

Klasterisasi Penjualan Produk Bayi Organik Abata Bendul Menggunakan Algoritma K-Means

Annisa Aulia Arafah^{*1)}

¹⁾Teknologi Informasi, Fakultas Teknologi Informasi dan Bisnis, Institut Teknologi Telkom Surabaya, Jl.
Ketintang No.156, Ketintang, Kec. Gayungan, Surabaya, 60231, Indonesia
arap@student.ittelkom-sby.ac.id

Abstrak

Penelitian bertujuan untuk membantu Abata Bendul, sebuah merek barang MPASI di Surabaya, serta untuk memaksimalkan manajemen stok mereka. Mempertahankan pelanggan adalah cara menjaga stok agar tidak kosong. Untuk mencapai hal tersebut penjual perlu menganalisis data penjualan, mengetahui produk mana yang memberikan laba tinggi dan penjualan tinggi serta mana yang memberikan laba rendah dan penjualan rendah. Dalam penelitian ini, masalah tersebut dapat diselesaikan menggunakan algoritma K-means Clustering, salah satu teknik dalam data mining. Penelitian ini fokus pada pengelompokan data penjualan barang Abata Bendul menggunakan variabel seperti nama barang, laba, banyaknya transaksi, dan jumlah produk yang terjual dalam periode tertentu, kemudian diolah menggunakan algoritma K-means dan ditampilkan pada website. Hasil akhir penelitian ini adalah empat kelompok yang terbentuk; 1 produk laris menguntungkan, 8 produk laris tidak menguntungkan, 9 produk tidak laris menguntungkan, dan 12 produk tidak laris dan tidak menguntungkan. Hasil penelitian ini dapat dimanfaatkan oleh manajemen Abata Bendul untuk meningkatkan efektivitas kampanye pemasaran dan mengoptimalkan hasil penjualan.

Kata kunci: *Data mining, K-means, Pengelompokan*

1. Pendahuluan

Di Indonesia, masalah malnutrisi pada bayi dimulai ketika mereka membutuhkan lebih banyak Makanan Pendamping Air Susu Ibu (MPASI) sekitar usia enam bulan hingga lima tahun. MPASI bervariasi, termasuk biskuit, bubur, nasi saring, puding, dan oatmeal. Bubur bayi organik umum digunakan karena kandungan gizinya yang tinggi. Perkembangan MPASI pesat, dengan persaingan ketat di Surabaya, termasuk brand Abata Bendul. Mereka menghadapi kendala dalam manajemen stok, pencatatan manual yang rawan kesalahan, ketidakstabilan stok akibat permintaan fluktuatif, dan pengelompokan barang yang tidak efisien. Tantangan utama adalah menentukan stok minimum berdasarkan minat konsumen. Untuk mengatasi masalah ini, Abata Bendul akan menggunakan algoritma K-means Clustering untuk mengelompokkan barang berdasarkan popularitas dan potensi laba. Hasilnya akan membantu mereka mengoptimalkan stok dan meningkatkan efisiensi manajemen penjualan.

Penerapan metode *K-means* memiliki tujuan untuk mengelompokkan produk menjadi empat kategori berbeda berdasarkan tingkat popularitas dan tingkat profitabilitas: produk laris menguntungkan, laris namun tidak menguntungkan, tidak laris namun menguntungkan, dan tidak laris dan tidak menguntungkan. Manfaat yang dapat diperoleh dari sistem yang akan dikembangkan termasuk membantu meningkatkan efisiensi dalam beberapa aspek. Hal ini mencakup pengurangan kesalahan dalam pencatatan data transaksi, peningkatan efisiensi waktu dalam pengolahan data transaksi, dan penghasilan pengelompokan barang yang harus diproduksi dalam jumlah tertentu.

