# Perbandingan Algoritma Klastering dalam Pengelompokan Penjualan Produk Komputer

Naufal Aditya Rahman[1], Wiyli Yustanti[2]

[1,2] *Program Studi Sistem Informasi, Fakultas Teknik, Universitas Negeri Surabaya*
[1]naufalrahman16051214033@mhs.unesa.ac.id
[2]wiyliyustanti@unesa.ac.id

***Abstrak*— Penelitian ini membandingkan tiga algoritma clustering (K-means, K-Medoids, dan Agglomerative) untuk mengelompokkan data penjualan dari CV. Media Karya Komputindo. Untuk pemilihan jumlah klaster yang paling optimum menggunakan elbow dan divalidasi dengan Silhouette Coefficient. Statistik deskriptif dan *Word Cloud* digunakan untuk mengevaluasi interpretabilitas dari hasil algoritma *clustering* yang paling optimum. Efisiensi komputasi juga dipertimbangkan, mengingat sumber daya yang terbatas. Hasil penelitian menunjukkan bahwa semua algoritma menggunakan sumber daya yang sedikit dan mudah digunakan dengan PyCaret. Dari hasil Silhouette Coefficient, algoritma yang paling optimal untuk kumpulan data ini adalah K-Means, hasil clustering menunjukkan bahwa ada 4 kategori cluster, dimana cluster 1 berisi permintaan rendah dengan harga tinggi, cluster 2 permintaan rendah dengan harga rendah, cluster 3 permintaan tinggi dengan harga rendah, dan cluster 4 adalah permintaan cukup dengan harga rendah. Hasil clustering berhasil membentuk 4 cluster dalam segi harga maupun kuantitas penjualan.**

***Kata Kunci*— Studi Perbandingan, Clustering, Silhouette Coefficient, Validasi, Tanpa Supervisi, Nilai Masukan Optimal**

## I. PENDAHULUAN

Data mining merupakan suatu metode ekstraksi informasi dari suatu data dengan hasil informasi yang bermanfaat dan belum diketahui dan penggunaan data mining telah menjadi bagian integral dari berbagai disiplin ilmu, termasuk ilmu komputer, statistik, dan bidang-bidang lainnya [1]. Clustering adalah suatu identifikasi dan pengelompokkan menjadi suatu klaster dengan basis ukuran kesamaan [2]**.** Metode ini termasuk bagian dari *unsupervised* learning karena tidak ada ketentuan khusus dari setiap klaster [3]**.** Algoritma Clustering dibagi menjadi tiga kategori. *hierarchical*, *partitioning* dan *density*-based. Algoritma *hierarchical* membuat dekomposisi hirarkis terhadap database yang dipresentasikan dalam bentuk dendogram, Algoritma *partitioning* ditujukan dalam penentuan k-cluster yang mengoptimalkan dengan berbasis jarak dan kriteria lainnya, dan algoritma *desnsity*-based mencari daerah padat yang terpisah dengan daerah dengan kepadatan dan noise yang rendah [4]**.** Dalam melakukan *Clustering*, dalam setiap klaster yang terbentuk, juga dapat memiliki tolak ukur, seperti indeks validitas dari cluster [5]**.**

Dalam penelitian ini, tempat studi kasus yang digunakan adalah Toko Komputer di Kabupaten Lamongan yang bernama CV. Media Karya Komputindo merupakan salah satu UMKM yang bergerak di bidang penjualan komputer, laptop, aksesoris, dan jasa perbaikan komputer. Data yang digunakan adalah data penjualan. Selama ini, CV. Media Karya Komputindo masih melakukan analisa secara manual sehingga kurang efisiensinya waktu yang diperlukan, biaya dan objektivitas dalam menelaah isi dari data penjualan tersebut. Dengan menerapkan data mining dengan metode *clustering*, dapat melakukan analisa data dengan lebih cepat, efisien dan akurat.

