

PENERAPAN ALGORITMA *K-MEANS CLUSTERING* BERDASARKAN HASIL PENJUALAN PRODUK *INEZ COSMETIC* DI TOKO ROSALIA

Nilam Kumala¹⁾, Raditya Danar Dana²⁾

^{1,2)} Prodi Manajemen Informatika, STMIK IKMI Cirebon

Jl. Perjuangan No 10B Majasem, Kota Cirebon 4513, Jawa Barat, Indonesia

E-mail: ¹⁾ nilamkumala2000@gmail.com, ²⁾ radith_danar@yahoo.com

ABSTRAK

Produk kosmetik telah menjadi bagian penting dalam kehidupan sehari-hari bagi wanita dari remaja hingga dewasa. Toko Rosalia merupakan salah satu toko yang menyediakan dan menjual berbagai macam produk kosmetik dari berbagai brand salah satunya ialah produk *Inez Cosmetic*. Produk *Inez Cosmetic* memiliki penjualan yang cukup tinggi. Namun transaksi dan penyediaan produk yang belum terkomputerisasi menyebabkan pemilik toko kesulitan dalam menentukan produk *Inez Cosmetic* mana yang laku dan kurang laku di pasaran. Sehingga menyebabkan terjadinya penumpukan stok produk dan data penjualan tidak bisa dimanfaatkan secara optimal. Berdasarkan permasalahan tersebut dilakukan beberapa cara yaitu dengan menerapkan teknik *data mining* dengan menggunakan algoritma *K-Means Clustering*, data yang di gunakan berasal dari historis data penjualan produk *Inez Cosmetic* selama 6 bulan yaitu dari bulan Juni 2022 sampai dengan bulan Desember 2022. Dari penelitian yang telah dilakukan dengan menggunakan metode algoritma *K-Means Clustering* pada penjualan produk *Inez Cosmetic* didapatkan hasil 3 kelompok atau 3 *cluster* yaitu *Cluster 0* sebanyak 37 *items*, *Cluster 1* sebanyak 186 *items*, *Cluster 2* sebanyak 1 *items*, dan hasil nilai *Davies-Bouldin Index* yang di dapatkan sebesar 0.128.

Kata kunci : pengelompokan, *RapidMiner*, *K-Means Clustering*, penjualan

ABSTRACT

Cosmetics products have become an important part of everyday life for women from teenagers to adults. Rosalia Store is a store that provides and sells various kinds of cosmetic products from various brands, one of which is Inez Cosmetic products. Inez Cosmetic products have quite high sales. However, transactions and supply of products that have not been computerized have caused difficulties for shop owners in determining which Inez Cosmetic products are selling well and which are not selling well on the market. So that it causes a buildup of product stock and sales data cannot be used optimally. Based on these problems, several methods are carried out, namely by applying data mining techniques using the K-Means Clustering algorithm, the data used comes from historical sales data for Inez Cosmetic products for 6 months, namely from June 2022 to December 2022. From research that has been carried out using the K-Means Clustering algorithm method for selling inez cosmetic products, the results are 3 groups or 3 clusters, namely Cluster 0 of 37 items, Cluster 1 of 186 items, Cluster 2 of 1 item, and the results of the Davies-Bouldin Index value are get 0.128.

Keywords: *Clustering, RapidMiner, K-Means Clustering, Sales*

1. PENDAHULUAN

Produk kosmetik telah menjadi bagian penting dalam kehidupan sehari-hari bagi wanita

dari remaja hingga dewasa. Produk kosmetik dibedakan kedalam beberapa jenis sesuai dengan kegunaan dan fungsinya masing- masing

yaitu *make up* atau riasan wajah, dan ada juga produk kosmetik yang difokuskan untuk bagian perawatan saja seperti halnya *skincare*, *bodycare*, *haircare*, *lipcare* dan lain sebagainya. Toko Rosalia yang terletak di JL. Ki Hajar Dewantara Desa Junjang Kecamatan Arjawinangun, merupakan salah satu toko yang menyediakan dan menjual berbagai macam produk kosmetik dari berbagai brand salah satunya ialah produk *Inez Cosmetic*. Penjualan produk *Inez Cosmetic* mendapatkan hasil yang cukup tinggi, namun penyediaan stok produk dan transaksi penjualan yang belum terkomputerisasi menyebabkan data penjualan tersebut tidak bisa dimanfaatkan secara optimal, sehingga menyebabkan terjadinya penumpukan stok produk *Inez Cosmetic*.