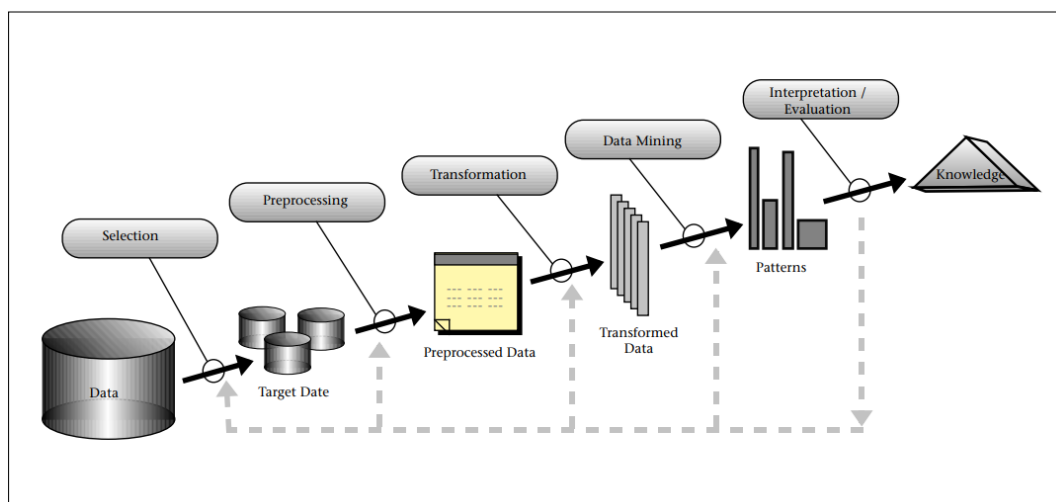
Beberapa penelitian berikut ini menggunakan metode K-means Clustering dalam berbagai konteks:

- a. Penjualan Produk Outdoor: Penelitian ini menerapkan K-means Clustering untuk mengelompokkan data penjualan di toko Genta Corp di Bogor menjadi tiga kluster

berdasarkan tingkat keliharannya. Hasilnya membantu manajemen stok dan strategi penjualan mereka (Indriyani & Irfiani, 2019).

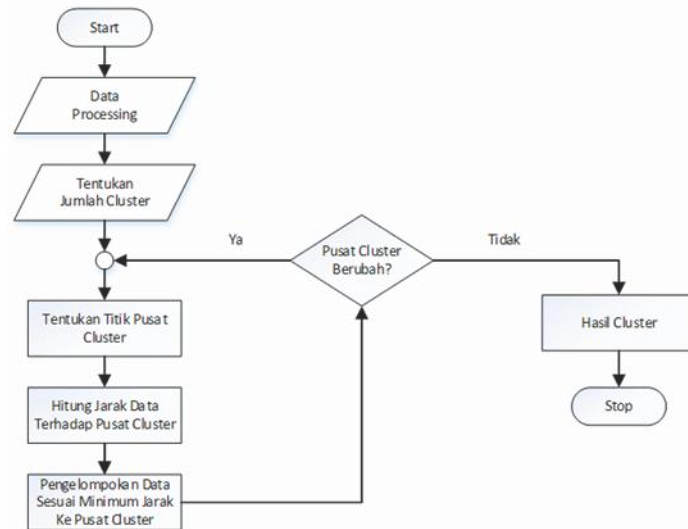
- b. Ujian Nasional SMP: Penelitian ini menggunakan K-means Clustering untuk mengelompokkan data capaian Ujian Nasional SMP tahun 2018/2019 di Indonesia menjadi tiga kelompok berdasarkan nilai, membantu pemahaman tentang capaian pendidikan di berbagai provinsi (Aditya et al., 2020).
- c. Penjualan Produk dengan Konsep Down-Selling: Penelitian ini menggunakan K-means Clustering untuk mengelompokkan produk dan memberikan rekomendasi untuk konsep down-selling. Hasilnya menciptakan 5 kluster produk yang dapat diterapkan dalam konsep tersebut (Putra Samudra et al., n.d.).
- d. Penjualan Produk Frozen Food: Penelitian ini menggunakan K-means Clustering untuk mengelompokkan penjualan makanan beku menjadi dua kluster, membantu mengidentifikasi minat konsumen terhadap produk (Amalina et al., 2022).
- e. Peminatan Konsumen Terhadap Smartphone: Penelitian ini menerapkan K-means Clustering untuk menentukan peminatan terhadap merek smartphone di dua gerai Tio Cell. Hasilnya akan membantu dalam pengambilan keputusan stok di masa depan (Bomanthara & Nasrul Hali, 2023).
- f. Penyebaran Kasus COVID-19: Penelitian ini menggunakan K-means Clustering untuk mengelompokkan provinsi-provinsi di Indonesia berdasarkan tingkat penyebaran COVID-19 menjadi tiga cluster. Hasilnya memberikan wawasan tentang perbedaan karakteristik antar provinsi (Sari & Sukestiyarno, 2021).

