

Penerapan Rapidminer Dengan K-Means Pada Clustering Daerah Terjangkit AIDS

Ninuk Wiliani, Anang Martoyo, Angel Tiarma Sipahutar,

Alfi Prabowo, Eksa Manda Pramaswari, Novita Maharani Suparta

Sistem dan Teknologi Informasi, Kewirausahaan, Teknik Informatika,

Institut Teknologi dan Bisnis Bank Rakyat Indonesia Jakarta Indonesia

ninukwiliani15@gmail.com, anangmrt@gmail.com, sipahutarangel16@gmail.com,

alfiprabowo05@gmail.com, mandaeksa5@gmail.com, novitamaharani150@gmail.com

Abstract— Aids merupakan sekumpulan gejala akibat kekurangan atau kelemahan sistem kekebalan tubuh yang dibentuk setelah manusia lahir. Aids disebabkan oleh virus yang disebut HIV atau *Human Immunodeficiency Virus*. Apabila seseorang terkena HIV, maka tubuh manusia akan mencoba menyerang infeksi. Sistem kekebalan manusia yang disebut dengan antibody akan menyerang HIV tersebut. Pada kasus AIDS terdapat banyak data mentah yang dapat diolah dan dikembangkan untuk membantu melihat daerah mana yang mempunyai potensi AIDS terburuk. Metode yang digunakan pada penelitian ini menggunakan metode Klustering K-Mean untuk memperoleh gambaran dari setiap wilayah di Indonesia.

Kata Kunci : AIDS, Klustering, Data

I. PENDAHULUAN

AIDS adalah kependekan dari (*Acquired Immune Deficiency Syndrome* 'Acquired') berarti didapat, bukan keturunan. Immune terkait dengan sistem kekebalan tubuh kita. *Deficiency* berarti kekurangan. *Syndrome* atau sindrom berarti penyakit dengan kumpulan gejala, buka gejala tertentu. Jadi AIDS berarti kumpulan gejala akibat kekurangan atau kelemahan sistem kekebalan tubuh yang dibentuk setelah kita lahir. AIDS disebabkan oleh virus yang disebut HIV atau *Human Immunodeficiency Virus*. Bila manusia terinfeksi HIV, Tubuh manusia akan mencoba menyerang infeksi. Sistem kekebalan dari manusia kana membuat *antibody* molekul khusus yang menyerang HIV itu[1]. Hal tersebut biasanya diikuti dengan adanya gejala yang cukup berkepanjangan. Indonesia merupakan salah satu negara yang memiliki potensi kenaikan jumlah kasus HIV/AIDS setiap tahunnya. Hal ini dapat dilihat datanya dari tahun 2008 – 2021. Kenaikan yang cukup signifikan terjadi di hamper seluruh provinsi yang ada di Indonesia. Bukan hanya itu, kasus penderita AIDS yang meninggal menjadi faktor penentu kenaikan potensi tersebut[2].

Data Mining merupakan suatu disiplin ilmu dalam *computer* yang digunakan untuk tujuan menggali informasi tersembunyi

dari sekumpulan data dan kemudian digunakan dalam pengambilan keputusan maupun kebijakan yang berguna dalam bidang-bidang tertentu. Pada kasus AIDS ini dengan tersedianya data mentah yang dapat diolah, penelitian ini dapat dikembangkan untuk membantu dalam melihat daerah mana atau provinsi dengan potensi AIDS terburuk. Oleh karena itu, dengan menggunakan metode *clustering* K-Mean, diperoleh gambaran dari setiap wilayah di Indonesia[3].

Data Mining

Data Mining adalah suatu proses pengerukan atau pengumpulan informasi penting dari suatu data yang besar. Proses data mining seringkali menggunakan metode statistika, matematika, hingga memanfaatkan Teknologi *Artificial Intelligence*. Nama alternatifnya yaitu *Knowledge discovery (mining)* in databases (KDD), *knowledge extraction*, *data/pattern analysis*, *data archeology*, *data dredging*, *information harvesting*, *business intelligence*, dan lain-lain[4].

