Journal of Informatics Management and Information Technology
ISSN 2774-4744 (Media Online)
Vol 5, No 3, July 2025 | Hal 167-176
DOI: 10.47065/jimat.v5i3.610
https://hostjournals.com/jimat

Implementasi K-Means Untuk Pengelompokan Kategori Penjualan Barang Berbasis Web

Shania Risky Agustin*, Intan Purnamasari , Betha Nurina Sari

Fakultas Ilmu Komputer, Informatika, Universitas Singaperbangsa Karawamg, Karawang, Indonesia
Email: 1,*shania.risky18097@student.unsika.ac.id, 2 Intan.Purnamasari@staff.unsika.ac.id, 3Betha.Nurina@staff.unsika.ac.id
Email Penulis Korespondensi: shania.risky18097@student.unsika.ac.id

Abstrak−Toko Harto Joyo memiliki kendala dalam pengelolaan stok barang akibat pencatatan yang masih manual, Permasalahan utama dalam pengelolaan data penjualan adalah kesulitan dalam mengelompokkan produk berdasarkan pola penjualan untuk mendukung strategi pemasaran. Penelitian ini bertujuan untuk merancang dan membangun sistem berbasis web yang dapat mengelompokkan produk menggunakan algoritma K-Means Clustering. Pengembangan sistem dilakukan dengan metodologi System Development Life Cycle (SDLC) model Waterfall dan menggunakan data penjualan sebanyak 1.630 record data periode Februari 2023 hingga Januari 2024. Dalam prosesnya data diolah terlebih dahulu menggunakan KDD (Knowledge Discovery in Databases) yaitu Data Selection, Data Preprocessing, Data Transformation, Data Mining, dan Knowledge Interpretation/Evaluation. Proses clustering menghasilkan 3 kelompok data (Cluster 1 sebanyak 9 data, Cluster 2 sebanyak 191 data, dan Cluster 3 sebanyak 28 data) dengan jumlah cluster yang ditentukan menggunakan metode Elbow. Cluster 1 merupakan kelompok produk terlaris yang memerlukan stok tinggi, Cluster 2 adalah produk tidak laris dengan kebutuhan stok minimal, dan Cluster 3 merupakan produk dengan tingkat kelarisan standar dimana stok barang pada kelompok ini harus tidak terlalu banyak dan tidak terlalu sedikit (sedang). Evaluasi menggunakan metode Davies Bouldin Index (DBI) menunjukkan nilai DBI sebesar 0,356 pada K=3, yang menunjukkan hasil clustering optimal karena mendekati nilai 0. Hasil pengujian awal menunjukkan bahwa sistem mampu mengelompokkan produk ke dalam beberapa klaster penjualan secara akurat dan dapat digunakan sebagai dasar pengambilan keputusan bisnis. Penelitian ini memberikan kontribusi dalam bentuk sistem analisis penjualan yang dapat diintegrasikan dengan sistem informasi perusahaan. Hasil analisis menunjukkan bahwa produk tidak laris mendominasi data, sehingga toko Harto Joyo perlu menerapkan strategi pemasaran seperti diskon atau promosi untuk meningkatkan penjualan.

Kata Kunci: Clustering; CodeIgniter; K-Means; Pengelompokan Barang; SDLC

Abstract−Harto Joyo Store faces challenges in inventory management due to manual recording processes. The main issue in managing sales data lies in the difficulty of grouping products based on sales patterns to support marketing strategies. This research aims to design and develop a web-based system capable of clustering products using the K-Means Clustering algorithm. The system was developed using the System Development Life Cycle (SDLC) with the Waterfall model, utilizing a dataset of 1,630 sales records from February 2023 to January 2024. The data was processed through the Knowledge Discovery in Databases (KDD) stages, which include Data Selection, Data Preprocessing, Data Transformation, Data Mining, and Knowledge Interpretation/Evaluation. The clustering process resulted in three data groups (Cluster 1 with 9 records, Cluster 2 with 191 records, and Cluster 3 with 28 records), with the number of clusters determined using the Elbow method. Cluster 1 represents best-selling products that require a high stock level, Cluster 2 includes low-demand products with minimal stock needs, and Cluster 3 consists of moderately-selling products that require a balanced inventory level. Evaluation using the Davies-Bouldin Index (DBI) yielded a DBI score of 0.356 for K=3, indicating optimal clustering results as it approaches zero. Initial testing shows that the system accurately classifies products into several sales clusters and can be used as a basis for business decision-making. This research contributes by providing a sales analysis system that can be integrated with the company's information system. The analysis reveals that low-selling products dominate the data, suggesting that Harto Joyo Store should implement marketing strategies such as discounts or promotions to boost sales.

Keywords: Clustering; CodeIgniter; K-Means; Product Grouping; SDLC

1. PENDAHULUAN

Teknologi informasi saat ini sudah sangat berkembang pesat dimana kecepatan dan ketepatan dalam mengolah dan menyimpan informasi penting dalam dunia berbisnis. Dengan banyaknya data ataupun informasi yang diperoleh akan tidak efektif bila diolah dengan cara manual karena membutuhkan waktu yang lama dan tingkat ketelitian yang rendah. Maka dari itu, dibutuhkan sebuah teknologi sistem yang dapat membantu dalam strategi bisnis.

Toko Harto Joyo adalah toko *online* yang berjualan di beberapa *marketplace* yaitu Tokopedia, Shopee dan GrabMart yang menjual produk kesehatan dan berbagai macam jenis produk lainnya. Selain itu, toko ini juga menjual produk secara luring. Berdasarkan observasi dan wawancara, terdapat permasalahan yaitu pada persediaan stok barang. Jumlah permintaan *customer* yang fluktuatif sering mengakibatkan persediaan stok barang menjadi tidak stabil.

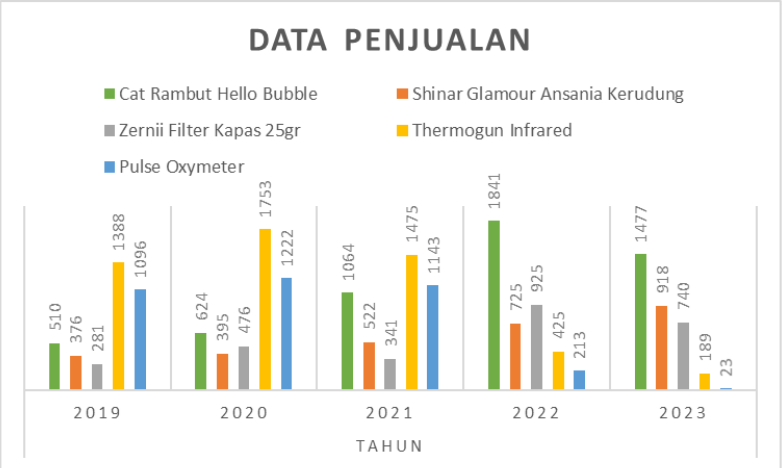
Dapat dilihat pada Gambar 1 bahwa permintaan permintaan *customer* fluktuatif dari tahun ke tahun. Pada tahun 2019 sampai 2021 terlihat bahwa produk kesehatan yaitu *thermogun* dan *oxymeter* memimpin, sedangkan pada 2022 dan 2023 terlihat bahwa produk cat rambut dan kerudung memiliki angka penjualan yang terbanyak.