Berdasarkan penelitian terdahulu dengan tema pengelompokan menggunakan metode algoritma *K-Means Clustering* [1]. Dalam jurnal ini masalah yang difokuskan yaitu data penjualan dan pembelian yang hanya berfungsi sebagai arsip toko dan tidak terstruktur dengan baik sehingga tidak bisa digunakan sebagai acuan strategi pemasaran. Untuk mengatasi permasalahan tersebut digunakan penerapan metode *K-Means Clustering*. Dengan mengelompokkan data stok produk digital menjadi 3 grup. Hasil pengolahan yang di dapatkan berupa 3 kelompok yaitu kelompok produk dengan kategori laku sebanyak 114, produk dengan katerori kurang laku sebanyak 5, dan produk dengan kategori tidak laku sebanyak 14. Berdasarkan penelitian yang telah dijelaskan diatas perbedaan dalam penelitian ini ialah terletak pada algoritma *K-Means Clustering* yang di implementasikan dengan pola data penjualan yang di ambil dari bulan Juni 2022 sampai dengan bulan Desember 2022.

2. METODE PENELITIAN

Pembahasan pada bagian ini yaitu tentang landasan teori teknik pengumpulan data dan tahapan perancangan.

2.1.Landasan Teori

Landasan teori yang akan dibahas adalah :

1.Data Mining

Data mining adalah sebuah proses pengumpulan, pengolahan dan pola penggalian informasi penting dari data [2]. *Data mining* merupakan proses pengekstrasian informasi dari data yang besar [3]. *Data mining* menerapkan sistem *artificial intelligence* dalam menguraikan permasalahan [4]. Tujuan yang utama dari *data mining* yaitu memperluas pengetahuan informasi dari data [5]. *Data mining* memiliki beberapa kelompok yaitu: deskripsi, estimasi, prediksi, klasifikasi, *cluster*, asosiasi [6].

2.Knowledge Discovey In Data Base

Knowlegde Discovery In Data Base merupakan metode yang digunakan untuk memperoleh pengetahuan dari data yang bisa dijadikan penentu keputusan [7]. Berikut ini merupakan proses KDD yaitu : *Data selection*, *pre processing* data, transformasi data, datamining, evaluasi, dan *knowledge* [8].

3.Clustering

Clustering merupakan teknik data mining yang memiliki sifat sederhana yaitu membagi kumpulan data ke dalam beberapa kelompok yang sama [9]. Sehingga didalam satu kelompok terdapat banyak kesamaan dan perbedaan dengan kelompok lain [10]. *Clustering* memiliki dua tujuan, yaitu *clustering* yang berfungsi untuk pemahaman dan *clustering* yang berfungsi untuk penggunaan [11].

4.Algoritma K-Means

K-Means termasuk kategori algoritma clustering dengan proses yang di lakukan secara berulang-ulang. Huruf K diartikan sebagai jumlah *cluster* pemilihan nilai K ditetapkan secara acak. Sedangkan means merupakan nilai pusat dari klaster sementara atau disebut dengan *centroid* [12]. Secara historisnya *K-Means* merupakan salah satu algoritma yang paling penting dalam bidang pengolahan *data mining*. Algoritma *K-Means Clustering* merupakan algoritma yang hanya biasa bekerja dengan atribut *numeric* atau angka.

5. Davies Bouldin Index (DBI)

Davies-Bouldin Index merupakan metode evaluasi internal yang yang di gunakan untuk mengukur hasil *cluster* atau pengelompokkan [13]. Rumus untuk menentukan nilai *Davies – Bouldin Index* :

$$DBI = \frac{1}{k} \sum_{i=1}^k \max_{i \neq j} (R_{ij}) \quad (1)$$

K merupakan jumlah cluster yang di gunakan semakin kecilnya nilai *Davies Bouldin Index*, maka semakin baik *cluster* yang di kelompokkan menggunakan *K-Means* [13].

