

Terbit online pada laman : <http://teknosi.fti.unand.ac.id/>

Jurnal Nasional Teknologi dan Sistem Informasi

| ISSN (Print) 2460-3465 | ISSN (Online) 2476-8812 |



Artikel Penelitian

Dynamic Time Warping Pada Metode K-Means Untuk Pengelompokan Data Trend Penjualan Produk

Fairizal Aaron Wahyu Tanamas Satria Wijaya, Eko Prasetyo*, Rahmawati Febrifyaning Tias

Program Studi Teknik Informatika, Fakultas Teknik, Universitas Bhayangkara Surabaya, Jl. Ahmad Yani No.114, Ketintang, Kec. Gayungan, Kota Surabaya, Jawa Timur 60231, Indonesia

INFORMASI ARTIKEL

Sejarah Artikel:

Diterima Redaksi: 28 Maret 2024

Revisi Akhir: 02 Juni 2024

Diterbitkan Online: 31 Agustus 2024

KATA KUNCI

Trend,
Pengelompokan,
Penjualan,
K-Means,
Dynamic Time Warping

KORESPONDENSI

E-mail: eko@ubhara.ac.id *

ABSTRACT

Pasar swalayan menghasilkan transaksi penjualan dengan pelanggan dengan laju pertumbuhan data tinggi. Data transaksi sejumlah barang memiliki pola berbeda, ada yang tinggi dan ada yang rendah. Perbedaan ini mengakibatkan pola penjualan antar satu produk dengan produk lainnya berbeda. Pola ini juga mengakibatkan pengelolaan stok dan promosi juga berbeda. Maka, untuk memudahkan pengelolaan barang, perlu dilakukan pengelompokan trend penjualan otomatis menggunakan K-Means. Namun, metode perhitungan jarak (dissimilarity) standar tidak tepat digunakan pada data periodik. Penelitian ini bereksperimen mengelompokkan data periodik trend penjualan produk menggunakan metode K-Means dan DTW (Dynamic Time Warping) untuk perhitungan jarak. DTW melakukan perhitungan jarak dengan memperhatikan data histori periode sebelum dan setelah data yang sedang dihitung. Pelibatan data disekitarnya akan membentuk jarak sesuai dengan pola data periodik. Hasil pengelompokan data tahun 2021, Cluster 1 dengan 417 data memiliki pola naik dari awal tahun kemudian menurun, selanjutnya perlahan naik kembali di pertengahan hingga akhir tahun. Cluster 2 dengan 2403 data memiliki pola stabil di awal tahun, ketika di pertengahan tahun mulai naik, kemudian di akhir tahun turun namun kembali naik. Cluster 3 dengan 289 data memiliki pola menurun saat awal tahun, namun terlihat stabil hingga akhir tahun. Hasil ekeperimen pengelompokan trend penjualan menggunakan metode K-Means dengan jarak DTW menunjukkan hasil yang baik dan memiliki pola data dalam kelompok yang sesuai karakteristik alami. Hal ini didukung oleh validitas Davies-Bouldin index -1.215 dimana nilai yang semakin kecil mendekati nol menunjukkan kualitas cluster yang dicapai baik.

1. PENDAHULUAN

Swalayan merupakan salah satu bentuk pasar modern yang menjembatani produk antara produsen dan konsumen. Dengan pelayanan dan pembayaran yang semakin mudah dan cepat maka transaksi penjualan juga semakin banyak. Semakin banyak transaksi penjualan yang dilakukan maka semakin banyak pula data yang dibangkitkan. Maka, pengolahan data penjualan dengan bantuan teknologi penting dilakukan untuk mendapatkan informasi baru yang berguna bagi keberlanjutan bisnis [1]. Pada kenyataannya, penjualan produk antara satu produk dengan produk lainnya tidak selalu sama. Ada yang penjualannya tinggi dan ada yang rendah. Perbedaan tingkat penjualan produk ini