Toko ini juga masih menggunakan cara manual untuk melakukan pemenuhan stok barang, yaitu dengan cara melakukan pencatatan pada buku. Sehingga menyebabkan kesalahan pencatatan data dan membutuhkan waktu yang lama untuk mengolahnya. Di samping itu, toko ini tidak dapat mengelompokkan produk yang terlaris dan tidak laris. Sehingga sering menyebabkan kekosongan barang yang laris dan penumpukan barang yang tidak laris.

Persaingan bisnis yang sangat ketat khususnya di bidang belanja *online* mengharuskan para pelaku bisnis memiliki strategi dalam mengelola bisnis mereka agar usaha mereka dapat berada di depan para pesaing. Seiring bertambah banyaknya tranksaksi penjualan harian, maka data juga akan terus bertambah dari waktu ke waktu. Jika data dibiarkan

Journal of Informatics Management and Information Technology
ISSN 2774-4744 (Media Online)
Vol 5, No 3, July 2025 | Hal 167-176
DOI: 10.47065/jimat.v5i3.610
https://hostjournals.com/jimat

tanpa kejelasan, maka data tersebut menjadi tak berarti. Karena itu, pelaku bisnis perlu menggunakan sistem komputerisasi untuk memaksimalkan bisnis mereka. Salah satu cara untuk mengatasi permasalahan yang telah dijelaskan sebelumnya yaitu dengan membuat sistem penentuan persediaan stok barang berbasis web yang memanfaatkan algoritma *K-Means* untuk mengelompokkan datanya.



Gambar 1. Data Penjualan Toko Harto Joyo 2019 - 2023

Beberapa penelitian telah mengimplementasikan algoritma K-Means untuk berbagai tujuan analisis data, seperti pengelompokan data penjualan, penentuan penerima beasiswa, analisis potensi wilayah, hingga manajemen stok barang. Misalnya, Aditya Agusti Achray (2020) menggunakan K-Means untuk mengelompokkan data penjualan mobil guna merumuskan strategi promosi yang lebih tepat di PT Honda Arista Mangga Dua [1]. Sementara itu, Sri Astuti memanfaatkan algoritma ini untuk menentukan kelayakan mahasiswa penerima beasiswa UPZ UIN Sumatera Utara [2]. Kedua sistem dibangun berbasis web dengan PHP dan MySQL, menggunakan pendekatan SDLC dan KDD secara berturut-turut. Adapun Ajie Prasetya dkk. (2023) memanfaatkan metode Elbow untuk menentukan jumlah cluster optimal dalam menganalisis pola penjualan produk di Traffic Room Summarecon Bekasi, dan hasilnya menunjukkan performa clustering yang cukup baik [3].

Penelitian lain oleh Muhammad Fikry Adiansyah (2020) mengembangkan aplikasi berbasis Laravel untuk menentukan persediaan barang, dan menggunakan validasi Davies Bouldin Index untuk menentukan jumlah cluster terbaik. Hasil menunjukkan bahwa dua cluster lebih optimal, dengan satu cluster untuk barang yang paling dibutuhkan dan satu untuk barang yang jarang diperlukan [4]. Sementara itu, Elly Muningsih dkk. (2021) menerapkan K-Means dengan optimasi jumlah cluster menggunakan DBI untuk mengelompokkan provinsi di Indonesia berdasarkan potensi desa dalam industri kecil dan mikro. Penelitian ini menggunakan data dari BPS dan alat bantu RapidMiner, dengan hasil tiga cluster yang menggambarkan tingkat potensi industri desa dari rendah hingga tinggi [5].

Penelitian ini bertujuan membangun sistem pengelompokan kategori penjualan barang berbasis web menggunakan algoritma K-Means Clustering, dengan metode pengembangan SDLC model waterfall dan perancangan sistem menggunakan diagram UML (use case, sequence, activity, dan class diagram). Aplikasi dikembangkan menggunakan PHP dan MySQL dengan framework CodeIgniter, serta diuji menggunakan metode Black Box Testing. Data yang digunakan sebanyak 1.630 record penjualan dari Toko Harto Joyo periode Februari 2023 hingga Januari 2024, yang diolah melalui tahapan KDD, yang membedakan dengan penelitian sebelumnya adalah penggunaan framework CodeIgniter dan penerapan dua metode evaluasi, yaitu Elbow dan Davies Bouldin Index, untuk menentukan jumlah cluster yang optimal.

Atas dasar permasalahan yang dipaparkan sebelumnya maka penulis ingin membuat suatu sistem pengelompokan persediaan stok barang berbasis web menggunakan *K-Means* yang dapat membantu usaha tersebut dalam mengatasi permasalahan persediaan stok barang serta Menganalis dan mengevaluasi hasil clustering algoritma *K-Means* dalam mengelompokkan kategori penjualan barang pada Toko Harto Joyo dengan metode *elbow* dan *Davies Bouldin Index* (DBI).

2. METODOLOGI PENELITIAN

2.1 System Development Life Circle

Merupakan metodologi klasik dalam pengembangan perangkat lunak yang digunakan untuk membangun atau memodifikasi sistem informasi melalui tahapan terstruktur. Model SDLC yang umum digunakan antara lain Waterfall, Agile, Prototype, dan RAD [6]. Model Waterfall adalah pendekatan linier dan berurutan yang terdiri dari enam tahap: analisis kebutuhan, desain sistem, implementasi, integrasi dan pengujian, penerapan, serta pemeliharaan. Setiap tahap harus diselesaikan sebelum melanjutkan ke tahap berikutnya, cocok untuk sistem berskala kecil yang membutuhkan integritas tinggi.

Journal of Informatics Management and Information Technology
ISSN 2774-4744 (Media Online)
Vol 5, No 3, July 2025 | Hal 167-176
DOI: 10.47065/jimat.v5i3.610
https://hostjournals.com/jimat

2.2 Data Mining

Proses untuk menemukan pola tersembunyi dalam data berukuran besar guna mendukung pengambilan keputusan [7]. Data mining dibagi menjadi dua jenis utama: *descriptive mining* (misalnya clustering, association) untuk menggambarkan karakteristik data, dan *predictive mining* (misalnya klasifikasi, regresi) untuk memprediksi nilai masa depan. Berdasarkan fungsinya, data mining dapat dikelompokkan ke dalam enam kategori: deskripsi, klasifikasi, prediksi, estimasi, clustering, dan asosiasi. Clustering, seperti algoritma K-Means, digunakan untuk mengelompokkan data berdasarkan kemiripan tanpa menggunakan variabel target [8].

2.3 Tahap Penelitian

Metode pengumpulan data dilakukan melalui tiga cara, di awali dengan wawancara dengan pihak toko untuk memahami permasalahan secara langsung, kemudian dilakukan observasi langsung terhadap proses keluar masuk barang di toko, dan terakhir melakukan studi literatur dari buku, jurnal, e-book, dan artikel ilmiah yang relevan untuk mendukung pengembangan sistem. Dengan menggunakan objek penelitian yaitu sistem penentuan persediaan stok barang pada Toko Harto Joyo di Mustika Jaya, Bekasi. Dimana toko ini menghadapi masalah penumpukan produk yang tidak laku dan kekurangan stok produk yang laris. Untuk mengatasi hal tersebut, dikembangkan sistem pengelompokan kategori penjualan berbasis web menggunakan algoritma K-Means Clustering dengan framework CodeIgniter. Data yang digunakan sebanyak 1.630 data penjualan periode Februari 2023 hingga Januari 2024.

