

## OPTIMALISASI STOK BARANG MELALUI ALGORITMA K-MEANS CLUSTERING ANALISIS UNTUK MANAJEMEN PERSEDIAAN DALAM KONTEKS BISNIS MODERN

Risma'ruf Samsudin<sup>1</sup>, Martanto<sup>2</sup>, Umi hayati<sup>3</sup>

<sup>1,3</sup> Program Teknik Informatika, STMIK IKMI Cirebon

<sup>2</sup> Program Studi Manajemen Informatika, STMIK IKMI Cirebon

Jl. Perjuangan No.10B, Karyamulya, Kec. Kesambi, Kota Cirebon, Jawa Barat 45135

samsudinrismaruf@gmail.com

### ABSTRAK

Penelitian ini mengeksplorasi hasil data penjualan untuk di analisis dengan penerapan metode K-Means dalam mengelompokkan transaksi penjualan untuk mengoptimalkan strategi manajemen persediaan barang. Analisis data menjadi kunci utama dalam mengoptimalkan strategi persediaan barang, terutama dengan akses data yang semakin besar. Sistem manajemen persediaan konvensional memiliki kelemahan dalam akurasi dan waktu, sehingga diperlukan pendekatan modern dan efektif. Algoritma K-Means clustering memungkinkan perusahaan untuk mengkategorikan barang berdasarkan pola permintaan dan karakteristik lainnya, sehingga mempermudah penetapan tingkat persediaan ideal untuk setiap kategori barang. Penelitian ini menunjukkan bahwa analisis data dengan algoritma K-Means clustering dapat membantu mengidentifikasi pola permintaan dan karakteristik barang, sehingga membantu perusahaan dalam mengambil keputusan yang terarah dan efektif terkait strategi persediaan barang. Hal ini dapat meminimalisir ketidakpastian dan risiko, serta memastikan pengambilan keputusan persediaan barang secara personal dan efektif.

**Kata kunci :** Analisis data, K-Means clustering, manajemen persediaan barang, strategi persediaan barang, pengambilan keputusan.

### 1. PENDAHULUAN

Seiring dengan berkembang teknologi dan akses terhadap data yang lebih besar, dengan mengelola data dapat memiliki potensi untuk mengumpulkan, menyimpan, dan menganalisis sejumlah besar informasi terkait stok barang produk. Analisis data sebagai cara yang paling kompleks untuk menentukan penyediaan stok barang yang tepat dengan banyaknya produsen dan perubahan cepat dalam perilaku konsumen dalam membeli barang kebutuhan logistik. Dengan menggunakan teknologi dan analisis data yang canggih, dapat dengan cepat menyesuaikan diri terhadap perubahan pasar dan mengantisipasi kebutuhan persediaan barang dengan lebih tepat. Melalui pemahaman yang mendalam terhadap perilaku konsumen dan tren pasar, sehingga dalam berbisnis dapat mengenali pola-pola yang mungkin tidak terdeteksi secara manual, Sehingga memungkinkan untuk meningkatkan strategi manajemen persediaan dan efisiensi operasional secara keseluruhan. memahami tren pasar, referensi konsumen, dan dinamika pesaing untuk mengoptimalkan strategi dalam memanejemen penyediaan stok barang dengan mengelola data untuk penyediaan stok barang produk akan lebih optimal, oleh karena itu, penelitian ini bertujuan mengeksplorasi penerapan metode K-Means dalam mengelompokkan transaksi penjualan, dengan harapan dapat memberikan kontribusi dalam mengoptimalkan strategi manajemen persediaan stock barang.

Sistem manajemen persediaan konvensional umumnya mengandalkan metode manual, seperti penghitungan stok fisik dan perkiraan permintaan

berdasarkan pengetahuan dan pengalaman. Kelemahan sistem ini terletak pada akurasi yang rendah dan prosesnya yang memakan waktu, sehingga dapat memicu terjadinya kelebihan atau kekurangan stok. Perkembangan teknologi terhadap pengolahan data memungkinkan pengelolaan data untuk mengumpulkan dan menganalisis data dalam skala besar, analisis data dapat membantu untuk mengidentifikasi pola-pola yang tidak terlihat sebelumnya, dan memberikan wawasan yang mendalam tentang manajemen persediaan stock barang. Untuk mengatasi kekurangan metode konvensional, diperlukan pendekatan yang lebih modern dan efektif dalam mengoptimalkan persediaan barang. Salah satu solusi yang potensial adalah dengan memanfaatkan algoritma K-Means clustering analysis. Algoritma ini memungkinkan perusahaan untuk mengkategorikan barang berdasarkan pola permintaan dan karakteristik lainnya, sehingga mempermudah penetapan tingkat persediaan ideal untuk setiap kategori barang.

Beberapa penelitian sebelumnya telah menyelidiki clustering data persediaan stock barang menggunakan metode K-Means. sebagai contohnya, studi yang dilakukan [1] (Ramadhan et al., n.d.), Persediaan barang, sebagai aset penting dalam siklus bisnis, memerlukan pengelolaan yang cermat. Data mining, dengan teknik statistik, matematika, kecerdasan buatan, dan machine learning, menawarkan solusi untuk mengekstraksi dan mengidentifikasi informasi yang relevan dari data persediaan. Penelitian tersebut bertujuan untuk memanfaatkan data mining dalam pengelompokan

barang dan meningkatkan akurasi klusterisasi data persediaan barang melalui metode K-Means Clustering, ini memungkinkan pembagian data menjadi kelompok-kelompok berdasarkan karakteristik yang serupa. Data dengan karakteristik yang berbeda dikelompokkan secara terpisah, dengan menggunakan centroid (rata-rata) sebagai model kluster.

Studi kedua yang dilakukan [2](Kurnia Alvisan, 2021) menggambarkan penelitian penerapan metode clustering K-Means dalam menentukan jumlah kebutuhan pembelian makanan dan minuman di Minimarket. Analisis data penjualan dilakukan untuk mengidentifikasi pola konsumsi dan mengelompokkan produk berdasarkan tingkat penjualan. Hasil penelitian menunjukkan bahwa metode clustering K-Means efektif dalam mengelompokkan produk berdasarkan tingkat penjualan. Studi ketiga dilakukan oleh [3](Syafrial & Febrianti, 2023) membahas tentang penerapan algoritma k-means pada aplikasi Data mining untuk menentukan pola Penjualan (studi kasus: zahra mart), hasil studi tersebut menghasilkan Persaingan bisnis di supermarket, minimarket, dan swalayan semakin ketat. Untuk meningkatkan penjualan dan stok barang barang, Zahra Mart dapat menggunakan aplikasi data mining yang dirancang dengan Algoritma K-Means, sehingga dapat mencegah penumpukan barang di gudang. Walaupun dari beberapa literatur sebelumnya telah mengurai tuntas aspek aspek penting dalam clustering dan analisis persediaan, serta penelitian yang spesifik pada analisis persediaan stock barang khususnya menggunakan K-Means clustering, masih terbatas oleh karena itu, studi ini akan melengkapi literatur sebelumnya yang akan memberikan pandangan dalam tentang cara optimalisasi stok barang melalui algoritma k-means clustering.

