

Pengembangan Datamining Klustering Pada Kasus Pencemaran Lingkungan Hidup

Mawaddah Anjelita*, Agus Perdana Windarto, Anjar Wanto, Irfan Sudahri

Prodi Sistem Informasi STIKOM Tunas Bangsa, Pematangsiantar, Indonesia
Email: mawaddah@amiktunasbangsa.ac.id, agus.perdana@amiktunasbangsa.ac.id

Abstrak—Penelitian ini bertujuan untuk melihat performa terbaik dari dua metode *datamining clustering* yaitu metode *K-Means Clustering* dengan metode *K-Medoids Clustering* menggunakan *software Rapidminer* versi 5.3. Data yang digunakan pada penelitian ini yaitu banyaknya desa/kelurahan menurut jenis pencemaran lingkungan hidup pada tahun 2018 yang terdiri dari 34 provinsi di Indonesia yang diperoleh melalui situs resmi Direktorat Badan Pusat Statistik(BPS). Variabel yang digunakan yaitu pencemaran air. Data dikelompokkan menjadi 2 *cluster* yaitu C1 dan C2. Dari hasil *software Rapidminer* model *K-Means Clustering* lebih baik performanya dibandingkan dengan *K-Medoids*. Hal ini dapat dibuktikan dengan hasil yang didapat pada *Cluster Distance Performance*. Untuk *K-Means* diperoleh -59084.803, sedangkan untuk *K-Medoids* sebesar -66471.853. Dalam hal ini, semakin kecil jarak maka semakin baik pula *clustering* terbentuk.

Kata Kunci: *K-Means, K-Medoids, Clustering, Datamining, Rapidminer.*

1. PENDAHULUAN

Lingkungan hidup merupakan segala sesuatu baik makhluk hidup, benda maupun kondisi termasuk manusia dan perilakunya yang mempunyai pengaruh dalam keberlangsungan hidup manusia maupun makhluk hidup lain. Semakin meningkat hidup, semakin meningkat pula populasi dan eksploitasi terhadap sumber daya alam guna untuk mencukupi berbagai kebutuhan manusia, hal ini tidak hanya terjadi di Indonesia, tetapi juga dibelahan dunia lainnya. Sehingga banyak permasalahan lingkungan yang timbul, diantaranya seperti pemanasan global, berkurangnya sumber daya alam, pencemaran lingkungan baik air, tanah maupun udara, dan masalah lainnya. Lingkungan yang terjaga akan menciptakan kehidupan yang sehat pula, baik dari segi kualitas maupun kuantitas. Akan lebih banyak masalah yang akan ditimbulkan oleh kerusakan lingkungan hidup jika tidak segera ditangani. Salah satu yang akan dibahas yaitu tentang pencemaran lingkungan hidup terhadap air. Air memiliki banyak manfaat, dan manusia tidak bisa hidup tanpa air. Air yang telah tercemar memiliki ciri-ciri seperti : (a) air keruh, (b) air memiliki rasa yang aneh serta berbau, (c) air yang tercemar akan meninggalkan noda, (d) terdapat endapan, (e) pH air tidak netral. Untuk mengukur sehat atau tidaknya air dapat menggunakan indikator seperti warna air, bau air, rasa air, derajat keasaman air, suhu air[1].

Saat ini ada banyak penelitian yang menggunakan cabang ilmu komputer untuk memecahkan masalah yang kompleks. Diantaranya adalah Artificial Intelligence seperti datamining [2]–[8], Sistem Pendukung Keputusan [9]–[22], sistem pakar [23], Jaringan Saraf Tiruan [24]–[27][28], [29], logika fuzzy [30] dan lain-lain.. Pada penelitian sebelumnya dengan judul Pemanfaatan Datamining Pada Pengelompokan Provinsi Terhadap Pencemaran Lingkungan Hidup yang ditulis oleh Mawaddah Anjelita. Penelitian tersebut membahas tentang pengelompokan provinsi yang ada di Indonesia menggunakan Algoritma *Datamining K-means Clustering*, data tersebut dikelompokkan menjadi 2 bagian yaitu pencemaran air tinggi dan pencemaran air rendah. Dari penelitian tersebut dapat menjadi masukan pemerintah untuk segera menanggulangi daerah-daerah yang memiliki tingkat pencemaran air tertinggi.

