

Penerapan Clustering Menggunakan Algoritma K-Means Sebagai Analisis Produksi Komoditas Perikanan Provinsi di Indonesia

Akbar Maulana¹, Khahari Nur Akbar², Nurahman³

^{1 2 3}Program Studi : Sistem Informasi, Universitas Darwan Ali

Email : akbarquipper68@gmail.com¹, nurkahariakbar@gmail.com², nurrahman.ikhtiar@gmail.com³

ABSTRACT— In recent years, aquaculture commodities in Indonesia have experienced a slowdown in production growth and tend to stagnate. Therefore, current research only focuses on the accumulation factor of aquaculture production in units (tons), therefore the author makes a journal concerning the percentage of the total (tons) in various types of fish commodities in all provinces. Of the 310 fishery production data, the main commodities from 34 provinces of Indonesia and 10 main fishery commodities are contained in the database from the BPS (National Statistics Center). Then the data is analyzed by implementing the k-means method clustering through the RapidMiner Studio software. The results of the analysis showed that the inter-provincial fishery commodity production was grouped into 3 clusters, namely the largest, medium, and smallest. So it was found that the results of the analysis showed the frequency level of each province related to its aquaculture commodities so that from that frequency it could be a benchmark in the program to generate superior potential in terms of aquaculture in each province.

Keywords— Data mining, Cultivation, Clustering, K-means

ABSTRAK—Komoditas budidaya perikanan di Indonesia pada beberapa tahun ini mengalami perlambatan pertumbuhan produksi dan cenderung mengalami stagnasi. Maka dari itu penelitian pada saat ini hanya berfokus pada faktor akumulasi jumlah produksi budidaya dalam satuan (ton), maka dari itu penulis membuat jurnal dengan acuan persentase jumlah (ton) di berbagai macam komoditas jenis ikan pada semua provinsi. Dari 310 data produksi perikanan komoditas utama dari 34 provinsi Indonesia dan 10 komoditas utama perikanan yang terdapat pada basis data dari BPS (Badan Pusat Statistik Nasional). Kemudian dari data tersebut dilakukan analisis dengan mengimplementasikan clustering metode k-means melalui alat (software) RapidMiner Studio. Hasil analisis didapatkan bahwa produksi komoditas perikanan antar provinsi dikelompokkan menjadi 3 buah kelompok(cluster) yaitu dari yang terbesar, sedang, dan terkecil. Sehingga didapatkan bahwa hasil analisis yang menunjukkan tingkat frekuensi dari setiap provinsi terkait dengan komoditas budidaya perikananannya sehingga dari frekuensi tersebut dapat menjadi tolak ukur dalam program membangkitkan potensi unggulan dari segi budidaya perikanan pada setiap provinsi.

Kata kunci— Data mining, Budidaya, Clustering, K-means

I. PENDAHULUAN

Persoalan terkait dengan tingkat produksi pembudidayaan perikanan di setiap provinsi di Indonesia, khususnya yang menjadi komoditas utama seperti gurame, lele, patin, nila, ikan mas, kakap, bandeng, kerapu, udang dan rumput laut. Saat ini bisa tercermin dengan jelas bahwa tingkat fluktuatif kebutuhan pasar akan komoditas tersebut berbanding lurus dengan tingkat produksi yang bisa dihasilkan pada suatu daerah. Pada beberapa tahun terakhir sering terdengar informasi tentang bagaimana suatu komoditas perikanan tertentu melonjak naik di beberapa daerah di Indonesia, hal ini sangat disayangkan jika bisa melihat potensi yang bisa dianalisis dan dikembangkan dari setiap daerah tersebut dalam lini budidaya perikananannya.

Analisa yang dilakukan pada permasalahan terhadap pembudidayaan perikanan yang terdapat dalam 34 provinsi di Indonesia, diperlukan perhitungan pada

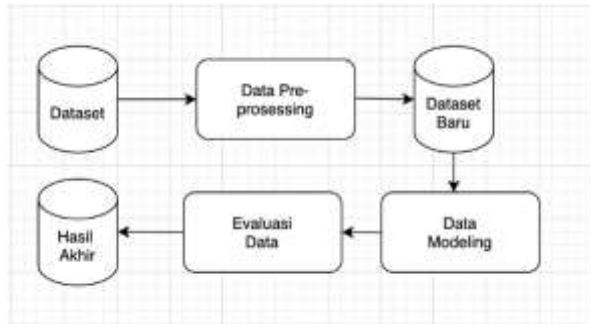
jumlah produksi pada setiap daerah dan dibagi dalam akumulasi total dari setiap komoditas yang bisa dihasilkan.

