

Penerapan Metode K-Means Dalam Mengelompokkan Persebaran Lahan Kritis Di Indonesia Berdasarkan Provinsi

Putra Pratama Siregar*, Solikhun, Zulia Almaida Siregar

STIKOM Tunas Bangsa, Pematangsiantar, Indonesia

Jalan Jendral Sudirman Blok A, No.1,2 & 3, Pematangsiantar, Sumatera Utara, Indonesia

Email: ¹tamasiregar46@gmail.com,

Email Penulis Korespondensi: ¹tamasiregar46@gmail.com

Abstrak—Penelitian bertujuan untuk mengelompokkan persebaran lahan kritis di Indonesia berdasarkan provinsi. Untuk menyelesaikan permasalahan ini peneliti menerapkan metode Algoritma K-Means. Dimana sumber data penelitian dikumpulkan berdasarkan dokumen – dokumen keterangan Luas dan Penyebaran Lahan Kritis Menurut Provinsi yang dihasilkan oleh Badan Pusat Statistik (BPS). Data yang digunakan dalam penelitian adalah data dari tahun 2011, 2013 dan 2018 yang terdiri dari 34 provinsi. Data akan diolah dengan melakukan clustering dalam 2 cluster yaitu cluster tingkat persebaran lahan kritis tinggi dan cluster tingkat persebaran lahan kritis rendah. Cluster tinggi berjumlah 4 data yaitu provinsi Sumatera Utara, Jambi, Jawa Timur, dan Kalimantan Tengah. Dengan dilakukannya penelitian dapat memberikan kontribusi dalam meningkatkan kinerja Balai Pengelolaan Daerah Aliran Sungai dan Hutan Lindung (BPDASHL) tentang proses membenahi dan menanggulangi lahan kritis pada provinsi yang ada di Indonesia.

Kata Kunci: Data Mining; Lahan Kritis; Clustering; K-Means; Pengelompokan

Abstract—The study aims to group the distribution of critical land in Indonesia by province. To solve this problem, researchers applied the K-Means Algorithm method. Where the source of research data is collected based on documents - documents of Information on The Extent and Dissemination of Critical Land By Province produced by the Central Statistics Agency (BPS). The data used in the study was data from 2011, 2013 and 2018 consisting of 34 provinces. Data will be processed by clustering in 2 clusters, namely clusters of high critical land distribution rates and clusters of low critical land distribution rates. The high cluster amounted to 4 data, namely the provinces of North Sumatra, Jambi, East Java, and Central Kalimantan. With the conduct of research can contribute in improving the performance of Balai Pengelolaan Daerah Aliran Sungai dan Hutan Lindung (BPDASHL) on the process of fixing and tackling critical land in the provinces in Indonesia.

Keywords: Data Mining; Critical Land; Clustering; K-Means

1. PENDAHULUAN

Keberadaan lahan tidak dapat dipisahkan dalam kehidupan manusia karena manusia memiliki ketergantungan terhadap pemanfaatan lahan. Manusia dapat memanfaatkan lahan untuk meningkatkan kualitas hidupnya sesuai dengan kebutuhan. Lahan merupakan sumber daya yang sangat penting untuk memenuhi segala kebutuhan hidup, sehingga dalam pengelolaannya harus sesuai dengan kemampuannya agar tidak menurunkan produktivitas lahan. Dalam penggunaan lahan sering tidak memperhatikan kelestarian lahan terutama pada lahan yang memiliki keterbatasan baik secara fisik maupun kimia. Lahan kritis didefinisikan sebagai lahan yang mengalami proses kerusakan fisik, kimia dan biologi karena tidak sesuai dengan penggunaan dan kemampuannya, yang akhirnya membahayakan produksi pertanian, pemukiman, kehidupan sosial ekonomi dan lingkungan [1].