Clusteirng

Clustering atau klasterisasi biasanya digunakan pada *business intelligence*, pengenalan pola citra, *web search*, bidang ilmu biologi, dan untuk keamanan (*security*). *Clustering* adalah proses pengelompokan data ke dalam beberapa *cluster* sehingga data-data di suatu *cluster* memiliki kemiripan maksimum. Cara kerja teknik ini ialah mengelompokkan sekumpulan data ke dalam kelas-kelas atau kluster-kluster, yang mana objek-objek yang ada pada kelas tersebut memiliki similaritas yang tinggi jika dibandingkan dengan objek lain yang ada dalam kelas tersebut, namun memiliki similaritas yang rendah jika dibandingkan dengan objek yang ada di kelas/kluster lain[5]

K-Means Clustering

K-Means Clustering adalah suatu metode penganalisaan data atau metode *Data Mining* yang melakukan proses

pemodelan tanpa supervisi (*unsupervised*) dan merupakan salah satu metode yang melakukan pengelompokan data dengan sistem partisi. Pada algoritma pembelajaran ini, komputer mengelompokkan sendiri data-data yang menjadi masukannya tanpa mengetahui terlebih dulu target kelasnya. Pembelajaran ini termasuk dalam *unsupervised learning*. Masukan yang diterima adalah data atau objek dan k buah kelompok (*cluster*) yang diinginkan. Algoritma ini akan mengelompokkan data atau objek ke dalam k buah kelompok tersebut. Pada setiap *cluster* terdapat titik pusat (*centroid*) yang merepresentasikan *cluster* tersebut. Algoritma ini menggunakan masukan berupa jumlah k dan sekumpulan data set dari sekian objek untuk dimasukkan ke dalam k kelas/*cluster*[6]. Cara kerja *k-means* adalah sebagai berikut. Pertama, algoritma ini akan memilih sejumlah k objek dari beberapa objek yang ada pada sekumpulan data. Masing-masing objek yang terpilih, merepresentasikan nilai rata-rata dari sejumlah k kluster. Kemudian objek yang tersisa, masing-masing akan di-assign ke dalam kluster yang sudah ditentukan berdasarkan similaritasnya dengan nilai rata-rata masing-masing kluster. Setelah itu, dilakukan perubahan nilai rata-rata kluster yang sudah menjadi beberapa anggota kluster. Ketiga langkah di atas dilakukan secara iteratif (berulang) hingga tidak ada lagi perubahan nilai rata-rata dan seluruh data sudah didistribusikan ke masing-masing kluster yang ada.

II. METODOLOGI PENELITIAN

Dalam kegiatan pengumpulan data untuk penelitian ini digunakan metode pengumpulan studi pustaka yang mana pada metode ini kegiatan dilakukan adalah mempelajari, mencari dan mengumpulkan data yang berhubungan dengan penelitian ini. Data yang digunakan dalam pengelompokan kasus penyakit AIDS ini diperoleh dari dokumen keterangan jumlah kasus penyakit AIDS berdasarkan provinsi oleh Badan Pusat Statistik pada tahun 2018 – 2021[4].

Analisis data merupakan proses mencari dan menyusun secara sistematis data yang telah diperoleh dari wawancara, *dokumentasi*, dokumen pribadi, *observasi*, catatan lapangan, gambar foto dan sebagainya, dengan cara mengorganisasikan data tersebut ke dalam kategori, menjabarkan ke dalam unit - unit, melakukan sintesa, menyusun ke dalam pola, memilih mana yang penting dan mana yang akan dipelajari dan kemudian membuat kesimpulan agar dapat dipahami diri sendiri dan orang lain. Dalam penulisan penelitian ini menggunakan

analisis data yang bersifat kualitatif, penelitian kualitatif adalah analisis yang dilakukan dengan mengelompokkan data untuk mencari suatu pola dari hal yang dipelajari dan membandingkan konsep-konsep yang ada dalam sumber[7].