2.2. Teknik Pengumpulan Data

Teknik pengumpulan data yang di gunakan berupa :

1. Observasi

Observasi merupakan salah satu teknik metode pengumpulan data yang dilakukan secara langsung atau dilakukan langsung di tempat penelitian [3] untuk mengetahui proses pengolahan data produk *Inez Cosmetic* di Toko Rosalia.

2. Wawancara

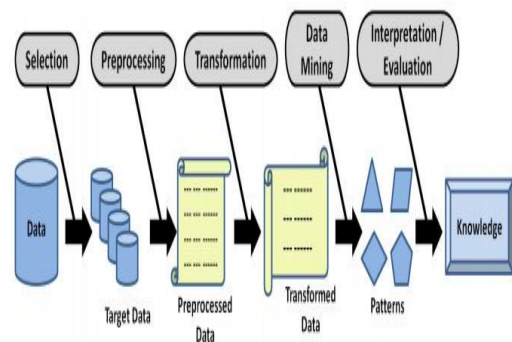
Wawancara merupakan suatu proses untuk mendapatkan informasi dari narasumber secara langsung. Adapun proses wawancara dilakukan langsung dengan pemilik Toko Rosalia Karyawan, dan Bagian Pemasaran (*Beauty Advisor*).

3. Studi Literature

Studi Literature merupakan proses pengumpulan informasi yang berasal dari sumber-sumber tertulis baik berupa buku ataupun artikel jurnal di internet yang bisa dijadikan sebagai referensi [3].

2.3. Tahapan Perancangan

Tahapan perancangan digunakan ialah *Knowledge Discovery In Database (KDD)*. Berikut merupakan proses *Knowledge Discovery in Database (KDD)* :



Gambar 1. Proses Knowledge Discovery InDatabase (KDD)

1. Data Selection

Tahapan pertama yang dilakukan ialah *data selection* atau pemilihan data yang relevan. Data yang yang digunakan adalah data historis penjualan produk *Inez Cosmetic* dari bulan Juni 2022 sampai dengan bulan Desember 2022 yang terdiri dari 224 produk *Inez Cosmetic* dengan 6 atribut. Atribut tersebut diantaranya no, kode produk, nama produk, stok awal, penjualan, stok akhir, total traksaksi penjualan.

2. Preprocessing/Cleaning

Preprocessing data merupakan proses yang bertujuan untuk memeriksa data, membuang dan menghapus duplikasi data serta memperbaiki kesalahan pada data yang akan di gunakan pada tahap selanjutnya. Pada tahap ini di lakukan dengan cara mengaktifkan *Replace Error With Missing Value* dan menggunakan *Select Atribut* pada *RapidMiner*.

3. Tranformasi Data

Tranformasi Data merupakan tahap yang dilakukan setelah bagian *Pre proessing* atau setelah proses pembersihan dan perbaikan data. Pada tahap ini dilakukan perubahan data diantaranya dengan menggunakan atribut *Nominal to Numerical* dan atribut *Normalize* pada *RapidMiner*.

4. Interpretasi/Evaluasi

Interprestasi atau Evaluasi merupakan tahap terkahir yang akan membahas hasil dari *proses data selection, preprocessing / cleaning, tranformasi data, dan data mining* atau pemodelan data.

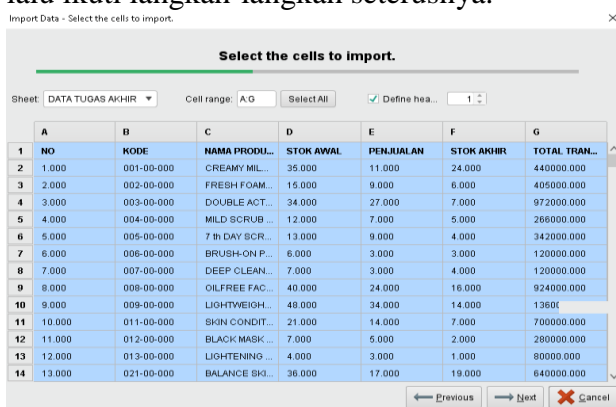
3. HASIL DAN DISKUSI

Bagian ini berisikan mengenai hasil-hasil penerapan algoritma *K-Means Clustering* berdasarkan hasil penjualan produk *Inez Cosmetic*.