menghasilkan perbedaan pola penjualan antara satu produk dengan produk lainnya. Pola yang terjadi ini juga mengakibatkan perlakuan pengelolaan stok dan promosi juga berbeda, sehingga untuk memudahkan pemilahan data tersebut, perlu dilakukan pengelompokan trend penjualan yang ada. Masalah penting lainnya adalah pola penjualan selama periode tertentu dapat menjadi early warning bagi kepentingan bisnis untuk mempersiapkan stok pada periode tersebut. Misalnya penjualan alat tulis kantor akan menunjukkan peningkatan penjualan dibulan Pebruari-Maret maka akan menjadi perhatian pihak manajemen agar mempersiapkan stok menyambut datangnya bulan tersebut. Penyelesaian masalah tersebut dapat dibantu dengan pengelompokan data penjualan selama periode tertentu. Pekerjaan tersebut juga akan mengalami kendala ketika jumlah

barang penjualannya sangat banyak. Akibatnya, proses pengelompokan tersebut membutuhkan waktu lama jika dilakukan secara manual. Penelitian ini menyelesaikan masalah pengelompokan trend penjualan untuk mendapatkan kelompok trend penjualan barang yang dijual.

Pengelompokan menjadi pilihan yang tepat digunakan untuk menyelesaikan masalah bisnis seperti pengolahan data pariwisata untuk meningkatkan strategi pemasaran [2], pengelompokan data penjualan untuk mendapatkan koherensi antar kelompok [3] dan pengelompokan data penjualan berdasarkan lokasi, jenis industry dan faktor yang mempengaruhi untuk mendukung peningkatan level bisnis [4]. Namun permasalahan data penjualan tidak semata-mata data matrik yang umum digunakan, data ini memiliki pola tren yang dapat berubah dari waktu ke waktu sesuai dengan periode tertentu atau periodik.

Pengelompokan data bertujuan untuk melakukan partisi data sesuai dengan karakter alami data. Umumnya, data yang mirip akan tergabung dalam cluster yang sama, sedangkan data berbeda akan terpisah dalam cluster yang lain. Metode yang populer digunakan adalah K-Means, seperti penelitian oleh Gustientiedina, M. Hasmil Aditya, dan Yenny Desnelitab (2019) mengenai pengelompokan data obat-obatan pada RSUD [5], Wakhid Afifi, Dhiya'an Ramadhanty Nastiti dan Qurrotul Aini (2020) mengelompokkan data ekspor [6], Adi Supriyatma, Irmawati Carolina Suhar Janti, dan Ali Haidir (2020) melakukan pengelompokan koridor Transjakarta berdasarkan jumlah penumpang [7]. Penelitian lain yang mereapkan K-Means juga dilakukan oleh R. A. Indraputra dan R. Fitiana (2020) yaitu tentang pengelompokan data covid-19[8], Achmad Bahauddin, Agustina Fatmawati, dan Febrianti Permata Sari (2021) melakukan analisis pengelompokan provinsi di Indonesia berdasarkan tingkat kemiskinan [9]. Penelitian lain berikutnya yang menggunakan K-Means adalah Ninaria Purba, Poningsih, dan Heru Satria Tambunan (2021) melakukan pengelompokan penyakit ispa di provinsi Riau [10], Dimas Galang Ramadhan, Indri Prihatini, dan Febri Liatoni (2021) pengelompokan penjualan paket data [11]. Penelitian yang dilakukan oleh Muhammad Irfan Rizkia, Teguh Ammar Taqqiyunnis, dan Jessica Jesslyn Cereliad (2021) juga melakukan pengelompokan provinsi di Indonesia berdasarkan kasus aktif covid-19 [12]. Varian K-Means juga sudah digunakan peneliti sebelumnya, Sani Putriana, Ernawati, dan Desi Andreswari (2021) melakukan pengelompokan data titik gempa menggunakan metode Fuzzy Possibilistic C-Means [13], Muhamad Budiman Johra (2021) tentang pengelompokan desa di kota Tidore, menggunakan metode Fuzzy K-Means [14]. Semua penelitian tersebut menggunakan data standar dimana tidak ada urutan nilai pada variabelnya, sehingga penggunaan jarak standar seperti Euclidean masih dapat digunakan. Data dengan nilai variabel urutan memiliki makna biasanya disebut data periodik (*time series*).

