

## PENGELOMPOKAN KATALOG PADA TOKO MAHABBATAIN STORE MENGUNAKAN ALGORITMA K-MEANS

Siti Farikah<sup>1</sup>, Rini Astuti<sup>2</sup>, Fadhil M Basysyar<sup>3</sup>

<sup>1</sup>Program Studi Teknik Informatika, STMIK IKMI Cirebon

Jl. Perjuangan No.10B, Karyamulya, Kesambi, Kota Cirebon, Jawa Barat, Indonesia

<sup>2</sup>Program Studi Sistem Informasi, STMIK LIKMI Bandung

Jl. Ir. H. Djuanda No. 96, Bandung, Jawa Barat, Indonesia

<sup>3</sup>Program Studi Sistem Informasi, STMIK IKMI Cirebon

Jl. Perjuangan No.10B, Karyamulya, Kesambi, Kota Cirebon, Jawa Barat, Indonesia

Sitifarikah74@gmail.com

### ABSTRAK

Di Indonesia perubahan sosial yang disertai kemajuan ini salah satunya adalah sandang. Pakaian menjadi salah satu fungsi dari diferensiasi sosial dalam kehidupan sosial di masyarakat modern. Banyak toko menjual berbagai Katalog mulai dari fashion, sepatu, krudung produk lainnya yang dijual belikan. Permasalahannya adalah belum adanya klasifikasi terhadap menentukan jumlah produk terhadap pengelompokan katalog pada Toko Mahabbatain store dan bagaimana algoritma k-means dapat melakukan pengelompokan katalog pada Toko Mahabbatain Store. Oleh karena itu, dengan pengaruh tersebut maka penulis ingin mengetahui tentang bagaimana algoritma k-means dalam pengelompokan katalog. Tujuan penelitian ini untuk mengetahui hasil klasifikasi terhadap jumlah produk terhadap pengelompokan katalog pada Toko Mahabbatain store dan untuk menarapkan algoritma k-means sebagai pengelompokan katalog pada Toko Mahabbatain Store. Metode yang digunakan dalam penelitian ini adalah algoritma *k-means* dan metode *Knowledge Discovery in Database (KDD)* karena metode ini dapat mengolah data tanpa diketahui labelnya. Dari penelitian yang sudah dilakukan secara observasi dan dokumentasi menunjukkan bahwa dataset dari pembukuan Toko Mahabbatain Store yang dengan jumlah 500 data penjualan pada tahun 2023. Tujuan penulisan akhir ini untuk mengetahui penerapan algoritma *K-means* dan mengetahui hasil k optimum dalam penerapan menggunakan *RapidMiner*. Hasil uji coba cluster distane performance bahwa nilai k optimum adalah k=2 dengan nilai DBI=0,047.

**Kata Kunci :** Pengelompokan, K-means, Produk, cluster.

### 1. PENDAHULUAN

Sejalan dengan perkembangan zaman, kebutuhan manusia di era modern saat ini tidak hanya terdiri atas sandang, papan dan pangan. Hal tersebut disebabkan dengan perkembangan dan perubahan yang berlangsung di masyarakat sehingga menciptakan banyak kebutuhan yang beragam. Di Indonesia perubahan sosial yang disertai kemajuan ini salah satunya adalah sandang. Pakaian menjadi salah satu fungsi dari diferensiasi sosial dalam kehidupan sosial di masyarakat modern.[1] pengelompokan K-Means merupakan metode analisis kelompok yang mengarah pada pemartisian N objek pengamatan ke dalam K kelompok (cluster) dimana setiap objek pengamatan dimiliki oleh sebuah kelompok dengan mean (rata-rata) terdekat, mirip dengan algoritma Expectation-Maximization untuk Gaussian Mixture dimana keduanya mencoba untuk menemukan pusat dari kelompok dalam data sebanyak iterasi perbaikan yang dilakukan oleh kedua algoritma [2]

Penelitian sebelumnya yang berjudul “Pengelompokan Kelas Menggunakan Self Organizing Map Neural Network Pada SMK N 1 Depok” membahas bahwa pengelompokan atau grouping adalah pengelompokan peserta didik berdasarkan karakteristiknya. Karakteristik demikian perlu digolongkan, agar mereka berada dalam kondisi yang sama. Adanya kondisi yang sama ini bisa

memudahkan pemberian layanan yang sama. Oleh karena itu, pengelompokan (*grouping*) ini lazim yang dengan istilah pengklasifikasikan (*clasifition*). Dengan diadakan pengelompokan agar pelaksanaan kegiatan proses belajar mengajar disekolah bisa berjalan lancar, tertib, dan bisa Stercapai tujuan-tujuan pendidikan yang telah diprogramkan.[3]