Namun salah satu tantangan dalam menerapkan algoritma clustering adalah mencari algoritma *clustering* yang paling tepat dan juga cara penentuan jumlah klaster yang terbaik pada *dataset* yang akan digunakan. Oleh karena itu, algoritma clustering yang akan digunakan adalah algoritma yang pernah diterapkan pada data yang memiliki kemiripan dengan data yang akan digunakan pada penelitian ini. Penelitian terdahulu yang dijadikan acuan adalah implementasi metode *Agglomerative* pada data penjualan toko elektronik [6]**,** analisa data penjualan produk smartphone dengan metode K-medoids [7]**,** dan penerapan metode k-means dalam data penjualan produk digital konter [8]. Hasil dari beberapa penelitian tersebut adalah hasil kelompok-kelompok produk berdasarkan harga dan kuantitas penjualan dari produk-produk tersebut. *K-Means* merupakan algoritma klastering yang melakukan pengelompokan objek berdasarkan jarak terdekat dengan pusat klaster ke kelompok yang memiliki kesamaan satu sama lain, *K-Medoids* menggunakan objek representatif sebagai titik acuan, bukan mengambil nilai rata-rata dari objek dalam setiap klaster [9], dan *Agglomerative* merupakan algoritma klastering yang menetapkan setiap elemen sebagai klaster terpisah dan menggabungkannya dalam cluster yang lebih besar secara berurutan [2].

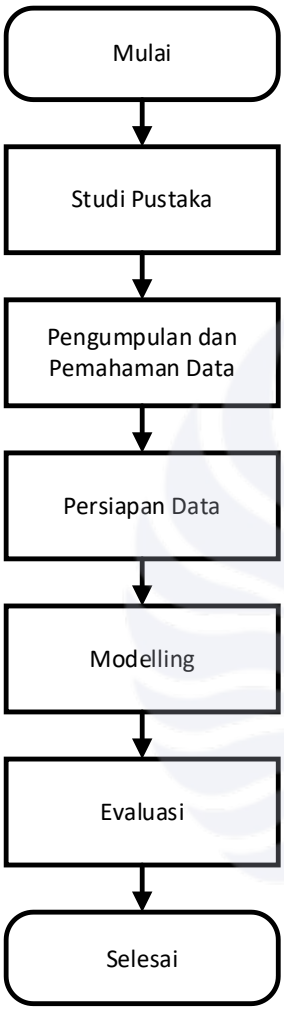
Untuk pemilihan klaster yang paling optimum digunakan metode *elbow* yang mana akan membandingkan selisih SSE (*Squared Sum of Errors*) dari setiap klaster. perbedaan yang paling tinggi yang membentuk sudut siku adalah jumlah klaster yang terbaik [10]. Lalu untuk indeks validasi digunakan *Silhouette Coefficient* dengan menghitung selisih dari jarak rata-rata di dalam klaster dan jarak minimum antar klaster [11]. Hasil dari *Silhouette Coefficient* adalah skor dalam rentang -1 dan 1, dimana semakin mendekati 1, semakin baik klaster tersebut. Kedua metode tersebut digunakan karena sudah ada beberapa penelitian terdahulu yang menggunakan metode indeks validasi dan pemilihan klaster terbaik pada algoritma *clustering* yang akan dipakai pada penelitian ini [12] [13] [14].

Hasil dari penelitian ini diharapkan dapat menemukan metode clustering yang dapat melakukan segmentasi

kelompok produk penjualan dari seberapa laku atau minat dari suatu produk dan juga harga dari produk tersebut yang mana dapat membantu CV. Media Karya Komputindo dalam menyusun strategi manajemen stok barang.

## II. METODOLOGI PENELITIAN

Berikut adalah tahapan penelitian yang akan dilakukan pada penelitian ini.




Gbr. 1 Tahapan Penelitian

### A. *Studi Literatur*

Sebelum melakukan penelitian, peneliti akan melakukan studi literatur untuk dijadikan referensi penelitian yang akan dilakukan, khususnya pada algoritma dan metode yang pernah digunakan pada data penjualan.

### B. *Pengumpulan dan Pemahaman Data*

Pada penelitian ini, peneliti akan mengambil data penjualan pada CV. Media Karya Komputindo pada tahun 2020 hingga 2021.