Dalam pengelompokan, ada langkah perhitungan jarak antar data yang harus dilakukan, Namun pada data periodik, perhitungan jarak antar data memiliki keunikan karena memperhitungkan urutan nilai, sehingga perlu metode khusus untuk melakukan seperti *Dynamic Time Warping* (DTW). DTW menjadi metode yang dibuktikan penggunaannya untuk menyelesaikan perhitungan jarak pada data periodik, seperti dilakukan dalam penelitian kecenderungan degradasi bantalan [15], deteksi

anomali [16], evaluasi perubahan ekosistem [17], penilaian distorsi skala [18], dan klasifikasi catatan aktivitas vulkanologi [19]. Penggunaan tersebut memang tepat karena DTW mempertimbangkan hubungan satu variabel dengan variabel lain yang berurutan.

Pada penelitian yang sudah dilakukan sebelumnya, pengelompokan dilakukan menggunakan metode seperti K-Means, namun data yang diselesaikan adalah data matrik, bukan data periodik (*time series*) sehingga penggunaan metode dissimilarity standar seperti Euclidean masih dapat digunakan. Sedangkan pada data periodik mempunyai sifat berbeda dimana kolom sudah terurut waktu sehingga tidak boleh dipertukarkan. Maka dari itu, penggunaan dissimilarity standar seperti K-Means tidak dapat digunakan. Data penjualan dalam penelitian ini juga merupakan data periodik, sehingga untuk menyelesaikan pengelompokannya digunakan K-Means dan DTW sebagai metrik jarak.

Maka dari itu, penelitian ini mengusulkan penggunaan metode K-Means dan DTW untuk mengelompokan data trend penjualan. DTW digunakan sebagai metode untuk perhitungan jarak (*dissimilarity*) dalam K-Means. Dataset yang digunakan dalam penelitian ini adalah data penjualan di swalayan Rumah Indra selama 2 tahun mulai Januari 2020 hingga Desember 2021. Data penjualan yang digunakan sudah melalui proses agregasi menjadi data penjualan tiap bulan. Metode K-Means merupakan salah satu metode pengelompokan data yang bukan *time series* karena pengukuran jarak menggunakan metode *Euclidean*, maka diperlukan metode DTW sebagai pengganti *Euclidean* untuk menghitung pengukuran jarak dari data *time series*.

Hasil pengelompokan pada data tahun 2021, Cluster 1 dengan 417 data yang naik dari awal tahun yang sempat turun kemudian perlahan naik kembali saat berjalan dari pertengahan hingga akhir tahun. Cluster 2 dengan 2403 data yang stabil di awal tahun, menuju pertengahan tahun mulai terjadi kenaikan, kemudian di akhir tahun sempat turun namun kembali naik. Cluster 3 dengan 289 data yang menurun saat awal tahun, namun terlihat stabil hingga akhir tahun.

2. METODE

2.1. *Dynamic Time Warping*

Data berfluktuasi dan bervariasi dari waktu ke waktu, sehingga metode pengklasteran tidak lagi dapat menggunakan jarak sederhana Euclidean untuk mengukur kedekatan antar objeknya. Jarak Euclidean ini cocok digunakan untuk fungsi data yang tidak berubah dari waktu ke waktu atau data statis. Sementara itu, dalam pembentukan matriks kedekatan jarak antar objek pada data deret waktu juga dibutuhkan perhitungan jarak antar deret waktunya yang bersifat dinamis, sehingga diperlukan cara pengelompokan yang berbeda dari data statis. Oleh karena itu, tindakan pengukuran alternatif yang dapat digunakan adalah *Dynamic Time Warping* (DTW) [12].