Dampak lahan kritis mengakibatkan penurunan fungsi lahan yang tidak mampu lagi untuk menjaga tata air dan sumber daya tanah. Lahan kritis juga dipandang tidak mampu lagi sebagai media tumbuh dan berkembang tanaman, pertanian, perkebunan, kehutanan, perikanan, pemukiman, industri dan pariwisata. Berdasarkan kondisi tersebut dapat mempengaruhi kehidupan sosial dan ekonomi masyarakat yang memanfaatkan lahan tersebut. Maka perlu dilakukan pengelompokan untuk mengetahui provinsi yang memiliki tingkat persebaran lahan kritis yang masih tinggi. Solusi yang dapat digunakan untuk memecahkan permasalahan tersebut yaitu penulis menggunakan metode *Data Mining* algoritma *K-Means*.

“*Data Mining* merupakan proses menemukan korelasi baru, pola dan trend dengan menggali sejumlah *repository* data dalam jumlah besar, menggunakan teknologi pengenalah pola seperti statistik dan matematika” [2]. *K-Means* merupakan salah satu algoritma yang digunakan dalam pengelompokan secara partisi yang memisahkan data ke dalam kelompok yang berbeda. Pada dasarnya penggunaan algoritma ini dalam *clustering* tergantung pada data yang didapatkan dan konklus yang ingin dicapai diakhir proses [3]. Diharapkan menggunakan metode ini didapatkan hasil *clustering* yang terbukti akurat dalam kasus mengelompokkan persebaran lahan kritis berdasarkan provinsi yang ada di Indonesia.

Algoritma *K-Means* penggunaannya sudah diterapkan di beberapa penelitian yang sudah pernah dilakukan seperti penelitian yang dilakukan oleh [2] berjudul “Implementasi *Data Mining* dalam Mengelompokkan Rumah Tangga Kumuh di Perkotaan Berdasarkan Provinsi Menggunakan Algoritma *K-Means*”. Berdasarkan penelitian yang dilakukan, bisa diambil kesimpulan bahwa algoritma *K-means* telah berhasil mengelompokkan rumah tangga kumuh yang ada di provinsi Indonesia dengan ditentukannya nilai *centroid* dalam 3 *cluster* yaitu *cluster* tinggi (C1), *cluster* sedang (C2), dan *cluster* rendah (C3). Diharapkan hasil penelitian dapat memberikan masukan kepada pemerintah agar dilakukan penanganan lebih dan antisipasi untuk memperbaiki permasalahan tersebut.

2. METODOLOGI PENELITIAN

2.1. Teknik pengumpulan data

Sumber data penelitian ini diperoleh dari Badan Pusat Statistik (BPS) dengan situs <http://www.bps.go.id> untuk menyelesaikan permasalahan mengenai persebaran lahan kritis. Pengelompokan data persebaran lahan kritis di Indonesia kedalam 2 *cluster* yakni *cluster* rendah dan *cluster* tinggi. Dimana *cluster* tinggi menjadi pusat perhatian kepada pihak pemerintah untuk dapat meningkatkan kepedulian terhadap kelestarian lahan.

2.2. Algoritma K-Means

Algoritma *K-Means Clustering* merupakan algoritma yang berulang serta cukup sederhana dan cepat dalam pekerjaan pengelompokan data (*clustering*). Prinsip utama dari teknik ini adalah dimulai dengan pemilihan secara acak K, K disini merupakan banyaknya *cluster* yang ingin dibentuk. Kemudian tetapkan nilai-nilai K secara *random*, untuk sementara nilai tersebut menjadi pusat dari *cluster* atau biasa disebut dengan *centroid* (rata-rata) atau “mean”. Hitung jarak setiap data yang ada terhadap masing-masing *centroid* menggunakan rumus “*Euclidean*” hingga ditemukan jarak yang paling dekat dari setiap data dengan *centroid*. Klasifikasikan setiap data berdasarkan kedekatannya dengan *centroid*. Lakukan langkah tersebut hingga nilai *centroid* tidak berubah atau stabil[8]. *K-means* juga merupakan pengelompokan data non hirarki (sekatan) yang mempartisi data ke dalam bentuk dua atau lebih kelompok sehingga data yang berkarakteristik sama dimasukkan ke dalam satu kelompok yang sama [9]. Berikut langkah – langkah melakukan *clustering* dengan menggunakan algoritma *K-means* [10]:

