

Implementasi Algoritma K-Means Clustering Untuk Mengetahui Minat Pembeli Buah Melon

Nanda Ghina Syawali¹, Najmul Laila², Marrylinteri Istoningtyas^{3,*}, Mayang Ruza⁴,
Errissya Rasywir⁵, Maria Rosario Borroek⁶, yovi pratama^{7*}

^{1,2,3}Fakultas Ilmu Komputer, Program Studi Teknik Informatika, Universitas Dinamika Bangsa, Jambi, Indonesia
Email : ¹errissya.jurnal@gmail.com, ²diamar_ros@yahoo.com ³nandaghina8@gmail.com, ⁴mayangruza@gmail.com
⁵najmullaila840@gmail.com

Email penulis korespondensi : yovi.pratama@gmail.com

Abstrak - Penelitian ini dilakukan untuk mempelajari K-Means Clustering dalam pencarian knowledge (pengetahuan). Tujuan dari penelitian ini kemungkinan dapat membantu pihak Agen Buah Yudi untuk menentukan cluster minat pembeli banyak, sedang, dan sedikit diletakkan sesuai urutan minat pembeli buah yang banyak, sedang, dan sedikit di Agen Buah Yudi. Untuk itu dalam metode K-Means Clustering dimungkinkan adanya solusi dan analisa terhadap pengolahan data dan parameter-parameter yang menjadi acuan untuk mengambil keputusan. Di dalam metode ini terdapat langkah-langkah penyelesaian masalah. Adapun tools bantu untuk mengimplementasikan metode tersebut adalah Weka akan mengolah data secara tersusun atas operator-operator yang langsung didapatkan hasil secara akurat selanjutnya pada tahapan terakhir akan didapatkan knowledge baru.

Kata kunci : Buah Melon, klasifikasi, Algoritma, K-means Clustering, Weka

Abstract - This research was conducted to study K-Means Clustering in the search for knowledge (knowledge). The purpose of this research may be to help the Yudi Fruit Agent to determine the interest of many, medium, and few buyers according to the order of high, medium, and little interest of the Yudi Fruits Agent. For this reason, in the K-Means Clustering method, it is possible to provide solutions and analyzes for data processing and the parameters that become the reference for making decisions. In this method there are steps to solve the problem. As for the tools to implement this method, Weka will process data in an organized manner consisting of operators which immediately get accurate results, then in the last stage new knowledge will be obtained.

Keywords : Buah Melon, klasifikasi, Algoritma, K-means Clustering, Weka

1. PENDAHULUAN

Analisis cluster merupakan teknik multivariat yang mempunyai tujuan utama untuk mengelompokkan objek-objek berdasarkan karakteristik yang dimilikinya [1]–[3]. Analisis cluster mengklasifikasi objek sehingga setiap objek yang paling dekat kesamaannya dengan objek lain berada dalam cluster yang sama [4].

Objek-objek/data yang dikelompokkan ke dalam suatu grup memiliki ciri-ciri yang sama berdasarkan kriteria tertentu, salah satu aktifitas analisis data adalah klasifikasi atau pengelompokan data kedalam beberapa kategori atau cluster. Untuk melakukan pengelompokan terhadap data pengelompokan minat pembeli buah yang banyak, sedang, dan sedikit menggunakan teknik clustering [5]–[7]. Metode yang digunakan K-Means Clustering. Dengan menggunakan metode ini data-data yang didapatkan dapat dikelompokkan ke dalam beberapa cluster, berdasarkan minat beli konsumen, sehingga data-data yang dimiliki karakteristik yang sama dikelompokkan dalam satu cluster yang lain [8].

Solusi analisis cluster bersifat tidak unik, anggota cluster untuk tiap penyelesaian/solusi tergantung pada beberapa elemen prosedur dan beberapa solusi yang berbeda dapat diperoleh dengan mengubah satu elemen atau lebih [6], [7], [9]. Solusi cluster secara keseluruhan bergantung pada variabel-variabel yang digunakan sebagai dasar untuk menilai kesamaan. Penambahan atau pengurangan variabel-variabel yang relevan dapat mempengaruhi substansi hasil analisis cluster [4].