Data berfluktuasi dan bervariasi dari waktu ke waktu, sehingga metode pengklasteran tidak lagi dapat menggunakan jarak sederhana Euclidean untuk mengukur kedekatan antar objeknya. Jarak Euclidean ini cocok digunakan untuk fungsi data yang tidak

berubah dari waktu ke waktu atau data statis. Sementara itu, dalam pembentukan matriks kedekatan jarak antar objek pada data deret waktu juga dibutuhkan perhitungan jarak antar deret waktunya yang bersifat dinamis, sehingga diperlukan cara pengelompokan yang berbeda dari data statis. Oleh karena itu, tindakan pengukuran alternatif yang dapat digunakan adalah Dynamic Time Warping [12].

a. Baris pertama

$$D(1, j) = |c(x_1, y_j)| + D(i, j - 1) \quad (1)$$

b. Kolom pertama

$$D(i, 1) = |c(x_i, y_1)| + D(i - 1, j) \quad (2)$$

c. Elemen lainnya

$$D(i, j) = |c(x_i, y_j)| + \min \left\{ \begin{array}{l} D[i - 1, j - 1] \\ D[i - 1, j] \\ D[i, j - 1] \end{array} \right\} \quad (3)$$

Dimana:

x_i = baris matriks

y_j = kolom matriks

$c(x_i, y_j)$ = hasil baris dikurangi kolom

$D(i, j)$ = hasil matriks sebelumnya

2.2. K-Means

Algoritma K-Means merupakan algoritma klusterisasi yang mengelompokkan data berdasarkan titik pusat kluster (centroid) terdekat dengan data. Tujuan K-Means adalah pengelompokkan data dengan memaksimalkan kemiripan data dalam satu kluster dan meminimalkan kemiripan data antara kluster. ukuran kemiripan yang digunakan dalam kluster adalah fungsi jarak. sehingga pemaksimalan kemiripan data didapatkan berdasarkan jarak terpendek antara data titik centroid [10].

$$d(x_i - c_j) = \sqrt{\sum_{i=1}^N (x_i - c_j)^2} \quad (3)$$

Dimana:

d = jarak

X_i = data

C_j = centroid

i = variable data

N = dimensi data

Dalam K-Means, setiap data menjadi anggota saah satu cluster dengan formula (4).

$$a_{ij} = \begin{cases} 1 & \arg \min \{d(x_i, c_j)\} \\ 0 & \text{lainnya} \end{cases} \quad (4)$$

Dimana:

a_{ij} = nilai kenggotaan data ke- i terhadap cluster j dengan nilai 1 jika anggota cluster tersebut dan 0 jika bukan.

Setiap cluster dilakukan perhitungan centroid berdasarkan rata-rata data pada setiap fitur menggunakan formula (5).

$$c_j = \frac{1}{N_k} \sum_{l=1}^{N_k} x_{jl} \quad (5)$$

Dimana:

N_k adalah jumlah data pada setiap cluster.

Pengelompokan dengan K-Means diselesaikan secara iterative dengan algoritma sebagai berikut:

1. Inisialisasi: Tahap awal ditentukan nilai K sebagai jumlah cluster yang diharapkan, menentukan metrik ketidakmiripan (jarak). Jika perlu, ditetapkan juga ambang batas perubahan fungsi objektif, dan ambang batas perubahan posisi centroid.
2. Pilih K data dari set data secara acak sebagai centroid awal
3. Alokasikan semua data ke centroid terdekat dengan metrik jarak yang sudah ditetapkan (memperbarui cluster ID setiap data)
4. Lakukan perhitungan kembali centroid C berdasarkan data yang tergabung dalam cluster masing-masing.
5. Ulangi langkah 3 dan 4 hingga kondisi konvergen terpenuhi salah satu sebagai berikut: (1) Perubahan fungsi objektif sudah dibawah ambang batas yang diinginkan; atau (2) tidak ada data yang berpindah cluster; atau (3) perubahan posisi centroid sudah dibawah threshold yang ditetapkan.