Saat ini perkembangan kecerdasan manusia telah merambah ke dalam berbagai lini, salah satunya pengembangan teknologi sebagai problem solving akan masalah yang dihadapi. Data mining menjadi penyelesai masalah yang dapat menangani data-data dengan melakukan pemrosesan pada sumber data menggunakan berbagai teknik dan metode. Dalam tahapan penyelesaian suatu permasalahan terdapat beberapa pilihan yang digunakan dan disesuaikan dengan konteks yang sedang dianalisis.

Analisis yang digunakan dalam permasalahan ini mengolah data dan memprosesnya sehingga menghasilkan kelompok-kelompok (cluster) dari kemiripan yang bisa menjadi ciri khas dari setiap data dengan metode k-means. Oleh sebab itu penggunaan

metode di atas sehingga dari penelitian ini, dapat mengklusterisasikan wilayah atau provinsi dengan frekuensi tertinggi dalam nilai akumulasi produksi budidaya perikanan.

II. METODE PENELITIAN



Gambar 1. Metode Penelitian

Dari serangkaian proses yang dilakukan untuk mengolah data tersebut diperlukan lagi beberapa proses untuk pengolahan data agar menjadi sebuah hasil yang mengacu pada produksi komoditas budidaya perikanan antar provinsi.

A. Dataset

Dataset yang penulis pakai yaitu data dari Badan Pusat Statistik yang mendata pada tahun 2020 dengan jumlah data 34 provinsi dan label jenis komoditas dengan total 340 value, dataset tersebut sudah dalam tahap penyelesaian. Hasil yang didapat berkemungkinan bisa dipakai untuk rujukan dalam membandingkan proses evaluasi data.

Gambar 2. Dataset Mentah

B. Pre-Processing

Pada gambar 3. Memperllihatkan tahapan pre processing yaitu pemilihan antara atribut yang tidak diperlukan dalam penelitian, dari 35 label pada atribut provinsi yang diperlukan menjadi 34 label, dikarenakan label INDONESIA itu merupakan total dari semua label. Oleh karena itu label yang digunakan berjumlah 34 pada atribut provinsi karena label tersebut tidak mempunyai pengaruh pada proses evaluasi data.

Beserta atribut yang mempunyai value kosong dilakukan pengolahan kembali dengan menggantikan value tersebut dengan 0, karena proses komputasi akan melibatkan setiap value di dalam atribut. Olehkarena itu setiap value yang missing/kosong bisa digantikan dengan angka 0.

Gambar 3. Data Set Sesudah Pre-Processing

C. K-Means

K-Means adalah sebuah metode pemeriksaan pada data mining yang dimana metode ini melakukan proses pemodelan unsupervised yang salah satu metode yang melakukan pengelompokan berdasarkan kecocokan / kemiripan. data yang berkarakter sama akan dikelompokkan pada suatu cluster sedangkan data dengan karakter berbeda akan dimasukan ke kelompok Cluster lain Menurut McQueen JB, langkah-langkah dari algoritma K-Means :

1. Memilih banyak k-cluster yang akan dibentuk.
2. Memunculkan nilai acak untuk pusat awal cluster (centroid) sebanyak k-cluster.
3. Mengukur jarak pada data masukan untuk masing-masing centroid yang memakai rumus jarak (Euclidean Distance) sehingga didapat jarak yang paling dekat dari tiap data dengan centroid. Berikut persamaan Euclidean Distance:

$$d(xi, \mu_i) = \sqrt{(xi - \mu_i)}$$

keterangan:

d = distance atau jarak dari centroid

xi,ui = sebagai titik data antara x dan i

4. Mengkategorikan tiap data dari seberapa dekat dengan centroid (jarak terkecil).
5. Memperbaharui nilai centroid. Nilai centroid baru didapat dari rata-rata cluster yang bertaut dengan memakai rumus :