- Tentukan nilai k sebagai jumlah *cluster* yang ingin dibentuk.
- Membangkitkan k *centroid* (titik pusat *cluster*) awal secara random/acak, kemudian untuk menghitung *centroid cluster* ke-i berikutnya, digunakan rumus sebagai berikut :

$$v = \sqrt{\frac{\sum_{i=1}^n x_i}{n}} \quad (1)$$

Dimana :

i : 1,2,3,...n

v : *centroid* pada *cluster*

x_i : objek ke-i

n : banyaknya jumlah objek yang menjadi anggota *cluster*

- Hitung jarak setiap data ke masing-masing *centroid* menggunakan rumus.

$$d(x,y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (2)$$

Keterangan :

d_{ij} : Jarak objek antara objek x dan y

n : Jumlah Atribut

x_i : Objek Data

y_i : Data *Cluster*

- Kelompokkan setiap data berdasarkan jarak terdekat antara data dengan *centroidnya* (C).
- Kembali ke langkah 3 jika posisi *centroid* baru dengan *centroid* lama tidak sama.

3. HASIL DAN PEMBAHASAN

3.1. Pengolahan Data Manual K-means

Untuk mendapatkan hasil dari penelitian yang dilakukan, berikut uraian perhitungan manual proses *clustering* penyebaran lahan kritis menggunakan algoritma *K-means*. Proses *clustering* dilakukan dengan penentuan data yang ingin di *cluster*. Dalam hal ini variabel data yang ingin di *cluster* adalah data tahun 2011, 2013 dan 2018 dan peneliti mengambil nilai rata-rata dari tahun 2011, 2013 dan 2018 sebagai data yang akan di *cluster*. Berikut adalah langkah-langkah penyelesaian yang penulis lakukan dalam mengelompokkan penyebaran lahan kritis menggunakan algoritma *K-means* :

- Menentukan Data Yang Akan Di *Cluster*

Dimana sampel data penyebaran lahan kritis yang akan digunakan dalam proses *clustering* adalah data yang berasal dari Badan Pusat Statistik (BPS) pada tahun 2014, 2013 dan 2018 dengan jumlah data sebanyak 34 provinsi. Berikut data yang akan digunakan :

Tabel 1. Data Luas dan Penyebaran Lahan Kritis

Provinsi	Luas dan Penyebaran Lahan Kritis Menurut Provinsi			
	2011	2013	2018	Rata – Rata
Aceh	121290	150694	190399	154128

Provinsi	Luas dan Penyebaran Lahan Kritis Menurut Provinsi			
	2011	2013	2018	Rata – Rata
Sumatera Utara	280731	478523	502280	420511
Sumatera Barat	90453	144788	477221	237487
Riau	100723	151813	275164	175900
Jambi	1078917	264582	102130	481876
Sumatera Selatan	217707	13692	130818	120739
Bengkulu	111117	135648	28289	91685
Lampung	77061	84602	21977	61213
Kep. Bangka Belitung	26624	60720	261	29202
Kep. Riau	230230	114177	3961	116123
Dki Jakarta	0	0	0	0
Jawa Barat	68139	40952	679748	262946
Jawa Tengah	9877	5210	217359	77482
Di Yogyakarta	471	845	790	702
Jawa Timur	102577	736877	345289	394914
Banten	10750	3716	94083	36183
Bali	2940	2910	2284	2711
Nusa Tenggara Barat	23248	23219	10641	19036
Nusa Tenggara Timur	35162	17878	9585	20875
Kalimantan Barat	325357	106864	19090	150437
Kalimantan Tengah	1138854	359405	92484	530248
Kalimantan Selatan	78781	132645	225135	145520
Kalimantan Timur	4372	63230	118433	62012
Kalimantan Utara	0	29125	36214	21780
Sulawesi Utara	33702	79395	92034	68377
Sulawesi Tengah	24131	104277	100569	76326
Sulawesi Selatan	109948	144152	294019	182706
Sulawesi Tenggara	267944	313477	15031	198817
Gorontalo	72524	247244	258855	192874
Sulawesi Barat	8414	55749	14336	26166
Maluku	271803	257761	30	176531
Maluku Utara	163438	97153	14949	91847
Papua Barat	76742	50997	9784	45841
Papua	105235	266064	169481	180260