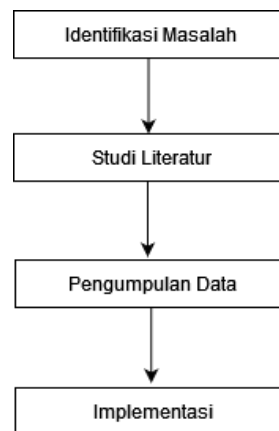
Dari permasalahan tersebut di atas, maka penulis tertarik melakukan penelitian ini dengan judul: "Implementasi Algoritma K-means Clustering Untuk Mengetahui minat Pembeli di agen buah melon (Studi Kasus Di Agen Buah Yudi)".

2. METODOLOGI PENELITIAN

Pada penelitian ini dilakukan di lingkungan Agen buah Yudi untuk mengklasterisasi tingkat penjualan pada buah melon. Data yang digunakan adalah data hasil penjualan buah melon dengan format tabulasi data excel. Teknik pengolahan data yang digunakan dalam penelitian ini adalah dengan metode K-Means Clustering . Tools yang akan digunakan dalam pengolahan data adalah Aplikasi WEKA dan selanjutnya hasil perhitungan dari aplikasi weka akan dibandingkan dengan hasil perhitungan manual untuk melihat presentase akurasi dari penggunaan weka terhadap data yang di olah [10].

2.1 KERANGKA KERJA PENELITIAN

Kerangka kerja penelitian merupakan tahapan-tahapan yang dilakukan selama mengerjakan penelitian. Kerangka kerja penelitian dibuat agar Mempermudah pencapaian hasil penelitian, dapat menyelesaikan penelitian tepat waktu dan penelitian dapat berjalan sesuai dengan yang diharapkan. Adapun kerangka kerja penelitian yang digunakan dapat dilihat pada gambar.[11].



Gambar 1 Kerangka Kerja Penelitian[12]

2.1.1 Identifikasi Masalah

Sebagai tahap awal dalam penelitian, penulis melakukan identifikasi masalah yang terjadi pada Penjualan buah melon di agen buah yudi. Dalam tahap ini, penulis menemukan permasalahan pada Penilaian minat pembelian buah melon. Hal ini mengakibatkan penjual kesulitan untuk menemukan pembeli yang minat belinya paling tinggi [13].

2.1.2 Studi Literatur

1. Clustering

Clustering merupakan suatu proses pengelompokkan suatu record, observasi, atau pengelompokan kelas yang memiliki kesamaan objek [14]. Clustering atau klasifikasi adalah metode yang digunakan untuk membagi rangkaian data menjadi beberapa group berdasarkan kesamaan-kesamaan yang telah ditentukan sebelumnya [15].

2. K-Means

K-means merupakan salah satu metode data clustering non hirarki yang berusaha berpartisipasi data yang ada ke dalam bentuk satu atau lebih cluster atau kelompok sehingga data yang memiliki karakteristik sama

akan dikelompokkan ke dalam satu cluster dan data yang mempunyai karakteristik berbeda akan dikelompokkan ke cluster lainnya. [16].

3. K-Means Clustering

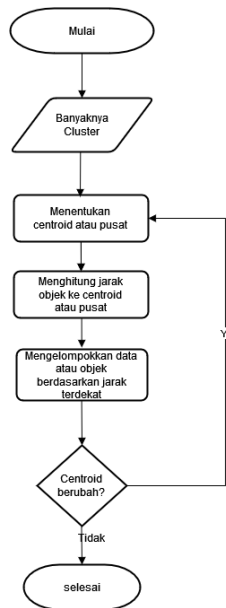
K-Means Clustering merupakan salah satu metode data clustering non-hirarki yang mengelompokkan data dalam bentuk satu atau lebih cluster/kelompok. Data-data yang memiliki karakteristik yang sama dikelompokkan dalam satu cluster/kelompok dan data yang memiliki karakteristik yang berbeda dikelompokkan dengan cluster/kelompok yang lain sehingga data yang berada dalam satu cluster/kelompok memiliki tingkat variasi yang kecil.[14]

4. Algoritma K-Means

Algoritma K-Means merupakan salah satu dari algoritma yang banyak digunakan dalam pengelompokan karena kesederhanaan dan efisiensi [17].