Yang secara nyata bahwa Analisis data membantu dalam manajemen persediaan stock barang serta meningkatkan peluang dan memperkecil ancaman kerugian di pasar perdagangan dan banyaknya yang blum memahami menggunakan analisis data penjualan yang dikelola sebagai pedoman dalam manajemen ketersediaan barang maka penelitian ini memiliki potensi implikasi yang signifikan dalam memberikan informasi dan cara manajemen ketersediaan barang sehingga bisa optimal dalam mengelola persediaan stock barang, membuat penulis tertarik mengangkatnya dalam penelitian.

## 2. TINJAUAN PUSTAKA

Dalam penelitian ini, metode literature review diterapkan dengan merinci artikel-artikel terkait [tuliskan topik / judul penelitian]. Proses literature review terdiri dari empat tahapan, yaitu (1) formulasi permasalahan, (2) pencarian literatur, (3) evaluasi data, dan (4) analisis serta interpretasi.

Formulasi permasalahan dilakukan memilih topik. Pada penelitian ini topik yang dipilih adalah mengenai Data penjualan “Algoritma K-means

clustering”, or “Davies Bouldin Index” . Langkah selanjutnya adalah pencarian literatur yang relevan dengan topik penelitian. Langkah ini dapat memberikan gambaran mengenai Clustering Data Penjualan Menggunakan Metode K-Means Proses pencarian dan pengumpulan artikel dengan menggunakan tools publish or perish, yang diambil dari database yaitu Google Scholar dengan kata kunci “Data Penjualan” dan “K-Means”.

Dari hasil pencarian ditemukan sebanyak 15 artikel jurnal. Langkah ketiga adalah evaluasi data yaitu dengan melakukan penyaringan dan pemilihan serta pemilihan artikel jurnal yang benar benar relevan dengan topik penelitian serta mempertimbangkan keterbaruannya. Keterbaruan dibatasi dengan memilih artikel 5 tahun terakhir yaitu mulai dari tahun 2023, 2022, 2021, 2020 dan 2019. Sedangkan kesesuaian dilihat dari kesesuaian judul paper, sinta index, atau hal lain yang dianggap memiliki reputasi. Berdasarkan hasil evaluasi data, kemudian dipilih 15 artikel jurnal untuk di review. Setelah keempat tahapan tersebut dilakukan, proses selanjutnya adalah pelaksanaan literature review. Adapun teknik yang dilakukan dalam literature review yaitu mencari kesamaan (compare), mencari ketidaksamaan (contrast), memberikan pandangan (criticize), membandingkan (synthesize), dan meringkas (summary).

### 2.1. Hasil Literatur Review

Hasil literature review yang telah dilakukan pada jurnal-jurnal penelitian terkait topik analisis clustering data penjualan dapat dijabarkan sebagai berikut :

Paper 1. Dari hasil penelitian [4] (Amalina et al., 2022) Analisis minat konsumen terhadap produk frozen food menggunakan Algoritma K-Means Clustering mengungkap bahwa data penjualan produk frozen food terbagi menjadi dua kategori berdasarkan tingkat minat konsumen. Kategori pertama, Cluster 1, menampung 3 produk dengan minat konsumen paling rendah. Sedangkan kategori kedua, Cluster 2, berisi 42 produk dengan minat konsumen tinggi.

Paper 2. Berdasarkan dari penelitian [5] (Setyorin et al., 2021) Analisis pola pembelian konsumen di PT. Citra Mustika Pandawa menggunakan algoritma K-Means clustering berhasil mengkategorikan konsumen menjadi 5 kelompok optimal. Validasi dengan Davies Bouldin Indeks (DBI) menghasilkan nilai 0.500, menunjukkan pengelompokan yang efektif penelusuran pola pembelian dengan Association Rules, dengan nilai minimum support 30% dan confidence 50%, menemukan pola unik di setiap kelompok. Cluster 1 memiliki 3 pola, cluster 2 & 3 memiliki 1 pola, cluster 4 memiliki 5 pola, dan cluster 5 memiliki 6 pola dengan memperketat nilai minimum support dan confidence menjadi 50%, pola pembelian hanya ditemukan di cluster 4 (3 pola) dan cluster 5 (1 pola). Menariknya, setiap kelompok menunjukkan 8 produk yang paling sering dibeli oleh pelanggan PT. Citra Mustika Pandawa.

Paper 3. Berdasarkan [6](Murpratiwi et al., 2021) Penelitian ini memanfaatkan algoritma K-means clustering, salah satu teknik machine learning unsupervised learning, untuk mengelompokkan pelanggan berdasarkan pola pembelian mereka. Algoritma ini telah terbukti efektif dalam segmentasi pelanggan sejak tahun 2018 hingga 2022. Sebelum pengelompokan, data dibersihkan dan dianalisis melalui proses data mining dan preprocessing. Analisis awal (exploratory data analysis) mengungkap tren penjualan dan perilaku pelanggan selanjutnya, algoritma K-means diterapkan dengan metode elbow untuk menentukan jumlah kluster yang optimal. Metode ini menghitung dan menggambarkan nilai WCSS (Within Cluster Sum of Squares) dalam grafik untuk menemukan titik siku (elbow) yang menunjukkan jumlah kluster terbaik. Hasilnya, diperoleh 4 kluster pelanggan yang berbeda pengelompokan K-means menggunakan 3 variabel data: quantity (jumlah barang yang dibeli), unit price (harga satuan barang), dan customer ID (kode pelanggan). Analisis menghasilkan 4 karakteristik pelanggan dengan ciri khasnya masing-masing: Kluster 0: Pelanggan dengan pembelian sedikit, harga satuan tinggi, dan kode pelanggan tinggi. Kluster 1: Pelanggan dengan pembelian sedang, harga satuan rendah, dan kode pelanggan tinggi. Kluster 2: Pelanggan dengan pembelian banyak, harga satuan rendah, dan kode pelanggan menengah. Kluster 3: Pelanggan dengan pembelian sedikit, harga satuan tinggi, dan kode pelanggan menengah.

