

Penerapan Algoritma K-Means untuk Pengelompokan Produksi Telur Ayam Ras Petelur di Indonesia

Dinda Giantika Utami¹, Fitriyani²

^{1,2}Program Studi Sistem Informasi, Universitas Adhirajasa Reswara Sanjaya
e-mail: ¹dindagiantikautami@gmail.com, ²fitriyani@ars.ac.id

Abstrak

Di Indonesia telur ayam sangat mudah di temukan dalam kehidupan masyarakat sehari-hari. Namun banyaknya tingkat kebutuhan konsumsi masyarakat terhadap telur ayam tidak di imbangin dengan banyaknya produksi telur ayam terutama telur ayam ras petelur yang ada di setiap provinsinya, maka dari itu harga telur ayam di beberapa provinsi terus melonjak naik. Penelitian ini bertujuan untuk mendapatkan informasi mengenai provinsi yang masih kurang dalam produksi telur ayam ras petelur menggunakan metode algoritma k-means. Algoritma k-means adalah algoritma yang dapat mengelompokkan data pada beberapa kluster sehingga data yang mempunyai kesamaan berada pada kelompok kluster yang sama dan data yang memiliki ketidaksamaan berada pada kelompok kluster lain. Sumber data penelitian ini di kumpulkan berdasarkan beberapa dokumen keterangan produksi telur ayam ras petelur yang di hasilkan oleh Badan Pusat Statistik Nasional. Data yang digunakan dalam penelitian ini adalah data dari tahun 2018-2020 yang berjumlah 102 data terdiri dari 34 provinsi. Proses kluster di bagi menjadi 3 cluster. Pada cluster 0 (cluster 1) menjadi cluster terendah dalam jumlah produksi telur ayam ras petelur yang terdiri dari 29 provinsi, cluster 1 (cluster 2) menjadi cluster sedang yang terdiri dari 4 provinsi, sedangkan pada cluster 2 (cluster 3) menjadi cluster tertinggi yang terdiri dari 1 provinsi.

Kata kunci—Telur Ayam, *Data Mining*, *Clustering*, *K-Means*

Abstract

In Indonesia, chicken eggs are very easy to find in public life the day-to-day. However the number of levels of the needs of the domestic consumption of chicken eggs is not balanced by the number of the production of chicken eggs, especially egg-laying chicken in every province, and therefore the price of chicken eggs in some provinces continue to soar up. This study aims to obtain information about the province that is still lacking in the production of egg-laying chicken using the method of k-means algorithm. K-means algorithm is an algorithm that can classify the data on some of the clusters so that data that has a similarity within the cluster groups is the same and the data that have the inequality within the cluster group to another. The source of this research data collected based on some document description production of egg-laying chicken that is generated by the national bureau of Statistics. The Data used in this research is data from the years 2018-2020 amounted to 102 data consisting of 34 provinces. The process of a cluster into 3 clusters. In cluster 0 (cluster 1) be a cluster low in the amount of production of egg-laying chicken, which consists of 29 provinces, cluster 1 (cluster 2) be the cluster is composed of 4 provinces, while in cluster 2 (cluster 3) into clusters, the highest of which consists of 1 province.

Keywords—*Chicken Eggs, Data Mining, Clustering, K-Means*

Corresponding Author:

Fitriyani

Email: fitriyani@ars.ac.id

1. PENDAHULUAN

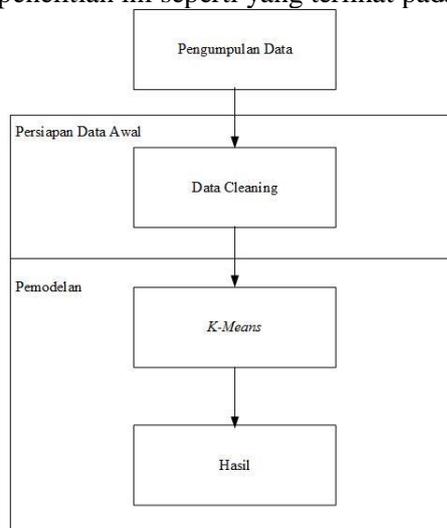
Telur yang memiliki tingkat produksi yang tinggi terdapat pada telur ayam ras petelur. Ayam ras petelur menjadi salah satu komoditi ternak yang mampu menghasilkan telur dengan gizi yang tinggi [1].