Proses untuk clustering menggunakan algoritma k-means adalah sebagai berikut :

1. Tentukan nilai k dalam total cluster yang ingin dibentuk.
2. Pilih nilai centroid cluster (centroid) berawal dari k.
3. Hitung jarak setiap data input dan setiap centroid menggunakan rumus jarak Euclidean untuk menemukan jarak terdekat antara setiap titik data dan centroid.
4. Mengklasifikasikan/mengelompokkan setiap item data berlandaskan jarak (jarak minimum) ke centroid.
5. Perbarui mediannya. Nilai pusat baru.
6. Mengulangi langkah ke 3 sampai 5 anggota setiap cluster tetap tidak berubah. [16].



Gambar.2 Flowchart proses algoritma K-Means

Berikut ini merupakan rumus terhadap perhitungan algoritma K-Means Clustering[18] :

$$de = \sqrt{(xi - si)^2 + (yi - ti)^2}$$

5. WEKA

WEKA (Waikato Environment for Knowledge Analysis) merupakan aplikasi data mining yang bersifat open source berbasis Java. WEKA pertama kali dikembangkan oleh Universitas Waikato Selandia Baru sebelum menjadi bagian di Pentanho. Weka terdiri dari koleksi algoritma machine learning yang dapat digunakan untuk melakukan generalisasi atau formulasi dari sekumpulan data [19].

2.1.3 Pengumpulan Data

Pada tahap ini yang dikumpulkan berupa data-data penjualan buah melon sebagai sample pada bulan april-mei 2022 pada agen buah yudi.[20].

a. Studi Literatur

Studi literatur yang digunakan dalam penulisan Penelitian ini adalah mengumpulkan bahan referensi mengenai metode klasifikasi dari berbagai buku, jurnal, dan beberapa referensi lainnya.

b. Observasi

Metode pengumpulan data yang dilakukan dengan mengadakan pengamatan atau peninjauan langsung terhadap sumber permasalahan serta berkomunikasi langsung dengan pihak yang bersangkutan yaitu Pemilik Agen Buah tersebut.

c. Wawancara

Metode pengumpulan data yang dilakukan melalui tanya jawab yang diajukan secara langsung kepada narasumber untuk mendapatkan data atau informasi yang berkaitan dengan objek penelitian [21].

2.1.4 Implementasi

Implementasi K-Means Clustering dilakukan dengan menggunakan metode Simple *K-Means Clustering* pada perangkat lunak data mining WEKA [22]. WEKA (*Waikato Environment for Knowledge Analysis*) merupakan perangkat lunak data mining yang dikembangkan oleh Universitas Waikato, New Zealand. Diimplementasikan pertama kali pada tahun 1997 dan mulai menjadi open source pada tahun 1999 [16].

3. HASIL DAN PEMBAHASAN

3.1. Implementasi Algoritma K-Means

Data Pembeli Melon yang akan diimplementasikan pada algoritma K-Means, dapat dilihat pada table 1 pada di bawah ini:

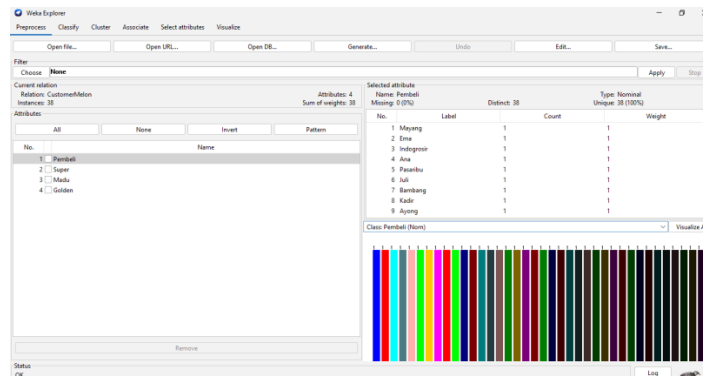
Tabel 1.Data Pembeli Melon Agen Buah Yudi

Pembeli	Melon (kg)		
	Super	Madu	Golden
Mayang	60	200	180
Ema	70	110	150
Indogrosir	150	150	200
Ana	150	300	100
Pasaribu	200	150	175
Kak Juli	150	150	120
Bambang	55	70	80
Kadir	100	100	150
Ayong	150	100	100
Daus	200	300	300
Ginting	500	1200	500
Sri	300	200	300
Novel	30	30	30
Sembiring	100	150	250
Bude	50	50	233
Am Angso	100	200	200
Petro	110	100	90
Trona	260	300	200
Jamtos	150	200	100
Uda Iin	50	170	110
Pir	400	115	200
Divi Fres	200	190	160
Pulo Mas	100	250	110