Temuan ini menunjukkan bahwa kuantitas dan harga satuan barang memiliki pengaruh signifikan terhadap perilaku pelanggan. Perbandingan metode elbow dengan metode lain untuk menentukan jumlah kluster optimal dapat menjadi bahan penelitian selanjutnya.

Paper 4. Pada penelitian [7](Syahputra, 2022) Proses clusterisasi mampu menyajikan pengetahuan baru dari data yang ada. Kasus tingkat penjualan menu pada restoran mama. Hasil tersebut mampu memberikan hasil yang cukup baik dalam mengkluster menu makanan yang terbaik. Hasil penelitian ini juga menyajikan knowledge base untuk dijadikan dasar dalam manajemen penjualan yang ada di restoran mama.

Paper 5. Berdasarkan penelitian [3](Syafrinal & Febrianti, 2023) Aplikasi ini membantu swalayan menentukan produk yang terjual dengan lebih mudah dan efisien melalui proses klasifikasi data produk. Dari proses klasifikasi, ditemukan pola pemilihan produk yang digunakan sebagai aturan dalam menentukan produk yang terjual di Zahra Mart.

Manfaat utama aplikasi ini adalah memudahkan penentuan produk terjual: Dengan memasukkan data barang ke dalam sistem, aplikasi memproses data penjualan dan menghasilkan keputusan dengan cepat, sehingga swalayan dapat menentukan produk mana yang terjual dengan lebih cepat dan mudah dan menentukan produk yang diminati dan tidak diminati:

Informasi tentang produk yang diminati dan tidak diminati membantu swalayan dalam mengambil keputusan konsisten untuk menambah stok produk laris dan mengurangi stok produk kurang laris. Hal ini membantu swalayan dalam mengoptimalkan stok produk dan meningkatkan keuntungan.

Secara keseluruhan, aplikasi ini membantu swalayan dalam mengelola stok produk secara lebih efisien dan efektif, sehingga meningkatkan keuntungan dan kepuasan pelanggan. Swalayan dapat menyediakan produk yang lebih sesuai dengan kebutuhan pelanggan, sehingga meningkatkan kepuasan pelanggan dan mendorong pembelian kembali.

Paper 6. Berdasarkan penelitian [8](Abdul Muni., 2020) PT. Alfa Scorpii, perusahaan penjualan sepeda motor, masih memiliki potensi besar yang belum digali. Laporan penjualan yang dihasilkan hanya berisi informasi dasar, tanpa analisis mendalam untuk mendapatkan wawasan berharga. Strategi promosi yang diterapkan fokus pada peningkatan pendapatan, namun tanpa mempertimbangkan efisiensi biaya dengan penerapan data mining melalui teknik knowledge discovery dan pattern recognition dapat membuka peluang baru bagi perusahaan. Dengan menggali pengetahuan terpendam dari kumpulan data yang besar, perusahaan dapat memahami pelanggannya dengan lebih baik. Salah satu metode data mining yang efektif adalah algoritma K-Means Clustering. Algoritma ini memungkinkan pengelompokan data pelanggan berdasarkan berbagai faktor, sehingga bagian pemasaran dapat merumuskan strategi promosi yang lebih terarah dan efektif untuk menarik pelanggan baru dan meningkatkan pendapatan perusahaan secara berkelanjutan.

Dengan memanfaatkan data secara maksimal, PT. Alfa Scorpii dapat meningkatkan efisiensi dan efektivitas strategi promosinya, sehingga mencapai target penjualan dan profitabilitas yang lebih tinggi. Hal ini akan membuka peluang baru bagi perusahaan untuk berkembang dan mencapai kesuksesan yang lebih besar di masa depan..

Paper 7. Dari penelitian [9] (Royal, 2024) Aplikasi ini membantu Rima Market menentukan produk yang terjual dengan lebih mudah dan efisien melalui proses klasifikasi data produk. Dari proses klasifikasi, ditemukan pola pemilihan produk yang digunakan sebagai aturan dalam menentukan produk terjual.

Manfaat aplikasi ini adalah memudahkan penentuan produk terjual Dengan memasukkan data barang ke dalam sistem, aplikasi memproses data penjualan dan menghasilkan keputusan dengan cepat, sehingga Rima Market dapat menentukan produk mana yang terjual dengan lebih cepat dan mudah dan menentukan produk yang diminati dan tidak diminati: Informasi tentang produk yang diminati dan tidak diminati membantu Rima Market dalam mengambil keputusan konsisten untuk menambah stok produk

laris dan mengurangi stok produk kurang laris. Hal ini membantu Rima Market dalam mengoptimalkan stok produk dan meningkatkan keuntungan.

Paper 8. Berdasarkan penelitian [2] (Kurnia Alvisan, 2021) Setelah melakukan analisis, perancangan, dan pengujian maka dapat diperoleh kesimpulan sebagai Penerapan Data Mining dengan menggunakan aplikasi yang dibangun dapat membantu PT. SINAR NIAGA SEJAHTERA sebagai gambaran bagi pengambilan keputusan perusahaan dalam rangka mendapatkan pola penjualan produk. Aplikasi yang dibangun dapat mengurangi penumpukan data yang kurang dimanfaatkan sebelumnya serta lebih efisiensi dalam kinerja.

Paper 9. Berdasarkan penelitian [10] (Widiyanto & Witanti, 2021) PT. Alfa Scorpii, perusahaan swasta di bidang penjualan sepeda motor, Data mining hadir sebagai solusi jitu untuk mengatasi tantangan ini. Dengan memanfaatkan metode seperti algoritma K-Means Clustering, PT. Alfa Scorpii dapat mengolah kumpulan data penjualan yang besar untuk menggali pola dan wawasan berharga. Wawasan ini membuka pintu bagi:

Pemahaman mendalam karakteristik pelanggan: Pengelompokan pelanggan berdasarkan pola pembelian mereka memungkinkan perusahaan untuk memahami preferensi, kebiasaan, dan kebutuhan mereka dengan lebih baik dengan pengembangan strategi promosi terarah: Pemahaman karakteristik pelanggan menjadi kunci untuk merancang strategi promosi yang efektif, menjangkau pelanggan yang tepat dengan pesan yang sesuai serta peningkatan efisiensi biaya pemasaran: Penargetan promosi yang tepat dapat meminimalisir pemborosan anggaran pemasaran dan meningkatkan ROI (Return on Investment).