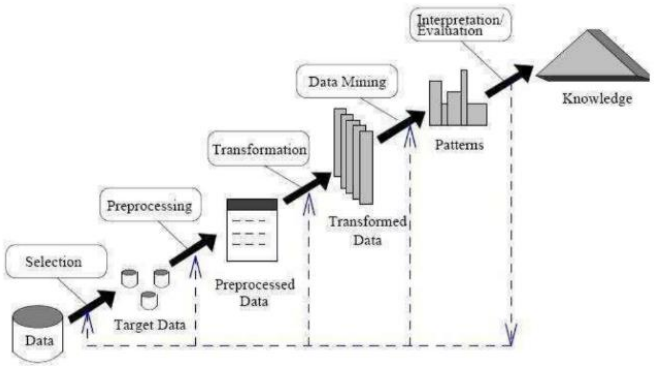
Rancangan penelitian ini menggunakan metode SDLC (System Development Life Cycle) dengan model waterfall yang terdiri dari beberapa tahapan berurutan. Tahap pertama adalah Requirements Analysis, yaitu analisis kebutuhan dan permasalahan sistem. Analisis ini mencakup identifikasi sistem berjalan untuk mengetahui kekurangan, serta pengusulan sistem baru untuk menyelesaikan permasalahan tersebut. Kebutuhan sistem dibagi menjadi fungsional, seperti fitur dan proses dalam sistem, serta non-fungsional, seperti kebutuhan perangkat keras dan lunak. Proses ini juga melibatkan tahapan data mining dengan pendekatan KDD (Knowledge Discovery in Databases), yang meliputi data selection, preprocessing (dengan RapidMiner), transformation, data mining menggunakan algoritma K-Means Clustering, dan evaluation menggunakan metode Elbow dan Davies Bouldin Index.

Tahap berikutnya adalah Design, yaitu menerjemahkan kebutuhan sistem ke dalam perancangan menggunakan UML (Unified Modeling Language) seperti use case diagram, activity diagram, dan sequence diagram. Desain database juga dilakukan menggunakan class diagram sebagai dasar pembentukan tabel dan relasinya. Selanjutnya, pada tahap Coding, sistem dikembangkan menggunakan Visual Studio Code dan Laragon, dengan bahasa pemrograman PHP dan MySQL. Framework yang digunakan adalah CodeIgniter untuk backend dan Bootstrap untuk antarmuka pengguna.

Setelah proses coding selesai, dilakukan Testing dengan metode Black Box Testing, yang menguji fungsionalitas sistem berdasarkan input dan output tanpa melihat implementasi kode. Pengujian dilakukan terhadap tampilan, fitur, dan akurasi sistem untuk memastikan kesesuaian dengan kebutuhan pengguna. Jika ditemukan kesalahan, dilakukan perbaikan pada sistem.

Tahap akhir adalah Maintenance, yaitu pemeliharaan sistem yang mencakup pembaruan teknologi, perbaikan kesalahan dari fase sebelumnya, serta penyesuaian sistem sesuai kebutuhan di masa mendatang agar tetap optimal dan relevan digunakan.

Lima tahapan tersebut adalah bagian dari KDD, yaitu serangkaian proses yang bertujuan untuk menemukan, menganalisis, serta mengekstrak informasi dan pengetahuan yang berguna dari data dalam jumlah besar [9] seperti digambarkan pada Gambar 2 dibawah ini.



Gambar 2. Tahapan Data Mining pada KDD

2.4 Clustering

Clustering termasuk kedalam kelompok *unsupervised learning. Clustering* (klasterisasi) adalah suatu teknik atau metode untuk mengelompokkan data. *Clustering* merupakan salah satu teknik penting dalam analisis data yang berfungsi untuk mengelompokkan sekumpulan objek ke dalam *cluster* berdasarkan kemiripan karakteristik mereka. Dalam proses ini, objek-objek yang berada dalam *cluster* yang sama memiliki tingkat kesamaan yang lebih tinggi dibandingkan dengan objek di *cluster* lainnya. Konsep ini mencerminkan prinsip dasar *clustering*, yaitu menjaga "keseragaman di dalam *cluster* dan perbedaan antar *cluster*" [10].

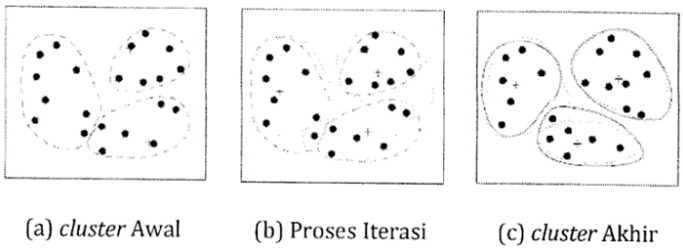
Journal of Informatics Management and Information Technology
ISSN 2774-4744 (Media Online)
Vol 5, No 3, July 2025 | Hal 167-176
DOI: 10.47065/jimat.v5i3.610
https://hostjournals.com/jimat

Tujuan utama dari teknik *clustering* adalah untuk mengungkap struktur tersembunyi dalam data yang bersifat tidak terstruktur atau semi-terstruktur. Dengan mengelompokkan data berdasarkan kemiripan fitur, *clustering* memungkinkan pengenalan pola dan hubungan yang tidak tampak secara langsung. Alasan utama penggunaan teknik klastering yaitu untuk membuat kelompok berdasarkan kesamaan data. Beberapa jenis *clustering* yang paling umum, yaitu:

a. *Partition Clustering*
   Ide inti dari metode *clustering* ini yaitu menganggap titik pusat data sebagai pusat *cluster*. Metode *K-Means* merupakan contoh representatif dari teknik ini. Metode *K-Means* mengambil n observasi dan sebuah bilangan bulat, k, dan mengeluarkan partisi dari n observasi ke dalam k set sedemikian rupa sehingga setiap observasi termasuk dalam *cluster* dengan *mean* terdekat. Metode *K-Means* menunjukkan kompleksitas waktu yang rendah dan efisiensi komputasi yang tinggi.
b. *Hierarchical Clustering*
   Algoritma ini menguraikan kumpulan data secara hierarki untuk memfasilitasi pengelompokan selanjutnya. Algoritma umum untuk *hierarchical clustering* mencakup BIRCH, CURE, dan ROCK.
c. *Clustering According to Density*
   Metode ini bertujuan untuk menemukan *cluster* yang berubah-ubah (arbitrer), dan DBSCAN adalah algoritma yang paling representatif. DBSCAN tidak perlu memasukkan jumlah *cluster* yang akan dipartisi dan dapat menangani *cluster* dengan berbagai bentuk. Namun, kompleksitas waktu dari algoritma ini tinggi. Selain itu, kepadatan data yang tidak teratur mengurangi kualitas *cluster*. Oleh karena itu, DBSCAN tidak dapat menangani data berdimensi tinggi.
d. *Clustering According to a Grid*
   Partition dan hierarchical clustering tidak dapat mengidentifikasi cluster dengan bentuk non-konveks, dan algoritma berbasis dimensi memiliki kompleksitas waktu yang tinggi. Keuntungan utama metode ini adalah kecepatan pemrosesan tinggi dan ketergantungan hanya pada jumlah unit di setiap dimensi ruang terkuantisasi [11].