Sumber : (Badan Pusat Statistik)

b) Menentukan Nilai k Jumlah Cluster

Jumlah cluster penyebaran lahan kritis sebanyak 2 cluster. Cluster yang dibentuk yaitu cluster tinggi (C1) dan cluster rendah (C2).

c) Menentukan Nilai Centroid (Pusat Cluster)

Penentuan pusat cluster awal ditentukan secara random yang diambil dari data yang ada dalam range. Adapun nilai untuk cluster tinggi (cluster 1) diambil dari nilai tertinggi yang terdapat pada tabel 1. dan nilai untuk cluster rendah (cluster 2) diambil dari nilai terendah yang terdapat pada tabel 1.

Tabel 2. Centroid Data Awal Iterasi 1

Cluster	Nilai
C1	530248
C2	0

d) Menghitung Jarak Setiap Data Terhadap Centroid (Pusat Cluster)

Setelah data nilai pusat cluster awal ditentukan, maka langkah selanjutnya adalah menghitung jarak masing-masing data terhadap pusat cluster. Untuk menghitung jarak setiap data penyebaran lahan kritis terhadap pusat cluster dapat menggunakan rumus yang perhitungannya dapat kita lihat sebagai berikut :

Dilakukan perhitungan jarak terhadap data penyebaran lahan kritis dengan titik pusat (centroid) pada cluster pertama. Berikut tabel 3 hasil perhitungan jarak data dengan titik pusat cluster pada iterasi 1 menggunakan *Eulidean Distance*.

e) Menentukan Posisi Cluster atau Pengelompokan

Dalam menentukan posisi cluster masing-masing data penyebaran lahan kritis berdasarkan jarak minimum data terhadap pusat cluster. Data yang memiliki jarak terkecil dengan centroid akan menjadi anggota pada kelompok

tersebut. Berikut ini merupakan tabel 3 posisi data dengan tiap *cluster* pada iterasi 1 dengan menggunakan bantuan angka 1 sebagai tanda yang mengatakan data menjadi anggota pada *cluster*

Tabel 3. Jarak Centroid Iterasi 1

C1	C2	Jarak Terdekat	C1	C2
376120	154128	154128		1
109736	420511	109736	1	
292760	237487	237487		1
354348	175900	175900		1
48371	481876	48371	1	
409509	120739	120739		1
438563	91685	91685		1
469034	61213	61213		1
501046	29202	29202		1
414125	116123	116123		1
530248	0	0		1
267301	262946	262946		1
452766	77482	77482		1
529546	702	702		1
135333	394914	135333	1	
494065	36183	36183		1
527536	2711	2711		1
511212	19036	19036		1
509373	20875	20875		1
379811	150437	150437		1
0	530248	0	1	
384727	145520	145520		1
468236	62012	62012		1
508468	21780	21780		1
461871	68377	68377		1
453922	76326	76326		1
347541	182706	182706		1
331430	198817	198817		1
337373	192874	192874		1
504081	26166	26166		1
353716	176531	176531		1
438401	91847	91847		1
484407	45841	45841		1
349988	180260	180260		1
			4	30

Selanjutnya dalam metode *K-means*, perhitungan berhenti apabila *cluster* pada iterasi yang dihasilkan sama pada iterasi sebelumnya. Maka selanjutnya mencari *cluster* pada iterasi selanjutnya sampai nilai iterasinya sama.

f) Menghitung *Centroid* Baru Menggunakan Hasil Dari Setiap Anggota Pada Masing-Masing *Cluster*

Untuk mencari nilai *centroid* selanjutnya dengan menggunakan *centroid* baru pada iterasi 1 dengan menjumlahkan nilai sesuai yang tertera pada *cluster* yang terdapat pada tabel 4.4. Adapun *centroid* baru untuk mencari *cluster* selanjutnya adalah dengan menjumlahkan nilai yang terpilih pada *cluster* tersebut kemudian membagikannya sebanyak jumlah nilai.