Di Indonesia permintaan telur ayam meningkat setiap tahunnya. Hal ini yang menjadikan para peternak dituntut untuk meningkatkan produksinya [2], namun dibandingkan dengan negara tetangga Indonesia seperti Malaysia dan Thailand pada tingkat konsumsi masyarakat di Indonesia masih tergolong sangat rendah. Penyebab hal tersebut adalah produksi ayam ras petelur pada peternakan yang masih belum mencapai kapasitas produksi sebenarnya sehingga antara ketersediaan dan permintaan tidak seimbang [3]. Pada data jumlah produksi telur ayam ras petelur yang di peroleh dari Badan Pusat Statistik Nasional menunjukkan bahwa masih belum meratanya produksi telur ayam ras petelur di setiap provinsi yang juga dapat menyebabkan adanya lonjakan kenaikan harga setiap tahunnya di setiap provinsi yang ada di indonesia [4]. Berdasarkan permasalahan yang ada oleh karena itu dengan penerapan datamining sebagai proses untuk mendapatkan informasi yang berguna dari penyimpanan database yang besar dan dapat dipahami sebagai ekstraksi informasi baru yang diekstraksi dari kumpulan data besar yang membantu untuk membuat suatu keputusan [5], serta untuk menentukan suatu pola ketergantungan suatu hubungan dengan memakai teknik tertentu[6].

Penelitian ini menggunakan Metode clustering untuk mengetahui cluster pengelompokan provinsi di indonesia yang berpotensi tinggi hingga rendah dalam produksi telur ayam ras petelur karena didasarkan pada penelitian terdahulu bahwa *clustering* adalah proses pembagian data yang memiliki kesamaan sifat yang di kelompokkan pada klaster yang serupa sedangkan data yang berbeda sifat dimasukan pada klaster yang lain [7]. Metode ini dapat mengelompokkan objek yang memiliki kesamaan karakteristik [8]. Algoritma yang dipakai ialah algoritma *K-Means* penggunaan algoritma ini pada dasarnya dalam proses clustering bergantung pada data yang didapatkan dan kesimpulan yang ingin dicapai di proses akhir [9]. Algoritma ini terkenal dengan kemampuan dan kemudahannya untuk mengelompokkan data yang besar dengan sangat cepat [10]. Sehingga algoritma ini sangat berguna untuk pengolahan objek dengan jumlah data yang besar yang dapat menghasilkan hasil yang lebih terukur, efektif dan efisien [7].

2. METODE PENELITIAN

Pada tahap ini, akan dijelaskan gambaran metode penelitian secara menyeluruh. Berikut adalah langkah-langkah metode penelitian ini seperti yang terlihat pada gambar 1.



Gambar 1. Tahapan Penelitian

2.1. Pengumpulan Data

Data yang digunakan pada penelitian ini adalah data *public* yaitu data jumlah produksi telur ayam ras petelur dari tahun 2018 – 2020 yang penulis dapatkan dari Badan Pusat Statistik Nasional melalui website www.bps.go.id. Data yang digunakan berjumlah 103 data. Terdiri dari atribut Nama Provinsi, Jumlah Produksi Tahun 2018, Tahun 2019, serta Tahun 2020. Pada Tabel 1 merupakan atribut keseluruhan.

Tabel 1. *Data Attribute*

No	Atribut	Ket.
1.	Provinsi	Nama setiap provinsi yang ada di Indonesia
2.	Jumlah Produksi Setiap Tahun	Jumlah produksi yang diambil setiap tahun

2.2. Persiapan Data Awal

Tahap ini adalah tahap awal untuk memastikan bahwa data produksi telur dari ayam petelur terpilih layak untuk diproses..

A. *Data Cleaning*

Selama pemrosesan data ini, pembersihan data dilakukan untuk data yang akan digunakan dalam perhitungan. Maka pada dataset ini data yang diambil berjumlah 102 data dari 103 data karena adanya data yang tidak bisa diolah datanya. Maka pada proses *cleaning* ini dilakukan penghapusan untuk menghilangkan *missing value* pada *dataset*.