Data mining adalah proses penemuan pengetahuan tersembunyi dalam database. Ini digunakan oleh ahli statistik, analis data, dan komunitas SIM. Melibatkan teknik analisis data untuk mengidentifikasi hubungan dan pola tersembunyi. Data mining merangkum data yang relevan dan tidak diketahui sebelumnya. Ini mengungkap tren yang tidak terlihat dalam analisis sederhana dan digunakan untuk prediksi dan deskripsi, seperti klasifikasi, regresi, pengelompokan, dan asosiasi.



Gambar 1. Tahapan Data mining (Fayyad et al., 1996)

K-means adalah algoritma Clustering dalam Unsupervised learning yang digunakan untuk mengelompokkan data tanpa label kelas sebelumnya. Ini termasuk dalam kategori Non-Hierarchical atau Partitional Clustering. Tujuannya adalah mengelompokkan data sehingga data dalam satu kelompok mirip dan berbeda dari kelompok lain.

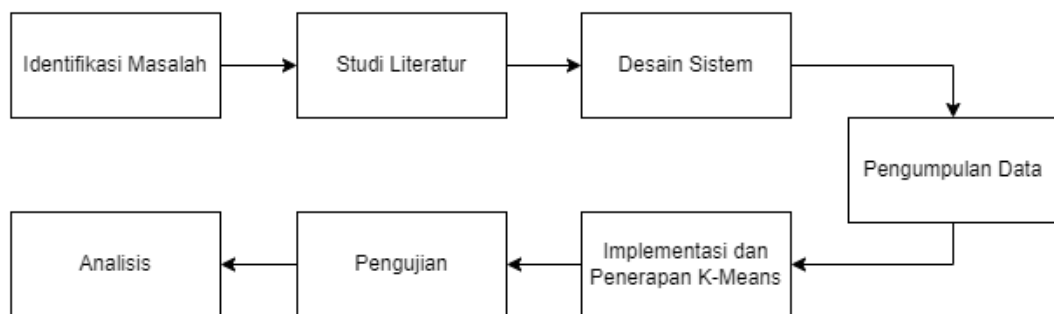


Gambar 2. Diagram alir dari algoritma Clustering K-means

Unified Modeling Language (UML) adalah bahasa visual untuk memodelkan dan berkomunikasi tentang sistem. Ditemukan oleh Grady Booch, Ivar Jacobson, dan James Rumbaugh, versi terbarunya adalah 2.5. UML digunakan secara luas untuk pemodelan sistem dan tidak terbatas pada satu metodologi tertentu (Denis et al., 2015). Ini mencakup 12 diagram yang dikelompokkan dalam 2 kategori.

2. Metode Penelitian

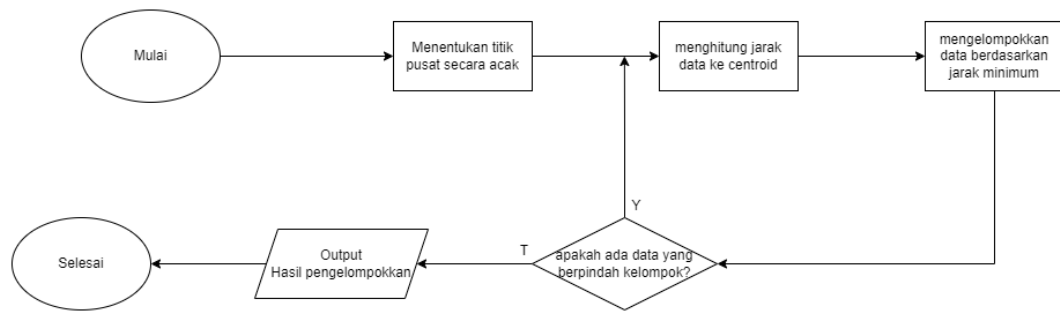
Penelitian ini menguraikan serangkaian langkah yang dimulai dengan mengidentifikasi masalah, melakukan studi literatur, merancang sistem, mengumpulkan data, mengimplementasikan dan menerapkan metode K-means, serta melakukan pengujian dan analisis. Diagram alur penelitian ini tersedia dalam gambar untuk memberikan gambaran visual tentang urutan langkah-langkah yang dilakukan.