Dalam melakukan *clustering*, data yang diperoleh akan dihitung terlebih dahulu berdasarkan jumlah Kasus Penyakit AIDS berdasarkan Provinsi. Hasil rata-rata berdasarkan 2 kriteria penilaian yakni 1). Rata-rata jumlah kasus penderita AIDS berdasarkan provinsi dan 2). Rata-rata jumlah kasus meninggal penderita AIDS seperti yang ditunjukkan pada tabel 1.

Tabel 1. Data Penderita AIDS

Provinsi	Jumlah Kasus				Rata Rata	Meninggal			
	2018	2019	2020	2021		2018	2019	2020	2021
Aceh	22	36	53	63	43,5	8	9	12	14
Sumatera Utara	670	485	507	509	542,75	135	93	94	94
Sumatera Barat	184	293	410	428	328,75	61	75	99	104
Riau	364	371	477	477	422,25	116	117	132	132
Jambi	106	165	268	291	207,5	30	50	62	66
Sumatera Selatan	124	219	219	219	195,25	29	38	38	38
Bengkulu	33	85	131	137	96,5	10	18	29	30
Lampung	143	144	144	144	143,75	42	42	42	42
Kep. Bangka	67	117	120	120	106	4	18	18	18
Kep. Riau	271	333	374	390	342	114	130	143	148
DKI Jakarta	2727	2811	3995	3997	3383	440	425	576	577
Jawa Barat	2603	3233	3728	3809	3233	503	588	665	678
Jawa tengah	409	669	944	1336	674	171	238	289	370
DI	129	247	505	673	388,5	46	70	108	134
Yogyakarta	2525	3133	3771	3775	3771	575	680	779	779
Jawa Timur	71	275	401	403	287,5	12	51	67	68
Banten	869	1506	1747	1747	869	145	275	311	311
Bali	114	107	142	206	142,25	47	56	69	110
NTB	110	138	242	385	218,75	23	25	36	50
NTT	730	730	1125	1125	730	110	103	138	138
Kalbar	7	15	57	69	37	2	2	4	4
Kalteng	22	27	27	28	26	5	5	5	5
Kalsel	11	11	11	12	11,25	10	10	10	10
Kaltim	117	173	173	557	255	44	62	62	125
Sulawesi Utara	8	12	12	12	11	4	6	6	6
Sulawesi Tengah	143	143	591	995	468	62	62	62	167
Sulsel	9	20	22	26	19,25	1	4	5	5
Sulteng	3	3	3	3	3	1	1	1	1
Gorontalo	-	-	-	0	0	-	-	-	0
Sulawesi Barat	186	192	192	195	191,25	68	70	70	70
Maluku	7	10	17	17	12,75	7	8	8	8
Maluku Utara	58	58	397	397	142,75	19	19	19	152
Papua Barat	2294	2681	3938	3938	3665	335	358	580	602
Papua	15136	18442	26483	26483	24131	3197	3708	4539	5056
Indonesia									

III. HASIL DAN PEMBAHASAN

Data tersebut kemudian diakumulasikan dan diambil nilai rata-rata dari setiap kriteria seperti yang ditunjukkan pada tabel 2 :