2.3. Pemodelan Algoritma *K-Means*

Data yang sudah melalui tahap *cleaning* kemudian, proses menggunakan metode algoritma *k-means* untuk dikelompokkan kedalam beberapa kluster. Berikut tahapan melakukan Clustering dengan metode *K-Means* [11] :

1. Menentukan k untuk menjadi jumlah kluster yang akan di buat
2. Inisialisasi k sebagai *centroid* yang bisa dipilih dengan acak.
3. Hitung jarak setiap data terhadap masing – masing *centroid* dengan rumus jarak *Euclidean (Euclidean Distance)* [12]. Rumus untuk menentukan jarak data dari masing-masing *centroid* sebagai berikut :

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

Keterangan:

() = jarak data

X_i = data pada pusat *cluster*

Y_i = data *testing*

n = jumlah variabel data

4. Kelompokan data yang telah dihitung dengan jarak terdekat pada *centroid* nya
5. Menentukan posisi *centroid* baru (k) dengan cara menghitung nilai rata-rata dari data yang terletak pada *centroid* yang sama
6. Kembali ke langkah 3, apabila posisi *centroid* baru dengan *centroid* lama tidak sama
7. Iterasi berhenti dan dianggap stabil jika hasil iterasi baru sama dengan hasil iterasi sebelumnya.

3. HASIL DAN PEMBAHASAN

3.1. Perhitungan Manual Algoritma *K-Means*

Pada tahap ini, penetapan jumlah kluster k pada penelitian ini yaitu berjumlah 3 kluster Tabel 2 merupakan *centroid* awal.

Tabel 2. *Centroid* Awal

Centroid	Nama Provinsi	2018	2019	2020
Cluster 0	Kalimantan Utara	7.335	607	644
Cluster 1	Sumatera Utara	12.312	512.432	543.804
Cluster 2	Jawa Timur	1.320.413	1.632.492	1.732.437

Setelah menentukan *centroid* awal, kemudian hitung jarak setiap data yang ada ke *centroid* terdekat untuk menetapkan kluster dengan rumus *Euclidean distance*. Persamaan yang digunakan adalah :

$$(p, q) = \sqrt{(p1 - q1)^2 + (p2 - q2)^2 + (p3 - q3)^2 + (p4 - q4)^2 + (p5 - q5)^2} \quad (2)$$

Berikut contoh perhitungan jarak ke setiap *centroid*

➤ Data Ke-1 kluster Ke-0

$$\begin{aligned} (1,0) &= \sqrt{(57.072 - 7.335)^2 + (12.312 - 607)^2 + (13.066 - 644)^2} \\ &= \sqrt{(49.737)^2 + (11.705)^2 + (12.422)^2} \\ &= \sqrt{2.473.769.169 + 137.007.025 + 154.306.085} \\ &= \sqrt{2.765.082.279} \\ &= 52.584 \end{aligned}$$

Hasil perhitungan data ke-1 dengan kluster ke-0 maka dihasilkan jarak 52.584

➤ Data Ke-1 kluster Ke-1

$$\begin{aligned} (1,1) &= \sqrt{(57.072 - 12.312)^2 + (12.312 - 512.432)^2} \\ &\quad + (13.066 - 543.804)^2 \\ &= \sqrt{(44.760)^2 + (-500.120)^2 + (-530.738)^2} \\ &= \sqrt{2.003.457.600 + 250.120.014.400 + 281.682.824.644} \\ &= \sqrt{533.806.296.644} \\ &= 730.620 \end{aligned}$$

Hasil perhitungan data ke-1 dengan kluster ke-1 maka dihasilkan jarak 730.620

➤ Data Ke-1 kluster Ke-2

$$\begin{aligned} (1,2) &= \sqrt{(57.072 - 1.320.413)^2 + (12.312 - 1.632.492)^2} \\ &\quad + (13.066 - 1.732.437)^2 \\ &= \sqrt{(-1.263.341)^2 + (-1.620.180)^2 + (-1.719.371)^2} \\ &= \sqrt{1.596.030.482.281 + 2.624.983.232.400 + 2.956.236.635.641} \\ &= \sqrt{7.177.250.350.322} \\ &= 2.679.040 \end{aligned}$$

Hasil perhitungan data ke-1 dengan kluster ke-2 maka dihasilkan jarak 2.679.040

Kemudian lakukan perhitungan yang sama sampai data terakhir dan kelompokkan berdasarkan kluster terdekat. Berikut perhitungan lengkapnya pada tabel 3