2.5 K-Means

K-Means adalah salah satu algoritma clustering yang cukup populer karena kemampuannya dalam mengelompokkan data dalam jumlah besar secara efisien dan cepat. Algoritma ini termasuk dalam metode partitional *clustering* dan berbasis pada *centroid* atau titik pusat. Proses *K-Means* dimulai dengan menentukan jumlah klaster awal yang diinginkan dan menetapkan titik centroid awal. Algoritma ini kemudian akan secara acak memilih k data sebagai *centroid* awal. Proses iterasi dilakukan hingga tercapai kondisi stabil, yaitu ketika tidak ada lagi data yang berpindah dari satu klaster ke klaster lainnya. Jumlah iterasi yang dibutuhkan bergantung pada pemilihan *centroid* awal secara acak. *Output* dari algoritma ini adalah kumpulan klaster dengan titik *centroid* akhirnya dapat dilihat pada **Gambar 3.** dibawah ini yang menampilkan proses clustering menggunakan K-Means [12].



Gambar 3. Proses *Clustering* Objek menggunakan Metode *K-Means*

Algoritma K-Means memiliki kelebihan dibandingkan algoritma lain seperti cukup mudah untuk di pahami dan diterapkan, proses pembelajaran memahami *K-Means* terbilang membutuhkan waktu yang *relative* cepat dan algoritma ini cukup popular dan sering digunakan untuk teknik *clustering*. kekurangan dari algoritma *K-Means Clustering* menurut Nursyafitri, 2022 adalah untuk mendapatkan nilai *k* yang optimal, harus menggunakan metode lain untuk menginisialisasi nilai *k*, jika nilai acak (random) titik *centroid* tidak memberikan hasil yang baik, maka hasil clustering yang diperoleh tidak akan maksimal dan mencari jarak dari data dengan banyak dimensi menggunakan *K-Means* terbilang cukup sulit [13].

2.5.1 Tahapan Alur *K-Means*

Adapun tahapan-tahapan alur algoritma *K-Means Clustering* menurut Muhima sebagai berikut:

a. Menentukan nilai *k*, dimana *k* merepresentasikan jumlah *cluster* yang digunakan untuk mengelompokkan data pada penelitian.
b. Menentukan *centroid*, yaitu menentukan titik pusat awal *cluster* secara acak. Sementara itu, untuk *centroid* berikutnya pada cluster ke-i dapat dihitung menggunakan rumus berikut:

$$v = \frac{\sum_{i=1}^{n} x_j}{n} j = 1,2,3, \ldots n \quad (1)$$

Keterangan:

Journal of Informatics Management and Information Technology
ISSN 2774-4744 (Media Online)
Vol 5, No 3, July 2025 | Hal 167-176
DOI: 10.47065/jimat.v5i3.610
https://hostjournals.com/jimat

v = Titik *centroid* pada *cluster*
j = Objek ke-*j*
n = Banyaknya jumlah objek atau data dalam *cluster*

c. Menghitung jarak antara objek (data) dan *centroid*. Untuk menghitung jarak tersebut dapat menggunakan metode *Euclidian Distance*. Berikut adalah rumus dari metode *Euclidian Distance*:

$$d(x_i, y_j) = \sqrt{\sum(x_i, y_j)^2} \quad (2)$$

Keterangan:
$d(x_i, y_j)$ = Jarak data-*i* ke titik *centroid- j*
x_i = Data ke-i
y_j = Titik *centroid* ke-*j*

d. Kelompokkan masing-masing data berdasarkan jarak *centroid* terdekat.
e. Ulangi langkah 3 ke sampai ke 5 sampai proses iterasi berhenti. Iterasi akan berhenti apabila:
   1. Terjadi konvergensi (tidak mendapatkan nilai yang lebih baik pada iterasi berikutnya).
   2. Telah mencapai iterasi maksimal yang telah ditentukan sebelumnya.
   3. Apabila selama proses *clustering* tidak mengubah *cluster* data yang sebelumnya (data tidak berpindah *cluster*), proses dapat diakhiri [14].

2.6 Evaluasi Model Clustering

Pengelompokan data yang ideal ditentukan dengan mengevaluasi model *clustering*. Proses evaluasi digunakan untuk mengetahui akurasi terbaik dari proses *clustering* dan memberikan rekomendasi hasil *clustering* yang akan digunakan [15]. Untuk memilih jumlah k kluster yang optimal, dapat menggunakan metode *Elbow*, *Davies Bouldin Index* atau *Silhouette Coeffiecient* [16].

2.6.1 Metode *Davies Bouldin Index* (DBI)

Validasi clustering adalah evaluasi kuantitatif terhadap hasil analisis klaster. Salah satu metode yang sering digunakan adalah Davies-Bouldin Index (DBI), yang mengukur kualitas clustering berdasarkan jarak antar cluster dan jarak antar data dalam satu cluster. DBI memaksimalkan jarak antar cluster (SSB) dan meminimalkan jarak dalam cluster (SSW). Semakin rendah nilai DBI, semakin baik clustering tersebut, karena menunjukkan homogenitas internal tinggi dan heterogenitas eksternal yang baik. Nilai DBI berkisar antara 0 hingga 1, dengan nilai lebih mendekati 0 menandakan kemiripan tinggi antar sampel dalam cluster, sementara nilai mendekati 1 menunjukkan kemiripan rendah [17].

Sum of Square Within Cluster (SSW) adalah rumus yang digunakan untuk mengukur tingkat kohesi atau kekompakan data dalam suatu cluster ke-i. Persamaan ini menunjukkan seberapa dekat data-data dalam *cluster* tersebut terhadap pusatnya (*centroid*):

$$SSW_i = \frac{1}{m_i} \sum_{j=1}^{m_i} d(x_j, c_i) \quad (3)$$

Keterangan:
SSW_i = *Sum of square within cluster* ke-*i*
m_i = Jumlah data dalam cluster ke-*i*
c_i = *Centroid cluster* ke-*i*
$d(x_j, c_i)$ = Jarak dari data-*i* ke titik *cluster-i* (dihitung menggunakan *Euclidean Distance*)

Setelah mendapatkan nilai rasio, maka dilakukan perhitungan *Davies Bouldin Index* (DBI) dengan persamaan:

$$DBI = \frac{1}{k} \sum_{i=1\ \&\ j=1}^{k} max_{i \neq j}(R_{i,j}) \quad (4)$$

Keterangan:
DBI = *Davies Bouldin Index*
k = Jumlah *cluster*

2.6.2 Metode Elbow

Metode *Elbow* adalah suatu metode pada titik tertentu akan terjadi grafik penurunan secara drastis yang disebut dengan kriteria siku, yaitu membentuk seperti sebuah lekukan. Metode *elbow* bertujuan untuk menentukan nilai *k* yang optimal, yaitu nilai *k* terkecil yang masih menghasilkan nilai *within-cluster sum of squares* (*withinss*) yang rendah. metode ini divisualisasikan melalui grafik yang menunjukkan hubungan antara jumlah cluster dan penurunan tingkat error. Nilai *k* terbaik biasanya terlihat pada titik siku grafik tersebut. Evaluasi jumlah *cluster* dilakukan menggunakan teknik SSE (*Sum of Squared Error*), yaitu perhitungan yang digunakan untuk menilai seberapa besar perbedaan antara data aktual dan hasil klasterisasi yang dihasilkan. Dibawah ini adalah rumus dari SSE :

$$SSE = \sum_{k=1}^{k} \sum_{x_i} |x_i - c_k|^2 \quad (5)$$

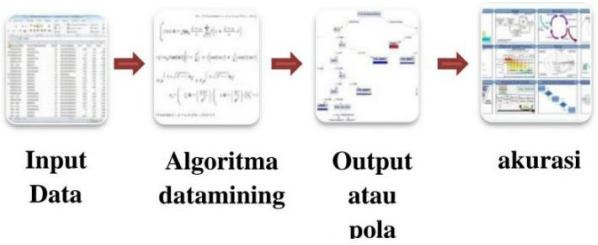
Journal of Informatics Management and Information Technology
ISSN 2774-4744 (Media Online)
Vol 5, No 3, July 2025 | Hal 167-176
DOI: 10.47065/jimat.v5i3.610
https://hostjournals.com/jimat