Tabel 2. Rata Rata Penderita AIDS

Provinsi	Jumlah Kasus				Rata Rata	Meninggal				Rata Rata
	2018	2019	2020	2021		2018	2019	2020	2021	
Aceh	22	36	53	63	43,5	8	9	12	14	10,75
Sumatera Utara	670	485	507	509	542,75	135	93	94	94	104
Sumatera Barat	184	293	410	428	328,75	61	75	99	104	84,75
Riau	364	371	477	477	422,25	116	117	132	132	124,25
Jambi	106	165	268	291	207,5	30	50	62	66	52
Sumatera Selatan	124	219	219	219	195,25	29	38	38	38	35,75
Bengkulu	33	85	131	137	96,5	10	18	29	30	21,75
Lampung	143	144	144	144	143,75	42	42	42	42	42
Kep. Bangka	67	117	120	120	106	4	18	18	18	14,5
Kep. Riau	271	333	374	390	342	114	130	143	148	133,75
DKI Jakarta	2727	2811	3995	3997	3383	440	425	576	577	504,5
Jawa Barat	2603	3233	3728	3809	3233	503	588	665	678	608,5
Jawa tengah	409	669	944	1336	674	171	238	289	370	267
DI	129	247	505	673	388,5	46	70	108	134	89,5
Yogyakarta										
Jawa Timur	2525	3133	3771	3775	3771	575	680	779	779	703,25
Banten	71	275	401	403	287,5	12	51	67	68	49,5
Bali	869	1506	1747	1747	869	145	275	311	311	260,5
NTB	114	107	142	206	142,25	47	56	69	110	260,5
NTT	110	138	242	385	218,75	23	25	36	50	33,5
Kalbar	730	730	1125	1125	730	110	103	138	138	122,25
Kalteng	7	15	57	69	37	2	2	4	4	3
Kalsel	22	27	27	28	26	5	5	5	5	5
Kaltim	11	11	11	12	11,25	10	10	10	10	10
Sulawesi Utara	117	173	173	557	255	44	62	62	125	73,25
Sulawesi Tengah	8	12	12	12	11	4	6	6	6	5,5
Sulse	143	143	591	995	468	62	62	62	167	88,25
Sulteng	9	20	22	26	19,25	1	4	5	5	3,75
Gorontalo	3	3	3	3	3	1	1	1	1	1
Sulawesi Barat	-	-	-	0	0	-	-	-	0	0
Maluku	186	192	192	195	191,25	68	70	70	70	69,5
Maluku Utara	7	10	17	17	12,75	7	8	8	8	7,75
Papua	58	58	397	397	142,75	19	19	19	152	52,25
Papua Barat	2294	2681	3938	3938	3665	335	358	580	602	473,25
Indonesia	15136	18442	26483	26483	24131	3197	3708	4539	5056	3708

Setelah diakumulasikan dan dicari nilai rata-rata maka akan didapatkan nilai dari setiap *variable*. Kemudian data tersebut akan masuk ke tahapan *clustering* dengan menerapkan algoritma *K-means* untuk *mengcluster* data

Centroid Data

Dalam penerapan algoritma *K-means* dihasilkan nilai titik tengah atau *centroid* dari data yang didapat dengan ketentuan bahwa clusterisasi yang diinginkan adalah 3, Penentuan *cluster* dibagi atas tiga bagian yakni *cluster* tinggi (C1), *cluster* sedang (C2) dan *cluster* tingkat rendah (C3). maka nilai titik tengah atau *centroid* juga terdapat 3 titik. Penentuan titik *cluster* ini dilakukan dengan mengambil nilai terbesar (maksimum) untuk *cluster* tinggi (C1), nilai rata-rata (*average*) untuk *cluster* sedang (C2) dan nilai terkecil (minimum) untuk *cluster* rendah (C3). Nilai titik tersebut dapat diketahui pada tabel 3

Tabel 3. Centroid Data (Iterasi 1)

Centroid		
Max (C1)	3771	703,25
Average (C2)	635,3636	125
Min (C3)	0	0

Clustering Data

Dengan menggunakan *centroid* tersebut maka dapat dicluster data yang telah didapat menjadi 3 *cluster*. Proses *cluster* dengan mengambil jarak terdekat dari setiap data yang diolah. Dari data jumlah desa/kelurahan yang memiliki sarana kesehatan menurut provinsi didapatkan pengelompokan pada

iterasi 1 untuk 3 cluster tersebut. *Cluster* penderita Kasus AIDS tinggi (C1) yakni DKI Jakarta, Jawa Barat, Jawa Timur dan Papua. *Cluster* penderita Kasus AIDS sedang (C2) yakni 9 Provinsi dan *cluster* penderita Kasus AIDS rendah (C3) yakni 20 provinsi lainnya. Proses pencarian jarak terdekat, pengelompokan data pada iterasi 1 dan Clustering atas dapat digambarkan pada tabel 4.