Tambak	250	110	75
Lian	200	190	300
Ate	50	50	50
Buluran	62	130	75
Uni Rat	70	70	50
Dani	100	90	90
Irul	100	250	200
Mamang Pir	250	203	150
Dely	462	100	270
Indah	580	120	50
Agus Handil	75	25	43
Tupang	200	100	150
Anton	100	100	100
Darsito	40	120	40
Da Am	200	300	100

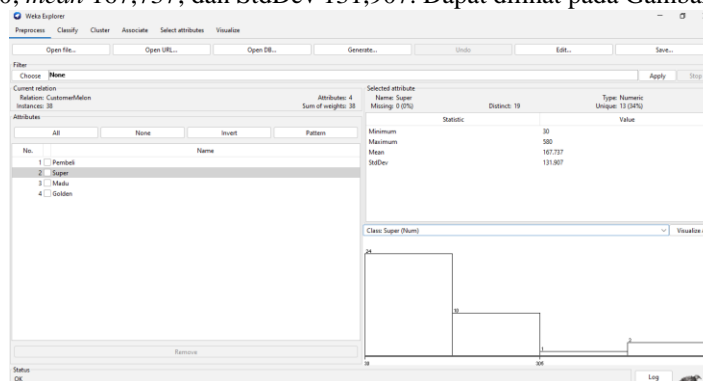
3.2. Implementasi perhitungan Metode K-Means Clustering dengan WEKA

Pertama, buka file yang akan di hitung pada WEKA, setelah dimasukkan ke WEKA maka dapat menampilkan atribut yang digunakan. Pada atribut Pembeli melon dengan *type* nominal terdapat missing 0(0%); distinct 38; unique 38(100%). Dapat dilihat pada Gambar 3 dibawah ini:



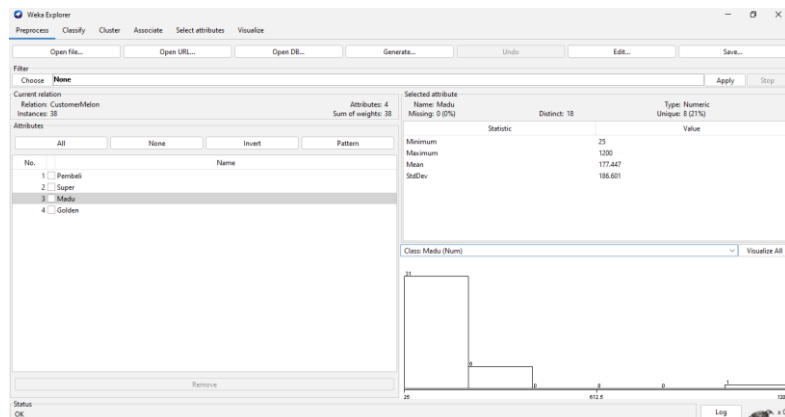
Gambar 3. Bagian-bagian dari atribut pembeli

Pada atribut melon Super dengan *type numeric* terdapat *missing* 0(0%); *distinct* 19; *unique* 13(34%); *minimum* 30; *maximum* 580; *mean* 167,737; dan *StdDev* 131,907. Dapat dilihat pada Gambar 4 dibawah ini:



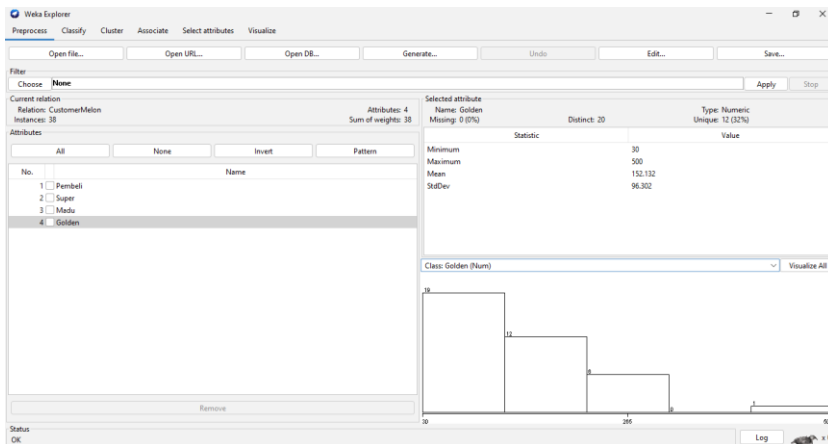
Gambar 4. Bagian-bagian dari atribut melon super