$$C_k = \frac{1}{nk} \sum d_i$$

keterangan:

n_k = jumlah data dalam cluster

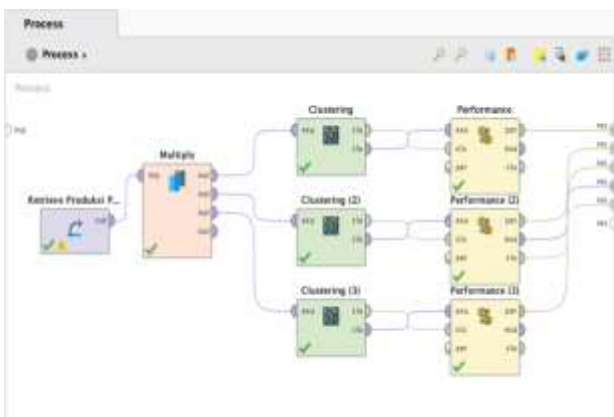
d_i = jumlah dari nilai jarak yang masuk dalam masing-masing cluster

6. Melakukan looping pada langkah 2 hingga 5 hingga anggota tiap cluster tidak ada yang berubah.
7. Jika langkah 6 telah terpenuhi, maka nilai rata-rata pusat cluster (μ_j) pada iterasi terakhir akan digunakan sebagai parameter untuk menentukan klasifikasi data.

III. HASIL DAN PEMBAHASAN

Rapid miner merupakan sebuah tools/perangkat lunak yang berfungsi sebagai alat untuk pengolahan data. tools ini menggunakan esensi serta pengimplementasian algoritma data mining, RapidMiner menganalisa secara otomatis dataset yang kita berikan lalu memberikan solusi dalam penganalisaan data tersebut. Selain itu RapidMiner juga memudahkan user dalam melakukan perkiraan dalam data yang banyak dengan menggunakan operator-operator yang dibutuhkan.

Permodelan dari setiap kasus dapat dilakukan simulasi menggunakan Rapid Miner, dalam komputasi tingkat lanjut dan juga data yang kompleks. Penelitian ini menjabarkan setiap langkah dengan beberapa pertimbangan. Penentuan untuk optimasi terbaik dan juga penyelesaian masalah yang paling efektif dituangkan kedalam konfigurasi Rapid Miner.



Gambar 4. Konfigurasi RapidMiner

Operator tersebutlah yang berperan dalam transformasi data. Data ini dihubungkan dengan node-node pada operator dan kemudian tinggal menghubungkan ke node Final untuk memperlihatkan hasil akhir nya. dari hasil ini dapat divisualisasikan dengan berbagai grafik yang mudah dipahami. karena itulah Tools RapidMiner ini bisa menjadi salah satu alternatif untuk proses ekstrak data dengan teknik data mining penulis memakai konfigurasi data dan operator.

Pada data yang sebelumnya sudah dilakukan pemrosesan sebelum dimasukkan ke dalam konfigurasi, sudah tidak terdapat atribut dan label yang tidak perlu

pada data tersebut. Setelah itu dilakukan proses multiply terhadap data yang mana tujuannya agar dapat menggunakan beberapa operator (k-means) dengan satu buah data. Setelah itu sebelum data diputuskan untuk dilakukan cluster terlebih dahulu mengecek untuk melihat potensi terbaik dari setiap jumlah kelompok atau cluster untuk diputuskan, disini untuk pemrosesan pertama diputuskan untuk 2 buah cluster, selanjutnya 3 buah cluster dan terakhir 4 buah cluster.

Pemrosesan terakhir yang ditetapkan dari 3 buah operator dengan setiap jumlah cluster nya dilakukan analisis dengan menetapkan pemrosesan segmentasi melalui cluster distance performance. Tujuannya sebagai penetapan berapa jumlah cluster terbaik yang perlu diputuskan dalam permasalahan ini dengan melihat nilai minimum paling kecil dari Davies Bouldin pada performance vektor.