Gambar 3. Prosedur Penelitian

3. Hasil dan Pembahasan

Penelitian ini mengadopsi metode K-means dengan menggunakan langkah-langkah prosedur algoritma K-means sebagai berikut:



Gambar 4. Metode K-means

Tabel 1. Data Penjualan

Kode	Nama Produk	Laba Produk	Total transaksi	Total Quantity
K1	Abon 30ml	1.700	88	1393
K2	Abon 60ml	2.500	88	1443
K3	Bubur Ati	1.500	85	1310
K4	Bubur ayam	1.500	82	1234
K5	Bubur Daging	1.500	87	1266
K6	Bubur Kakap	1.500	89	1439
K7	Bubur Salmon	1.500	89	1403
K8	Bubur Tuna	1.500	87	1296
K9	Nasi Tim Ayam	1.500	90	1331
K10	Nasi Tim Daging	1.500	87	1177

Adapun variabel yang digunakan dalam pengelompokkan atau clustering yaitu Nama Produk, Laba Produk, Total Transaksi, dan Total Quantity. Pada Tabel 1 adalah tabel data penjualan yang akan dikelompokkan.

3.1. Normalisasi

Normalisasi adalah transformasi data untuk memastikan bahwa setiap fitur memiliki rentang nilai yang seragam. Tujuannya adalah untuk mengatasi perbedaan skala antar fitur sehingga kontribusi setiap fitur dalam analisis data seimbang. Dalam penelitian ini, normalisasi menggunakan Min-Max Scaling untuk mengubah nilai setiap fitur menjadi rentang [0, 1]. Normalisasi penting karena variasi skala antar variabel dapat memengaruhi hasil analisis dan interpretasi data. Perbedaan sebelum dan setelah normalisasi dapat dilihat pada **Tabel 2**.

$$X_{scaled} = (x - min)/(max - min) \quad (1)$$

Tabel 2. Data Setelah Normalisasi

Kode	Nama Produk	Laba Produk	Total transaksi	Total Quantity
K1	Abon 30ml	0,2	0,8	0,7
K2	Abon 60ml	1,0	0,8	0,8

Kode	Nama Produk	Laba Produk	Total transaksi	Total Quantity
K3	Bubur Ati	0,0	0,4	0,4
K4	Bubur ayam	0,0	0,0	0,2
K5	Bubur Daging	0,0	0,6	0,3
K6	Bubur Kakap	0,0	0,9	0,8
K7	Bubur Salmon	0,0	0,9	0,7
K8	Bubur Tuna	0,0	0,6	0,4
K9	Nasi Tim Ayam	0,0	1,0	0,5
K10	Nasi Tim Daging	0,0	0,6	0,0

3.2. Menentukan nilai K

Metode Silhouette menghitung koefisien silhouette untuk setiap titik data dalam setiap kluster. Nilai rata-rata koefisien siluet untuk berbagai nilai k akan membantu dalam memahami kualitas pengelompokan.

Koefisien Siluet ($s(i)$) untuk suatu objek i dihitung menggunakan rumus berikut:

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \quad (2)$$

Tabel 3. Nilai Skor Siluet

K	Hasil
2	0,332347
3	0,326070
4	0,345686
5	0,342332

Dalam hal ini, nilai 0. 345686 (k=4) lebih mendekati 1 dibandingkan dengan nilai k lainnya. Meskipun tidak mencapai 1, nilai yang lebih tinggi mendekati 1 menunjukkan kualitas clustering yang lebih baik.

3.3. Penerapan Metode K-Means Clustering

Pada tahap ini, dilakukan proses utama yaitu segmentasi atau pengelompokkan data penjualan produk yang diakses dari database menggunakan metode clustering K-means. Berikut adalah algoritma K-means dengan parameter input berupa jumlah dataset sebanyak n data dan jumlah inialisasi centroid K=4, sesuai dengan penelitian. Dari banyaknya data penjualan yang diperoleh, diambil 10 jenis produk sebagai sampel untuk menerapkan algoritma K-means. Percobaan dilakukan dengan menggunakan parameter sebagai berikut:

$$\text{Jumlah kluster} = 4$$

Jumlah data = 30

Iterasi ke-1

Pusat awal cluster atau centroid diperoleh secara acak/random. Proses penentuan awal cluster adalah sebagai berikut:

- Centroid Cluster 1 = (0; 0,875; 0,799)
- Centroid Cluster 2 = (0,2; 1; 0,265)
- Centroid Cluster 3 = (1; 0,75; 0,811)
- Centroid Cluster 4 = (0,2; 0,275; 0,101)

Perhitungan jarak pusat cluster

Untuk mengukur jarak antara data dengan pusat cluster, digunakan metode Euclidean distance.