Paper 10. Berdasarkan penelitian [11] (Hasugian, 2018) Data transaksi penjualan bagaikan jendela yang membuka wawasan tentang perilaku konsumen, pola pembelian, dan tren pasar. Dengan menganalisis data ini, toko dapat memahami mengapa konsumen memilih berbelanja di tempat mereka, apa yang mereka inginkan, dan bagaimana mereka dapat meningkatkan pengalaman berbelanja dengan algoritma K-Means, atau metode data mining lain, telah digunakan untuk mengelompokkan konsumen berdasarkan berbagai faktor seperti kebiasaan berbelanja, nilai indeks massa tubuh (BMI), dan ukuran rangka. Pengelompokan ini memungkinkan toko untuk menargetkan promosi dan penawaran secara lebih efektif kepada kelompok konsumen yang berbeda serta penerapan data mining dalam industri retail masih tergolong baru, namun potensinya sangat besar. Dengan memanfaatkan data secara cerdas, toko dapat membuka peluang baru untuk meningkatkan penjualan, meningkatkan kepuasan pelanggan, dan membangun keunggulan kompetitif di pasar.

Paper 11. Berdasarkan hasil penelitian [12] (Silalahi, 2018) Penerapan algoritma K-Means menghasilkan temuan penting yang membuka peluang

bagi PT Batamas Niaga Jaya untuk meningkatkan kinerja bisnisnya:

1. Pengelompokan Produk:  
Teridentifikasi 2 kelompok produk yang berbeda berdasarkan tingkat penjualan kelompok C0: Terdiri dari 657 produk yang tidak laris kelompok C1: Terdiri dari 70 produk yang laris.
2. Perbedaan Signifikan:  
Kelompok C0 dan C1 memiliki perbedaan yang signifikan dalam hal tingkat penjualan.
3. Peluang Strategi:  
Analisis ini membuka peluang bagi PT Batamas Niaga Jaya untuk merumuskan strategi yang lebih efektif dalam mengelola produknya.

Paper 12. Dalam penelitian [13] (Siregar et al., n.d.) Analisis penjualan Toko Adi Bangunan yang dilakukan pada periode 1 Juli 2017 hingga 9 Agustus 2017 menggunakan metode k-means clustering menghasilkan beberapa kesimpulan penting yang berimplikasi pada strategi pengelolaan stok barang di tok efektivitas Metode Clustering dalam Penentuan Stok Barang penerapan metode clustering terbukti efektif dalam membantu Toko Adi Bangunan menentukan item barang yang perlu diprioritaskan untuk pembelian stok. Dari penelitian ini, ditemukan bahwa terdapat 10 item barang yang termasuk dalam kelompok barang laris. Hal ini menunjukkan bahwa pembelian stok barang perlu difokuskan pada 10 item tersebut untuk memaksimalkan keuntungan dan efisiensi toko.

Paper 13. Berdasarkan penelitian [14] (Retno Lestari, 2019) hasil dan pembahasan, maka kesimpulan bahwa penelitian tersebut menghasilkan sebagai berikut:

1. Pengklasteran dengan menggunakan metode K-Means didapatkan untuk produk yang laku dengan yang tidak laku.
2. Semakin banyak data penjualan barang yang diinput, maka semakin banyak clustering centroid yang terbentuk di proses metode kmeans.
3. Produk yang laku terdiri dari makanan, minuman dan produk yang tidak laku terdiri dari kebanyakan dari kosmetik.

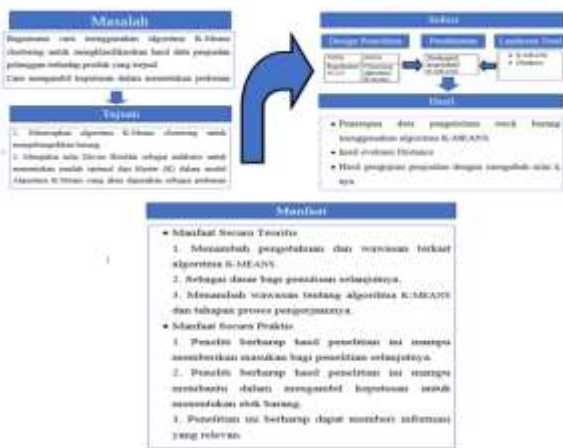
Paper 14. Berdasarkan hasil penelitian [1] (Ramdhan et al., n.d.) Metode ini menghasilkan dua kluster: kluster dengan penjualan terbanyak dan kluster dengan penjualan terendah. Algoritma K-means menunjukkan akurasi 100% dengan dataset 249 data. Penelitian clustering data persediaan barang dengan K-Means dapat ditingkatkan dengan menggunakan dataset yang lebih besar dan melakukan pengujian dengan atribut yang lebih lengkap untuk meningkatkan penelitian ini gunakan dataset yang lebih besar dan lakukan pengujian dengan atribut yang lebih lengkap. Bandingkan K-means dengan metode clustering lain seperti Fuzzy C-Means.

Paper 15. Dalam penelitian [15] Penelitian ini menggunakan metode pengelompokan K-Means, teknik data mining yang ampuh, untuk mengungkap

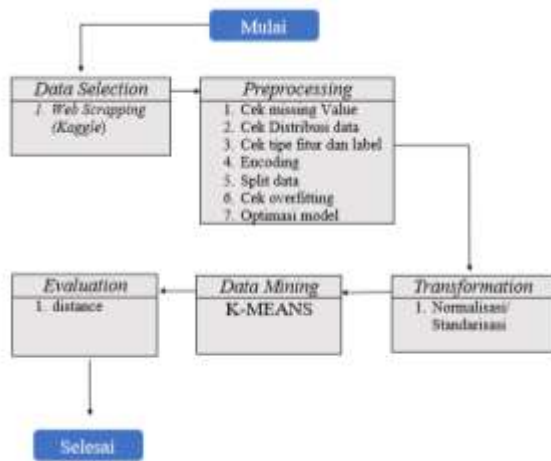
pola penjualan tersembunyi dan memandu pengambilan keputusan strategis. K-Means berhasil mengelompokkan data penjualan menjadi tiga kategori berbeda: penjualan rendah, sedang, dan tinggi. Hebatnya, metode ini secara akurat mengidentifikasi Vivo Y12 3+32GB sebagai produk tunggal di kelompok penjualan tertinggi, sesuai dengan angka penjualan aktual.

**3. METODE PENELITIAN**

Metode penelitian menggunakan pendekatan kuantitatif dengan memanfaatkan berbagai teknik, termasuk pengumpulan data melalui studi pustaka, analisis statistik, dan penerapan algoritma K-Means untuk mengklasifikasikan data penjualan.



Gambar 1. Kerangka Berfikir



Gambar 2. Flowchart Metode/algorithm K-MEAN

Tahapan dalam penelitian ini untuk melakukan Clustering terhadap data hasil Data Penjualan berikut ini beberapa langkah dalam tahapan penelitian yang akan dilakukan.