Tabel 4. Pengelompokan Data Klastering

Provinsi	X	Y	Iterasi 1			Jarak Terpendek
			C1	C2	C3	
Aceh	43,5	10,75	3791,28	602,79	44,81	44,81
Sumatera Utara	542,75	104,00	3283,40	94,96	552,62	94,96
Sumatera Barat	328,75	84,75	3497,37	309,24	339,50	309,24
Riau	422,25	124,25	3398,44	213,11	440,15	213,11
Jambi	207,5	52,00	3622,52	434,05	213,92	213,92
Sumatera Selatan	195,25	53,75	3637,52	449,07	198,50	198,50
Bengkulu	96,5	21,75	3737,16	548,67	98,92	98,92
Lampung	143,75	42	3687,03	498,57	149,76	149,76
Kep. Bangka	106	14,5	3729,16	540,27	106,99	106,99
Kep. Riau	342	133,75	3475,97	293,49	367,22	293,49
DKI Jakarta	3383	504,50	436,39	2773,23	3419,92	436,39
Jawa Barat	3233	608,50	546,28	2642,25	3289,77	546,28
Jawa tengah	674	267,00	3127,57	147,16	724,96	147,16
DI	388,5	89,50	343,73	249,40	398,68	249,40
Yogyakarta						
Jawa Timur	3771	703,25	0,00	3188,51	3836,01	0,00
Banten	287,5	49,50	3544,31	355,96	291,73	291,73
Bali	869	260,50	2935,58	270,09	907,21	270,09
NTB	142,25	270,50	3683,50	496,12	158,76	158,76
NTT	218,75	33,50	3614,84	426,54	221,30	221,30
Kalbar	730	122,25	3096,00	94,68	740,17	94,68
Kalteng	37	3	3799,09	610,67	37,12	37,12
Kalsel	26	5	3809,54	621,07	26,48	26,48
Kaltim	11,25	10	3823,13	634,62	15,05	15,05
Sulawesi	255	73,25	3572,00	383,87	265,31	265,31
Sulawesi Utara						
Sulawesi Tengah	11	5,50	3824,19	635,70	12,30	12,30
Sulse	468	88,25	3359,77	171,35	476,25	171,35
Sulteng	19,25	3,75	3816,40	627,93	19,61	19,61
Gorontalo	3	1,00	3832,88	644,41	3,16	3,16
Sulawesi Barat	0	0,00	3836,01	647,54	0,00	0,00
Maluku	191,25	69,50	3653,42	447,57	203,49	203,49
Maluku Utara	12,75	7,75	3822,06	633,56	14,92	14,92
Papua Barat	142,75	52,25	3686,19	497,96	152,01	152,01
Papua	3665	473,25	253,25	3049,59	3695,43	253,25

Hasil Pengelompokan data dapat dilihat di dalam Tabel 5

Tabel 5. Hasil Pengelompokan pada Iterasi pertama

Provinsi	C1	C2	C3
Aceh			1
Sumatera Utara		1	
Sumatera Barat		1	
Riau		1	
Jambi			1
Sumatera Selatan			1
Bengkulu			1
Lampung			1
Kep. Bangka			1
Kep. Riau		1	
DKI Jakarta	1		
Jawa Barat	1		
Jawa tengah		1	
DI		1	
Yogyakarta			
Jawa Timur	1		
Banten			1
Bali		1	
NTB			1
NTT			1
Kalbar		1	
Kalteng			1
Kalsel			1
Kaltim			1
Sulawesi			1
Sulawesi Utara			
Sulawesi Tengah			1
Sulse		1	
Sulteng			1
Gorontalo			1
Sulawesi Barat			1
Maluku			1
Maluku Utara			1
Papua Barat			1
Papua	1		
Jumlah	4	9	20