### C. *Persiapan Data*

Pada tahap ini akan dilakukan pembersihan data. Produk-produk abnormal, misal dengan harga atau kuantitas kosong atau nol, akan dihapus.

### D. *Modelling*

Pada tahap *modelling*, dilakukan pemilihan jumlah klaster terbaik menggunakan *elbow*, lalu algoritma *K-means*, *K-medoid*, dan *Agglomerative*, dijalankan dengan menggunakan *PyCaret*.

### E. *Evaluasi*

Pada tahap ini, validasi dari hasil model akan dilakukan. Metode indeks validasi clustering tersebut akan dijadikan acuan dalam penentuan klaster yang paling optimal. Metode atau indeks validasi klaster yang akan digunakan sebagai metrik kualitas hasil clustering, dalam tahapan ini, adalah *Silhouette Coefficient*. Setelah klaster di validasi dan di uji, akan dicari model yang paling optimal dengan indeks skor tertinggi, lalu akan dilakukan penggalian informasi dari setiap klaster.

## III. HASIL DAN PEMBAHASAN

Paragraf harus teratur. Semua paragraf harus rata, yaitu sama-sama rata kiri dan dan rata kanan.

### A. *Pengumpulan dan Pemahaman Data*

Data yang diambil adalah sekitar 5899 rekor data penjualan yang berisi tanggal, nama barang, kuantitas, dan harga.

. TABEL I
CUPLIKAN DATA PENJUALAN

| Tanggal | Nama Barang | Kuantitas | Harga |
|---|---|---|---|
| ... | ... | ... | ... |
| 02/01/2020 | Flashdisk Sandisk 16gb | 1 | 90000 |
| 02/01/2020 | Tinta Epson 664 Black | 2 | 85000 |
| 03/01/2020 | Tinta Canon F1 Black | 1 | 30000 |
| 03/01/2020 | Tinta Epson 664 Black | 1 | 85000 |
| 03/01/2020 | Flashdisk HP 16gb | 1 | 145000 |
| 03/01/2020 | Flashdisk Toshiba 16gb | 1 | 90000 |
| 04/01/2020 | Tinta Canon 810 Black | 1 | 210000 |
| .. | ... | ... | ... |

Kuantitas pada tabel ini adalah jumlah penjualan dari produk tersebut.

### B. *Persiapan Data*

Pada tahap ini, dilakukan proses pembersihan data. Penulis melakukan ekspor file excel menjadi CSV (*Comma separated value*), agar mudah dalam melakukan pembersihan data. Untuk pembersihan awal dilakukan secara manual. Tanggal dihapus karena data akan digabung dari nama produk yang sama dengan harga yang berbeda. Pada saat pembersihan, ditemukan produk yang satu paket dengan harga yang sangat

tinggi, produk tersebut adalah gabungan dari beberapa produk. Namun spesifikasi dari kumpulan produk tersebut tidak tertera atau bisa berbeda dari setiap produk, karena produk tersebut tidak dapat dipecah maka akan dihapus. Produk dengan Harga atau Kuantitas nol juga akan dihapus.

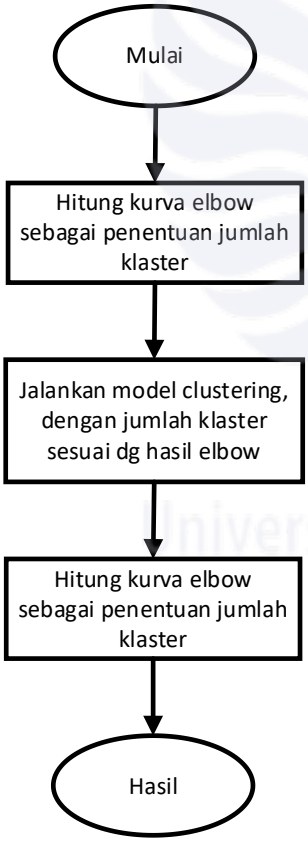
Setelah pembersihan data, dilakukan penggabungan data dengan nama produk sama dengan harga yang berbeda. Hasil setelah pembersihan dan penggabungan data adalah sebesar 2327 rekor. Setelah data dibersihkan, dilakukan proses impor data menuju PyCaret, fitur data numerik yang dimasukkan ke PyCaret adalah kuantitas dan harga, dan data yang dimasukkan namun diabaikan atau tidak digunakan sebagai fitur data adalah nama produk.