1. Data selection Web scapping  
Web scraping adalah proses otomatis pengumpulan informasi dari halaman web menggunakan program komputer atau bot.
2. Preprocessing  
Preprocessing adalah sebuah proses yang melalui tahapan-tahapan sebagai berikut:

- a. Cek missing Value  
Cek missing Value Melakukan pengecekan terhadap nilai yang hilang pada dataset untuk mengetahui keberadaan data yang kosong.
  - b. Cek Distribusi data  
Cek Distribusi data Memeriksa sebaran atau distribusi data untuk memahami pola atau karakteristik dari dataset yang digunakan.
  - c. Cek tipe fitur dan label  
Cek tipe fitur dan label Memeriksa jenis fitur dan label yang digunakan untuk memastikan pemahaman yang tepat terhadap variabel yang akan dimodelkan.
  - d. Encoding  
Encoding Mengaplikasikan encoding pada variabel kategorikal untuk memungkinkan penggunaan dalam algoritma machine learning yang memerlukan data numerik.
  - e. Split data  
Split data Memisahkan data menjadi subset pelatihan dan pengujian untuk evaluasi model.
  - f. Cek overfitting  
Cek overfitting Menyelidiki tanda-tanda overfitting, yaitu kondisi ketika model terlalu beradaptasi dengan data pelatihan dan tidak umum untuk data baru.
  - g. Optimasi model  
Optimasi model Mengoptimalkan model dengan menyesuaikan parameter atau struktur untuk meningkatkan kinerja model pada data yang belum pernah dilihat sebelumnya.
3. Data Transformasi  
Normalisasi adalah proses mengubah skala nilai fitur ke rentang tertentu, sedangkan standarisasi mengatur nilai rata-rata menjadi 0 dan deviasi standar menjadi 1 pada data
  4. Data Mining  
Data mining adalah proses mengeksplorasi dan menganalisa data dalam jumlah yang besar yang bertujuan untuk menemukan suatu pola atau informasi yang menarik dari data yang tersimpan dalam jumlah yang besar. Dalam penelitian ini akan diterapkan teknik data mining menggunakan metode algoritma *K-Means Clustering*.
  5. Evaluasi  
Beberapa tahapan untuk evaluasi yang akan dilalui sebagai berikut:
    - a. Accuracy  
Accuracy adalah metrik evaluasi yang mengukur seberapa sering model memberikan prediksi yang benar secara keseluruhan
    - b. Precision  
Precision adalah metrik evaluasi yang mengukur seberapa akurat model dalam memprediksi kelas positif dari total prediksi positif yang diberikan.
    - c. Recall  
Recall adalah metrik evaluasi yang mengukur seberapa baik model dapat mendeteksi kelas positif dari total data yang sebenarnya positif.

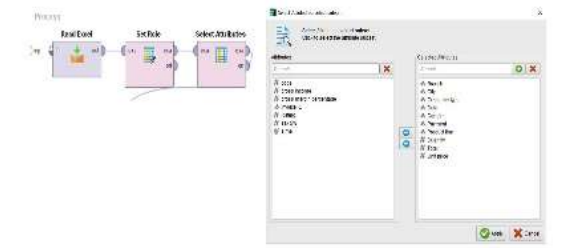
Kemudian beberapa parameter yang akan digunakan dalam penelitian dengan menggunakan K-Means Clustering yaitu mencari nilai K terbaik, mencari nilai max runs dan mencari type parameter numerical measure yang menghasilkan K terbaik dengan Cluster Distance Performance berdasarkan Davies Bouldin.

Tahapan selanjutnya memasukkan operator clustering yang berfungsi untuk menentukan cluster atau pengelompokan data.

| Instansi      | Prov       | Udara  | Kelembaban | Kecepatan Angin | Waktu  | Temp   | Pres   | Puasa  | Ang    | gsm    | Ang    | gsm    | Ang    | gsm    | Ang    |
|---------------|------------|--------|------------|-----------------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| 208-01-2081-C | Yogyakarta | Normal | Normal     | Normal          | Normal | Normal | Normal | Normal | Normal | Normal | Normal | Normal | Normal | Normal | Normal |
| 208-01-2081-D | Yogyakarta | Normal | Normal     | Normal          | Normal | Normal | Normal | Normal | Normal | Normal | Normal | Normal | Normal | Normal | Normal |
| 208-01-2081-E | Yogyakarta | Normal | Normal     | Normal          | Normal | Normal | Normal | Normal | Normal | Normal | Normal | Normal | Normal | Normal | Normal |
| 208-01-2081-F | Yogyakarta | Normal | Normal     | Normal          | Normal | Normal | Normal | Normal | Normal | Normal | Normal | Normal | Normal | Normal | Normal |
| 208-01-2081-G | Yogyakarta | Normal | Normal     | Normal          | Normal | Normal | Normal | Normal | Normal | Normal | Normal | Normal | Normal | Normal | Normal |
| 208-01-2081-H | Yogyakarta | Normal | Normal     | Normal          | Normal | Normal | Normal | Normal | Normal | Normal | Normal | Normal | Normal | Normal | Normal |
| 208-01-2081-I | Yogyakarta | Normal | Normal     | Normal          | Normal | Normal | Normal | Normal | Normal | Normal | Normal | Normal | Normal | Normal | Normal |
| 208-01-2081-J | Yogyakarta | Normal | Normal     | Normal          | Normal | Normal | Normal | Normal | Normal | Normal | Normal | Normal | Normal | Normal | Normal |
| 208-01-2081-K | Yogyakarta | Normal | Normal     | Normal          | Normal | Normal | Normal | Normal | Normal | Normal | Normal | Normal | Normal | Normal | Normal |
| 208-01-2081-L | Yogyakarta | Normal | Normal     | Normal          | Normal | Normal | Normal | Normal | Normal | Normal | Normal | Normal | Normal | Normal | Normal |
| 208-01-2081-M | Yogyakarta | Normal | Normal     | Normal          | Normal | Normal | Normal | Normal | Normal | Normal | Normal | Normal | Normal | Normal | Normal |
| 208-01-2081-N | Yogyakarta | Normal | Normal     | Normal          | Normal | Normal | Normal | Normal | Normal | Normal | Normal | Normal | Normal | Normal | Normal |
| 208-01-2081-O | Yogyakarta | Normal | Normal     | Normal          | Normal | Normal | Normal | Normal | Normal | Normal | Normal | Normal | Normal | Normal | Normal |
| 208-01-2081-P | Yogyakarta | Normal | Normal     | Normal          | Normal | Normal | Normal | Normal | Normal | Normal | Normal | Normal | Normal | Normal | Normal |
| 208-01-2081-Q | Yogyakarta | Normal | Normal     | Normal          | Normal | Normal | Normal | Normal | Normal | Normal | Normal | Normal | Normal | Normal | Normal |
| 208-01-2081-R | Yogyakarta | Normal | Normal     | Normal          | Normal | Normal | Normal | Normal | Normal | Normal | Normal | Normal | Normal | Normal | Normal |
| 208-01-2081-S | Yogyakarta | Normal | Normal     | Normal          | Normal | Normal | Normal | Normal | Normal | Normal | Normal | Normal | Normal | Normal | Normal |
| 208-01-2081-T | Yogyakarta | Normal | Normal     | Normal          | Normal | Normal | Normal | Normal | Normal | Normal | Normal | Normal | Normal | Normal | Normal |
| 208-01-2081-U | Yogyakarta | Normal | Normal     | Normal          | Normal | Normal | Normal | Normal | Normal | Normal | Normal | Normal | Normal | Normal | Normal |
| 208-01-2081-V | Yogyakarta | Normal | Normal     | Normal          | Normal | Normal | Normal | Normal | Normal | Normal | Normal | Normal | Normal | Normal | Normal |
| 208-01-2081-W | Yogyakarta | Normal | Normal     | Normal          | Normal | Normal | Normal | Normal | Normal | Normal | Normal | Normal | Normal | Normal | Normal |
| 208-01-2081-X | Yogyakarta | Normal | Normal     | Normal          | Normal | Normal | Normal | Normal | Normal | Normal | Normal | Normal | Normal | Normal | Normal |
| 208-01-2081-Y | Yogyakarta | Normal | Normal     | Normal          | Normal | Normal | Normal | Normal | Normal | Normal | Normal | Normal | Normal | Normal | Normal |
| 208-01-2081-Z | Yogyakarta | Normal | Normal     | Normal          | Normal | Normal | Normal | Normal | Normal | Normal | Normal | Normal | Normal | Normal | Normal |