Dimana k merupakan Cluster ke-c, x_i adalah Jarak data pada objek ke-i dan c_k adalah Pusat kelompok ke-i. Apabila nilai *cluster* semakin besar, maka nilai SSE akan semakin kecil [18].

2.7 Rapid Miner

RapidMiner adalah platform open-source untuk data mining dan analisis bisnis yang memungkinkan pengguna mengekstraksi pengetahuan dari data dan membangun model prediktif tanpa memerlukan pengetahuan pemrograman. Dengan antarmuka visual yang mudah digunakan, RapidMiner menyediakan sekitar 500 operator untuk berbagai proses seperti input, output, praproses data, dan visualisasi. Aplikasi ini dapat berjalan di berbagai sistem operasi berkat penggunaan bahasa Java dan juga dapat diintegrasikan dengan aplikasi lain [19].

RapidMiner juga menawarkan antarmuka grafis (GUI) yang memudahkan pengguna dalam merancang alur analisis data, yang kemudian disimpan dalam format XML untuk dieksekusi secara otomatis dalam analisis data. Langkah utama dalam Rapid Miner dapat dilihat pada Gambar 4 sebagai berikut :



Gambar 4. Tahapan Pemrosesan Data pada RapidMiner

2.8 Framework CodeIgniter

CodeIgniter adalah framework PHP ringan dan populer yang menggunakan pola desain Model View Controller (MVC) untuk pengembangan aplikasi web. Framework ini cepat, mudah diinstal, dan memiliki dokumentasi lengkap yang memudahkan pengembang, menjadikannya pilihan utama untuk aplikasi PHP dinamis [20]. Beberapa fitur utamanya termasuk kecepatan tinggi, kemudahan konfigurasi, dukungan untuk berbagai database (MySQL, SQLite, PostgreSQL), dan fitur keamanan built-in seperti perlindungan terhadap SQL injection dan XSS. CodeIgniter juga mendukung berbagai library dan helper, serta memiliki sistem error handling yang memudahkan *debugging* [21].

3. HASIL DAN PEMBAHASAN

Dalam pembuatan sistem ini, metodologi penelitian yang akan digunakan adalah metode *System Development Life Cycle* (SDLC) dengan model *waterfall*. SDLC yaitu siklus yang digunakan untuk membuat atau mengembangkan sistem informasi untuk memecahkan masalah secara efektif yang akan terbagi menjadi 5 tahapan utama dalam penelitian ini yaitu :

3.1 Requirement Analysis

Tahapan ini dilakukan untuk mengumpulkan kebutuhan yang diperlukan dalam pembuatan aplikasi pengelompokan kategori penjualan barang berbasis web pada toko Harto Joyo. Pengumpulan ini dilakukan dengan dua cara yaitu wawancara dan observasi.

Toko Harto Joyo menghadapi beberapa masalah dalam pengelolaan data dan strategi pemasaran, di antaranya, data penjualan tidak dikelola dengan baik, hanya disimpan sebagai arsip, mengakibatkan penumpukan data yang tidak berguna, tidak ada analisis produk yang diminati, sehingga promosi iklan dilakukan secara merata untuk semua produk, mengakibatkan biaya iklan yang berlebihan dan fluktuasi permintaan menyebabkan ketidakstabilan stok, dengan kekurangan produk laris dan penumpukan produk yang kurang diminati.

3.1.1 Penerapan Tahapan Data Mining

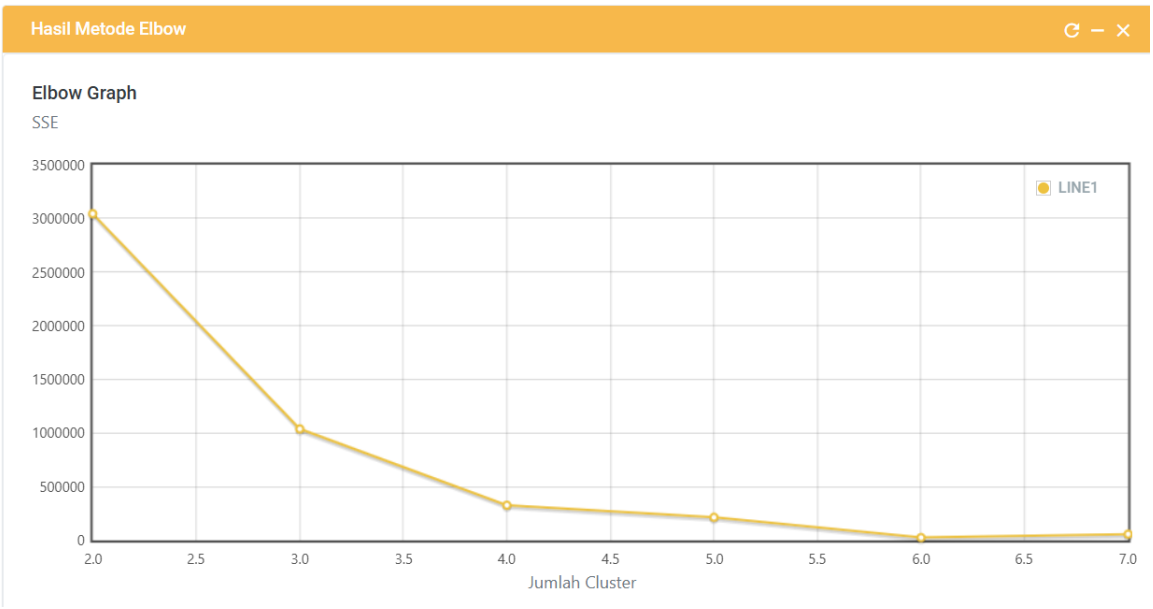
Tahapan data mining diperlukan untuk menyederhanakan proses memperoleh data yang dapat digunakan sebagai informasi yang berguna dari data skala besar.

a. *Data Selection*
   1. Pengumpulan Data
      Pengumpulan data merupakan tahap pengumpulan dataset awal hasil observasi yang didapatkan dari data primer toko Harto Joyo. Data yang diperoleh berupa data soft file dengan format excel, data tersebut adalah data penjualan produk sebanyak 1630 record data pada periode Februari 2023 sampai Januari 2024 (1 tahun). Pada dataset awal terdiri dari 10 atribut yaitu No, Nama Produk, Kategori, Periode, Pendapatan, Produk Terjual, Produk Dilihat, Keranjang, Wishlist dan Pesanan.
   2. *Data Preprocessing*
      Pembersihan Data Tidak Bermakna dengan Data dengan nama produk "Produk Tidak Tersedia" dihapus karena tidak relevan, sehingga jumlah data berkurang dari 250 menjadi 228 record kemudian pengecekan missing value

Journal of Informatics Management and Information Technology
ISSN 2774-4744 (Media Online)
Vol 5, No 3, July 2025 | Hal 167-176
DOI: 10.47065/jimat.v5i3.610
https://hostjournals.com/jimat

dan duplikasi menggunakan RapidMiner, dilakukan pengecekan terhadap data kosong dan data duplikat. Hasilnya, tidak ditemukan missing value maupun duplikasi pada 228 data yang tersisa.