Proses *K-means* akan terus beriterasi sampai pengelompokan data sama dengan pengelompokan data iterasi sebelumnya. Dengan kata lain, proses akan terus melakukan iterasi sampai data pada iterasi terakhir sama dengan iterasi sebelumnya. Pada iterasi 1 diperoleh *cluster* data jumlah kasus penyakit AIDS berdasarkan Provinsi. Proses iterasi tersebut berhenti pada iterasi berikutnya, pada iterasi berikutnya akan

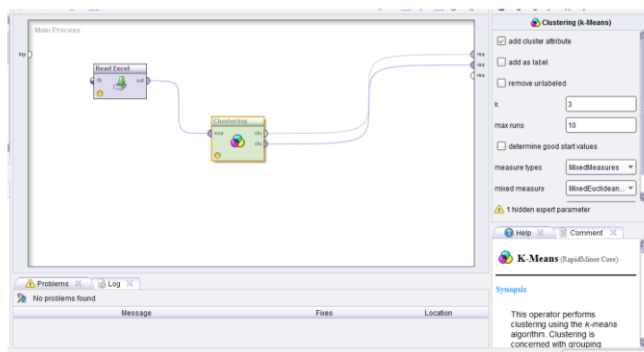
dilakukan proses mencari nilai titik tengah atau *centroid*. Setelah mendapatkan nilai titik tengah atau *centroid*, proses sama dilakukan dengan mencari jarak terdekat. Proses pencarian jarak terdekat.

Analisa Data

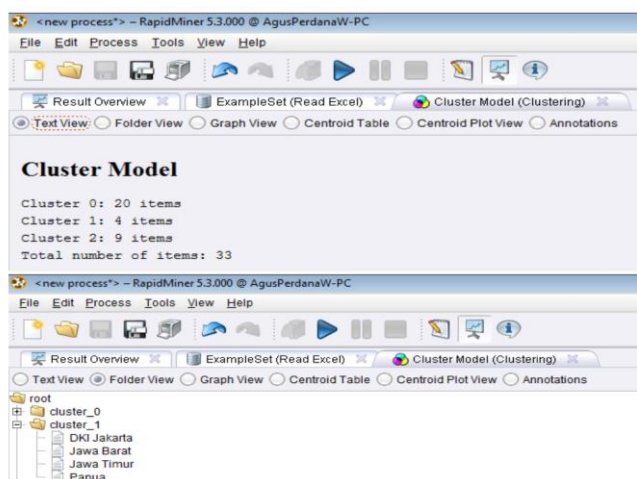
Pengelompokan data yang dilakukan terhadap 3 *cluster* dengan iterasi 1 didapatkan hasil yang sama. Dari 34 data jumlah kasus penderita AIDS berdasarkan provinsi dapat diketahui, 4 provinsi *cluster* tingkat tinggi untuk penderita kasus AIDS yakni Papua, DKI Jakarta, Jawa Barat dan Jawa Timur, 9 provinsi *cluster* tingkat sedang dan 20 provinsi lainnya termasuk *cluster* tingkat rendah

Implementasi Tool RapidMiner

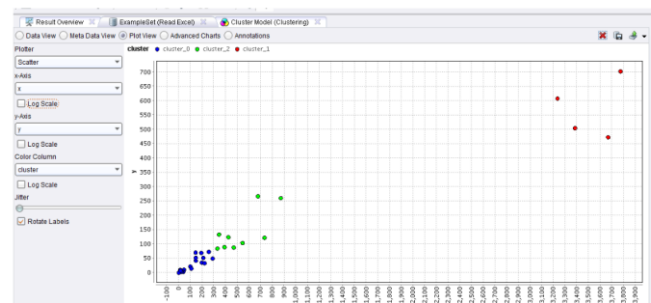
Dalam pengelompokan/*clustering*, peneliti menggunakan *K-means* untuk mengelompokkan data berdasarkan pada atribut. *K-Means* adalah teknik yang cukup sederhana dan cepat dalam pekerjaan pengelompokkan data (*clustering*) seperti yang ditunjukkan pada Gambar 1. Prinsip utama dari teknik ini adalah menyusun *k* buah prototipe/pusat massa (*centroid*)/rata-rata (*mean*) dari sekumpulan data berdimensi *n*. Aplikasi Rapidminer akan membantu proses pengelompokan/*clustering* dan ditunjukkan pada Gambar 2.