Pada atribut melon Madu dengan *type numeric* terdapat *missing* 0(0%); *distinct* 18; *unique* 8(21%); *minimum* 25; *maximum* 1200; *mean* 177,447; dan *StdDev* 186,601. Dapat dilihat pada Gambar 5 dibawah ini:



Gambar 5. Bagian-bagian dari atribut melon madu

Pada atribut melon Golden dengan *type numeric* terdapat *missing* 0(0%); *distinct* 20; *unique* 12(32%); *minimum* 30; *maximum* 500; *mean* 152,132; dan *StdDev* 96,302. Dapat dilihat pada Gambar 6 dibawah ini:



Gambar 6. Bagian-bagian dari atribut melon golden

- Selanjutnya pilih metode yang akan digunakan pada data yang ada. Disini penulis menggunakan metode K-Means Clustering. Dapat di lihat pada gambar 6 di bawah ini:

Pada perhitungan menggunakan WEKA dapat diperoleh bahwa dalam perhitungan ini dilakukan 6 iterasi dengan menggunakan 3 *centroid* (Daus, Tupang, dan Agus Handil). Cluster0 : Daus, 200, 300, 300; Cluster1 : Tupang, 200, 100, 150; Cluster2 : Agus Handil, 75, 25,43.

```
kMeans
=====

Number of iterations: 6
Within cluster sum of squared errors: 37.179515588245934

Initial starting points (random):

Cluster 0: Daus,200,300,300
Cluster 1: Tupang,200,100,150
Cluster 2: AgusHandil,75,25,43

Missing values globally replaced with mean/mode

Final cluster centroids:

Attribute      Full Data      Cluster#
                (38.0)        (6.0)         (18.0)        (14.0)
=====
Pembeli        Mayang         Daus          Mayang        Ema
Super          167.7368      407           159.4444     75.8571
Madu           177.4474      339.1667     182.9444     101.0714
Golden         152.1316      270          169.0556     79.8571
```

Gambar 7. Perhitungan K-Means Clustering menggunakan WEKA

Pada Gambar 8 terdapat banyaknya pembeli setiap cluster. Cluster0 (pembeli banyak) memiliki 6 pembeli dengan persentase 16%, Cluster1 (pembeli lumayan banyak) memiliki 18 pembeli dengan persentase 47%, dan Cluster2 (pembeli sedikit) memiliki 14 pembeli dengan persentase 37%.

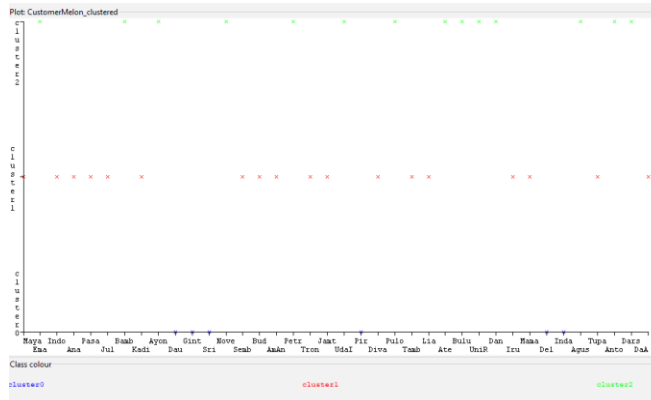
```
=== Model and evaluation on training set ===

Clustered Instances

0          6 ( 16%)
1         18 ( 47%)
2         14 ( 37%)
```

Gambar 8. Hasil Evaluasi perhitungan K-Means

- b. Jika selesai melakukan perhitungan dapat di lihat hasil perhitungan tersebut, maka perhitungan pun selesai. Pada Gambar 9 terdapat anggota pembeli setiap cluster, berikut ini anggota pembeli setiap Cluster0 : Daus, Ginting, Sri, Pir, Dely, dan Indah. Cluster1 : Mayang, Indogrosir, Ana, Pasaribu, Juli, Kadir, Sembring, Bude, Am Angso, Trona, Jamtos, Diva, Tambak, Lian, Irul, Mamang Pir, Tupang, Da Am. Cluster2 : Ema, Bambang, Ayong, Novel, Petro, Uda Iin, Pulo Mas, Ate, Buluran, Uni Rat, Dani, Agus Handil, Anton dan Darsito.