3. *Data Transformation*
   Transformasi data dilakukan untuk mengubah tipe data nominal menjadi numerik agar dapat diproses dengan algoritma data mining. Namun, karena seluruh atribut data sudah sesuai setelah tahap seleksi dan pembersihan, tidak diperlukan lagi perubahan tipe data.
4. *Data mining*
   Salah satu kelemahan algoritma K-Means yakni saat menentukan nilai k atau jumlah cluster yang optimal, maka dari itu peneliti menggunakan metode Elbow agar dapat mencari nilai k yang terbaik. Pada tahap ini akan dilakukan pengujian jumlah cluster K = 1 sampai K = 7.

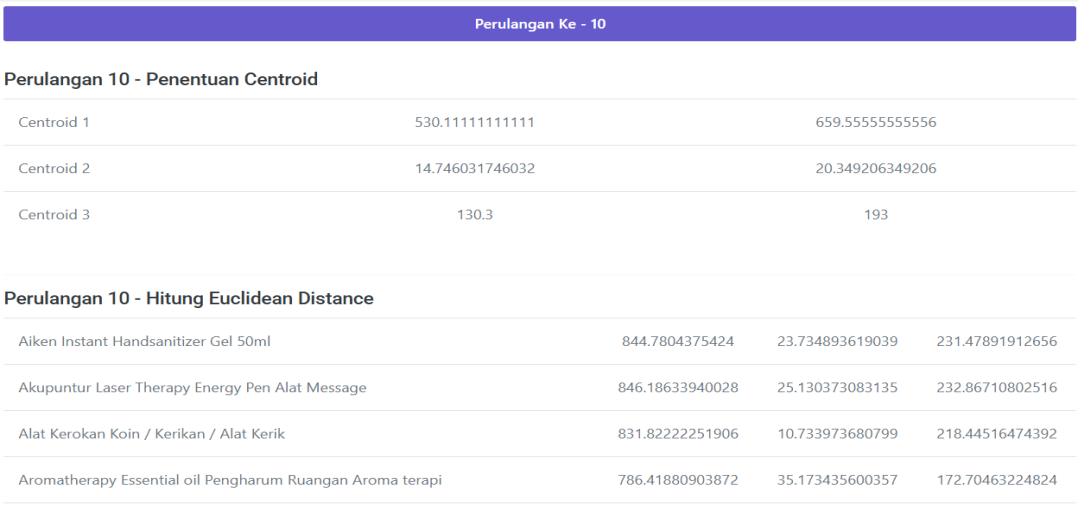


Gambar 5. Hasil Elbow pada tahap Data mining

Dapat dilihat pada Gambar 5 diatas hasil dari implementasi menunjukkan bahwa jumlah cluster yang optimal dan membentuk siku adalah 3 cluster. Maka data produk akan dibagi menjadi kelompok yaitu terlaris, standar, dan tidak laris.

Setelah mendapatkan nilai k yang optimal dari hasil elbow, didapatkan hasil 3 cluster. Jumlah titik centroid (jumlah kelompok data) dimasukan ke dalam sistem dan batas maksimal perhitungan. Selain itu dalam tahap ini titik centroid awal ditentukan secara acak atau random, titik diambil dari sample data penjualan produk.

proses perhitungan data mining dengan Clustering K-Means dengan aplikasi pengelompokan kategori penjualan barang di toko Harto Joyo berbasis web, Pada data hasil clustering diatas terdapat 10 iterasi yang mehasilkan Cluster 1 (Produk Terlaris) sebanyak 9 data, Cluster 2 (Produk Tidak Laris) sebanyak 191 data, dan Cluster 3 (Produk Standar) sebanyak 28 data. Dalam prosesnya, hasil cluster di jabarkan pada Gambar 6, Gambar 7 dan Gambar 8 dibawah ini:



Gambar 6. Data Hasil Perhitungan Clustering

Journal of Informatics Management and Information Technology
ISSN 2774-4744 (Media Online)
Vol 5, No 3, July 2025 | Hal 167-176
DOI: 10.47065/jimat.v5i3.610
https://hostjournals.com/jimat

Hasil Cluster KMeans

nama_produk	pesanan	produk_terjual	Cluster
Neurobion / Neurobion Putih / Neurobion Pink / Neurobion Forte	534	554	1
Thermo hygrometer digital thermometer thermo hygro	408	500	1
NICE Facial Tissu isi 180 sheets 2 ply	419	458	1
Mise En Scene Cat Rambut Hello Bubble | Perfect Color | Hello Cream	1246	1373	1
Shinar Glamour Laser Cut Glitter Hijab Jilbab Kerudung Segi Empat	510	535	1
Shinar Glamour Ansania Square - Jilbab Hijab Kerudung Segiempat Sinar	738	777	1
Korek Api / Korek Gas	262	502	1

Gambar 7. Data Hasil Pengelompokan Clustering

Jumlah Cluster

Cluster	Jumlah	Deskripsi
1	191	Produk Tidak Laris
2	9	Produk Terlaris
3	28	Produk Standar

Gambar 8. Jumlah Data pada Setiap Cluster

Setelah proses data mining, Selanjutnya adalah melakukan tahap evaluasi untuk mengetahui kualitas cluster. Metode evaluasi DBI (Davies Bouldin Index) digunakan untuk mengetahui clustering terbaik dengan mengetahui nilai DBI yang mendekati 0. Pada proses evaluasi ini diuji nilai K = 2 sampai K = 7. Proses DBI dilakukan menggunakan tools RapidMiner dan bantuan operator Cluster Distance Perfomance.

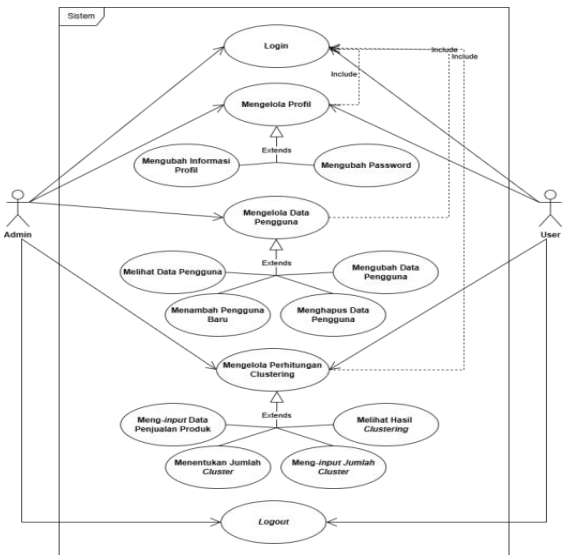
Tabel 1. Nilai DBI pada proses hasil clustering

No.	Jumlah K *Cluster*	Nilai DBI
1	K = 2	0.488
2	K = 3	0.356
3	K = 4	0.451
4	K = 5	0.524
5	K = 6	0.478
6	K = 7	0.505

Dapat dilihat dari Tabel 1 diatas bahwa nilai DBI yang paling mendekati nilai 0 adalah K = 3 dengan nilai DBI 0.356. Hasil tersebut dapat dikatakan hasil yang terbaik dibandingkan dengan nilai jumlah K yang lainnya.