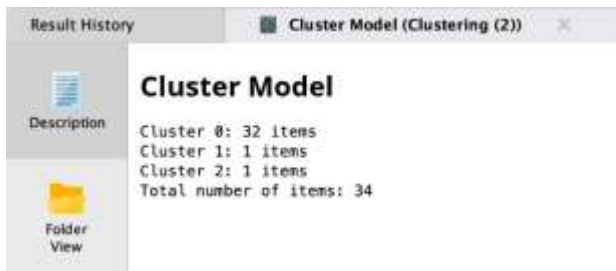


Gambar 5. Hasil analisis performa K-Means

Komputasi yang dilakukan sebelumnya terdapat 3 model performance, pada setiap model tersebut di tujukkan untuk mengetahui nilai berapa yang menjadi pilihan terbaik. Bisa diketahui ketika melihat hasil dengan membandingkan nilai minimum pada setiap cluster. Sehingga akan menghasilkan nilai Davies Bouldin yang berbeda-beda, pengambilan keputusan bisa dilanjutkan setelah menentukan pilihan terbaik dari cluster yang akan digunakan.

Dilakukan pemrosesan untuk melihat performa terbaik berdasarkan dengan jumlah cluster yang dipilih. Menghasilkan nilai untuk 2 buah cluster sebesar -0.417 lalu 4 cluster sebesar -0.248 dan paling kecil dari operator k-means dengan 3 buah kelompok(cluster) dengan nilai Davies Bouldin sebesar -0.113. Sehingga penentuan cluster terbaik bisa di dapat bahwa penggunaan 3 buah cluster menjadi keputusan dalam kasus ini.

Tahapan ini bisa diketahui bahwa dari hasil komputasi menghasilkan beberapa cluster model, diantaranya nya untuk cluster 0 sebanyak 32 item, cluster 1 sebanyak 1 item dan cluster 2 sebanyak 1 item. Sehingga total cluster model ialah sebanyak 3 buah cluster model. Lalu dari total 34 item tersebut akan dilihat setiap rincian yang sudah terklasifikasi dalam cluster model.



Gambar 6. Cluster Model

Pada aplikasi Rapid Miner menampilkan hasil yang lebih mendetail dan spesifik terkait dengan setiap cluster, dalam hal ini komputasi yang telah ditentukan menghasilkan sejumlah data yang sudah diolah. Pada hasil diatas juga memberikan nilai 0 yang menjadi nilai eliminasi terhadap value dari atribut yang hilang atau missing.

Pada tahapan ini cluster pertama menjadi dominan karena pada cluster model dapat diketahui bahwa terdapat 34 item yang merupakan himpunan dari cluster, lalu terdapat juga atribut yang menyediakan data terkait. Dengan begitu penjabaran lebih lanjut untuk melihat titik dari setiap centroid di tahapan selanjutnya bisa dilakukan berdasarkan data yang tersedia pada atribut terkait.

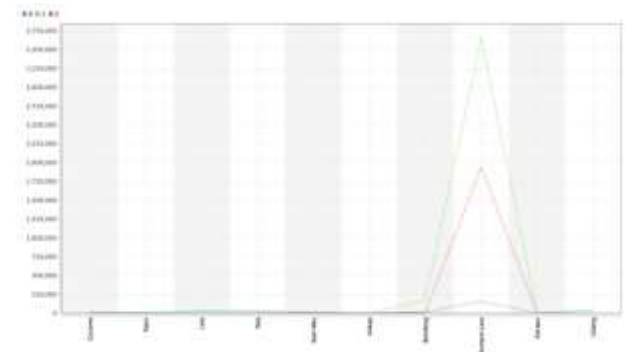
Dengan konfigurasi operator tersebut dapat diketahui bahwa dari jumlah 34 provinsi dan label jenis komoditas dengan total 340 value. Maka didapatkan hasil klasterisasi pada data tersebut sebanyak 3 cluster yang mana pada cluster pertama terdapat sebanyak 32 kasus dan pada cluster selanjutnya masing-masing terdapat 1 buah kasus. Dari setiap kasus yang sudah dikelompokkan pada cluster tersebut bisa dilihat bagaimana hubungannya terhadap setiap titik centroid.