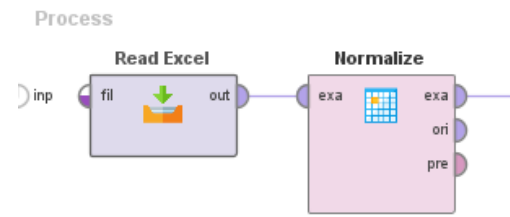
Gambar 3. Data Asli

Data Asli yang disajikan di atas merupakan data asli sebelum diproses untuk pengujian model K-Means. Data ini sangat penting bagi model dan pembersihan data serta seleksi atribut yang digunakan untuk mengcluster dan kelompok secara efektif dalam kumpulan data. Langkah-langkah pemrosesan yang diperlukan untuk menyiapkan proses analisis K-Means dapat, dinormalisasi, dan pemilihan fitur. Langkah-langkah ini memastikan bahwa data berada dalam format yang sesuai untuk model melakukan perhitungan dan menghasilkan hasil yang berarti. Model K-means, setelah itu pada data yang sudah diproses, kemudian dapat mengelompokkan data ke dalam kelompok yang berbeda berdasarkan pola yang cluster. Proses ini memberikan wawasan berharga tentang struktur dasar data dan dapat digunakan untuk berbagai aplikasi, seperti segmentasi pelanggan, analisis cluster berdasarkan Davies Bouldin Index.



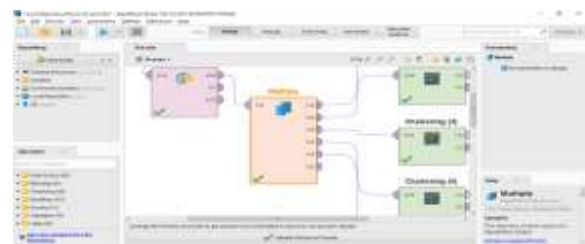
Gambar 4. Seleksi Atribut

Data seleksi atribut data diatas adalah data yang akan diolah agar sesuai dengan model dari Algoritma clustering K-Means clustering untuk pengujian dari model, merupakan representasi dari subkumpulan atribut yang dipilih dengan cermat dari kumpulan data asli. Pemilihan atribut ini dilakukan secara khusus agar sesuai dengan kebutuhan algoritma K-Means clustering.



Gambar 5. Normalize

Operator Normalize dalam RapidMiner adalah salah satu alat yang digunakan untuk mengubah nilai-nilai pada atribut dalam sebuah dataset agar memiliki rentang nilai yang seragam atau sesuai dengan skala yang diinginkan. Tujuannya adalah untuk menghilangkan disparitas dalam rentang nilai antar atribut, sehingga mencegah atribut dengan rentang nilai yang besar mendominasi proses analisis, seperti pada proses pembelajaran mesin. Dengan menggunakan operator Normalize, data dapat disesuaikan menjadi skala yang lebih mudah diinterpretasikan dan diproses, memungkinkan model yang lebih baik dalam analisis dan prediksi.



Gambar 6. Operator Multiply

Operator Multiply dalam RapidMiner digunakan untuk melakukan perkalian antara dua nilai atau dua kolom dalam dataset. Operasi ini berguna dalam melakukan perhitungan matematika sederhana, seperti mengalikan dua atribut numerik atau mengalikan nilai dengan konstanta tertentu. Misalnya, Anda dapat menggunakan operator ini untuk mengalikan nilai-nilai dari dua atribut yang berbeda untuk membuat atribut baru yang merepresentasikan hasil perkalian keduanya. Dengan demikian, operator Multiply membantu dalam mengembangkan fitur-fitur baru yang mungkin relevan atau berguna dalam analisis data lebih lanjut.



Gambar 7. Pengujian Algoritma K-Means

Hasil evaluasi kinerja model menunjukkan adanya variasi yang signifikan antara cluster-cluster yang terbentuk. Rata-rata jarak dalam centroid menunjukkan perbedaan yang mencolok, dengan nilai terendah tercatat pada cluster 0 sebesar -271,699,703,951.411, sementara cluster 1 memiliki rata-rata jarak yang jauh lebih besar, mencapai -3,015,804,046,125.860. Nilai Davies Bouldin Index sebesar -0.375 menunjukkan bahwa kecenderungan dalam setiap cluster relatif terdefinisi dengan baik, meskipun perbedaan yang signifikan antar-cluster menunjukkan bahwa ada variasi yang signifikan dalam pola-pola data di dalamnya.