3.2 Design

Dalam merancang sebuah program, desain dibagi menjadi dua bagian utama, yaitu desain antarmuka pengguna (interface) dan desain sistem. Pada tahap desain sistem, Use Case Diagram digunakan untuk menggambarkan fungsi-fungsi yang terdapat dalam sistem serta aktor-aktor yang memiliki akses terhadap fungsi tersebut. Dalam penelitian ini, perancangan sistem digambarkan dan dirangkum melalui sebuah Use Case Diagram yang mencakup keseluruhan proses sistem yang akan dibangun, Gambar 9 menunjukan bagaimana desain sistem dibuat.



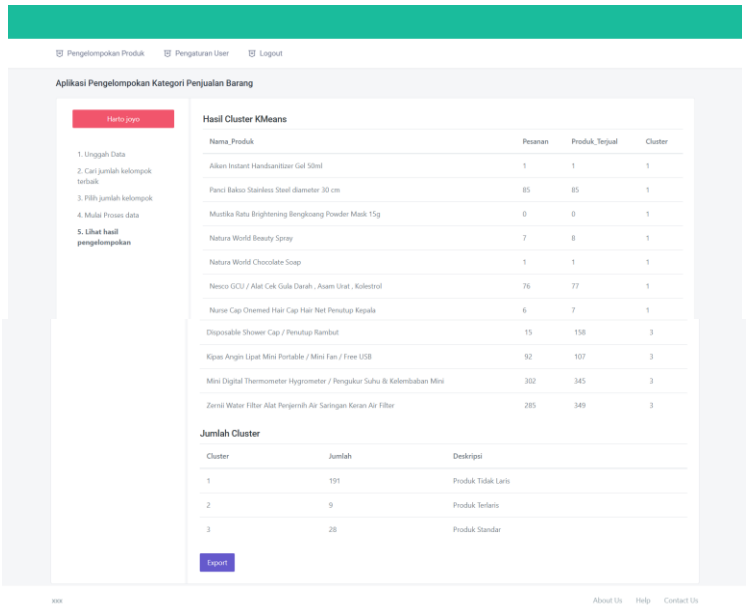
Gambar 9. Usecase Diagram Sistem Pengelompokan Barang

Journal of Informatics Management and Information Technology
ISSN 2774-4744 (Media Online)
Vol 5, No 3, July 2025 | Hal 167-176
DOI: 10.47065/jimat.v5i3.610
https://hostjournals.com/jimat

3.3 Coding

Pada tahap ini, desain yang telah dibuat diimplementasikan dengan menulis kode PHP menggunakan Visual Studio Code dan framework CodeIgniter. Laragon digunakan sebagai server localhost, sementara MySQL berfungsi sebagai database. Tampilan antarmuka pengguna dibuat dengan Bootstrap, menghasilkan Aplikasi Pengelompokan Kategori Penjualan Barang Berbasis Web.

Aplikasi ini memiliki beberapa halaman utama: halaman login untuk validasi akses; halaman unggah data untuk mengimpor file Excel berisi data penjualan; halaman elbow method untuk menentukan jumlah cluster optimal; halaman pilih jumlah kelompok yang menyediakan tiga metode penentuan centroid awal; halaman hasil perhitungan clustering yang menampilkan proses iterasi menggunakan metode euclidean; halaman hasil pengelompokan clustering yang menampilkan data hasil akhir pengelompokan; halaman edit profil bagi admin dan user untuk mengubah data pribadi serta password, dan halaman tambah user yang hanya dapat diakses oleh admin untuk menambahkan pengguna baru ke sistem.



Gambar 10. Halaman Hasil Pengelompokan Clustering

Gambar 10 menampilkan halaman hasil pengelompokan clustering dari proses perhitungan yang telah dilakukan pada halaman sebelumnya. Terdapat data hasil proses clustering dari masing-masing banyaknya iterasi yang dilakukan. Halaman ini menampilkan tabel yang berisi kolom nama produk, pesanan, produk terjual dan cluster. Selain itu, halaman ini juga menampilkan tabel yang memperlihatkan seberapa banyak data barang yang masuk ke kelompok suatu cluster.

3.4 Testing

Pada tahap pengujian, dilakukan evaluasi untuk memastikan aplikasi berjalan sesuai harapan menggunakan blackbox testing. Beberapa pengujian dilakukan untuk login, logout, pengelolaan profil, pengelolaan data pengguna, dan pengelolaan perhitungan clustering. Hasil pengujian menunjukkan bahwa setiap fungsi bekerja sesuai yang diharapkan, seperti login dengan akun yang valid, mengubah data profil, menambah atau menghapus pengguna, serta melakukan proses clustering dengan file yang sesuai. Semua pengujian berhasil dengan hasil yang sesuai, misalnya, jika input tidak valid atau tidak lengkap, aplikasi memberikan pesan kesalahan yang sesuai.

3.5 Maintenance

Tahap terakhir dalam SDLC model waterfall adalah pemeliharaan sistem (maintenance) untuk memperbaiki kesalahan atau bug yang ditemukan setelah implementasi aplikasi pengelompokan kategori penjualan barang di Toko Harto Joyo. Proses ini bertujuan untuk mengevaluasi dan memperbaiki fungsionalitas aplikasi agar lebih baik di masa mendatang.

3.6 Pembahasan

Aplikasi pengelompokan kategori penjualan barang berbasis web di Toko Harto Joyo dibuat dengan menggunakan framework CodeIgniter dan mengikuti model SDLC waterfall, meliputi tahapan analisis, desain, pembuatan kode program, dan pengujian. Aplikasi ini bertujuan untuk memudahkan pengelolaan persediaan stok barang di toko. Sebelum data diimpor ke sistem, data diproses dengan menggunakan teknik KDD (Knowledge Discovery in Databases) yang meliputi tahapan data selection, preprocessing, transformation, mining, dan evaluation. Hasil clustering dari aplikasi menunjukkan bahwa produk dibagi menjadi tiga kategori: produk tidak laris (Cluster 1), produk terlaris (Cluster 2), dan produk standar (Cluster 3).

ISSN 2774-4744 (Media Online)
Vol 5, No 3, July 2025 | Hal 167-176
DOI: 10.47065/jimat.v5i3.610
https://hostjournals.com/jimat

Berdasarkan hasil clustering, Cluster 1 terdiri dari produk yang kurang laris, sehingga tidak perlu persediaan stok yang banyak. Cluster 2 berisi produk terlaris yang memerlukan persediaan stok besar, sementara Cluster 3 mencakup produk standar dengan kebutuhan stok yang sedang. Hasil ini menunjukkan bahwa Toko Harto Joyo perlu memiliki strategi khusus, seperti promosi atau diskon, untuk meningkatkan penjualan produk yang ada di Cluster 1, mengingat jumlah produk dalam Cluster 1 paling banyak, sedangkan Cluster 2 yang paling sedikit.