Row No.	ID	Provinsi	Cluster	Gurame	Patin	Lele	Nila	Ikan Mas	Kakap	Bandeng
1	1	ACEH	Cluster 0	518	143	11141	28018	8241	148	6108
2	2	SUMATERA	Cluster 0	4986	1307	14536	51186	58147	886	0
3	3	SUMATERA	Cluster 0	21772	7761	31761	118081	75811	0	0
4	4	BALI	Cluster 0	487	21080	44161	13816	1461	11	58
5	5	JAWA	Cluster 0	591	11811	6148	18111	458	0	411
6	6	SUMATERA	Cluster 0	1188	121881	11848	18118	8114	0	111
7	7	BENGKULU	Cluster 0	10	114	18117	18114	1811	0	111
8	8	JAWA	Cluster 0	11847	7841	41111	18111	8111	0	1188
9	9	KOTA BANGKAL	Cluster 0	8	111	111	111	0	11	0
10	10	KOTA BANGKAL	Cluster 0	148	111	18111	118	1811	1111	0
11	11	DIYAKART	Cluster 0	111	11	111	111	0	11	1111
12	12	JAWA BARAT	Cluster 0	18111	11111	118111	118111	11811	11	18111
13	13	JAWA BARAT	Cluster 0	118111	11111	118111	118111	11811	0	118111
14	14	DIYAKART	Cluster 0	11811	1111	11811	11811	0	0	0
15	15	JAWA BARAT	Cluster 0	118111	11811	118111	11811	11811	11811	118111
16	16	BANTEN	Cluster 0	148	111	11811	11811	11811	11811	118111
17	17	BALI	Cluster 0	0	0	11811	11811	11811	11811	11811

Gambar 7. Rincian Cluster

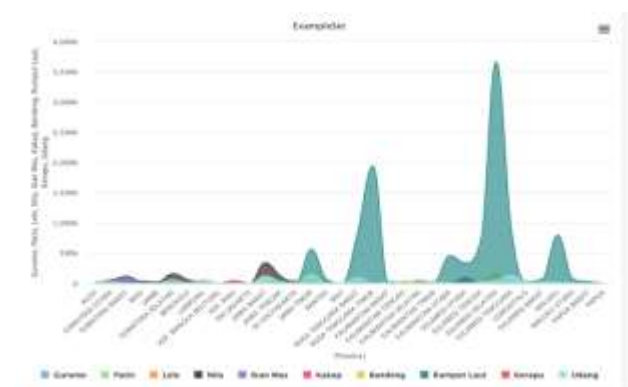
Analisa yang dilakukan dengan konfigurasi sebelumnya, dapat diketahui jumlah yang dihasilkan dalam satu tahun periode yang menjadi titik data dari

berbagai komoditas utama pembudidayaan perikanan. Untuk itu dari total sebanyak 3.750.000 berada pada centroid rumput laut, yang menjadikan atribut komoditas rumput laut sebagai tingkat puncak dalam plot hasil pada RapidMiner.



Gambar 8. Hubungan Titik Centroid Tiap Cluster

Pada analisis lanjutan yang dilakukan bukan hanya menentukan komoditas utama (Gurame, Patin, Lele, Nila, Ikan Mas, Kakap, Bandeng, Rumput Laut, Kerapu, Udang) saja tetapi juga hasil yang dapat menunjukkan relasi yang berhubungan dengan lokasinya. Dalam konteks ini dapat dilihat pada gambar dibawah yang mengimplikasikan bahwa Sulawesi Selatan menjadi atribut dengan tingkat tertinggi dengan nilai produksi rumput laut terbanyak dibandingkan dengan provinsi lain, lalu untuk peringkat kedua adalah Nusa Tenggara Timur sebagai penghasil rumput laut tertinggi.



Gambar 9. Akumulasi Produksi Perikanan Berdasarkan Provinsi

Dari hasil penelitian yang dapat diketahui bahwa pada setiap atribut memiliki keterkaitan atas setiap daerahnya. Komoditas yang menjadi titik puncak produksi pada setiap lokasi menunjukkan frekuensi yang berbeda. Untuk nilai tertinggi pada atribut temporal sekaligus relasi terhadap lokasi berada di Provinsi Sulawesi Selatan sebagai penghasil Rumput Laut terbanyak. Nusa Tenggara Timur berada di urutan kedua sebagai lokasi dengan produksi budidaya Rumput Laut terbesar dibandingkan provinsi lainnya.