3.1. Hasil Pre Processing Data

1.Data Selection

Langkah pertama yang dilakukan untuk menerapkan metode algoritma *K-Means* pada *RapidMiner* ialah memilih *Blank Process* lalu pilih *Import Data* selanjutnya pilih lokasi penyimpanan dari “*My Computer*”. Pilih file lalu ikuti langkah-langkah seterusnya.

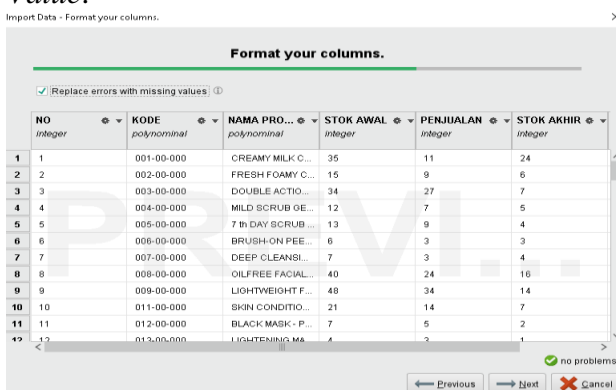


Gambar 2. Proses Import Data

2. Preprocessing Data

a. Cleaning Data

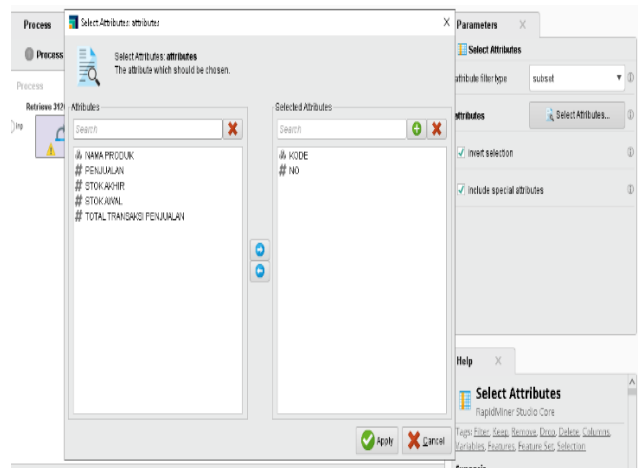
Langkah selanjutnya merupakan tahapan proses *preprocessing* data atau *cleaning* dengan mengaktifkan *Replace Error With Missing Value* yang berfungsi untuk melihat apakah ada duplikat data, data yang salah ataupun data yang bermasalah. Berikut adalah gambar proses *Replace Error With Missing Value*.



Gambar 3. Proses Cleaning Replace Error With Missing Value

b. Select Attribute

Setelah proses *import* data selesai, maka tahapan selanjutnya yaitu menggunakan operator *Select Attribute* dengan *parameters Select Atributes* sebagai berikut: atribut filter yang digunakan ialah *Subset* dengan *select attribute* yang di pilih untuk dihilangkan adalah *attribute NO* dan *attribute Kode*. Sedangkan untuk *attribute* yang digunakan ialah *attribute Nama Produk*, *attribute Stok Awal*, *attribute Penjualan*, *attribute Stok Akhir* dan *attribute Total Transaksi Penjualan*. Serta mengaktifkan *Invert Selection* dan *Include Special Attributes* agar *attribute Nama Produk*, *attribute Stok Awal*, *attribute Penjualan*, *attribute Stok Akhir*, dan *attribute Total Transaksi Penjualan* tidak dihilangkan. Berikut adalah gambar proses *Select Attribute*:



Gambar 4. Proses Select Attribute

3. Transformasi Data

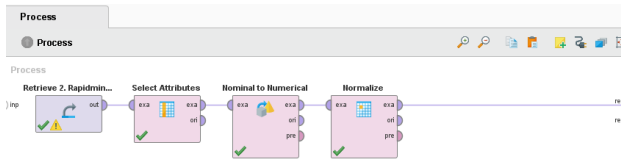
Proses selanjutnya yaitu melakukan tahapan perubahan atribut pada data di antaranya dengan menggunakan operator *Nominal to Numerical* dan operator *Normalize*.

a. Nominal to Numerical

Pada parameters *Nominal to Numerical*, atribut filter yang digunakan adalah atribut filter *type All* dan *Coding type* yang digunakan adalah *Unique Integer*.

b. Normalize

Pada parameters *Normalize*, *attribute* filter yang digunakan adalah *atribut filter type All* sedangkan *Method* yang digunakan adalah *Z-Transformation*.

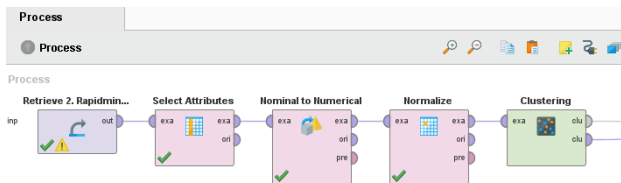


Gambar 5. Model Operator *Nominal to Numerical* dan Operator *Normalize*

3.2. Hasil Komputasi Data Mining Clustering

1. Implementasi Algoritma K-Means Clustering

Implementasi model *algoritma K-Means Clustering* pada *RapidMiner* adalah sebagai berikut:



Gambar 6. Implementasi *Algoritma K-Means*

Pada parameters *clustering K-Means* nilai parameter *K* ditentukan dengan nilai yang optimal yaitu nilai *K* 3. Nilai parameter *K* merupakan jumlah *cluster* yang akan dikelompokkan dengan proses *run max* 10 atau 10 kali percobaan dengan *Measure types Numerical Measures*, *Numerical Measure EuclideanDistance* dan *max optimization steps 100*.

2. Hasil Komputasi Clustering

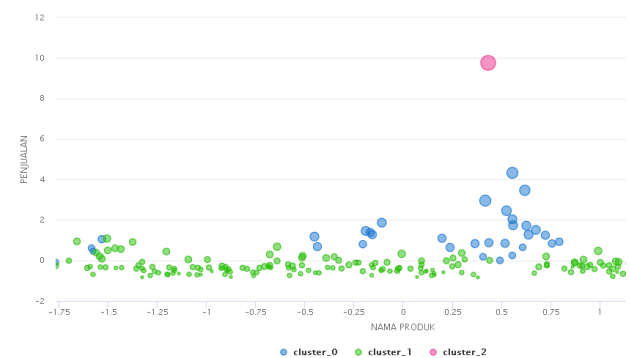
Pengelompokkan Produk *Inez Cosmetic* Berdasarkan Hasil Penjualan Toko Rosalia pada

RapidMiner menghasilkan 3 kelompok atau 3 *cluter* model di antaranya sebagai berikut :

Cluster Model

Cluster 0: 37 items
Cluster 1: 186 items
Cluster 2: 1 items
Total number of items: 224

Gambar 7. Hasil Komputasi *Clustering (Cluter Model)*



Gambar 8. Visualisasi Hasil Komputasi *Clustering*

Pada Gambar 8. Hasil Komputasi *Clustering* pada Pengelompokkan Produk *Inez Cosmetic* Menggunakan *RapidMiner* Berdasarkan hasil penjualan Toko Rosalia yang terbagi kedalam 3 model *cluster* diantaranya sebagai berikut : Berdasarkan nilai penjualan *Cluster 2* memiliki 1 items adalah kumpulan data dengan karakteristik penjualan produk tinggi namun dengan nilai transaksi yang rendah , sedangkan *Cluster 0* memiliki 37 items adalah kumpulan data dengan karakteristik penjualan produk sedang namun dengan nilai transaksi yang tinggi , dan *Cluster 1* memiliki 186 items adalah kumpulan data dengan karakteristik penjualan rendah namun dengan nilai transaksi yang sedang.