2.3. Dataset

Dataset yang digunakan dalam penelitian ini adalah data penjualan diperoleh selama 2 tahun dalam 12 bulan sebanyak lebih dari 3000 data yang telah diolah. Data tahun 2020 dan 2021 masing-masing berjumlah 3668 dan 3109 data. Data tersebut telah di olah dengan pre-processing sehingga menjadi bentuk tabulasi sehingga siap diolah menggunakan metode pengelompokan. Tabel 1 merupakan contoh data yang diambil dari data yang telah disebutkan.

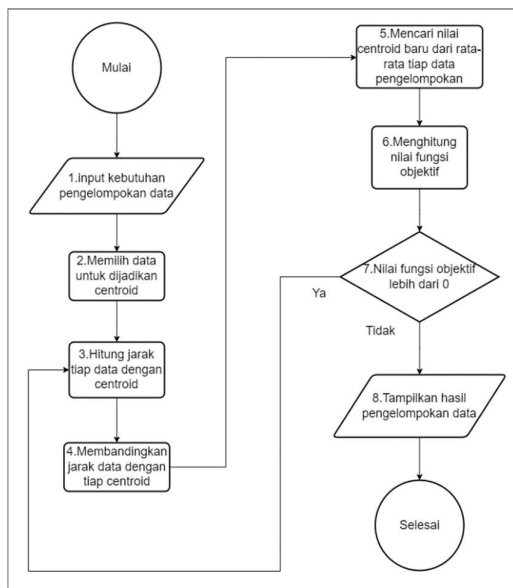
Tabel 1. Sampel data penjualan

Nama Item	Jan	Feb	Mar	Apr	Mei	Jun	Jul	Ags	Sep	Okt	Nov	Des
Frostbite Cookies Cream 60 ml	5	6	5	6	1	1	10	1	10	0	1	1
Paddle Pop Trico 60 ml	116	71	61	29	43	59	27	16	68	14	21	36
Paddle Pop Choco Magma 50 ml	44	54	79	60	37	25	34	91	51	17	62	31
Paddle Pop Rainbow 50 ml	67	58	42	58	57	46	38	58	53	23	50	63
Feast Coklat 65 ml	32	29	36	37	18	11	16	21	0	0	11	17
Feast Vanilla 65 ml	31	32	27	40	31	18	8	13	31	0	23	11
Cornetto Black and White 82 ml	38	53	50	51	49	48	49	41	31	4	29	54
Cornetto Strawberry 82 ml	27	42	37	18	35	16	19	30	1	3	17	4
Cornetto Disc Chocolate 120 ml	7	6	11	5	8	2	2	0	0	0	0	7
Cornetto Mini Eceran	22	63	49	44	95	84	122	62	10	1	74	55

2.4. Diagram alir sistem

Gambar 1 merupakan diagram alir aplikasi untuk mengelompokkan data menggunakan metode K-Means, seperti yang digunakan dalam penelitian ini, sebagai berikut:

1. Diawali dengan memasukkan kebutuhan pengelompokan seperti tahun, kategori item, dan jumlah cluster. Kami menggunakan data penjualan tahun 2020 dan 2021 berjumlah 6777 data. Proses clustering menggunakan K-Means.
2. Selanjutnya memilih data yang digunakan sebagai centroid awal.
3. Kemudian menghitung jarak tiap data dengan data centroid. Perhitungan jarak menggunakan DTW menggunakan 12 bulan data penjualan.
4. Lalu membandingkan hasil jarak tiap data.
5. Setelah itu mencari centroid baru dari hasil rata-rata data yang termasuk dalam kelompok centroid.
6. Lalu menghitung nilai fungsi objektif.
7. Jika nilai fungsi objektif lebih besar dari ambang batas maka dilakukan perulangan ke langkah menghitung jarak tiap data dengan centroid (kembali ke langkah 3).
8. Jika tidak maka tampilkan hasil pengelompokan.