Setelah langkah diatas penarikan kesimpulan akan bisa dilakukan dengan membentuk terlebih dahulu himpunan fuzzy dari setiap anggota. Sehingga setiap derajat keanggotaan dapat di ketahui dengan kata lain juga dapat dinyatakan kedalam fungsi keanggotaan.

Selanjutnya pembentukan himpunan fuzzy yang terdiri dari variabel input output. Kemudian dilanjutkan dengan fungsi implikasi yang di terapkan pada tiap-tiap aturan adalah fungsi MIN. Defuzzifikasi menggunakan software pendukung yaitu software matlab dengan fitur toolbox fuzzy, dan proses defuzzifikasi yang dilakukan menggunakan metode centroid. Langkah terakhir menarik kesimpulan yang didapat dari hasil pengolahan data.

IV. KESIMPULAN

Dari hasil penelitian ini dapat ditarik kesimpulan dengan hasil pengolahan data menghasilkan sebanyak 3 kelompok / cluster dalam tingkat produksi komoditas perikanan. Penjabaran dari hasil penelitian ini menggunakan RapidMiner didapati bahwa Provinsi Sulawesi Selatan sebagai wilayah penghasil rumput laut tertinggi dan juga bandeng tertinggi kedua. Lalu Jawa Timur sebagai penghasil bandeng tertinggi dan lele tertinggi kedua setelah Jawa Barat. Adapun hasil daripada komoditas yang lain menunjukkan bahwa frekuensi produksi dari setiap provinsi di Indonesia. Serta menghubungkan dari tiap-tiap jenis komoditas dengan tingkat produksinya pada setiap provinsi yang ada.

Pengecualian yang terdapat pada penarikan kesimpulan ini ialah penilaian dilakukan hanya dari tingkat akumulatif yang bisa dihasilkan pada setiap daerah. Tanpa memberikan penjabaran alternatif yang menjadi faktor eksternal utama, seperti penilaian untuk area geografis dan juga media yang menjadi sarana pendukung produksi suatu provinsi lebih dominan dibandingkan provinsi lainnya.