C. *Modelling*

Pada tahap ini, akan dilakukan penerapan dari beberapa algoritma *clustering*. Algoritma yang digunakan adalah algoritma yang akan digunakan pada beberapa penelitian sebelumnya. Algoritma yang akan dipakai dalam tahap *modelling* adalah sebagai berikut :

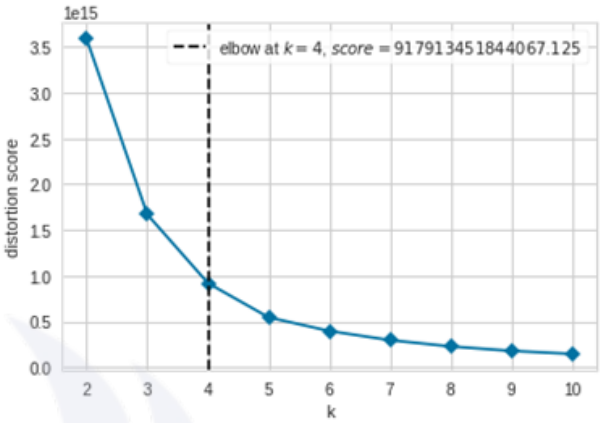
1) *K-means*
2) *K-medoids*
3) *Agglomerative*

Selanjutnya, alur proses pemodelan dengan algoritma-algoritma yang telah ditentukan, adalah sebagai berikut.




Gbr. 2 Tahapan Modelling

. Ketiga algoritma tersebut membutuhkan jumlah klaster yang akan digunakan sebelum dijalankan, oleh karena itu akan digunakan metode *elbow* dalam penentuan jumlah klaster yang paling optimum, yang mana metode tersebut mencari penurunan SSE (*Sum of Square Errors)* yang paling tinggi dari setiap nilai K. *K-value* dengan penurunan SSE paling tinggi akan digunakan sebagai jumlah klaster yang digunakan pada tahap modelling.




Gbr. 3 Grafik Kurva Elbow

Dalam grafik diatas, terlihat penurunan SSE terbesar terjadi pada k=3 dan k=4, dapat disimpulkan bahwa jumlah klaster yang akan digunakan sebagai jumlah klaster pada tahap *modelling* adalah 4.

Setelah jumlah klaster yang paling optimum ditetapkan, dijalankan semua algoritma clustering yang akan dibandingkan menggunakan indek validitas *Silhouette Coefficient*. Berikut adalah interpertasi dari hasil *Silhouette Coefficient* [11].

TABEL II
INTERPRETASI SKOR SILHOUETTE COEFFICIENT

| ***Silhouette Coefficient*** | **Interpretasi** |
|---|---|
| 0.71 – 1.00 | Struktur Kuat |
| 0.51 – 0.70 | Struktur Baik/Wajar |
| 0.26 – 0.50 | Struktur Lemah |
| ≤ 0.25 | Tidak ditemukan struktur yang substansial. |

Untuk, algoritma *K-Medoids* tidak ada dalam PyCaret. Cara lain menggunakan *K-Medoids* pada PyCaret adalah dengan melakukan instalasi modul "*scikit-learn-extra*" yang berisi algoritma *K-Medoids*, lalu algoritma tersebut dimasukkan saat menjalankan fitur clustering pada PyCaret.

Setelah dijalankan, akan dipilih algoritma terbaik dari skor *Silhouette Coefficient* tersebut, dimana semakin mendekati 1, semakin baik algoritma clustering tersebut pada data ini. Berikut adalah hasil indeks validitas S*ilhoeutte Coefficient* dari ketiga model clustering yang telah dijalankan.