Penelitian ini bertujuan untuk mengetahui hasil klasifikasi terhadap jumlah produk terhadap pengelompokan katalog pada Toko Mahabbatain store dan untuk menarapkan algoritma k-means sebagai pengelompokan katalog pada Toko Mahabbatain Store. dengan menggunakan metode k-means, data dapat dikelompokkan ke dalam cluster-cluster yang memiliki kesamaan karakteristik berdasarkan jumlah produk, sehingga membantu dalam memahami preferensi pembeli, membedakan jenis produk, serta memungkinkan untuk memberikan rekomendasi atau strategi yang sesuai untuk masing-masing kelompok tersebut.

### 2. TINJAUAN PUSTAKA

#### 2.1. Penelitian Terdahulu

Penelitian menurut (Ika Anikah) membahas pengendalian ketersediaan merupakan bagian penting dari perusahaan. Metode *K-means* tahapan penelitian ini yaitu identifikasi masalah, studi literature, pengumpulan data, pemrosesan data, data mining dan

evaluasi. Tujuan penelitian ini adalah memudahkan pt SAS Group dalam menentukan stok persediaan barang. Hasil dari penelitian yaitu setiap cluster\_0 sebanyak 1282 atau sebesar 98.6%, cluster\_1 18 atau 1.4%. berdasarkan dari kelompok stok barang pada cluster\_0 memiliki rata-rata penjualan yang sebesar yaitu mencapai 2368 barang, arunya barang kelompok barang cluster\_0 perlu menyimpan banyak barang jarena produk yang paling banyak diminati.[4]

Penelitian menurut (sulisto) menyimpulkan bahwa algoritma *x-means* mendapatkan dataset 333 jenis data penjualan dan transaksi pada tahun 2019. Dengan kategori accessories, fashion, sepatu, kebutuhan keluarga dan produk yang dijual belikan. Kajian yang diusulkan yaitu klasterisasi penjualan produk menggunakan algoritma *x-means* yaitu untuk mengetahui informasi dan identifikasi produk mana yang paling laku dan produk mana yang kurang laku. Adanya pengelompokan barang yang ada digudang tidak dapat menumpuk dan sesuai dengan kebutuhan.[5]

Penelitian menurut (Holwati Edi Widodo) penulis meneliti pengelompokan obat merupakan arus penempatan atau tata letak obat ini sesuai untuk diproses standar. Menggunakan algoritma *k-means* menunjukkan hasil sebuah wawasan baru yaitu pengelompokan data berdasarkan 2 cluster yaitu, cluster\_1 merupakan kebutuhan tinggi ketersediaan 71 dari 100 kategori jumlah data yang diuji, cluster\_2 merupakan kategori obat ketersediaan sedang yaitu 29 dari 100 kategori jumlah data obat yang diuji. Pengujian dengan rapidminer maenghasilkan wawasan masing klaster memiliki anggota kelompok klaster dengan perhitungan manual seperti cluster\_0 pada *Rapidminer* memiliki 72 anggota klaster yang disebut klaster medium, cluster\_1 memiliki 72 anggota klaster *representasi* klaster tinggi. Dan cluster\_3 memiliki 3 anggota dengan *representasi* rendah. [6]

## 2.2. Pengelompokan

Pengelompokan atau grouping adalah pengelompokan peserta didik berdasarkan karakteristiknya. Karakteristik demikian perlu digolongkan, agar mereka berada dalam kondisi yang sama. Adanya kondisi yang sama ini bisa memudahkan pemberian layanan yang sama. Oleh karena itu, pengelompokan (*grouping*) ini lazim yang dengan istilah pengklasifikasikan (*clasifition*). Dengan diadakan pengelompokan agar pelaksanaan kegiatan proses belajar mengajar disekolah bisa berjalan lancer, tertib, dan bisa Stercapai tujuan-tujuan pendidikan yang telah diprogramkan.[3]

## 2.3. Katalog

Katalog adalah carik kartu daftar, atau buku memuat nama benda atau informasi tertentu yang ingin disampaikan. Katalog pun memiliki karakter seperti buku, yaitu sesuai kebutuhan pengguna dan hampir tanpa batas waktu.[7]