V. REFERENSI

- [1] B. Rahmat C.T.I. et al., "Implementasi k-means clustering pada rapidminer untuk analisis daerah rawan kecelakaan," Semin. Nas. Ris. Kuantitatif Terap. 2017, no. April, pp. 58–60, 2017.
- [2] Bappenas, "Kajian Strategi Pengelolaan Perikanan Berkelanjutan," Kementerian. PPN/Bappenas Direktorat Kelaut. dan Perikan., p. 120, 2014.
- [3] MacQueen, J.B. 1967. Some Methods for classification and Analysis of Multivariate Observations. Proceedings of 5-th Berkeley Symposium on Mathematical Statistics and Probability, Berkeley, University of California Press, 1: 281 – 297
- [4] P. Windarto, "Penerapan Data Mining Pada Ekspor Buah-Buahan Menurut Negara Tujuan Menggunakan K-Means Clustering Method," Techno.Com, vol. 16, no. 4, pp. 348–357, 2017, doi: 10.33633/tc.v16i4.1447.
- [5] N. Rofiqo, A. P. Windarto, and D. Hartama, "Penerapan Clustering Pada Penduduk Yang Mempunyai Keluhan Kesehatan Dengan Data Mining K-Means," KOMIK (Konferensi Nas. Teknol. Inf. dan Komputer), vol. 2, no. 1, pp. 216–223, 2018, doi: 10.30865/komik.v2i1.929.
- [6] W. M. P. Dhuhita, "Clustering Menggunakan Metode K-Means Untuk," J. Inform., vol. 15, no. 2, pp. 160–174, 2016.
- [7] S. M. Dewi, A. P. Windarto, and D. Hartama, "Penerapan Datamining Dengan Metode Klasifikasi Untuk Strategi Penjualan Produk Di Ud.Selamat Selular," KOMIK (Konferensi Nas. Teknol. Inf. dan Komputer).
- [8] N. Azis et al., "Mapping study using the unsupervised learning clustering approach," IOP Conf. Ser. Mater. Sci. Eng., vol. 1088, no. 1, p. 012005, 2021, doi: 10.1088/1757-899x/1088/1/012005
- [9] A. Rahmat, H. Hardi, F. A. Syam, Z. Zamzami, B. Febriadi, and A. P. Windarto, "Utilization of the field of data mining in mapping the area of the Human Development Index (HDI) in Indonesia," J. Phys. Conf. Ser., vol. 1783, no. 1, 2021, doi: 10.1088/1742-6596/1783/1/012035.
- [10] M. A. Amri, A. P. Windarto, A. Wanto, and I. S. Damanik, "Analisis Metode K-Means Pada Pengelompokan Perguruan Tinggi Menurut Provinsi Berdasarkan Fasilitas Yang Dimiliki Desa," KOMIK (Konferensi Nas. Teknol. Inf. dan Komputer), vol. 3, no. 1, pp. 674–679, 2019, doi: 10.30865/komik.v3i1.1677.
- [11] K. Handoko, "Penerapan Data Mining Dalam Meningkatkan Mutu Pembelajaran Pada Instansi Perguruan Tinggi Menggunakan Metode K-Means Clustering (Studi Kasus Di Program Studi Tkj Akademi Komunitas Solok Selatan)," J. Teknol. dan Sist. Inf., vol. 02, no. 03, pp. 31–40, 2016, [Online]. Available: <http://teknosi.fti.unand.id/index.php/teknosi/article/view/70>.
- [12] A. K. Wardhani, "Implementasi Algoritma K-Means untuk Pengelompokan Penyakit Pasien pada Puskesmas Kajen Pekalongan," J. Transform., vol. 14, no. 1, pp. 30–37, 2016.
- [13] H. U. Sari, A. P. Windarto, and D. Hartama, "Analisa Metode Data Mining Pada Produksi Perikanan Laut Yang Dijual Di Tempat Perikanan Ikan (Tpi)," KOMIK (Konferensi Nas. Teknol. Inf. dan Komputer), vol. 3, no. 1, pp. 630–636, 2019, doi: 10.30865/komik.v3i1.1671.
- [14] A. Nurzahputra, M. A. Muslim, and M. Khusniati, "Penerapan Algoritma K-Means Untuk Clustering Penilaian Dosen Berdasarkan Indeks Kepuasan Mahasiswa," Techno.Com, vol. 16, no. 1, pp. 17–24, 2017, doi: 10.33633/tc.v16i1.1284.

- [15] I. Wahyudi, S. Bahri, and P. Handayani, "Aplikasi Pembelajaran Pengenalan Budaya Indonesia," vol. V, no. 1, pp. 135–138, 2019, doi: 10.31294/jtk.v4i2.
- [16] W. M. Baihaqi, K. Indartono, and S. Banat, "Penerapan Teknik Clustering Sebagai Strategi Pemasaran pada Penjualan Buku Di Tokopedia dan Shopee," *Paradig. - J. Komput. dan Inform.*, vol. 21, no. 2, pp. 243–248, 2019, doi: 10.31294/p.v21i2.6149.
- [17] F. Firza and S. Sarjono, "Penerapan Algoritma K-Means Dalam Metode Clustering Untuk Peminatan Jurusan Bagi Siswa Swasta Pelita Raya Kota Jambi," *J. Manaj. Sist. Inf.*, vol. 5, no. 3, pp. 371–382, 2020, [Online]. Available: <http://ejournal.stikom-db.ac.id/index.php/manajemensisteminformasi/article/view/907>.
- [18] D. R. S. P, A. P. Windarto, D. Hartama, and E. Irawan, "Analisis Datamining Pada Pengelompokan Penduduk Yang Menjadi Korban Kekerasan Fisik Menurut Wilayah," *Semin. Nas. Sains Teknol. Inf. SENSASI 2019 ISBN*, pp. 350–356, 2019, [Online]. Available: <https://www.prosiding.seminar-id.com/index.php/sensasi/article/view/325>.