Perkembangan zaman membuat banyak aplikasi atau *software* yang dapat digunakan untuk mempermudah penyelesaian masalah menggunakan *artificial intelligence* atau kecerdasan buatan. Salah satu aplikasi yang digunakan yaitu *Rapidminer*. *Software* ini mempunyai banyak model algoritma terutama penyelesaian *datamining* seperti *clustering*, *id3*, *naivebayes*, *c4.5* dan lainnya. Dalam hal ini peneliti menggunakan *Software Rapidminer* versi 5.3, untuk menguji performa pada kasus yang sama dengan 2 metode yang berbeda, yaitu *K-means* dan *K-medoids*. Dimana data dengan metode yang berbeda dikelompokkan menjadi 2 *cluster*.

2. METODOLOGI PENELITIAN

Dalam penelitian ini metode yang digunakan yaitu metode data mining sebagai berikut. (a) Tahap pengumpulan data, (b) Tahap pengolahan data, (c) Tahap Clustering dan (d) Tahap Analisis[31]. Tahapan-tahapan metode penelitian dapat dijelaskan sebagai berikut :

2.1 Tahap Pengumpulan Data

Dalam penerapan datamining terhadap kasus pencemaran lingkungan, diperlukan data. Data pada penelitian ini diambil dari data yang telah dipakai pada peneliti sebelumnya dengan judul Pemanfaatan Datamining Pada Pengelompokan Provinsi Terhadap Pencemaran Lingkungan Hidup yang ditulis oleh Mawaddah Anjelita.

2.2 Tahap Pengolahan Data

Data yang telah diperoleh akan diolah terlebih dahulu. Pada tahap sebelumnya, data tersebut telah diolah menjadi 2 bagian yaitu C1 dan C2. Dengan menggunakan 2 metode yang berbeda, kemudian peneliti melihat nilai performa terhadap kedua metode tersebut dengan menggunakan *software Rapidminer*.

2.3 Tahap Clustering

Clustering merupakan suatu metode untuk mencari dan mengelompokkan data yang memiliki kemiripan karakteristik (*similarity*) antara satu data dengan data yang lain. Tujuan utama dari metode *clustering* adalah pengelompokan sejumlah data atau obyek ke dalam *cluster (group)* sehingga dalam setiap *cluster* dapat berisi data yang semirip mungkin. Dalam *clustering* metode ini berusaha untuk menempatkan obyek yang mirip (jaraknya dekat) dalam satu *cluster* dan membuat jarak antar *cluster* sejauh mungkin. Ini berarti obyek dalam satu *cluster* sangat mirip satu dengan lain dan berbeda dengan obyek dalam *cluster-cluster* yang lain[32]. Pada *K-Means Clustering*, menggunakan nilai centroid tinggi, rendah dari data yang telah diperoleh. Sedangkan pada *K-Medoids Clustering*, menggunakan nilai centroid secara acak.

2.4 Tahap Analisis

Pada tahapan ini dilakukan analisis data pencemaran lingkungan hidup menggunakan aplikasi *tools Rapid Miner*. *RapidMiner* adalah sebuah lingkungan *machine learning data mining, text mining* dan *predictive analytics*. Data yang diperoleh diolah dengan menggunakan perhitungan bobot dari tiap indeks[31]

3. ANALISA DAN PEMBAHASAN

Data dianalisis berdasarkan banyaknya desa/kelurahan menurut jenis pencemaran lingkungan hidup pada tahun 2018, yang diperoleh dari Badan Pusat Statistik Nasional melalui situs <https://www.bps.go.id>. Hasil perhitungan berdasarkan variabel pencemaran air yang tertera pada tabel 1.