Contoh perhitungan titik pusat baru pada *cluster* x dan y adalah sebagai berikut :

$$C_1 = \frac{420511+481876+394914+530248}{4} = 456887$$

$$C_2 = \frac{154128+237487+175900+120739+91685+612113+29202+116123+0+262946+77482+702+36183+2711+19036+20875+150437+145520+62012+21780+68377+76326+182706+192874+26166+176531+91847+45841+180260}{30} = 100864$$

Maka, *centroid* data baru iterasi 1 adalah sebagai berikut :

Tabel 4. *Centroid* Data Baru Iterasi 1

<i>Cluster</i>	Nilai
C1	456887
C2	100864

g) Selanjutnya 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 permukiman di bantaran sungai 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.

Berikut tabel hasil *cluster* iterasi 2 :

Tabel 5. *Centroid* Data Iterasi 2

<i>Cluster</i>	Nilai
C1	456887
C2	100864

Setelah data nilai pusat *cluster* iterasi 1 didapat, maka langkah selanjutnya adalah melakukan perhitungan jarak terhadap data penyebaran lahan kritis dengan titik pusat (*centroid*) pada *cluster* pertama.

Berikut tabel 4.7 hasil perhitungan jarak data dengan titik pusat *cluster* pada iterasi 2 menggunakan *Eulidean Distance*.

Tabel 6. Jarak *Centroid* Iterasi 2

C1	C2	Jarak Terdekat	C1	C2
302760	53264	53264		1
36376	319648	36376	1	
219400	136624	136624		1
280987	75036	75036		1
24989	381013	24989	1	
336148	19875	19875		1
365203	9179	9179		1
395674	39650	39650		1
427686	71662	71662		1
340765	15259	15259		1
456887	100864	100864		1
193941	162083	162083		1
379405	23382	23382		1
456185	100162	100162		1
61973	294051	61973	1	
420704	64681	64681		1
454176	98152	98152		1
437851	81828	81828		1
436012	79989	79989		1
306450	49573	49573		1
73360	429384	73360	1	
311367	44657	44657		1
394876	38852	38852		1
435108	79084	79084		1
388510	32487	32487		1
380562	24538	24538		1
274181	81843	81843		1
258070	97954	97954		1
264013	92011	92011		1
430721	74697	74697		1
280356	75668	75668		1
365041	9017	9017		1
411046	55023	55023		1
276627	79396	79396		1
			4	30

Data yang memiliki jarak terkecil dengan *centroid* akan menjadi anggota pada kelompok tersebut. Berikut ini merupakan tabel 4.8 posisi data dengan tiap *cluster* pada iterasi 2 dengan menggunakan bantuan angka 1 sebagai tanda yang mengatakan data menjadi anggota pada *cluster*.

Perhitungan manual pada data penyebaran lahan kritis di atas didapatkan hasil akhir yang dimana pada iterasi 2 pengelompokan data yang dilakukan terhadap 2 *cluster* dengan iterasi 1 didapatkan hasil yang sama. Hasil dari kedua iterasi tersebut bernilai C1 = 4 dan C2 = 30. Sehingga posisi *cluster* pada data tersebut tidak mengalami perubahan lagi

maka proses iterasi berhenti. Berdasarkan posisi *cluster* masing-masing data penyebaran lahan kritis dan nilai *cluster* hasil iterasi 2 maka dapat disimpulkan bahwa :

- 1) *Cluster* tinggi (C1) dengan jumlah data penyebaran lahan kritis sebanyak 4 provinsi yaitu : Sumatera Utara, Jambi, Jawa Timur dan Kalimantan Tengah.
- 2) *Cluster* rendah (C2) dengan jumlah data penyebaran lahan kritis sebanyak 30 provinsi yaitu : Aceh, Sumatera Barat, Riau, Sumatera Selatan, Bengkulu, Lampung, Kepulauan Bangka Belitung, Kepulauan Riau, DKI Jakarta, Jawa Barat, Jawa Tengah, DI Yogyakarta, Banten, Bali, Nusa Tenggara Barat, Nusa Tenggara Timur, Kalimantan Barat, Kalimantan Selatan, Kalimantan Timur, Kalimantan Utara, Sulawesi Utara, Sulawesi Tengah, Sulawesi Selatan, Sulawesi Tenggara, Gorontalo, Sulawesi Barat, Maluku, Maluku Utara, Papua Barat, dan Papua.