TABEL III
SKOR SILHOUETTE COEFFICIENT

| **Algoritma** | ***Silhouette Coefficient*** |
|---|---|
| K-Means | 0,63510001 |

| | |
|---|---|
| K-Medoids | 0,63319999 |
| Agglomerative | 0,63029999 |

Dari tabel diatas, terlihat bahwa algoritma K-Means menghasilkan nilai *Silhouette Coefficient* yang paling tinggi, sementara algoritma *Agglomerative* mendapatkan skor yang paling rendah. Ketiga algoritma memiliki nilai *Silhoutte Coefficient* yang baik di angka 0.6, yang mana berarti ketiga klaster tersebut memiliki struktur yang baik.

TABEL IV
PENGUNAAN MEMORI DAN WAKTU EKSEKUSI

| **Algoritma** | ***Peak Memory* (mb)** | ***Peak Wall Time* (s)** |
|---|---|---|
| K-Means | 630 | 1,11 |
| K-Medoids | 441 | 0,961 |
| Agglomerative | 631 | 1,14 |

Dalam percobaan penggunaan memori dan waktu yang diperlukan saat menjalankan algoritma-algoritma tersebut, algoritma *K-Medoids* memiliki penggunaan memori dan waktu yang paling efisien, namun tidak terlalu signifikan dibanding yang lain. Dapat disimpulkan bahwa semua algoritma dapat berjalan pada suatu komputer dengan spesifikasi yang cukup atau terjangkau untuk perusahan dengan skala UMKM.

D. *Evaluasi*

Dari hasil pengamatan pada tahapan sebelumnya, *K-Means* 4 adalah model yang paling optimum menurut indeks validasi Silhouette Coefficient, oleh karena itu *K-means 4* adalah hasil *clustering* yang akan dievaluasi pada tahap ini. Digunakan statistik deskriptif dan metode *text mining Word Cloud* untuk mengevaluasi hasil dari hasil model yang paling optimum. Berikut adalah tabel statistika deskriptif untuk harga dan kuantitas penjualan pada model tersebut :

TABEL V
STATISTIKA HARGA PRODUK K-MEANS 4

| **Model** | **Klaster** | **Harga** | | | |
|---|---|---|---|---|---|
| | | **St.Dev** | **Avg** | **Min** | **Max** |
| K-Means (4) | 1 | 1957880 | 3887891 | 2100000 | 21400000 |
| | 2 | 548130 | 342205 | 2000 | 2100000 |
| | 3 | 50304 | 64893 | 6817 | 94913 |
| | 4 | 53732 | 58884 | 3000 | 235176 |

Pada hasil statistika harga produk, tampak bahwa klaster pertama memiliki jarak minimum dan maksimum yang jauh, dari 2 juta hingga 21 juta, dan klaster kedua dari 2 ribu hingga 21 juta. Klaster ketiga dan keempat memiliki angka maksimum di angka ratusan ribu, yang mana lebih rendah dibanding dengan klaster pertama dan kedua.

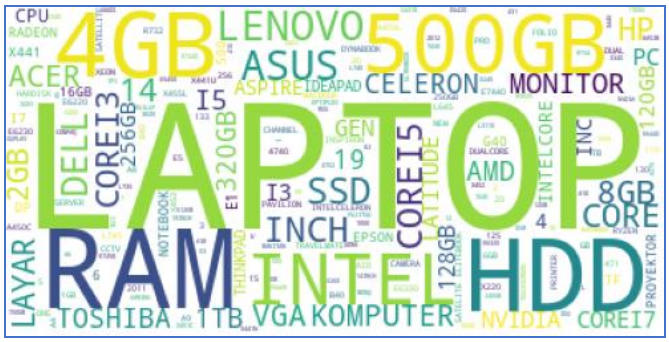
TABEL VI
STATISTIKA KUANTITAS PENJUALAN K-MEANS 4

| **Model** | **Klaster** | **Total Penjualan** | | | |
|---|---|---|---|---|---|
| | | **St.Dev** | **Avg** | **Min** | **Max** |
| K-Means (4) | 1 | 2,012 | 1,48 | 1 | 30 |
| | 2 | 3,73 | 2,66 | 1 | 28 |
| | 3 | 86,06 | 261 | 204 | 360 |
| | 4 | 27 | 61,72 | 35 | 145 |