Gambar 8. Graphic Cluster Root Set

Klaster 0, Klaster 1 merujuk pada hasil pengelompokan data yang ditemukan dalam analisis. Masing-masing klaster mencerminkan kelompok data yang memiliki karakteristik atau pola tertentu yang serupa di antara anggotanya. Pengelompokan ini memungkinkan untuk mengidentifikasi subset data yang memiliki perilaku atau atribut yang mirip, sehingga memungkinkan untuk melakukan analisis lebih lanjut secara spesifik terhadap masing-masing klaster. Dengan memahami perbedaan antara klaster-klaster ini, pengguna dapat mengeksplorasi dan memahami lebih dalam tentang struktur dan pola yang terdapat dalam dataset, serta potensi implikasi atau strategi yang dapat diambil berdasarkan hasil pengelompokan tersebut.

### Cluster Model

```
Cluster 0: 902 items
Cluster 1: 98 items
Total number of items: 1000
```

Gambar 9. Cluster Model

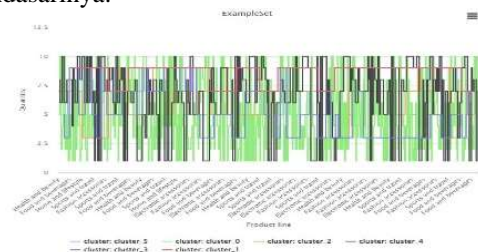
Dalam hasil pengelompokan data, terdapat empat klaster yang berhasil dibentuk, dengan masing-masing memiliki jumlah item yang berbeda. Klaster 0 memiliki 902 item, diikuti oleh Klaster 1 dengan 98 item. Jumlah total item dalam dataset adalah 1000. Distribusi item yang berbeda-beda ini menunjukkan variasi dalam karakteristik atau pola di antara klaster-

klasternya. Dengan demikian, hasil ini memberikan gambaran tentang struktur dan kompleksitas data serta memungkinkan untuk dilakukannya analisis lanjutan untuk memahami ciri khas dan kebutuhan setiap klaster dengan lebih baik.

```
PerformanceVector
PerformanceVector:
Avg. within centroid distance: -540621929484.507
Avg. within centroid distance_cluster_0: -271699703951.411
Avg. within centroid distance_cluster_1: -3015804046125.860
Davies Bouldin: -0.375
```

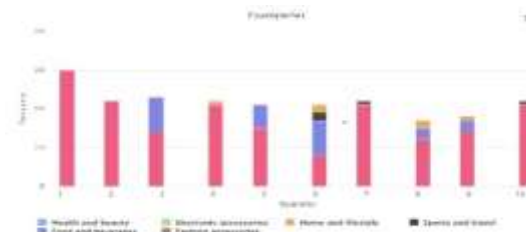
Gambar 10. Performance Vektor dari K-Means

Hasil analisis menunjukkan nilai rata-rata jarak dalam centroid secara keseluruhan sebesar -5406, yang mencerminkan tingkat kepadatan data terhadap centroid. Selanjutnya, nilai rata-rata jarak dalam centroid untuk masing-masing klaster adalah -5406 untuk Klaster 0, -2716 untuk Klaster 1, -3015. Selain itu, nilai Davies Bouldin yang sebesar -0.375 menunjukkan bahwa pengelompokan data menjadi klaster yang baik, dengan kepadatan setiap klaster yang seimbang dan terpisah dengan baik dari klaster lainnya. Dengan demikian, hasil evaluasi ini memberikan gambaran yang komprehensif tentang distribusi dan kepadatan data di dalam masing-masing klaster, memungkinkan untuk pemahaman yang lebih mendalam tentang pola dan karakteristik yang mendasarinya.



Gambar 11. Visualisasi Cluster Example Set

Hasil dari Visualisasi Example Set yang menggunakan plot yang memungkinkan membandingkan beberapa atribut menunjukkan bahwa cluster yang lebih unggul adalah cluster 0 menunjukkan warna hijau yang dominan dilihat pada keterangan gambar.



Gambar 12. Visualiasasi Barang Produk Example Set

Hasil dari visualisasi Example Set terhadap produk menghasilkan produk line yang paling banyak terjual adalah “Fashion accessories”. Ini yang akan

menjadi pedemon dalam pengambilan keputusan penyediaan stock barang produk.

Tabel 1. Hasil Davies bouldin index

| No | Nilai K | Hasil Davies Bouldin |
|----|---------|----------------------|
| 1  | K2      | 0,375                |
| 2  | K3      | 0,482                |
| 3  | K4      | 0,457                |
| 4  | K5      | 0,496                |
| 5  | K6      | 0,493                |

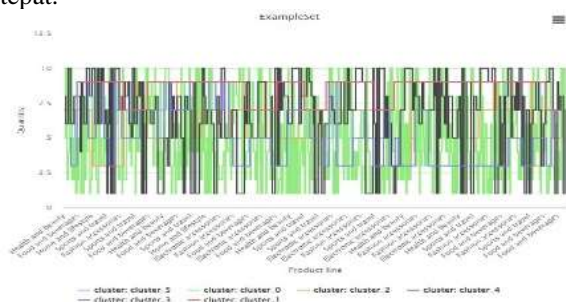
4. HASIL DAN PEMBAHASAN

Hasil dari penelitian ini adalah menghasilkan Pengelompokkan Data Penjualan yang berguna untuk mengoptimalkan rencana penyediaan stok barang dengan menyesuaikan tingkat persediaan dengan permintaan yang aktual menghindari kekurangan atau kelebihan stock, serta merencanakan pemesanan barang secara optimal, yang menjadi pedemoan dalam memanajemen stok barang dengan menggunakan algoritma clustering K-Means untuk evaluasi keakuratan model pada nilai Davies Bouldin Index untuk optimalisasi persediaan stock barang.



Gambar 13. Hasil DBI K2

Hasil evaluasi dari analisis menunjukkan bahwa pengelompokkan data menggunakan algoritma clustering secara keseluruhan sebesar -5406, yang mencerminkan tingkat kepadatan data terhadap centroid. Selanjutnya, nilai rata-rata jarak dalam centroid untuk masing-masing klaster adalah -5406 untuk Klaster 0, -2716 untuk Klaster 1, -3015. Selain itu, nilai Davies Bouldin yang sebesar -0.375 mengindikasikan kualitas pengelompokkan yang baik, dengan setiap klaster memiliki kepadatan yang seimbang dan terpisah dengan baik dari klaster lainnya. Hasil evaluasi ini memberikan pemahaman yang mendalam tentang distribusi data dalam setiap klaster, memberikan landasan yang kokoh untuk analisis lanjutan dan pengambilan keputusan yang tepat.