Gambar 1. Algoritma K-Means dengan nilai K = 3



Gambar 2. Hasil Klastering dengan K = 3



Gambar 3. Grafik Rapidminer dengan K = 3

Berdasarkan gambar 3, diperoleh hasil yang sama dalam pengelompokan cluster dimana *cluster* C1 terdiri 4 provinsi yakni: DKI Jakarta, Jawa Barat, Jawa Timur dan Papua

IV. KESIMPULAN

Untuk melakukan penilaian terhadap kasus penderita AIDS menurut provinsi dapat menerapkan metode *clustering K-means*. Data diolah untuk memperoleh nilai dari kasus penderita AIDS. Data tersebut diolah menggunakan Ms. Excel untuk ditentukan nilai *centroid* dalam 3 *cluster* yaitu *cluster* tinggi (C1), *cluster* sedang (C2) dan *cluster* rendah (C3). Sehingga diperoleh penilaian berdasarkan pengelompokan kasus penderita AIDS berdasarkan provinsi dengan 4 provinsi dalam *cluster* C1: Papua, DKI Jakarta, Jawa Barat dan Jawa Timur, 9 provinsi *cluster* C2, dan 20 provinsi lainnya *cluster* C3. Hasil yang didapat dari penelitian dapat menjadi masukan kepada pemerintah, khususnya provinsi yang menjadi *cluster* tertinggi dalam kasus penderita AIDS untuk mendapatkan tindakan lanjut.

V. DAFTAR PUSTAKA

- [1] O. Lesetedi *et al.*, "Incidence of Mortality among Under-Five-Year-Old Children Born to Women Living with HIV and Those Born to Women Not Living with HIV in Botswana : A 5-Year Retrospective Study," vol. 2022, 2022.
- [2] D. Fajardo-Ortiz *et al.*, "The emergence and evolution of the research fronts in HIV/AIDS research," *PLoS One*, vol. 12, no. 5, pp. 1–13, 2017, doi: 10.1371/journal.pone.0178293.
- [3] T. Javid, M. K. Gupta, and A. Gupta, "A hybrid-security model for privacy-enhanced distributed data mining," *J. King Saud Univ. - Comput. Inf. Sci.*, no. xxxx, 2020, doi: 10.1016/j.jksuci.2020.06.010.
- [4] Honakan, Adiwijaya, and S. AL-Faraby, "Analisis Dan Implementasi Support Vector Machine Dengan String Kernel Dalam Melakukan Klasifikasi Berita Berbahasa Indonesia Analysis and Implementation Support Vector Machine With String Kernel for Classification Indonesian news," vol. 5, no. 1, pp. 1701–1710, 2018.
- [5] N. Wiliani, A. Sani, and A. T. Andyanto, "Klasifikasi Kerusakan Dengan Jaringan Syaraf Backpropagation Pada Permukaan Solar Panel," *JITK (Jurnal Ilmu Pengetah. dan Teknol. Komputer)*, vol. 5, no. 1, pp. 89–

- 94, 2019, doi: 10.33480/jitk.v5i1.662.
- [6] L. Fadilla¹, A. Rizal², and E. Rachmawati³, “Implementasi dan Analisis Content-Based Image Retrieval Pada Citra X-Ray Menggunakan Algoritma Hierarki dan Algoritma Fast Genetic K-Means - Tugas Akhir-2011 Fakultas Teknik Informatika Program Studi S1 Teknik Informatika.”
- [7] S. A. Ningtyas, Moh.Halim, and A. Puspito, “Desain Sistem Informasi Akuntansi Penjualan Tunai Dan,” *J. Ris. Akunt. Keuang.*, vol. 4, no. 1, pp. 75–86, 2019.