Selanjutnya, dilakukan perhitungan jarak dari sisa sampel data ke pusat cluster dengan menggunakan Euclidean distance. Misalkan $M(a, b, c)$ adalah jarak Euclidean antara setiap data dengan pusat cluster, di mana a merupakan laba penjualan tiap produk, dan b adalah total transaksi dan c total produk terjual.

Hitung jarak Euclidean dari semua data ke titik pusat pertama:

$$d = \sqrt{(M_x - C_x)^2 + (M_y - C_y)^2 + (M_z - C_z)^2} \tag{3}$$

Tabel 4. Hasil Iterasi 1

	C1	C2	C3	C4	Kelompok
M1	0,27	0,47	0,81	0,67	C1
M2	1,01	1,00	0,00	1,13	C3
M3	0,64	0,67	1,14	0,36	C4
M4	1,08	1,02	1,40	0,43	C4
M5	0,58	0,43	1,14	0,36	C4
M6	0,00	0,58	1,01	0,88	C1
M7	0,11	0,48	1,02	0,80	C1
M8	0,50	0,44	1,10	0,41	C4
M9	0,35	0,29	1,09	0,75	C2
M10	0,27	0,47	0,81	0,67	C1

Iterasi ke-2

Untuk menentukan posisi centroid baru (C_k), kita dapat menghitung nilai rata-rata dari data-data yang terdapat pada centroid yang sama.

$$C_k = \left(\frac{1}{n_k}\right) \sum d_i \tag{3}$$

Di mana n_k adalah jumlah data dalam cluster k dan d_i adalah data dalam cluster k. sehingga didapatkan titik pusat atau centroid yang baru yaitu:

- C1 = (0,127; 0,716; 0,777)
- C2 = (0,15; 0,891; 0,328)

C3 = (1; 0,75; 0,811)

C4 = (0,08; 0,45; 0,317)

Tabel 5. Hasil Iterasi 2

	C1	C2	C3	C4	Kelompok
M1	0,14	0,36	0,81	0,47	C1
M2	0,87	0,99	0,00	1,09	C3
M3	0,52	0,54	1,14	0,14	C4
M4	0,94	0,92	1,40	0,48	C4
M5	0,53	0,31	1,14	0,20	C4
M6	0,20	0,49	1,01	0,65	C1
M7	0,22	0,39	1,02	0,57	C1
M8	0,44	0,31	1,10	0,20	C4
M9	0,44	0,23	1,09	0,58	C2
M10	0,79	0,45	1,29	0,37	C4

Iterasi Ke-3

Hitung Titik Pusat Baru

C1 = (0,14; 0,738; 0,78)

C2 = (0,15; 0,891; 0,328)

C3 = (1; 0,75; 0,811)

C4 = (0,073; 0,455; 0,338)

Tabel 6. Hasil Iterasi 3

	C1	C2	C3	C4	Kelompok
M1	0,15	0,36	0,81	0,45	C1
M2	0,86	0,99	0,00	1,08	C3
M3	0,55	0,54	1,14	0,13	C4
M4	0,98	0,92	1,40	0,49	C4
M5	0,56	0,31	1,14	0,20	C4
M6	0,20	0,49	1,01	0,63	C1
M7	0,23	0,39	1,02	0,55	C1
M8	0,47	0,31	1,10	0,19	C4
M9	0,44	0,23	1,09	0,57	C2
M10	0,82	0,45	1,29	0,39	C4

Iterasi Ke-4

Hitung Titik Pusat Baru

C1 = (0,133; 0,75; 0,83)

C2 = (0,15; 0,891; 0,328)

C3 = (1; 0,75; 0,811)

C4 = (0,083; 0,469; 0,353)