3. Penentuan Jumlah Cluster Optimal

Penentuan jumlah *cluster* optimal didapatkan dalam sembilan kali percobaan

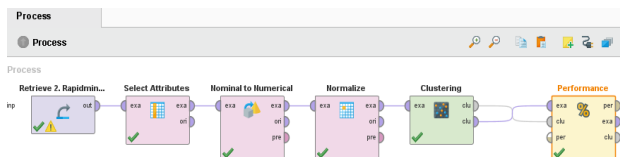
untuk menentukan nilai K. Pada percobaan kedua nilai K yang di tentukan ialah 3(tiga) dan menghasilkan nilai *Performance Vector (Davies Bouldin Index)* yang lebih optimal dibandingkan dengan percobaan pertama, ketiga dan seterusnya sampai ke percobaan sembilan. Dengan nilai *Davies Bouldin Index* yang dihasilkan berupa 0.128 yang mendekati 0. Seperti *table* di bawah ini:

Table 1. Penentuan Jumlah Cluster Optimal

| Cluster (K) | Davies Bouldin Index |
|-------------|----------------------|
| 2 | 0.171 |
| 3 | 0.128 |
| 4 | 0.132 |
| 5 | 0.134 |
| 6 | 0.169 |
| 7 | 0.164 |
| 8 | 0.162 |
| 9 | 0.164 |
| 10 | 0.177 |

4. Hasil Performa

Setelah semua proses *data selection, pre processing / cleaning*, tranformasi data, dan *Data Mining* atau pemodelan data telah selesai di laksanakan maka tahapan selanjutnya ialah proses evaluasi hasil dari *Clustering* menggunakan algoritma *K-Means* untuk mencari nilai K yang paling optimal. Maka dari itu dilakukan proses pengujian menggunakan operator *Performance (Cluster Distance Performance)* Pada *parameters Performance (Cluster Distance Performance)*, *main criterion* yang digunakan ialah *Davie Bouldin*, serta mengaktifkan *Normalize* dan *Maximize*.



Gambar 9. Model Performa (*Cluster Distance Performance*)

Davies Bouldin

Davies Bouldin: 0.128

Gambar 10. Hasil Performa (*Davies Bouldin Index*)

Pada Gambar 10. Hasil yang didapatkan berupa nilai *Davies-Bouldin Index*: 0.128. Yang merupakan nilai *Davies Bouldin Index* yang paling optimal.

4. KESIMPULAN

Berdasarkan pengujian data yang telah diolah menggunakan *RapidMiner* didapatkan hasil berupa :

1. *Preprocessing* menggunakan operator *select attribute*. *Attribute* yang digunakan ialah Nama Produk, Stok Awal, Penjualan, Stok Akhir, Total Transaksi penjualan.
2. Hasil nilai K yang optimal didapatkan setelah sembilan kali percobaan dengan menggunakan nilai *Davies-Bouldin Index* yang paling optimal yaitu nilai K=3, dengan menggunakan Algoritma *K-Means* didapatkan hasil 3 kategori *cluster*: *Cluster 0* memiliki 37 items, *Cluster 1* memiliki 186 items, *Cluster 2* memiliki 1 items.
3. Hasil *Performance Davies-Bouldin Index* yang optimal dengan nilai K yang paling mendekati nol adalah K = 3 sebesar 0.128.

5. DAFTAR PUSTAKA

[1] N. Astuti, J. N. Utamajaya, and A. Pratama, "Penerapan Data Mining Pada Penjualan Produk Digital Konter Leppangeng Cell Menggunakan Metode K-Means Clustering," *JURIKOM (Jurnal Riset Komputer)*, vol. 9, no. 3, p. 754, Jun. 2022.