Tabel 1. Data Pencemaran air tahun 2018

No.	Provinsi	Pencemaran Air
1	Aceh	729
2	Sumatera Utara	1205
3	Sumatera Barat	319
4	Riau	454
5	Jambi	614
6	Sumatera Selatan	673
7	Bengkulu	286
8	Lampung	572
9	Kep. Bangka Belitung	159
10	Kep. Riau	55
11	DKI Jakarta	126
12	Jawa Barat	1890
13	Jawa Tengah	1900
14	DI Yogyakarta	99
15	Jawa Timur	1643
16	Banten	513
17	Bali	130
18	Nusa Tenggara Barat	282
19	Nusa Tenggara Timur	122
20	Kalimantan Barat	915
21	Kalimantan Tengah	782
22	Kalimantan Selatan	714
23	Kalimantan Timur	318
24	Kalimantan Utara	139
25	Sulawesi Utara	327
26	Sulawesi Tengah	303
27	Sulawesi Selatan	400
28	Sulawesi Tenggara	227
29	Gorontalo	111
30	Sulawesi Barat	115
31	Maluku	105
32	Maluku Utara	216
33	Papua Barat	155
34	Papua	249

3.1 Input Data

Data yang telah diperoleh, kemudian diolah menggunakan *software Rapidminer*[33]. Data diinput dalam bentuk excel, kemudian masukan model data, lalu ketikkan *Cluster Distance Performance* untuk menghitung performa dari model yang telah diinputkan. Untuk lebih jelas dapat dilihat sebagai berikut :



Gambar 1. Input data *K-means Clustering* pada *Rapidminer*



Gambar 2. Input data *K-medoids Clustering* pada *Rapidminer*

3.2 Ouput Data

Setelah data diinput, maka hasil yang didapat dari *software Rapidminer* dapat dilihat sebagai berikut :



Gambar 3. Output data *K-means Clustering* pada *Rapidminer*



Gambar 4. Grafik *K-means Clustering*



Gambar 5. Output data *K-medoids Clustering* pada *Rapidminer*

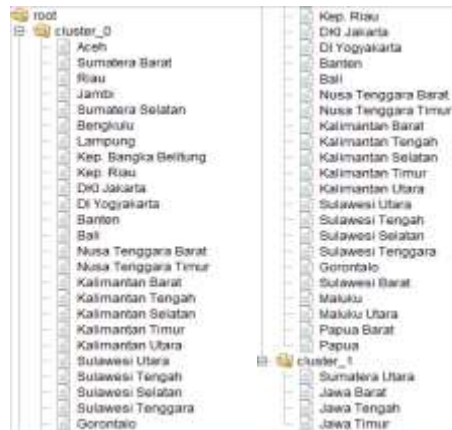


Gambar 6. Grafik *K-medoids Clustering*

Dapat dilihat perbedaan gambar 3 dan gambar 5, hasil *Cluster Model* menggunakan *tools Rapidminer* dengan menggunakan data kasus yang sama, tetapi dengan metode yang berbeda. Pembacaan *cluster* pada *Rapidminer* dimulai dari *cluster 0*(nol), sementara pada teori *clustering* dimulai dari *cluster 1*. Sehingga untuk *cluster_0* pada *Rapidminer* sama dengan *cluster 1*(C1) pada perhitungan secara teori. Hasil yang diperoleh menggunakan *K-Means* yaitu 4 provinsi untuk C1, dan 30 provinsi untuk C2. Sementara pada *K-Medoids*, hasil yang diperoleh C1 ada 30 provinsi dan C2 4 provinsi. Untuk lebih memperjelas item-item apa saja yang terdapat pada *cluster model* tersebut, dapat dilihat sebagai berikut :

cluster_0	cluster_1
Sumatera Utara	Sumatera Utara
Jawa Barat	Jawa Barat
Jawa Tengah	Jawa Tengah
Jawa Timur	Jawa Timur
	Aceh
	Sumatera Barat
	Riau
	Jambi
	Sumatera Selatan
	Bengkulu
	Lampung
	Kep. Bangka Belitung
	Kep. Riau
	DKI Jakarta
	DI Yogyakarta
	Banten
	Riau
	Nusa Tenggara Barat
	Nusa Tenggara Timur
	Kalimantan Barat
	Kalimantan Selatan
	Kalimantan Tengah
	Kalimantan Selatan
	Kalimantan Timur
	Kalimantan Utara
	Sulawesi Utara
	Sulawesi Tengah
	Sulawesi Selatan
	Sulawesi Tenggara
	Sulawesi Tenggara
	Sulawesi Barat
	Maluku
	Maluku Utara
	Papua Barat
	Papua