Tabel 7. Hasil Iterasi 4

	C1	C2	C3	C4	Kelompok
M1	0,18	0,36	0,81	0,43	C1
M2	0,87	0,99	0,00	1,06	C3
M3	0,58	0,54	1,14	0,14	C4
M4	1,01	0,92	1,40	0,51	C4
M5	0,59	0,31	1,14	0,20	C4
M6	0,19	0,49	1,01	0,61	C1
M7	0,23	0,39	1,02	0,53	C1
M8	0,50	0,31	1,10	0,18	C4
M9	0,46	0,23	1,09	0,55	C2
M10	0,85	0,45	1,29	0,40	C4

Dikarenakan hasil posisi cluster pada iterasi ke-4 sama dengan posisi iterasi ke-3, maka proses berhenti. Produk dengan kode produk K2 termasuk dalam klaster C3, yang menunjukkan tingkat penjualan tinggi dan memberikan keuntungan yang tinggi. Produk dengan kode K9 termasuk dalam klaster C2 yang menunjukkan tingkat penjualan tinggi namun memberikan keuntungan yang rendah. Produk dengan kode K1, K6, K7 termasuk dalam klaster C1 yang menunjukkan tingkat penjualan rendah namun memberikan keuntungan yang tinggi dan produk dengan kode K3, K4, K5, K8, K10 termasuk dalam klaster C4 yang menunjukkan tingkat penjualan rendah dengan keuntungan yang rendah. Analisis ini diperoleh dengan menghitung nilai rata-rata banyaknya transaksi dan keuntungan yang didapat dari tiap klaster.

4. Kesimpulan

Dalam analisis data penjualan Abata Bendul Januari-Maret 2023 menggunakan metode K-Means, ditemukan 4 klaster berdasarkan silhouete score sebesar 0,345686. Klaster tersebut adalah:

C1: 9 produk, tidak laris namun menguntungkan.

C2: 8 produk, laris namun tidak menguntungkan.

C3: 1 produk, laris dan menguntungkan.

C4: 12 produk, tidak laris dan tidak menguntungkan.

Pada C1, meskipun tidak laris, produk memiliki potensi menguntungkan dan perlu penyesuaian strategi pemasaran. C2, yang populer, perlu evaluasi biaya produksi dan strategi harga. C3 memerlukan pemeliharaan kualitas dan pertimbangan diversifikasi produk. Sedangkan C4 perlu dinilai secara mendalam untuk perbaikan atau penghapusan sesuai dampak terhadap profitabilitas keseluruhan.

Dengan demikian, metode K-Means membantu Abata Bendul dalam mengelompokkan produk dan merancang strategi yang sesuai.



Gambar 5 . Implementasi Website

Ucapan Terima Kasih

Saya berterima kasih kepada Ibu Oktavia Ayu Permata, S.T., M.T., dan Bapak Yohanes Setiawan, S.Si., M. Kom., sebagai pembimbing saya dalam penulisan tugas akhir ini. Dukungan mereka telah memberi inspirasi dan motivasi bagi saya.

Terima kasih kepada keluarga saya atas dukungan moral dan materi yang berharga. Kata-kata semangat mereka memotivasi saya melewati tantangan penelitian ini. Terima kasih juga kepada rekan-rekan, teman-teman, dan sesama mahasiswa atas kolaborasi dan diskusi berharga.

Saya mengucapkan terima kasih kepada semua yang berkontribusi dalam penelitian ini, baik langsung maupun tidak langsung. Kontribusi mereka penting bagi penyelesaian tugas akhir ini. Meskipun tugas akhir ini tidak sempurna, saya berharap hasilnya dapat berkontribusi positif pada ilmu bisnis dan bermanfaat bagi pembaca yang tertarik dengan topik ini.