:

Tabel 3. Hasil Perhitungan Pada Iterasi-1

Data ke-i	Jarak Ke Centroid			Jarak Terdekat	Cluster yang diikuti
	C0	C1	C2		
1	52.584	73.0620	2.679.040	5.2584	C0
2	746.332	0	2.092.490	0	C1
..
33	20.846	735.316	2.699.550	20.846	C0
34	18.383	729.492	2.706.467	18.383	C0

Selanjutnya data dikelompokkan berdasarkan jarak klaster paling dekat. Dari data yang sudah dikelompokkan akan diperoleh *centroid* baru dari hasil rata-rata setiap klaster. Kemudian hitung kembali pusat klaster yang baru berdasarkan rata-rata anggota yang ada pada klaster tersebut sepusat dari setiap klaster hingga tidak berubah lagi dan tidak ada lagi data yang berpindah dari klaster satu klaster ke klaster yang lain.

Pada perhitungan yang telah dilakukan, percobaan perhitungan berhenti pada iterasi ke-2. Berikut merupakan tabel *centroid* hasil iterasi ke-2 yang sudah tidak mengalami perubahan sehingga perhitungan berhenti pada iterasi tersebut.

Tabel 4. Hasil Perhitungan Pada Iterasi-2

Data ke-i	Jarak Ke Centroid			Jarak Terdekat	Kelompok
	C0	C1	C2		
1	52.584	73.0620	2.679.040	5.2584	C0
2	746.332	0	2.092.490	0	C1
...
33	20.846	735.316	2.699.550	20.846	C0
34	18.383	729.492	2.706.467	18.383	C0

Berikut merupakan anggota klaster 0 pada itersi-2

Tabel 5. Anggota Klaster 0 Pada Iterasi-2

Cluster 0				
No.	Nama Provinsi	Jumlah Produksi per Tahun		
		2018	2019	2020
1	Aceh	57.072	12.312	13.066
2	Riau	35.012	12.224	12.972
3	Jambi	17.500	22.702	24.092
4	Sumatera Selatan	179.467	136.807	145.182
5	Bengkulu	1.361	10.842	11.505
6	Lampung	113.298	130.824	138.834
7	Kep. Bangka Belitung	5.806	10.580	11.227
8	Kep. Riau	3.868	13.987	14.843
9	Dki Jakarta	28.904	0	0
10	Banten	218.054	204.376	216.888
11	Bali	155.161	186.311	197.717
12	Nusa Tenggara Barat	20.441	33.947	36.025
13	Nusa Tenggara Timur	18.968	9.187	9.749
14	Kalimantan Barat	80.389	115.402	122.468
15	Kalimantan Tengah	1.551	7.391	7.843
16	Kalimantan Selatan	108.111	86.777	92.089
17	Di Yogyakarta	111.613	62.845	66.693
18	Kalimantan Timur	8.592	33.346	35.388
19	Kalimantan Utara	7.335	607	644

20	Sulawesi Utara	14.267	26.588	28.216
21	Sulawesi Tengah	12.069	13.835	14.682
22	Sulawesi Selatan	140.662	194.650	206.598
23	Sulawesi Tenggara	4.578	2.832	3.005
24	Gorontalo	6.180	3.819	4.053
25	Sulawesi Barat	0	2.482	2.603
26	Maluku	5	619	657
27	Maluku Utara	0	88	94
28	Papua Barat	24.975	8.226	8.729
29	Papua	160	12.214	12.962
Jumlah Anggota : 29				

Tabel 6. Anggota Klaster 1 Pada Iterasi-2

Cluster 1				
No.	Nama Provinsi	Jumlah Produksi per Tahun		
		2018	2019	2020
1	Sumatera Barat	201.097	284.135	301.530
2	Sumatera Utara	12.312	512.432	543.804
3	Jawa Barat	802.860	468.872	497.577
4	Jawa Tengah	586.400	499.632	530.220
Jumlah Anggota : 4				

Tabel 7. Anggota Klaster 2 Pada Iterasi-2

Cluster 2				
No.	Nama Provinsi	Jumlah Produksi per Tahun		
		2018	2019	2020
1	Jawa Timur	1.320.413	1.632.492	1.732.437
Jumlah Anggota : 1				

Berikut ini merupakan tabel *centroid* baru pada iterasi ke-2 yang merupakan hasil perhitungan akhir.