- [2] Normah, S. Nurajizah, and A. Salbinda, “Penerapan Data Mining Metode K-Means Clustering Untuk Analisa Penjualan Pada Toko Fashion Hijab Banten,” *Jurnal Teknik Komputer AMIK BSI*, vol. 7, no. 2, Jul. 2021.
- [3] A. Fikri Sallaby, R. Tri Alinse, V. Novita Sari, and T. Ramadani, “Pengelompokan Barang Menggunakan Metode K-Means Clustering Berdasarkan Hasil Penjualan Di Toko Widya Bengkulu,” *Jurnal Media Infotama*, vol. 18, no. 1, 2022.
- [4] I. Nurrohmat, O. Nurdiawan, and A. Bahtiar, “IMPLEMENTASI ALGORITMA FP-GROWTH UNTUK MENUNJANG KEPUTUSAN PERSEDIAN BARANG DI CV INDOTECH JAYA SENTOSA KOTA CIREBON,” *JURSIMA Sistem Informasi dan Manajemen*, vol. 10, no. 2, Aug. 2022.
- [5] T. P. Sari, A. L. Hananto, E. Novalia, T. Tukino, and S. S. Hilabi, “Implementasi Algoritma K-Means dalam Analisis Klasterisasi Penyebaran Penyakit Hiv/Aids,” *Infotek : Jurnal Informatika dan Teknologi*, vol. 6, no. 1, pp. 104–114, Jan. 2023.
- [6] M. Noperia, Ishak, and V. Winda Sari, “Implementasi Data Mining Pengelompokan Data Nilai Untuk Menentukan Minat Belajar Seni Budaya,” *JURNAL SISTEM INFORMASI TGD*, vol. 2, no. 1, pp. 65–72, Jan. 2023.
- [7] G. Gustientiedina, M. H. Adiya, and Y. Desnelita, “Penerapan Algoritma K-Means Untuk Clustering Data Obat-Obatan Pada RSUD Pekanbaru,” *Jurnal Nasional Teknologi dan Sistem Informasi*, vol. 5, no. 1, pp. 17–24, Apr. 2019.
- [8] F. M. Basysyar, Y. Arie Wijaya, I. Ali, and S. Anwar, “CLUSTERING DATA DISABILITAS MENGGUNAKAN ALGORITMA K-MEANS DI KABUPATEN CIREBON,” *JURSIMA (Jurnal Sistem Informasi dan Manajemen)*, vol. 9, no. 3, Dec. 2021.
- [9] R. Amelia, M. Martanto, and A. Bahtiar, “PENERAPAN ALGORITMA K-MEANS CLUSTERING PADA PENGELOMPOKAN PASANGAN USIA SUBUR PESERTA KB DI KABUPATEN CIREBON,” *JURNAL TEKNOLOGI TECHNOSCIENTIA*, pp. 9–14, Sep. 2022.
- [10] R. Hidayat and H. Kusniyati, “Analisis Clustering Dalam Pengelompokan Penjualan Menggunakan Algoritma K-Means Pada Cafe 47°Coffee,” *CESS (Journal of Computing Engineering, System and Science)*, vol. 7, no. 2, pp. 420–434, Jul. 2022.
- [11] A. Muhidin and S. Kevin Alfandara, “ANALISIS DATA MINING MENGGUNAKAN ALGORITMA K-MEANS UNTUK CLUSTERING PENJUALAN STUDI KASUS DAPUR BU IPUNG,” *SIGMA - Jurnal Teknologi Pelita Bangsa*, vol. 13, no. 2, Jun. 2022.
- [12] L. M. Harahap, W. Fuadi, L. Rosnita, E. Darnila, and R. Meiyanti, “Klastering Sayuran Unggulan Menggunakan Algoritma K-Means,” *Jurnal Teknik Informatika dan Sistem Informasi*, vol. 8, no. 3, Dec. 2022.
- [13] Z. Nabila, A. Rahman Isnain, and Z. Abidin, “ANALISIS DATA MINING UNTUK CLUSTERING KASUS COVID-19 DI PROVINSI LAMPUNG DENGAN ALGORITMA K-MEANS,” *Jurnal Teknologi dan Sistem Informasi (JTSI)*, vol. 2, no. 2, p. 100, 2021.