Daftar Pustaka

- Aditya, A., Jovian, I., & Sari, B. N. (2020). Implementasi K-Means Clustering Ujian Nasional Sekolah Menengah Pertama di Indonesia Tahun 2018/2019. *JURNAL MEDIA INFORMATIKA BUDIDARMA*, 4(1), 51. <https://doi.org/10.30865/mib.v4i1.1784>
- Amalina, T., Bima, D., Pramana, A., & Sari, B. N. (2022). Metode K-Means Clustering Dalam Pengelompokan Penjualan Produk Frozen Food. *Jurnal Ilmiah Wahana Pendidikan*, 8(15), 574–583. <https://doi.org/10.5281/zenodo.7052276>
- Arista Rizki, N., & Fidia Deny Tisna Amijaya, Ms. (2019). Database System (Sistem Basis Data). <http://math.fmipa.unmul.ac.id>
- Bomanthara, & Nasrul Hali, R. M. (2023). Implementasi Metode K-Means Clustering Untuk Menentukan Tingkat Peminatan Konsumen Terhadap Type Smartphone Di Tio Cell Dan Tio Cell 2. *JUPITER (Jurnal Penelitian Ilmu Teknologi Komputer)*, 15(1), 49–60.
- Chusyairi, A., & Ramadar Noor Saputra, P. (2019). Pengelompokan Data Puskesmas Banyuwangi Dalam Pemberian Imunisasi Menggunakan Metode K-Means Clustering. *Telematika*, 12(2), 139–148. <https://doi.org/10.35671/telematika.v12i2.848>
- Denis, A., Wixom, B. H., & Tegarden, D. (2015). *SYSTEMS ANALYSIS & DESIGN An Object-Oriented Approach with UML* (B. L. Golub, Ed.; 5th ed.). John Wiley & Sons.
- Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From Data Mining to Knowledge Discovery in Databases. www.ffly.com/

- Hemeida, A. M., Hassan, S. A., Mohamed, A. A. A., Alkhalaf, S., Mahmoud, M. M., Senjyu, T., & El-Din, A. B. (2020). Nature-inspired algorithms for feed-forward neural network classifiers: A survey of one decade of research. In *Ain Shams Engineering Journal* (Vol. 11, Issue 3, pp. 659–675). Ain Shams University. <https://doi.org/10.1016/j.asej.2020.01.007>
- Indriyani, F., & Irfiani, E. (2019). Clustering Data Penjualan pada Toko Perlengkapan Outdoor Menggunakan Metode K-Means. *JUITA: Jurnal Informatika*, 7(2), 109–113. <https://doi.org/10.30595/juita.v7i2.5529>
- Data Mining, (April 8, 2022). <https://teknik-informatika-s1.stekom.ac.id/informasi/baca/Data-Mining/349a35c6f665191bb8fe901b91284feb29d1736d>
- Larose, D. T., & Larose, C. D. (2014). *DISCOVERING KNOWLEDGE IN DATA* (2nd ed.). John Wiley & Sons. <https://doi.org/10.1002/9781118874059>
- Parlina, I., Windarto, A. P., Wanto, A., & Lubis, M. R. (2018). MEMANFAATKAN ALGORITMA K-MEANS DALAM MENENTUKAN PEGAWAI YANG LAYAK MENGIKUTI ASESSMENT CENTER UNTUK CLUSTERING PROGRAM SDP. *CESS (Journal of Computer Engineering System and Science)*, 3(1).
- Putra Samudra, F., Hendro Pudjiantoro, T., Santikarama Jurusan Informatika, I., & Sains dan Informatika Universitas Jenderal Achmad Yani Jl Terusan Sudirman, F. (n.d.). Klasterisasi Tingkat Penjualan Produk Menggunakan Metode K-Means Untuk Penerapan Konsep Down-Selling Studi Kasus pada ARTCH Indonesia. *Seminar Nasional Informatika Dan Aplikasinya (SNIA)*, 2021.
- Sari, D. N. P., & Sukestiyarno, Y. L. (2021). Analisis Cluster dengan Metode K-Means pada Persebaran Kasus Covid-19 Berdasarkan Provinsi di Indonesia. *PRISMA, Prosiding Seminar Nasional Matematika*, 4, 602–610. <https://journal.unnes.ac.id/sju/index.php/prisma/>
- Simatupang, J., & Sianturi, S. (2019). PERANCANGAN SISTEM INFORMASI PEMESANAN TIKET BUS PADA PO. HANDOYO BERBASIS ONLINE. *Jurnal Intra-Tech*, 3(2).
- Sudirman, Windarto, A. P., & Wanto, A. (2018). Data mining tools | rapidminer: K-means method on clustering of rice crops by province as efforts to stabilize food crops in Indonesia. *IOP Conference Series: Materials Science and Engineering*, 420(1). <https://doi.org/10.1088/1757-899X/420/1/012089>
- UML 2.5 Diagrams Overview. (n.d.). Retrieved July 9, 2023, from <https://www.uml-diagrams.org/uml-25-diagrams.html>
- Wanto, A., & dkk. (2020). *Data Mining : Algoritma dan Implementasi* (T. Limbong, Ed.; 1st ed.). Yayasan Kita Menulis.