Tabel 8. *Centroid* Akhir Hasil Iterasi-2

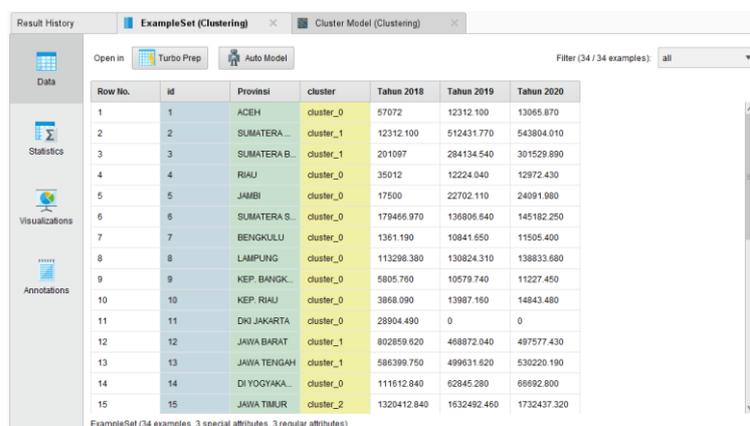
Centroid	2018	2019	2020
Cluster 0	47.428	46.752	49.615
Cluster 1	400.667	441.267	468.283
Cluster 2	1.320.413	1.632.492	1.732.437

3.2. Perhitungan Algoritma *K-Means* Dengan *Rapidminer*

Pengujian yang dilakukan menggunakan *Rapidminer* versi 9.3. Dengan langkah seperti berikut:

1. Buka *software rapidminer*
2. Klik menu *file* kemudian *new proses*
3. Masuk kedalam panel lalu ketik *read excel*, kemudian *double* klik atau *drag* pada *read excel* agar *operator* tersebut muncul di panel *process*
4. Pilih *button import configuration wizard* agar bisa memilih file *dataset* yang akan diproses di *rapidminer*.
5. Langkah selanjutnya adalah memasukan model yang akan digunakan untuk proses *clustering data mining* dengan cara masuk ke *operator*, kemudian pilih *K-Means*.
6. Kemudian pada proses *clustering* tentukan terlebih dahulu jumlah klasternya menjadi $K=3$ dan *mixed measure* menjadi *mixed Euclidean distance*

7. Langkah selanjutnya adalah sambungkan *conector* masing – masing proses pada *main process* kemudian klik *running*. Maka akan muncul hasil *clustering*. Berikut data hasil *clustering* :



Row No.	id	Provinsi	cluster	Tahun 2018	Tahun 2019	Tahun 2020
1	1	ACEH	cluster_0	57072	12312.100	13065.870
2	2	SUMATERA...	cluster_1	12312.100	512431.770	543804.010
3	3	SUMATERA B...	cluster_1	201097	284134.540	301529.890
4	4	RIAU	cluster_0	35012	12224.040	12972.430
5	5	JAMBI	cluster_0	17500	22702.110	24091.980
6	6	SUMATERA S...	cluster_0	179466.970	136806.640	145182.250
7	7	BENGGKULU	cluster_0	1361.190	10841.650	11505.400
8	8	LAMPUNG	cluster_0	113298.380	130824.310	138833.680
9	9	KEP. BANGK...	cluster_0	5805.760	10579.740	11227.450
10	10	KEP. RIAU	cluster_0	3868.090	13987.190	14843.480
11	11	DKI JAKARTA	cluster_0	28904.490	0	0
12	12	JAWA BARAT	cluster_1	802859.620	468872.040	497577.430
13	13	JAWA TENGAH	cluster_1	586399.750	499631.620	530220.190
14	14	DI YOGYAKA...	cluster_0	111612.840	62845.280	66992.800
15	15	JAWA TIMUR	cluster_2	1320412.840	1632492.460	1732437.320

Gambar 2. Hasil *Clustering*

Berdasarkan pemodelan tiga kluster maka diperoleh data yang tergabung ke kluster 0 sebanyak 29 provinsi, kluster 1 sebanyak 4 provinsi dan kluster 2 sebanyak 1 provinsi.