## 2.4. Algoritma K-means

*Algoritma X-means* merupakan algoritma yang digunakan untuk pengelompokan data. Algoritma *x-means* merupakan pengembangan dari *k-means*. *X-means clustering* digunakan untuk menyelesaikan salah satunya kelemahan utama dari *k-means clustering*, yaitu perlunya pengetahuan sebelumnya tentang jumlah cluster (K). Dalam metode ini, nilai sebenarnya dari K diperkirakan dalam suatu yang tidak diawasi cara dan hanya berdasarkan set data itu sendiri.[5]

## 2.5. Data Mining

Data mining adalah proses pencarian dan penggalian informasi dari tumpukan-tumpukan data dengan jumlah yang besar merupakan proses utama dari data mining, tujuan utama dari pengolahan data tersebut adalah menghasilkan informasi yang baru.[8] Data mining juga disebut proses *Knowlage Discovery Database (KDD)*, yaitu proses untuk menggali dan menganalisis sejumlah data dan mengeskrak informasi dan pengetahuan yang berguna. Sebagai satu ringkasan proses, data mining dapat dibagi menjadi beberapa tahapan yang bersifat interaktif, yaitu pengguna terlibat langsung dengan *Knowlage Discovery Database (KDD)*.

## 3. METODE PENELITIAN

Metode penelitian yang digunakan yaitu melalui pendekatan kuantitatif. dalam pengertiannya yaitu yang menggambarkan suatu keadaan atau kondisi yang terjadi disuatu daerah, yang pemecahan masalahnya secara tersistem berdasarkan data. Data kuantitatif adalah jenis data berupa angka yang digunakan dalam proses operasi matematika, data ini berasal data fakta yang ada. [9]

### 3.1. Teknik Pengumpulan Data

Teknik pengumpulan data adalah salah satu cara atau metode yang dapat digunakan untuk mengumpulkan data seperti:

#### a. Metode Observasi

Metode observasi merupakan suatu pengamatan sistematis yang dilakukan secara langsung ditempat penelitian tersebut. Bertujuan untuk memperoleh wawasan yang mendalam mengenai jumlah proses penting yang terkait.

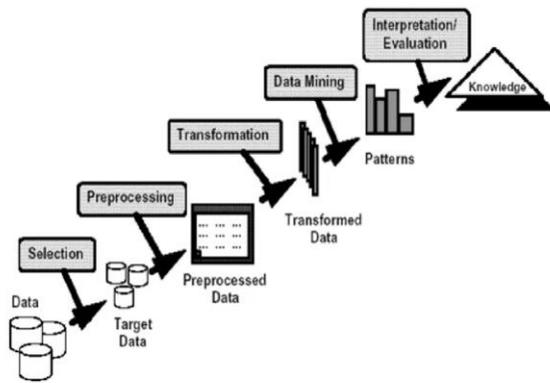
#### b. Metode Dokumentasi

Metode dokumentasi merupakan suatu Teknik pengumpulan data yang dilakukan dengan mempelajari pengolahan data yang bertujuan untuk memperoleh data dalam penelitian ini.

### 3.2. Tahapan Penelitian

*Data mining* merupakan tahap dari proses *Knowlage Discovery Database (KDD)*, yaitu proses untuk menggali dan menganalisis sejumlah data dan mengeskrak informasi dan pengetahuan yang berguna. Sebagai satu ringkasan proses, data mining dapat dibagi menjadi beberapa tahapan yang bersifat

interaktif, yaitu pengguna terlibat langsung dengan *Knowledge Discovery Database (KDD)*. [10]



Gambar 1. *Knowledge Discovery Database (KDD)*

a. Data Selection

Dalam tahap ini pemilihan data yang ada pada database merupakan tahapan KDD dimana database seringkali tidak semuanya dipakai. Oleh karena itu, hanya data yang sesuai dianalisis dilakukan seleksi atribut dan pengumpulan data, pengumpulan data diambil langsung dari Observasi dan Dokumentasi.

b. Preprocessing

Dalam tahap ini pembersihan data dan perbaikan. Pada saat pengumpulan data yang didapat biasanya data yang tidak berstruktur dan mengandung banyak karakter. Tujuan dari adanya preprocessing adalah untuk menghilangkan noise, dalam tahap preprocessing terdapat beberapa langkah yang harus dilakukan, seperti tahap cleaning, case folding, tokenize, dan filtering

c. Data Mining

Pada tahap data mining ini adalah proses klasifikasi analisis sentiment pada data pengetahuan kualitas produk terhadap pelanggan pada pengelompokan jenis brand dan katalog menggunakan metode K-means.

d. Evaluation

Pada tahap evaluation akan menggunakan confusion matriks untuk mengetahui hasil performa akurasi dari algoritma K-means dengan hasil beberapa accuracy, precision, dan recall. Dimana confusion matrix adalah salah satu metode untuk mengukur kinerja dari model klasifikasi.