Gambar 7. Data item pada model *K-means Clustering*



Gambar 8. Data item pada model *K-medoids Clustering*

Pada gambar 7 dan 8 dapat dilihat perbedaan item-item yang ada pada setiap *cluster*, baik item pada model *K-Means* maupun item pada model *K-Medoids*. Pada *K-Means*, item pada C1 yaitu Sumatera Utara, Jawa Barat, Jawa Tengah, Jawa Timur, untuk C2 ada 30 provinsi. Sementara pada *K-Medoids*, item C1 ada 30 provinsi dan untuk C2 ada 4 provinsi. Untuk mengetahui model performa terbaik pada kasus yang sama, yaitu kasus pencemaran lingkungan hidup, dapat dilakukan pengujian *performance* pada *tools Rapidminer*, dengan menggunakan *Cluster Distance Performance*. Sehingga didapat hasil sebagai berikut :

Avg. within centroid distance

Avg. within centroid distance: -59084.803

Gambar 9. Performance *K-means Clustering*

Avg. within centroid distance

Avg. within centroid distance: -66471.853

Gambar 10. Performance *K-medoids Clustering*

Avg. within centroid distance adalah rata-rata dalam jarak *cluster* dihitung dengan rata-rata jarak antara *centroid* dan semua contoh *cluster*. Semakin kecil jarak semakin baik karya *clustering*. *Avg. within centroid distance* pada *K-Means* yang diperoleh: -59084.803 , sedangkan untuk *K-Medoids* diperoleh: -66471.853.

4. KESIMPULAN

Berdasarkan penelitian ini didapat kesimpulan sebagai berikut:

- Telah berhasil dilakukan pengelompokan data banyaknya desa/kelurahan menurut jenis pencemaran lingkungan hidup pada tahun 2018, menggunakan algoritma *K-Means Clustering* menjadi 2 *cluster*, yaitu untuk *cluster 1* berjumlah 4 provinsi, dan 30 provinsi lainnya berada di *cluster 2*.
- Pengelompokan data banyaknya desa/kelurahan menurut jenis pencemaran lingkungan hidup pada tahun 2018, menggunakan algoritma *K-Medoids Clustering* menjadi 2 *cluster*, yaitu untuk *cluster 1* berjumlah 30 provinsi, dan 4 provinsi lainnya berada di *cluster 2*.
- Berdasarkan penelitian ini model *K-Means* lebih baik performanya dibandingkan dengan *K-Medoids*. Hal ini dapat dibuktikan dengan hasil yang didapat pada *Cluster Distance Performance*. Untuk *K-Means* diperoleh -59084.803, sedangkan untuk *K-Medoids* sebesar -66471.853. Dalam hal ini, semakin kecil jarak maka semakin baik pula *clustering* terbentuk.
- Penelitian ini dapat menjadi acuan bagi pemerintah terhadap daerah-daerah yang memiliki tingkat pencemaran air yang tinggi, yaitu daerah Sumatera Utara, Jawa Barat, Jawa Tengah, dan Jawa Timur.

REFERENCES

- M. Anjelita, A. P. Windarto, D. Hartama, and S. Informasi, "Pemanfaatan datamining pada pengelompokan provinsi terhadap pencemaran lingkungan hidup," vol. 3, pp. 659–666, 2019, doi: 10.30865/komik.v3i1.1675.
- 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," *J. Phys. Conf. Ser.*, vol. 1255, no. 12005, pp. 1–7, 2019, doi: 10.1088/1742-6596/1255/1/012005.
- 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," *J. Phys. Conf. Ser.*, vol. 1255, no. 12002, pp. 1–6, 2019, doi: 10.1088/1742-6596/1255/1/012002.
- 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 Conf. Ser. Mater. Sci. Eng.*, vol. 420, no. 1, 2018, doi: 10.1088/1757-