4. KESIMPULAN

Hasil penelitian yang telah diuji menggunakan Algoritma *K-Means Clustering*, menghasilkan beberapa kesimpulan diantaranya :

1. Data jumlah produksi telur ayam ras petelur di Indonesia dikelompokkan dengan membuat *design* pemodelan yang dibentuk menjadi 3 kluster. Didapatkan hasil pada kluster 0 terdiri dari 29 provinsi, kluster 1 sebanyak 4 provinsi dan kluster 2 sebanyak 1 provinsi. Maka dapat disimpulkan bahwa kluster tertinggi berada pada kluster 1, kluster 2 menjadi kluster sedang dan kluster 0 menjadi kluster terendah.
2. Berdasarkan perhitungan manual Algoritma *K-Means Clustering* dalam mengelompokkan produksi telur ayam ras petelur di Indonesia dan penerapan di *Rapidminer* menunjukkan hasil yang sama.
3. Hasil penelitian dapat dijadikan acuan oleh pemerintah pusat maupun provinsi untuk membuat suatu kebijakan dalam mengoptimalkan produksi telur ayam ras petelur di Indonesia, seperti: menambah jumlah peternakan ayam ras petelur di provinsi paling rendah dengan mengajak masyarakat untuk ikut andil.

DAFTAR PUSTAKA

- [1] P. O. A. Harmayanda, D. Rosyidi, and O. Sjoftan, "Evaluasi Kualitas Telur Dari Hasil Pemberian Beberapa Jenis Pakan Komersial Ayam Petelur," *J-Pal*, vol. 7, no. 1, p. 26, 2016.
- [2] F. H. Maulana, E. Prasetyo, and W. Sarenggat, "Analisis pendapatan usaha peternakan ayam petelur sumur banger farm Kecamatan Tersono Kabupaten Batang," *J. Chem. Inf. Model.*, vol. 13, no. 2, pp. 1–12, 2017.
- [3] N. Febrianto and J. aghniarahim putritamara, "Proyeksi Elastisitas Permintaan Telur Ayam Ras di Malang Raya," *J. Ilmu-Ilmu Peternak.*, vol. 27, no. 2, pp. 81–87, 2017.
- [4] "Badan Pusat Statistik." [Online]. Available: <https://www.bps.go.id/indicator/24/491/1/produksi-telur-ayam-petelur-menurut-provinsi.html>. [Accessed: 23-Oct-2021].
- [5] F. Fitriyani, "Implementasi Algoritma Fp-Growth Menggunakan Association Rule Pada Market Basket Analysis," *J. Inform.*, vol. 2, no. 1, 2016.

- [6] B. T. R. Doni, S. Susanti, and A. Mubarak, "Penerapan Data Mining Untuk Klasifikasi Penyakit Hepatocellular Carcinoma Menggunakan Algoritma Naïve Bayes," *J. Responsif Ris. Sains dan Inform.*, vol. 3, no. 1, pp. 12–19, 2021.
- [7] B. S. Praja, P. D. Kusuma, and C. Setianingsih, "Penerapan Metode K-Means Clustering Dalam Pengelompokan Data Penumpang Dan Kapal Angkutan Laut Di Indonesia," *e-Proceeding Eng.*, vol. 06, no. 1, p. 1442, 2019.
- [8] A. A. Rismayadi, N. N. Fatonah, and E. Junianto, "Algoritma K-Means Clustering Untuk Menentukan Strategi Pemasaran Di Cv. Integreet Konstruksi," *J. Responsif Ris. Sains dan Inform.*, vol. 3, no. 1, pp. 30–36, 2021.
- [9] 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.
- [10] I. S. Damanik, S. R. Andani, and D. Sehendro, "Teknik Data Mining Dalam Clustering Produksi Susu Segar Di Indonesia Dengan Algoritma K-Means," *BRAHMANA J. Penerapan Kecerdasan Buatan*, vol. 1, no. 1, pp. 31–39, 2019.
- [11] E. Irfiani and S. S. Rani, "Algoritma K-Means Clustering untuk Menentukan Nilai Gizi Balita," *J. Sist. dan Teknol. Inf.*, vol. 6, no. 4, p. 161, 2018.
- [12] F. Yunita, "Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Pada Penerimaan Mahasiswa Baru," *Sistemasi*, vol. 7, no. 3, p. 238, 2018.