4. HASIL DAN PEMBAHASAN

Penelitian ini menggunakan data dari pembukuan pada Toko Mahabbatain Store pengelompokan katalog pada tahun 2023 dengan jumlah record dataset 500 dataset yang digunakan dan hasil data dapat dilihat gambar 2

Gambar 2. Dataset

Penelitian ini melakukan proses clustering algoritma k-means menggunakan metode *Knowledge Discovery in Database (KDD)* metode tersebut ada 5 langkah. Langkah pertama data selection dengan mengelola data yaitu dengan menempatkan Port Read Excel untuk membaca dataset dalam format excel.



Gambar 3. Operator Red Excel

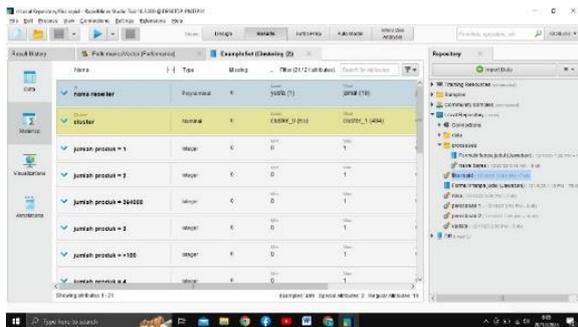
Langkah selanjutnya memilih import data untuk menganalisis dataset. Proses import melibatkan pemilihan lokasi file data, dalam dataset berada didokumen. Setelah itu dataset dipilih dan dimasukkan kedalam proses data sebagai berikut:

Gambar 4. Hasil Select Data

Tahap selanjutnya adalah data preprocessing atau pembersihan data dengan memeriksa data dari masing value dan pemilihan atribut yang relevan.

a. Missing value

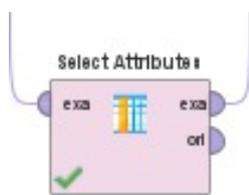
Langkah pertama dimulai dari menjalankan Red Excel di RapidMiner untuk mengakses dataset yang telah dimasukkan. Setelah dimasukkan data excel kemudian mengeksplorasi missing value yang terdapat dimana stastic. Berikut gambar proses missing value pada 5 dibawah ini:



Gambar 5. Missing Value

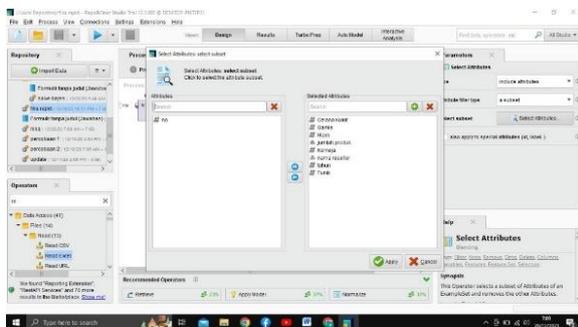
b. Select

Langkah selanjutnya dalam tahap preprocessing adalah menggunakan operator Select Atribut dengan menerapkan subset.



Gambar 6. Select Atribut

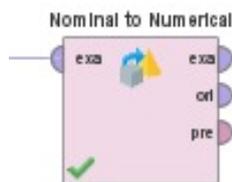
Setelah dalam proses ini, terjadi pemulihan atribut yang akan dihilangkan dari dataset. Atribut yang dipilih untuk dihilangkan adalah atribut No. sementara itu atribut yang akan digunakan adalah nama reseller, tahun, gamis, hijab, tunik, kemeja celana kulot, dan jumlah produk. Berikut gambar atribut yang akan diseleksi.



Gambar 7. Data Selection

Pada langkah transformation adalah kegiatan mengubah atribut ke dalam type data yang sesuai dengan Algoritma K-means.