Gambar 14. Hasil cluster yang dominan

Berdasarkan hasil pengelompokan data, direkomendasikan untuk menyusun strategi stok

barang yang disesuaikan dengan karakteristik masing-masing klaster. Untuk Klaster 0 yang dominan dalam kategori "Fashion accessories", upaya stok barang yang difokuskan pada busana dan benda benda pelengkap busana dapat meningkatkan respons dari segmen ini. Sementara itu, untuk Klaster 1 yang terkait dengan kategori "Food and beverages", dapat menjadi sasaran untuk promosi makanan dan minuman. Dengan memahami ciri khas dan preferensi setiap klaster, perusahaan dapat mengarahkan upaya manajemen stok barang dengan lebih efektif, meningkatkan keterlibatan pelanggan, dan meningkatkan hasil penjualan secara keseluruhan.

5. KESIMPULAN DAN SARAN.

Kesimpulan skripsi ini menegaskan bahwa analisis data, algoritma K-Means clustering yang digunakan menghasilkan nilai pengelompokan secara keseluruhan sebesar -5406. Nilai ini menunjukkan kepadatan data yang baik terhadap centroid, atau titik pusat klaster. Hal ini diperkuat dengan rata-rata jarak dalam centroid untuk setiap klaster yang menunjukkan distribusi data yang terpusat dengan baik. Lebih lanjut, nilai Davies Bouldin sebesar -0.375 semakin memperkuat indikasi kualitas pengelompokan yang baik. Nilai ini menunjukkan bahwa klaster-klaster yang dihasilkan terpisah dengan baik dan memiliki kepadatan data yang seimbang. Hasil analisis menunjukkan bahwa algoritma clustering berhasil mengelompokkan data dengan kualitas yang baik distribusi data yang terpusat dengan baik dan nilai Davies Bouldin yang rendah menunjukkan bahwa klaster-klaster yang dihasilkan valid dan dapat digunakan untuk pedoman manajemen dalam mengambil keputusan penyediaan stock barang produk.

DAFTAR PUSTAKA

[1] Ramdhan, D., Dwilestari, G., Dinar Dana, R., & Ajiz, A. (n.d.). Clustering Data Persediaan Barang dengan Menggunakan Metode K-Means. Clustering Data Persediaan Barang dengan Menggunakan Metode K-Means. 7(1). [http://ejournal.ust.ac.id/index.php/Jurnal\\_Means](http://ejournal.ust.ac.id/index.php/Jurnal_Means)

[2] Kurnia Alvisan, F. (2021). Clustering Minimarket Untuk Menentukan Jumlah Kebutuhan Pembelian Menggunakan Metode K-Means. Jurnal NOE, 4(2). <https://ojs.unpkediri.ac.id/index.php/noe>

[3] Syafrinal, I., & Febrianti, E. L. (2023). Penerapan Algoritma K-Means Pada Aplikasi Data Mining Untuk Menentukan Pola Penjualan (Studi Kasus: Zahra Mart) (Vol. 13, Issue 1).

[4] 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>

[5] Setyorini, S. G., Sari, E. K., Elita, L. R., & Putri, S. A. (2021). Market Basket Analysis with K-Means and FP-Growth Algorithm as Citra



- Mustika Pandawa Company Analisis Keranjang Pasar Menggunakan Algoritma K-Means dan. Institute of Research and Publication Indonesia, 1(April), 41–46.
- [6] Murpratiwi, S. I., Agung Indrawan, I. G., & Aranta, A. (2021). Analisis Pemilihan Cluster Optimal Dalam Segmentasi Pelanggan Toko Retail. *Jurnal Pendidikan Teknologi Dan Kejuruan*, 18(2), 152. <https://doi.org/10.23887/Jptk-Undiksha.V18i2.37426>
- [7] Syahputra, H. (2022). Clustering Tingkat Penjualan Menu (Food and Beverage) Menggunakan Algoritma K-Means. *Jurnal KomtekInfo*, 29–33. <https://doi.org/10.35134/komtekinfo.v9i1.274>
- [8] 1087-Article Text-1855-1-10-20200704. (n.d.).
- [9] Royal, S. (2024). Perancangan Aplikasi Data Mining Untuk Menentukan Tingkat Kelarisan Produk Menggunakan Metode Clustering Dengan Algoritma K-Means. In *Journal Of Science And Social Research (Issue 1)*. [Http://Jurnal.Goretanpena.Com/Index.Php/Jssr](http://Jurnal.Goretanpena.Com/Index.Php/Jssr)
- [10] Widiyanto, A. T., & Witanti, A. (2021). Segmentasi Pelanggan Berdasarkan Analisis RFM Menggunakan Algoritma K-Means Sebagai Dasar Strategi Pemasaran (Studi Kasus PT Coversuper Indonesia Global). *KONSTELASI: Konvergensi Teknologi Dan Sistem Informasi*, 1(1), 204–215. <https://doi.org/10.24002/konstelasi.v1i1.4293>
- [11] Hasugian, P. S. (2018). Penerapan Data Mining Untuk Klasifikasi Produk Menggunakan Algoritma K-Means (Studi Kasus : Toko Usaha Maju Barabai). *Jurnal Mantik Penusa*, 2(2), 191–198.
- [12] Silalahi, M. (2018). Computer Based Information System Journal Analisis Clustering Menggunakan Algoritma K-Means Terhadap Penjualan Produk Padapt Batamas Niaga Jaya. *Cbis Journal*, 06(02). <http://ejournal.upbatam.ac.id/index.php/cbishttp://ejournal.upbatam.ac.id/index.php/cbis>
- [13] Siregar, Mh., Kuantan Singingi Ji Gatot Subroto Km, I., & Kuantan, T. (N.D.). Klasterisasi Penjualan Alat-Alat Bangunan Menggunakan Metode K-Means (Studi Kasus Di Toko Adi Bangunan).
- [14] Retno Lestari, D. (2019). Implementasi Data Mining Untuk Pengaturan Layout Swalayan Delimas Lestari Kencana Dengan Menggunakan Dengan Menggunakan Algoritma K-Means Clustering. In *JURIKOM (Vol. 6, Issue 2)*. <http://ejurnal.stmik-budidarma.ac.id/index.php/jurikom|Page|112>
- [15] Hasibuan, F. P. A., Sumarno, S., & Parlina, I. (2021). Penerapan K-Means pada Pengelompokan Penjualan Produk Smartphone. *SATESI: Jurnal Sains Teknologi Dan Sistem Informasi*, 1(1), 15–20. <https://doi.org/10.54259/satesi.v1i1.3>