lebih mendalam dan akurat.

5. DAFTAR PUSTAKA

- Alhamidi, R., 2022. Membaca Grafik Kasus HIV/AIDS di Jabar: Naik 3 Tahun Terakhir!. Detikjabar.[online] Available at:<https://www.detik.com/jabar/berita/d6262009/membaca-grafik-kasus-hivaids-dijabar-naik-3-tahun-terakhir> [Accessed 18 Jan. 2023].
- D. Nugroho, D.Y., Wibowo, A., Fakultas Kesehatan & Universitas Muhammadiyah, M., 2019. Analisis Cluster K-Means Kabupaten/Kota di Jawa Timur Berdasarkan Level Epidemik HIV. Biometrika dan Kependudukan, 8(2), pp.108–117. doi:10.20473/jbk.v8i2.2019.108.
- Dodu, A.Y.E., Nugraha, D.W. & Ayyub, M.A.A., 2019. Penerapan Data Mining untuk Mendeteksi Tingkatan Stadium Penyakit HIV/AIDS Menggunakan Algoritma Naive Bayes Classifier (Studi Kasus pada RSUD Anutapura Palu). Scientific: Computer ScienceandnformaticsJournal,1(1),p.33.doi:10.22487/j26204118.2018.v1.i111900
- Herlinda, V. & Darwis, D., 2021. *Analisis Clustering untuk Recredesialing Fasilitas Kesehatan Menggunakan Metode Fuzzy C-Means*. Teknologi dan Sistem Informasi, 2(2), pp.94–99. [online] Available at: <http://jim.teknokrat.ac.id/index.php/ITSI>.
- Lesmana, A.A., et al., 2019. Implementasi Algoritma K-Means untuk Clustering Penyakit HIV/AIDS di Indonesia. eProceeding Engineering, 6(2), pp.5564–5580.
- Noor, H., Dharmawati, A. & Qur'ana, T.W., 2021. Penerapan Algoritma K-Means Clustering Analysis pada Kasus Penderita HIV/AIDS (Studi Kasus Kabupaten Banjar). Technologia: Jurnal Ilmiah, 12(2), p.72. doi:10.31602/tji.v12i2.4573
- Rohmatullailah, D. & Fikriyah, D., 2021. *Faktor Risiko Kejadian HIV pada Kelompok Usia Produktif di Indonesia*. Jurnal Biostatistik, Kependudukan, dan Informasi Kesehatan, 2(1), p.45. doi:10.51181/bikfokes.v2i1.4652.
- Samudra, A.W.P., Susanto, R.A., Putra, A.R., Kurniadi, F.I. & Juarto, B., 2022. Klasifikasi HIV AIDS dengan Aplikasi Rapid Miner. Sistem Komputer dan Kecerdasan Buatan, (1), pp.15–19.
- Sinaga, R.I.L., Saputra, W. & Qurniawan, H., 2021. Pengelompokan Jumlah Kasus Penyakit AIDS Berdasarkan Provinsi Menggunakan Metode K-Means. KESATRIA: Jurnal Penerapan Sistem Informasi (Komputer Manajemen), 2(2), pp.99–107.