Gambar. 1. Flowchart

3. HASIL

Pengujian yang dilakukan pada sistem pengelompokan trend penjualan adalah melakukan pengelompokan trend penjualan dengan menggunakan metode K-Means, untuk mengetahui produk apa saja yang memiliki intensitas penjualan yang sama per tahunnya, serta seperti apa trend penjualan tiap tahunnya. Dalam uji coba ini dilakukan identifikasi pada data penjualan, dengan data tahun, kategori produk, dan jumlah penjualan bulanan tiap produk pada tahun 2020 dan 2021. Eksperimen clustering dilakukan pada tiga kasus sebagai berikut: (1) Pengujian pada data kategori alat tulis; (2) Pengujian pada seluruh data. Data yang digunakan untuk masing-masing pengujian adalah data tahun 2020 dan 2021. Hasil pengelompokan dilakukan analisa dan penemuan pola data tertentu yang berbeda dari data umumnya.

3.1. Pengelompokan data kategori alat tulis

Dalam penentuan jumlah cluster, peneliti kami juga bereksperimen pada beberapa pilihan jumlah cluster. Validitas cluster menggunakan Davies-Bouldin Index (DBI). Kami menggunakan 3, 4, 5, 6, 10, 15. Hasil nya seperti disajikan pada Tabel 2. Nilai DBI yang baik adalah yang sekecil mungkin mendekati nol.

Tabel 2. Pemilihan jumlah cluster

K	Davies-Bouldin Index
3	-1.215
4	-2.192
5	-5.245
6	-6.192
10	-3.015
15	-4.205

Hasil yang dicapai pada tabel tersebut menunjukkan bahwa dari semua pilihan K (jumlah cluster) yang kami eksperimenkan, K = 3 memberikan hasil evaluasi yang mendekati nol dengan nilai DBI = -1.215. Untuk selanjutnya, dalam eksperimen clustering kami menggunakan K=3, sehingga didapatkan 3 kelompok data tren penjualan.