Gambar 9. Bagian dari cluster0, Cluster1, dan Cluster2

4. KESIMPULAN

Untuk melakukan penilaian terhadap minat pembeli buah melon dapat menerapkan metode *clustering K-Means*. Data diolah untuk memperoleh nilai dari minat berdasarkan penjualan buah melon. Data tersebut diolah menggunakan WEKA untuk ditentukan nilai *centroid* dalam 3 *cluster* yaitu *cluster* minat pembeli banyak, *cluster* minat pembeli sedang, dan *cluster* minat pembeli sedikit. *Centroid* data untuk *cluster* minat pembeli banyak 200, 300, 300; *Centroid* data untuk *cluster* minat pembeli sedang 200, 100, 150; dan *Centroid* data untuk *cluster* minat pembeli sedikit 75, 25,43. Sehingga diperoleh penilaian berdasarkan penjualan buah melon dengan pembeli *cluster* minat pembeli banyak pada buah melon super 200kg - 580kg; melon madu 100kg - 1200kg; dan melon golden 20kg -500kg, *cluster* minat pembeli sedang pada buah melon super 50kg - 260kg; melon madu 50kg - 300kg; dan melon golden 75kg -300kg, dan *cluster* minat pembeli sedikit pada buah melon super 30kg - 150kg; melon madu 25kg -250kg; dan melon golden 30kg -150kg.

REFERENSI

- [1] E. Rasywir, R. Sinaga, and Y. Pratama, "Analisis dan Implementasi Diagnosis Penyakit Sawit dengan Metode Convolutional Neural Network (CNN)," *J. Paradig. UBSI*, vol. 22, no. 2, pp. 117–123, 2020.
- [2] G. G. Anagnostopoulos and A. Kalousis, "A reproducible analysis of RSSI fingerprinting for outdoor localization using sigfox: Preprocessing and hyperparameter tuning," *2019 Int. Conf. Indoor Position. Indoor Navig. IPIN 2019*, pp. 1–8, 2019, doi: 10.1109/IPIN.2019.8911792.
- [3] M. R. Borroek, E. Rasywir, Y. Pratama, Fachruddin, and M. Istoningtyas, "Analysis on Knowledge Layer Application for Knowledge Based System," in *Proceedings of 2018 International Conference on Electrical Engineering and Computer Science, ICECOS 2018*, 2019, pp. 177–182, doi: 10.1109/ICECOS.2018.8605262.
- [4] T. Alfina and B. Santosa, "Analisa Perbandingan Metode Hierarchical Clustering, K-Means dan Gabungan Keduanya dalam Membentuk Cluster Data (Studi Kasus : Problem Kerja Praktek Jurusan Teknik Industri ITS)," *Anal. PerbandinganMetode Hierarchical Clust. K-means dan Gabungan Keduanya dalam Clust. Data*, vol. 1, no. 1, pp. 1–5, 2012.
- [5] D. Darwis, N. Siskawati, and Z. Abidin, "Penerapan Algoritma Naive Bayes Untuk Analisis Sentimen Review Data Twitter Bmkg Nasional," *J. Tekno Kompak*, vol. 15, no. 1, p. 131, 2021, doi: 10.33365/jtk.v15i1.744.
- [6] D. Zaenal Abidin and E. Rasywir, "Penerapan Data Mining Klasifikasi Untuk Memprediksi Potensi Mahasiswa Berprestasi Di Stikom Dinamika Bangsa Jambi Dengan Metode Naive Bayes," *J. Ilm. Mhs. Tek. Inform.*, vol. 3, no. 2, 2021.
- [7] E. Rasywir, R. Sinaga, and Y. Pratama, "Evaluasi Pembangunan Sistem Pakar Penyakit Tanaman Sawit dengan Metode Deep Neural Network (DNN)," vol. 4, pp. 1206–1215, 2020, doi: 10.30865/mib.v4i4.2518.
- [8] F. Fauzansyah, "Implementasi Algoritma K-Means Clustering Untuk Menentukan Arketipe Pembelian Suku Cadang Dan Asesoris Komputer(Studi Kasus Di Toko Laksamana Komputer Dumai)," *INFORMATIKA*, vol. 9, no. 1, p. 26, 2018, doi: 10.36723/juri.v9i1.60.
- [9] H. Hendrawan, A. Haris, E. Rasywir, and Y. Pratama, "Diagnosis Penyakit Tanaman Karet dengan Metode Fuzzy Mamdani," *J. Paradig. UBSI*, vol. 22, no. 2, pp. 132–138, 2020.
- [10] A. P. Windarto, "Penerapan Datamining 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.
- [11] P. Alam Jusia, F. Muhammad Irfan, and S. Dinamika Bangsa Jambi JI Jend Sudirman Thehok Jambi, "Clustering Data Untuk Rekomendasi Penentuan Jurusan Perguruan Tinggi Menggunakan Metode K-Means," *J. IKRA-ITH Inform.*, vol. 3, no. 3, p. 75, 2019.