4. KESIMPULAN

Penelitian ini menghasilkan aplikasi pengelompokan kategori penjualan barang untuk Toko Harto Joyo, dibangun dengan PHP, framework CodeIgniter, dan menggunakan MySQL untuk database. Aplikasi ini menerapkan metode SDLC model waterfall dan menggunakan algoritma K-Means Clustering untuk mengelompokkan produk berdasarkan kelarisan. Pengujian blackbox menunjukkan aplikasi berjalan sesuai dengan yang diharapkan. Proses clustering menghasilkan 3 cluster: Cluster 1 (191 data) untuk produk tidak laris, Cluster 2 (9 data) untuk produk terlaris, dan Cluster 3 (28 data) untuk produk standar. Hasil evaluasi dengan metode Davies-Bouldin Index (DBI) menunjukkan bahwa 3 cluster adalah yang terbaik dengan nilai DBI 0.356. Toko Harto Joyo perlu strategi khusus, seperti promosi, untuk meningkatkan penjualan produk di Cluster 1. Beberapa saran untuk pengembangan sistem di masa mendatang antara lain adalah dengan menggunakan algoritma lain atau mengombinasikan K-Means dengan algoritma lain guna memperoleh hasil pengelompokan yang lebih akurat dan efisien. Selain itu, pengembangan juga dapat difokuskan pada peningkatan desain tampilan sistem agar lebih menarik secara visual, serta penambahan fitur-fitur baru untuk meningkatkan fungsionalitas dan manfaat sistem. Penggunaan jumlah dataset yang lebih besar juga disarankan untuk meningkatkan akurasi hasil pengelompokan dan mendekatkan sistem pada tujuan yang diharapkan.

REFERENCES

[1] Achray, "Implementasi Algoritma K-Means Untuk Mengelompokkan Data Penjualan Mobil Di PT.Honda Arista Manga Dua," Universitas Satya Negara Indonesia, Jakarta, 2020.
[2] S. Astuti, "Algoritma K-Means Dalam Menentukan Penerima Beasiswa UPZ (Unit Pengumpulan Zakat) Pada Mahasiswa UIN Sumatera Utara Medan," UIN Sumatera Utara, Medan, 2020.
[3] Prasetya, R. Salkiawati, and A. D. Alexander, "Analisis Cluster K-Means Dengan Metode Elbow Untuk Menentukan Pola Penjualan Produk Traffic Room Summarecon Mal Bekasi," Journal of Students Research in Computer Science, vol. 4, no. 1, pp. 105-108, 2023.
[4] F. M. Adiansyah, "Aplikasi Algoritma K-Means Untuk Menentukan Persediaan Barang Berbasis Web Dengan Framework Laravel," Universitas Mercu Buana, Jakarta, 2020.
[5] E. Muningsih, I. Maryani, and V. R. Handayani, "Penerapan metode K-Means dan optimasi jumlah cluster dengan index Davies Bouldin untuk clustering propinsi berdasarkan potensi desa," *Jurnal Sains dan Manajemen*, vol. 9, no. 1, pp. 96, 2021.
[6] F. Hidayat, Konsep Pengembangan Sistem Informasi Kesehatan. Yogyakarta: Deepublish, 2020.
[7] Y. A. Mustika, A. Manuhutu, N. Ahmad, I. Hasbi, Guntoro, M. A. Manuhutu, M. Ridwan, Hozairi, A. K. Wardhani, S. Alim, I. Romli, Y. Religia, D. T. Ocafian, U. U. Sufandi, dan I. Ernawati, Data Mining dan Aplikasinya. Bandung: Widina Bhakti Persada Bandung, 2021.
[8] E. Bulolo, Data Mining untuk Perguruan Tinggi. Yogyakarta: Deepublish, 2020.
[9] F. Marisa, A. L. Maukaar, dan T. M. Akhriza, Data Mining: Konsep dan Penerapannya. Sleman: Deepublish, 2021.
[10] Rahayu, P. Wibawa, I. G. I. Sudipa, Suryani, A. Surachman, A. Ridwan, I. G. M. Daarmawiguna, M. N. Sutoyo, I. Slamet, S. Harlina, dan I. M. D. Maysanjaya, *Data Mining*. Jambi: PT Sonpedia Publishing Indonesia, 2024.
[11] W. T. Wu, Y. J. Li, A. Z. Feng, L. Li, T. Huang, A. D. Xu, dan J. Lyu, "Data Mining in Clinical Big Data: The Frequently Used Databases, Steps, and Methodological Models," *Military Medical Research*, vol. 8, pp. 1–12, 2021.
[12] C. Rianto dan S. Bunyamin, *Panduan Pembuatan Aplikasi Clustering Gangguan Jaringan Menggunakan Metode K-Means Clustering*. Bandung: Kreatif Industri Nusantara, 2020.
[13] G. D. Nursyafitri, "K-Means Clustering, Salah Satu Contoh Teknik Analisis Data Populer," *DQLab*, 2022. [Online]. Tersedia: https://dqlab.id/k-means-clustering-salah-satu-contoh-teknik-analisis-data-populer
[14] R. R. Muhima, M. Kurniawan, S. R. Wardhana, A. Yudhana, Sunardi, W. M. Rahmawati, dan G. E. Yuliastuti, *Kupas Tuntas Algoritma Clustering: Konsep, Perhitungan Manual, dan Program*. Yogyakarta: Penerbit ANDI, 2021.
[15] M. Sholeh dan K. Aeni, "Perbandingan Evaluasi Metode Davies Bouldin, Elbow dan Silhouette pada Model Clustering dengan Menggunakan Algoritma K-Means," *STRING (Satuan Tulisan Riset dan Inovasi Teknologi)*, vol. 8, no. 1, pp. 56–65, 2023.
[16] M. Orisa, "Optimasi Cluster Pada Algoritma K-Means," *SENIATI*, vol. 6, no. 2, pp. 430–437, 2022.
[17] Y. A. Mustika, A. Manuhutu, N. Ahmad, I. Hasbi, Guntoro, M. A. Manuhutu, M. Ridwan, Hozairi, A. K. Wardhani, S. Alim, I. Romli, Y. Religia, D. T. Ocafian, U. U. Sufandi, dan I. Ernawati, *Data Mining dan Aplikasinya*. Bandung: Widina Bhakti Persada Bandung, 2021.
[18] N. A. Maori dan E. Evanita, "Metode Elbow dalam Optimasi Jumlah Cluster pada K-Means Clustering," *Simetris: Jurnal Teknik Mesin, Elektro dan Ilmu Komputer*, vol. 14, no. 2, pp. 277–288, 2023. doi: 10.24176/simet.v14i2.9630.
[19] D. Gustian, S. Fahmi, dan N. D. Arianti, *Menggali Emas Terpendam Data Mining*. Tangerang: 3M Media Karya, 2020.
[20] D. R. Anamisa dan F. A. Mufarroha, *Dasar Pemrograman Web: Teori dan Implementasi (HTML, CSS, Javascript, Bootstrap, CodeIgniter)*. Malang: Media Nusa Creative, 2020.
[21] Salamun dan Sukri, *Buku Ajar Pengembangan Aplikasi Web*. Kuningan: Goresan Pena, 2024.