- 899X/420/1/012089.
- [5] 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, doi: 10.30865/komik.v2i1.930.
- [6] N. Rofiqo, A. P. Windarto, and D. Hartama, "Penerapan Clustering Pada Penduduk Yang Mempunyai Keluhan Kesehatan Dengan Datamining K-Means," *KOMIK (Konferensi Nas. Teknol. Inf. dan Komputer)*, vol. 2, no. 1, pp. 216–223, 2018, doi: 10.30865/komik.v2i1.929.
- [7] M. G. Sadewo, A. P. Windarto, and A. Wanto, "Penerapan Algoritma Clustering Dalam Mengelompokkan Banyaknya Desa/Kelurahan Menurut Upaya Antisipasi/ Mitigasi Bencana Alam Menurut Provinsi Dengan K-Means," *KOMIK (Konferensi Nas. Teknol. Inf. dan Komputer)*, vol. 2, no. 1, pp. 311–319, 2018, doi: 10.30865/komik.v2i1.943.
- [8] D. Hartama, A. Perdana Windarto, and A. Wanto, "The Application of Data Mining in Determining Patterns of Interest of High School Graduates," *J. Phys. Conf. Ser.*, vol. 1339, no. 1, 2019, doi: 10.1088/1742-6596/1339/1/012042.
- [9] 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," *J. Phys. Conf. Ser.*, vol. 1255, no. 1, 2019, doi: 10.1088/1742-6596/1255/1/012007.
- [10] R. W. Sari, A. P. Windarto, S. P. Keputusan, P. Kreatifitas, M. Pkm, and A. D. A. N. Pembahasan, "Penerapan Electree Pada Seleksi Proposal Program Kreativitas Mahasiswa (PKM) di STIKOM Tunas Bangsa," in *Seminar Nasional Teknologi Komputer & Sains (SAINTEKS) SAINTEKS 2019*, 2019, pp. 800–806.
- [11] F. Syahputra, M. Mesran, I. Lubis, and A. P. Windarto, "Sistem Pendukung Keputusan Pemilihan Guru Berprestasi Kota Medan Menerapkan Metode Preferences Selection Index (Studi Kasus : Dinas Pendidikan Kota Medan)," *KOMIK (Konferensi Nas. Teknol. Inf. dan Komputer)*, vol. 2, no. 1, pp. 147–155, 2018, doi: 10.30865/komik.v2i1.921.
- [12] P. P. P. A. N. W. 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," *J. Phys. Conf. Ser.*, vol. 1255, no. 1, 2019, doi: 10.1088/1742-6596/1255/1/012011.
- [13] P. P. P. A. N. W. F. I. R. H. Zer, D. Hartama, and A. P. Windarto, "Analisis Komparasi Metode AHP dan TOPSIS dalam Pemilihan Asuransi Kategori Kesehatan Terbaik PT . Prudential," in *Seminar Nasional Sains & Teknologi Informasi (SENSASI) SENSASI 2019*, 2018, pp. 427–432.
- [14] M. Widyastuti, F. R. S. Samosir, A. P. Windarto, and D. Hartama, "Implementasi Metode Promethee Dalam Pemilihan Kenaikan Jabatan Sous Chef Menjadi Chef," *Teknol. Komput. Sains*, vol. 1, no. 1, pp. 807–812, 2019.
- [15] 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," *J. Phys. Conf. Ser.*, vol. 1255, no. 1, pp. 1–8, 2019, doi: 10.1088/1742-6596/1255/1/012006.
- [16] 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," *J. Phys. Conf. Ser.*, vol. 1255, no. 1, 2019, doi: 10.1088/1742-6596/1255/1/012010.
- [17] D. N. Batubara, D. R. Sitorus P, and A. P. Windarto, "Penerapan Metode PROMETHEE II Pada Pemilihan Situs Travel Berdasarkan Konsumen," *J. Sisfokom (Sistem Inf. dan Komputer)*, vol. 8, no. 1, pp. 46–52, 2019, doi: 10.32736/sisfokom.v8i1.598.
- [18] K. Fatmawati *et al.*, "Analysis of Promethee II Method in the Selection of the Best Formula for Infants under Three Years," *J. Phys. Conf. Ser.*, vol. 1255, no. 1, 2019, doi: 10.1088/1742-6596/1255/1/012009.