3.2. Implementasi

Dapat dilihat dari hasil implementasi menggunakan software rapidminer didapat hasil yang sama dengan perhitungan manual yang sudah dihitung diawal seperti pada gambar 1 dibawah ini.

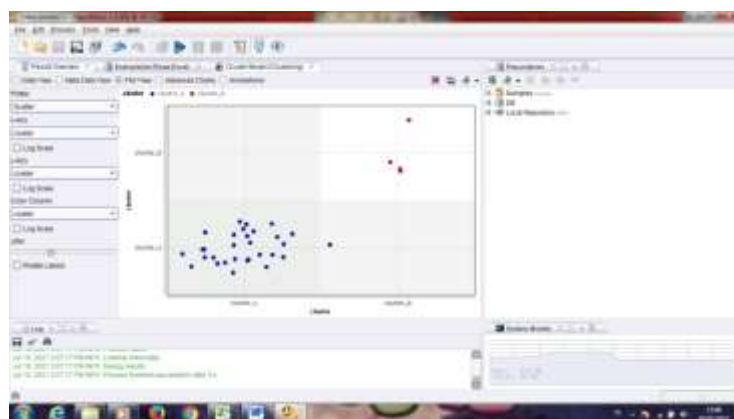


Gambar 1. Nilai *Cluster* Model

Keterangan :

- a) Jumlah *Cluster* 0 (Tinggi) berjumlah 4 *items*
- b) Jumlah *Cluster* 1 (Rendah) berjumlah 30 *items*
- c) Jumlah keseluruhannya 34 *items*

Sehingga dapat diketahui hasil pengelompokan dari *rapidminer* berikut ini dapat dilihat pada gambar 2 .



Gambar 2. Hasil Pengelompokan

Berdasarkan pada gambar 2. dapat diketahui bahwa pada kelompok tinggi memiliki sebanyak 4 *item*, sedangkan pada kelompok rendah memiliki sebanyak 30 *item*. Berisi hasil implementasi ataupun pengujian.

4. KESIMPULAN

Berdasarkan pembahasan sebelumnya dapat disimpulkan bahwa *data mining* dengan algoritma *K-means* dapat diterapkan untuk mengelompokkan banyaknya penyebaran lahan kritis di Indonesia berdasarkan provinsi. Hasil yang diperoleh dari metode *K-means* dapat diterapkan ke dalam *rapidminer* dengan nilai validasi yang sama. Sumber data yang digunakan pada penelitian ini adalah data yang diperoleh dari BPS. Jumlah data yang data uji sebanyak 34

provinsi dengan menggunakan dua *cluster*, yaitu untuk *cluster* tinggi berjumlah 4 data yaitu provinsi Sumatera Utara, Jambi, Jawa Timur dan Kalimantan Tengah. *Cluster* rendah berjumlah 30 provinsi