Dari hasil statistika kuantitas, dan juga hasil dari statistika harga produk, dapat disimpulkan bahwa klaster pertama dan kedua memiliki angka minimum dan maksimum yang sama, namun dalam segi harga, kedua klaster tersebut bertolak belakang. Klaster pertama memiliki angka minimum 21 juta dan klaster kedua memiliki angka maksimum 21 juta, menunjukkan adanya kontinuitas dari segmentasi harga untuk kategori klaster penjualan rendah. Klaster ketiga dan keempat walaupun dengan angka minimum dan maksimum yang mirip, perbedaan terletak pada kuantitas yang mana klaster ketiga memiliki penjualan yang lebih tinggi dibanding klaster keempat.

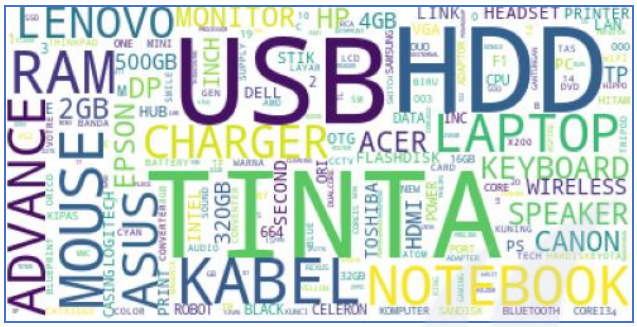
Secara keseluruhan, terlihat hampir tidak adanya *overlap* antar klaster dalam segi harga dan kuantitas. Klaster 2 dan Klaster 1 memiliki kuantitas penjualan yang mirip, namun dengan perbedaan di bagian harga. Klaster 3 dan Klaster 4 memiliki perbedaan di kuantitas penjualan, dengan perbedaan harga yang tidak banyak.

Setelah analisa statistika deskriptif, dilakukan data mining dengan menggunakan metode *Word Cloud*. Hal ini dilakukan agar dapat mencari karakteristik atau kategori dari setiap klaster berdasarkan nama-nama produk yang ada pada klaster tersebut. Dalam pemrosesan *word cloud*, semua nama produk dalam klaster gabung menjadi satu, semua kata dijadikan kapital, dan titik atau koma dihapus. Hal ini diperlukan untuk menghindari produk duplikat karena perbedaan huruf besar dan kecil, dan kata yang sama namun dengan simbol yang tidak bermakna. Kata-kata yang hanya terdiri dari angka juga dimasukkan, hal ini dikarenakan oleh adanya suatu produk yang menggunakan angka sebagai nama atau tipe dari produk tersebut. Hasil dari *word cloud* akan memberikan visualisasi dari nama atau tipe produk yang paling dominan pada klaster tersebut berdasarkan frekuensi, dimana semakin besar frekuensi dari kata tersebut, maka semakin besar kata tersebut pada gambar *word cloud*. Berikut adalah hasil gambar *word cloud* dari setiap klaster *K-means*.



Gbr. 4 WordCloud Klaster 1

Pada klaster pertama, tampak keyword dominan “Laptop”, “4GB”, “RAM, ”500GB” dan “HDD”. Dapat disimpulkan bahwa 4GB mengarah pada jumlah RAM (*Random Access Memory*, dan 500GB mengarah pada HDD (*Hard Disk Drive*). Namun, tidak dapat disimpulkan jika 4GB RAM dan 500GB. Dari rentang harga klaster pertama, kemungkinan besar 4GB RAM dan 500GB HDD mengarah pada kata-kata dalam produk Laptop atau Komputer.




Gbr. 5 WordCloud Klaster 2

Pada klaster kedua, tampak dominan “TINTA”, “USB”, “HDD”, “CHARGER”, “KABEL”, dan beberapa nama merek produk elektronik. Dari harga dan kuantitas klaster ini, dapat disimpulkan bahwa klaster ini berisi aksesoris printer, laptop dan komputer.