- [19] T. Imandasari, A. P. Windarto, and D. Hartama, "Seminar Nasional Teknologi Komputer & Sains (SAINTEKS) Analisis Metode MAUT Pada Pemilihan Deodorant," in *Seminar Nasional Teknologi Komputer & Sains (SAINTEKS) SAINTEKS 2019*, 2019, pp. 736–739.
- [20] K. F. Irnanda, F. N. Arifah, M. R. Raharjo, A. Arifin, and A. P. Windarto, "The selection of Calcium Milk Products that are appropriate for advanced age using PROMETHEE II Algorithm," *J. Phys. Conf. Ser.*, vol. 1381, no. 1, 2019, doi: 10.1088/1742-6596/1381/1/012070.
- [21] 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," *J. Phys. Conf. Ser.*, vol. 1255, no. 12008, pp. 1–7, 2019, doi: 10.1088/1742-6596/1255/1/012008.
- [22] T. Imandasari and A. P. Windarto, "Penerapan Metode VIKOR Pada Pemilihan Popok Bayi Berdasarkan Jenis Kulit," *Semin. Nas. Sains Teknol. Inf.*, pp. 215–220, 2018.
- [23] Hamdani, "Sistem Pakar Untuk Diagnosa Penyakit Mata Pada Manusia," vol. 5, no. 2, pp. 13–21, 2010.
- [24] A. P. Windarto, M. R. Lubis, and Solikhun, "IMPLEMENTASI JST PADA PREDIKSI TOTAL LABA RUGI KOMPRESIF BANK UMUM KONVENSIONAL DENGAN BACKPROPAGATION," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 5, no. 4, pp. 411–418, 2018, doi: 10.25126/jtiik.201854767.
- [25] A. P. Windarto, M. R. Lubis, and Solikhun, "Model Arsitektur Neural Network Dengan Backpropogation Pada Prediksi Total Laba Rugi Komprehensif Bank Umum Konvensional," *Kumpul. J. Ilmu Komput.*, vol. 5, no. 2, pp. 147–158, 2018.
- [26] A. P. Windarto, L. S. Dewi, and D. Hartama, "Implementation of Artificial Intelligence in Predicting the Value of Indonesian Oil and Gas Exports With BP Algorithm," *Int. J. Recent Trends Eng. Res.*, vol. 3, no. 10, pp. 1–12, 2017, doi: 10.23883/IJRTER.2017.3482.J5BBS.
- [27] Sumijan, A. P. Windarto, A. Muhammad, and Budiharjo, "Implementation of Neural Networks in Predicting the Understanding Level of Students Subject," *Int. J. Softw. Eng. Its Appl.*, vol. 10, no. 10, pp. 189–204, 2016.
- [28] Budiharjo, T. Soemartono, A. P. Windarto, and T. Herawan, "Predicting tuition fee payment problem using backpropagation neural network model," *Int. J. Adv. Sci. Technol.*, vol. 120, pp. 85–96, 2018, doi: 10.14257/ijast.2018.120.07.
- [29] Budiharjo, T. Soemartono, A. P. Windarto, and T. Herawan, "Predicting School Participation in Indonesia using Back-Propagation Algorithm Model," *Int. J. Control Autom.*, vol. 11, no. 11, pp. 57–68, 2018.
- [30] C. C. Lee, "Fuzzy Logic in Control Systems : Fuzzy," no. 2, 1990.
- [31] A. P. Windarto, P. Studi, S. Informasi, and D. Mining, "Penerapan Data Mining Pada Ekspor Buah-Buahan Menurut Negara Tujuan Menggunakan K-Means Clustering," vol. 16, no. 4, pp. 348–357, 2017.
- [32] H. Sulatri and acep irham Gufroni, "PENERAPAN DATA MINING DALAM PENGELOMPOKAN PENDERITA THALASSAEMIA," *J. Nas. Teknol. dan Sist. Inf.*, vol. 2, pp. 299–305, 2017.
- [33] M. Anjelita, A. P. Windarto, and A. Wanto, "Analisis Metode K-Means pada Kasus Ekspor Barang Perhiasan dan Barang Berharga Berdasarkan Negara Tujuan," pp. 476–482, 2019.