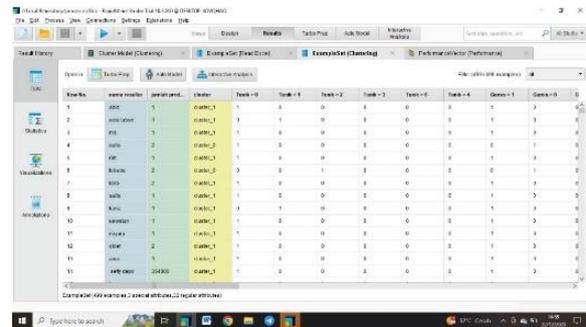
c. Nominal To Numerical



Gambar 8. Operator Nominal To Numeric

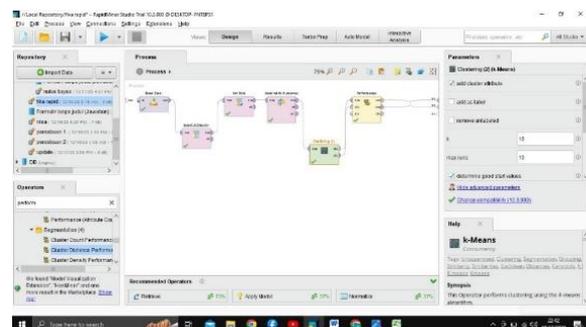
Langkah selanjutnya adalah melakukan transformation data dari tipe data nominal menjadi numeric yang sesuai. Seperti pada gambar 8.

Penggunaan operator nominal to numeric pada proses transformation data



Gambar 9. Operator Nominal To Numeric

Langkah selanjutnya melakukan pengelompokan data menggunakan algoritma K-means setelah proses transformation data menggunakan operator algoritma K-means clustering dalam RapidMiner.



Gambar 10. Operator K-Means Clustering

Langkah clustering K-means, dimana K 2 menandakan jumlah cluster yang ditentukan dalam proses clustering. Pengelompokan ini dilakukan proses maxrun sebanyak 10 kali percobaan, sebagai perbandingan untuk menentukan pengelompokan mana yang tepat untuk mengidentifikasi produk dengan kategori paling diminati, cukup diminati, dan kurang diminati dengan atribut yang berbeda-beda maka pengukuran Mixed Measure Interpretasi Data

Hasil dari data mining yang diperoleh dengan Rapidminer menggunakan algoritma k-means sebagai berikut:

Hasil pengelompokan menggunakan k-means diperoleh terbaik sebanyak 2 cluster, yaitu cluster 0 sebanyak 185 item, yaitu berdasarkan konsumen baru. dan cluster 1 sebanyak 314 item yaitu berdasarkan pelanggan.

**Cluster Model**

Cluster 0: 185 items  
Cluster 1: 314 items  
Total number of items: 499

Gambar 11. Cluster Model

Hasil cluster model tersebut ditampilkan juga pada example set yang berisi klasterisasi secara keseluruhan data yang telah di inputkan sebelumnya. .

Item No	Item Name	Item Price	Item Weight	Cluster 0	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7	Cluster 8	Cluster 9
1	pisang	1	0.5	0	1	0	0	0	0	0	0	0	0
2	jeruk	1	0.5	0	1	0	0	0	0	0	0	0	0
3	pisang	1	0.5	0	1	0	0	0	0	0	0	0	0
4	pisang	2	0.5	0	1	0	0	0	0	0	0	0	0
5	pisang	1	0.5	0	1	0	0	0	0	0	0	0	0
6	pisang	2	0.5	0	1	0	0	0	0	0	0	0	0
7	pisang	2	0.5	0	1	0	0	0	0	0	0	0	0
8	pisang	1	0.5	0	1	0	0	0	0	0	0	0	0
9	pisang	1	0.5	0	1	0	0	0	0	0	0	0	0
10	pisang	1	0.5	0	1	0	0	0	0	0	0	0	0
11	pisang	1	0.5	0	1	0	0	0	0	0	0	0	0
12	pisang	2	0.5	0	1	0	0	0	0	0	0	0	0
13	pisang	1	0.5	0	1	0	0	0	0	0	0	0	0
14	pisang	2	0.5	0	1	0	0	0	0	0	0	0	0

Gambar 12. Example Set (Clustering)

Langkah selanjutnya adalah evaluasi dari proses clustering tersebut. Untuk mengetahui hasil kualitas hasil kluster, pada proses pengujian menggunakan operator performance. Pada bagian parameter performance, kriteria utama yang digunakan adalah Davies Bouldin Index. Selain itu, parameter Maximize dan Normalize dipilih untuk mengoptimalkan proses evaluasi.

Hasil dari evaluasi gambar diatas menampilkan nilai Davies Bouldin Index sebesar 0,024 nilai ini menunjukkan bahwa hasil dari proses clustering dengan K=2 menunjukkan nilai Davies Bouldin Index yang paling optimal karena memiliki nilai DBI yang paling mendekati 0 diantara percobaan yang lain. Dengan penemuan nilai DBI sebesar 0,024 dinyatakan baik dalam penentuan jumlah Optimal cluster.