Gbr. 6 WordCloud Klaster 3

Pada klaster ketiga, tampak keyword dominan “TINTA”, “EPSON”, “BLACK” dan “RJ45”. Hasil *word cloud* juga tampak sedikit, yang menyatakan klaster ini memiliki produk yang sedikit. Dari analisa statistika deskriptif menyatakan bahwa klaster ini memiliki total penjualan yang tinggi atau barang yang paling laku.


Gbr. 7 WordCloud Klaster 4

Pada klaster ke empat, tampak keyword dominan “TINTA”, “KABEL”, “MOUSE”, “EPSON”, dan beberapa merek produk elektronik. Produk tinta pada klaster ini berbeda dengan klaster ketiga, dimana juga tampak keyword “Cyan”, “Magenta” dan “Yellow”, yang mana menyimpulkan bahwa produk tinta pada klaster ini adalah tinta warna.

Dari hasil statistika deskriptif dan *word cloud*, berikut adalah hasil analisa karakteristik dari setiap klaster tersebut:

TABEL VII
KARAKTERISTIK TIAP KLASTER

| Klaster | Karakteristik Klaster |
|---|---|
| Klaster 1 | Penjualan rendah, Harga mahal (diatas 2 juta)<br>Laptop, Komputer dan Part Komputer (VGA, Monitor, ...) |
| Klaster 2 | Penjualan rendah, Harga rendah (dibawah 2 juta)<br>Aksesoris Komputer dan laptop, Sparepart Komputer dan Laptop |
| Klaster 3 | Penjualan tinggi, Harga rendah (dibawah 100 ribu)<br>Tinta Hitam dan Kabel RJ45/Lan |
| Klaster 4 | Penjualan diantara klaster 2&3, Harga rendah (dibawah 235 ribu)<br>Aksesoris Komputer dan Laptop yang bernilai rendah seperti flashdisk, Tinta Warna, Mouse, Keyboard, Kabel, ... |

Dari hasil karakteristik yang ditemukan, hasil kategori dari setiap klaster adalah sebagai berikut :

1) *Klaster 1*: Barang harga tinggi, penjualan rendah.
2) *Klaster 2*: Barang harga rendah, penjualan rendah.
3) *Klaster 3*: Barang harga rendah, penjualan tinggi.
4) *Klaster 4*: Barang harga rendah, penjualan cukup.

Berdasarkan hasil kategori setiap klaster, toko dapat mengelola stok barang dengan lebih efektif yang memastikan persediaan yang memadai untuk memenuhi kebutuhan pelanggan dalam masing-masing kelompok produk tersebut.

## IV. KESIMPULAN DAN SARAN

### A. *Kesimpulan*

Dari ketiga Algoritma *clustering* yang digunakan, algoritma yang paling optimum pada penelitian ini adalah *K-Means*, dengan skor indeks *Silhouette Coefficient* tertinggi, dengan angka 0,635 yang berarti hasil *clustering* tersebut memiliki struktur yang baik atau wajar. Algoritma lain yang dibandingkan dalam penelitian ini memiliki skor *Silhouette*

*Coefficient* yang sedikit lebih rendah, dimana *K-Medoids* dengan skor 0,633 dan *Agglomerative* dengan skor 0,63.

Hasil analisa statistik deskriptif dan *Word Cloud* menghasilkan bahwa hasil dari algoritma *K-means* dapat melakukan pengelompokan dari segi harga dan kuantitas penjualan dengan baik, hasil dari setiap kelompok menunjukkan 4 kategori, yaitu kelompok harga tinggi dengan penjualan rendah dengan rentang harga 2 juta hingga 21 juta, kelompok harga rendah dengan penjualan rendah dengan rentang harga 2 ribu hingga 2 juta, dan kelompok harga rendah dengan penjualan tinggi dengan rentang harga 6 ribu hingga 94 ribu dan harga rendah dengan penjualan yang cukup dengan rentang harga 5 ribu hingga 235 ribu. Rentang harga dan kuantitas kelompok-kelompok produk tersebut memudahkan CV. Media Karya Komputindo untuk menemukan pola pengelompokan produk-produk untuk membantu dalam strategi manajemen stok barang.