REFERENCES

- [1] N. Bashit, "Analisis Lahan Kritis Berdasarkan Kerapatan Tajuk Pohon Menggunakan Citra Sentinel 2," vol. 2, no. 1, pp. 32–40, 2019.
- [2] R. A. Margolang, S. R. Andani, and M. R. Lubis, "Implementasi Data Mining dalam Mengelompokkan Rumah Tangga Kumuh di Perkotaan Berdasarkan Provinsi Menggunakan," no. September, pp. 602–609, 2019.
- [3] E. G. Sihombing, "KLASIFIKASI DATA MINING PADA RUMAH TANGGA MENURUT PROVINSI DAN STATUS KEPEMILIKAN RUMAH KONTRAK / SEWA MENGGUNAKAN K-MEANS CLUSTERING METHOD," vol. 2, no. 2, pp. 74–82, 2017.
- [4] Masitha, Solikhun, D. Suhendro, I. S. Damanik, and M. Fauzan, "Implementasi K-Means Clustering Untuk Mengelompokkan Hasil Pertanian Kacang Kedelai (Ha) Berdasarkan Provinsi," vol. 2, pp. 192–199, 2020.
- [5] M. Safii, "IMPLEMENTASI DATA MINING DENGAN METODE POHON KEPUTUSAN ALGORITMA ID3 UNTUK MENENTUKAN STATUS MAHASISWA," vol. 2, no. 1, pp. 82–86, 2018.
- [6] K. Fatmawati and A. P. Windarto, "DATA MINING : PENERAPAN RAPIDMINER DENGAN K-MEANS CLUSTER PADA DAERAH TERJANGKIT DEMAM BERDARAH DENGUE (DBD) BERDASARKAN PROVINSI," vol. 3, no. 2, pp. 173–178, 2018.
- [7] N. Erlangga, Solikhun, and Irawan, "PENERAPAN DATA MINING DALAM MENGELOMPOKAN PRODUKSI JAGUNG MENURUT PROVINSI MENGGUNAKAN ALGORITMA K-MEANS," vol. 3, pp. 702–709, 2019.
- [8] J. Candra and J. Hutabarat, "PERANCANGAN APLIKASI MENENTUKAN BERAT BADAN IDEAL DENGAN MENGGUNAKAN ALGORITMA K-MEANS CLUSTERING," pp. 339–345, 2016.
- [9] J. L. S. Sinaga, Solikhun, and D. Suhendro, "Penerapan Algoritma K-Means Dalam Mengelompokkan Rata-Rata Konsumsi Kalori Menurut Provinsi," vol. 6, pp. 75–88, 2020.
- [10] M. N. V. Waworuntu and M. F. Amin, "Penerapan Metode K-Means Pemetaan Calon Penerima JAMKESDA," *Kumpul. J. Ilmu Komput.*, vol. 5, no. 2, pp. 190–200, 2018.
- [11] Suendri, "Implementasi Diagram UML (Unified Modelling Language) Pada Perancangan Sistem Informasi Remunerasi Dosen Dengan Database Oracle (Studi Kasus: UIN Sumatera Utara Medan)," vol. 6341, no. November, pp. 1–9, 2018.
- [12] A. Hendini, "Pemodelan UML Sistem Informasi Monitoring Penjualan dan Stok Barang (Studi Kasus: Distri Zhezha Pontianak)," *J. Khatulistiwa Inform.*, vol. IV, no. 2, pp. 107–116, 2016.
- [13] R. Nurmalinga, "Perencanaan dan Pengembangan Aplikasi Absensi Mahasiswa Menggunakan Smart Card Guna Pengembangan Kampus Cerdas (Studi Kasus Politeknik Negeri Tanah Laut)," vol. 9, no. 1, pp. 84–91, 2017.
- [14] Y. Yamasari, "SISTEM INFORMASI SEKOLAH BERBASIS WEB (Studi Kasus : TK Kusuma Putra Kota Mojokerto) Alyyuddin," vol. 7, pp. 94–100, 2017.
- [15] L. Pujiastuti, M. Wahyudi, and Solikhun, "Penerapan Data Mining Dalam Mengelompokkan Data Impor Tembaga Menurut Negara Asal Menggunakan," vol. 2, pp. 424–431, 2020.
- [16] M. G. Sadewo, A. P. Windarto, and S. R. Andani, "Pemanfaatan Algoritma Clustering Dalam Mengelompokkan Jumlah Desa / Kelurahan Yang Memiliki Sarana Kesehatan," vol. I, pp. 124–131, 2017.
- [17] R. W. Sari, A. Wanto, and A. P. Windarto, "Implementasi Rapidminer Dengan Metode K-Means (Study Kasus: Imunisasi Campak Pada Balita Berdasarkan Provinsi)," *KOMIK (Konferensi Nas. Teknol. Inf. dan Komputer)*, vol. 2, no. 1, pp. 224–230, 2018.