5. KESIMPULAN DAN SARAN

Dari hasil penelitian yang dilakukan penulis dapat menarik beberapa kesimpulan sebagai berikut: Dapat disimpulkan bahwa algoritma k-means dapat menyelesaikan permasalahan terhadap pengelompokan katalog pada toko mahabbatain didapatkan hasil dari kluster dua dengan nilai DBi 0,047. Hasil yang didapatkan pada dua kluster, cluster 0 185 item konsumen baru dan cluster 1 314 item, yaitu pelanggan lama. Pada pengelompokan menjadi dua kluster (k=2) dengan nilai DBi 0,047 dengan rata-rata centroid dari tiapcluster terdapat cluster 3 dengan average wirthin centroid distane 0,000. Saran dari penulis penelitian ini adalah membandingkan dengan algoritma lain untuk mengklasterisasi dataset pengelompokan katalog sehingga diharapkan : Dalam klasterisasi yang dilakukan oleh penulis dapat memiliki nilai akurasi yang lebih baik dan akurat. Dengan analisa pengelompokan katalog dengan menggunakan algoritma k-means ini semoga selanjutnya dapat membandingkan dengan algoritma lain.

DAFTAR PUSTAKA

- [1] P. M. Guarango, "No Title הכי קשה לראות את מה הארץ", שבאמת לנגד העיניים, no. 8.5.2017, pp. 2003–2005, 2022.
- [2] 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," *J. Media Infotama*, vol. 18, no. 1, pp. 99–104, 2022.
- [3] E. Harli, A. Fauzi, and T. H. Kusmanto, "Pengelompokan Kelas Menggunakan Self Organizing Map Neural Network pada SMK N 1 Depok," *J. Edukasi dan Penelit. Inform.*, vol. 2, no. 2, pp. 90–95, 2016, doi: 10.26418/jp.v2i2.17574.
- [4] Ika Anikah, Agus Surip, Nela Puji Rahayu, Muhammad Harun Al- Musa, and Edi Tohidi, "Pengelompokan Data Barang Dengan Menggunakan Metode K-Means Untuk Menentukan Stok Persediaan Barang," *KOPERTIP J. Ilm. Manaj. Inform. dan Komput.*, vol. 4, no. 2, pp. 58–64, 2022, doi: 10.32485/kopertip.v4i2.120.
- [5] M. R. Sulistio, N. Suarna, and O. Nurdiawan, "Analisa Penerapan Metode Clustering X-Means Dalam Pengelompokan Penjualan Barang," *J. Teknol. Ilmu Komput.*, vol. 1, no. 2, pp. 37–42, 2023, doi: 10.56854/jtik.v1i2.49.
- [6] W. H. Holwati, Edi Widodo, "Pengelompokan Untuk Penjualan Obat Dengan Menggunakan Algoritma K-Means," vol. 4, no. 2, pp. 408–413, 2023, [Online]. Available: <https://journal.fkpt.org/index.php/BIT/article/view/848/468>
- [7] D. Waode Munaeni, Muhammad Aris, Ismi Musdalifah Darsan, Rusmawati Labenua, "Jurnal abdi insani," *J. Abdi Insa.*, vol. 9, no. 1, pp. 247–255, 2022.
- [8] R. Muliono and Z. Sembiring, "Data Mining Clustering Menggunakan Algoritma K-Means Untuk Klasterisasi Tingkat Tridarma Pengajaran Dosen," *CESS (Journal Comput. Eng. Syst. Sci.*, vol. 4, no. 2, pp. 2502–714, 2019.
- [9] C. Hutabarat, "Penerapan Data Mining Untuk Memprediksi Permintaan Produk Kartu Perdana Internet Menggunakan Algoritma C5.0 (Studi Kasus: Vidha Ponsel)," *Pelita Inform.*, vol. 6, no. April, pp. 419–424, 2018.
- [10] Danya Rizki Chaerunisa, Nining Rahaningsih, Fadhil Muhammad Basysyar, Ade Irma Purnamasari, and Nana Suarna, "Pengelompokan Penjualan Madu Menggunakan Algoritma K-Means," *KOPERTIP J. Ilm. Manaj. Inform. dan Komput.*, vol. 5, no. 1, pp. 23–28, 2021, doi: 10.32485/kopertip.v5i1.144.