## B. *Saran*

Saran yang dapat diberikan berdasarkan hasil pada penelitian ini adalah diharapkan pada penelitian selanjutnya dapat melakukan penelitian menemukan indeks validasi lain atau metode penentuan klaster lain yang dapat menghasilkan segmentasi yang lebih baik.

Adapun dimensi dan jumlah data yang digunakan pada penelitian ini relatif kecil, diharapkan penelitian selanjutnya dapat melakukan penelitian dengan dimensi dan/atau jumlah data yang lebih banyak.

## REFERENSI

[1] Witten, I. H., Frank, E., Hall, M. A., & Pal, C. (2016). *Data Mining: Practical Machine Learning Tools and Techniques*. Morgan Kaufmann.

[2] Madhulatha, T. S. (2012). An overview on clustering methods. *arXiv preprint arXiv:1205.1117*. https://doi.org/10.48550/arXiv.1205.1117

[3] Han, J., Kamber, M., & Mining, D. (2006). Concepts and techniques. *Morgan Kaufmann*, *340*, 94104–3205.

[4] Areed, M. F. (2020). Python-based graphical user interface for automatic selection of data clustering algorithm. *International Journal of Future Generation Communication and Networking*, *13*, 451–461.

[5] Brock, G., Pihur, V., Datta, S., & Datta, S. (2008). clValid: An R Package for Cluster Validation. *Journal of Statistical Software*, *25*, 1–22. doi:10.18637/jss.v025.i04

[6] Ariska, E., & others. (2018). Implementasi agglomerative hierarchical Clustering pada data produksi dan data penjualan perusahaan. Ph.D. dissertation. http://repositori.usu.ac.id/handle/123456789/4040.

[7] Yustika, M., Windarto, A. P., & Purba, Y. P. (2021). Analisis Metode K-Medoids Pada Penjualan Produk Smartphone Vivo Di Kota Pematangsiantar. Brahmana: Jurnal Penerapan Kecerdasan Buatan, 3(1), 27-33.

[8] Astuti, N., Utamajaya, J. N., & Pratama, A. (2022). Penerapan Data Mining Pada Penjualan Produk Digital Konter Leppangeng Cell Menggunakan Metode K-Means Clustering. JURIKOM (Jurnal Riset Komputer), 9(3), 754-760.

[9] Ramadhani, R. D., & AK, D. J. (2019, June). Evaluasi K-Means dan K-Medoids pada Dataset Kecil. In SNIA (Seminar Nasional Informatika dan Aplikasinya) (Vol. 3, pp. D-20).

[10] Umargono, E., Suseno, J. E., & Gunawan, S. (2019). K-Means clustering optimization using the elbow method and early centroid determination based-on mean and median. *Proceedings of the International Conferences on Information System and Technology*, (hal. 234–240).

[11] Wang, X., & Xu, Y. (2019, July). An improved index for clustering validation based on Silhouette index and Calinski-Harabasz index. *IOP Conference Series: Materials Science and Engineering, 569,* 052024. doi:10.1088/1757-899X/569/5/052024

[12] Utami, D. S., & Saputro, D. R. S. (2018). Pengelompokan data yang Memuat Pencilan dengan Kriteria Elbow dan Koefisien Silhouette (Algoritme K-Medoids).

[13] Samudra, F. P., Pudjiantoro, T. H., & Santikarama, I. (2021, October). Klasterisasi Tingkat Penjualan Produk Menggunakan Metode K-Means Untuk Penerapan Konsep Down-Selling: Studi Kasus pada Artch Indonesia. In SNIA (Seminar Nasional Informatika dan Aplikasinya) (Vol. 5, pp. B9-15).

[14] Patel, P., Sivaiah, B., & Patel, R. (2022, July). Approaches for finding optimal number of clusters using k-means and agglomerative hierarchical clustering techniques. In 2022 International Conference on Intelligent Controller and Computing for Smart Power (ICICCSP) (pp. 1-6). IEEE.