- [12] R. Habibi, *Penggunaan Framework Laravel Untuk Membuat Aplikasi Absensi Terintegrasi Mobile*. Bandung: Kreatif Industri Nusantara, 2019.
- [13] N. Ahlunaza, “Perancangan Aplikasi Sistem Informasi Geografis Lokasi Latihan Bolavoli Di Kota Jambi Berbasis Android,” *J. Inform. Dan Rekayasa Komput. ...*, vol. 1, no. April, pp. 118–130, 2022.
- [14] P. M. Silitonga Irene Sri, “Klusterisasi Pola Penyebaran Penyakit Pasien Berdasarkan Usia Pasien Dengan Menggunakan K-Means Clustering,” *J. TIMES*, vol. VI, no. Vol 6, No 2 (2017), pp. 22–25, 2017.
- [15] Y. D. Darmi and A. Setiawan, “Penerapan Metode Clustering K-Means Dalam Pengelompokan Penjualan Produk,” *J. Media Infotama*, vol. 12, no. 2, pp. 148–157, 2017, doi: 10.37676/jmi.v12i2.418.
- [16] J. Informatika *et al.*, “Implementasi Data Mining Untuk Menentukan Persediaan Stok Obat Jurnal Informatika Dan Rekayasa Komputer (JAKAKOM),” vol. 1, no. April, pp. 155–160, 2022.
- [17] G. Gustientiedina, M. H. Adiya, and Y. Desnelita, “Penerapan Algoritma K-Means Untuk Clustering Data Obat-Obatan,” *J. Nas. Teknol. dan Sist. Inf.*, vol. 5, no. 1, pp. 17–24, 2019, doi: 10.25077/teknosi.v5i1.2019.17-24.
- [18] A. Rohmah, F. Sembiring, and ..., “Implementasi Algoritma K-Means Clustering Analysis Untuk Menentukan Hambatan Pembelajaran Daring (Studi Kasus: Smk Yaspim ...,” ... *Sist. Inf. dan ...*, pp. 290–298, 2021.
- [19] T. M. Tamtelahitu, “Komparasi Algoritma Clustering dengan Dataset Penyebaran Covid-19 di Indonesia Periode Maret-Mei 2020,” *J. Teknol. Technoscintia*, vol. 13, no. 1, pp. 27–34, 2020.
- [20] M. Fauzi and Y. Yudi, “Penerapan Algoritma K-Means Clustering Untuk Mendeteksi Penyebaran Penyakit TBC (Studi Kasus: Di Kabupaten Deli Serdang),” *JTIK (Jurnal Tek. Inform. Kaputama)*, vol. 1, no. 2, pp. 1–7, 2017.
- [21] J. Informatika, D. Rekayasa, K. Jakakom, K. Telkomsel, R. Cell, and S. Candra, “Penerapan Data Mining Algoritma Naive Bayes Clasifier Untuk Mengetahui Minat Beli Pelanggan Terhadap Jurnal Informatika Dan Rekayasa Komputer (JAKAKOM),” vol. 1, no. April, pp. 149–154, 2022.
- [22] A. Asroni and R. Adrian, “Penerapan Metode K-Means Untuk Clustering Mahasiswa Berdasarkan Nilai Akademik Dengan Weka Interface Studi Kasus Pada Jurusan Teknik Informatika UMM Magelang,” *Semesta Tek.*, vol. 18, no. 1, pp. 76–82, 2016, doi: 10.18196/st.v18i1.708.