3.2. Pengelompokan data kategori alat tulis

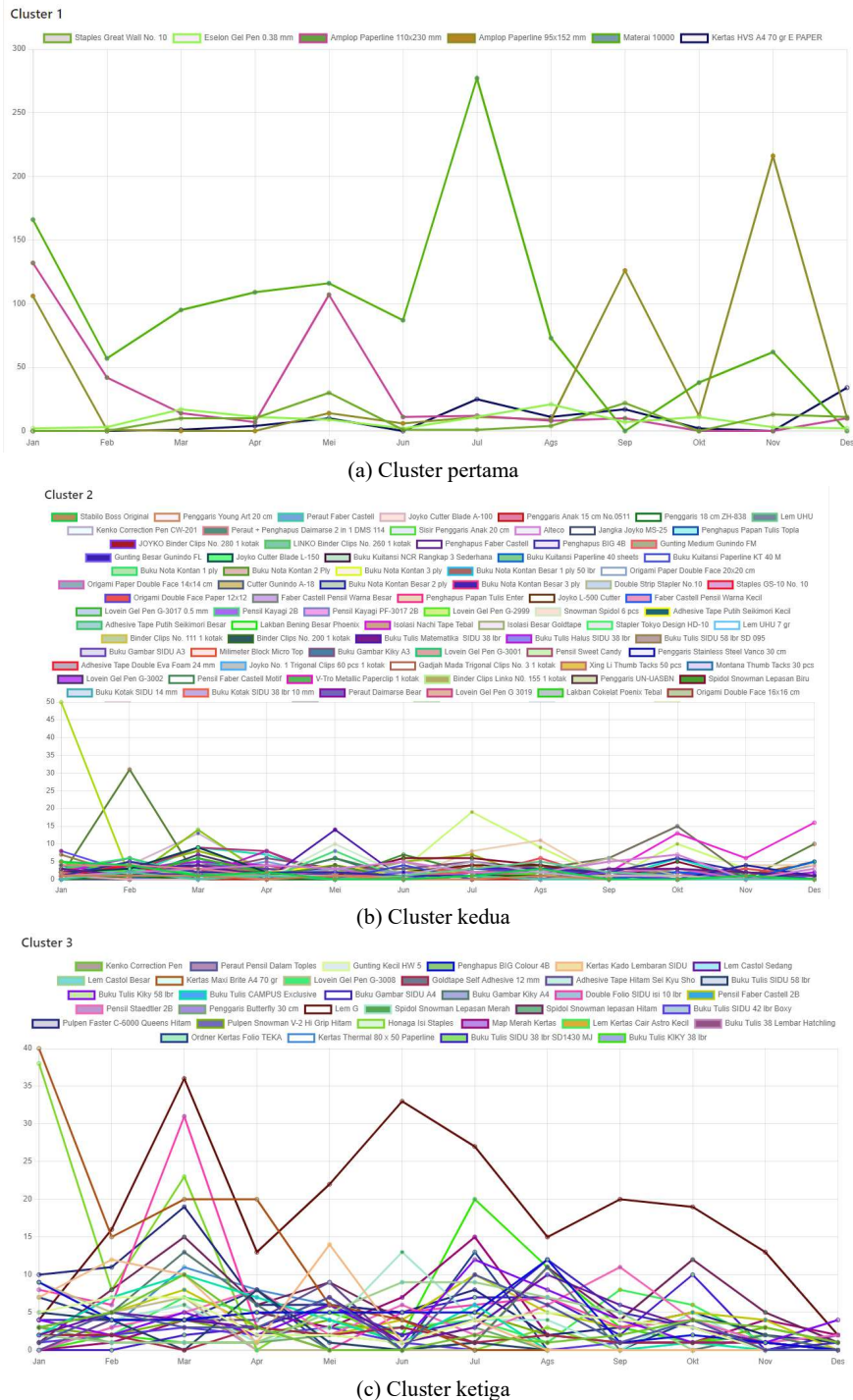
Data dengan kategori alat tulis tahun 2020 dan 2021 masing-masing berjumlah 164 dan 123 data. Kami melakukan pengelompokan dengan membaginya menjadi 3 cluster.

Pengelompokan dilakukan sendiri-sendiri pada data masing-masing tahun.

Hasil pengelompokan yang disajikan pada Gambar 2(a) untuk cluster pertama menunjukkan ada 6 data, dimana 4 diantaranya mempunyai pola volume penjualan yang berbeda dimana pada waktu tertentu memiliki volume penjualan sangat besar. Cluster kedua pada Gambar 2(b) menunjukkan pola yang hampir

semuanya sama dimana naik turun tiap bulan namun mayoritas volumenya dibawah 10 penjualan, ada 124 data yang tergabung di cluster ini.

Cluster ketiga Gambar 2(c) menunjukkan pola penjualan yang naik turun juga dengan jumlah 34 data. Namun ada volume yang cukup tinggi diawal tahun. Diantara semuanya ada 1 barang yang mempunyai volume penjualan tinggi.

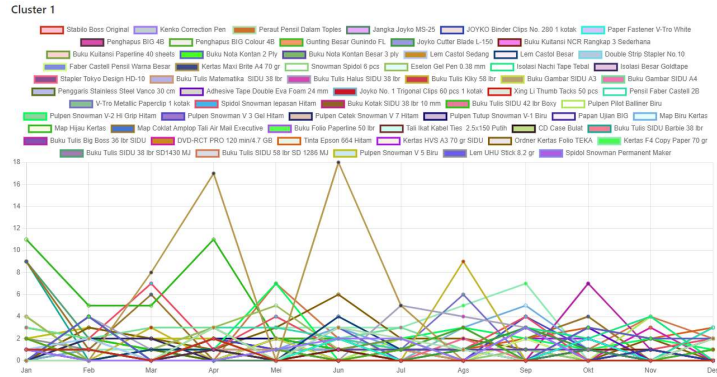


Gambar 2. Hasil pengelompokan data tahun 2020 kategori alat tulis

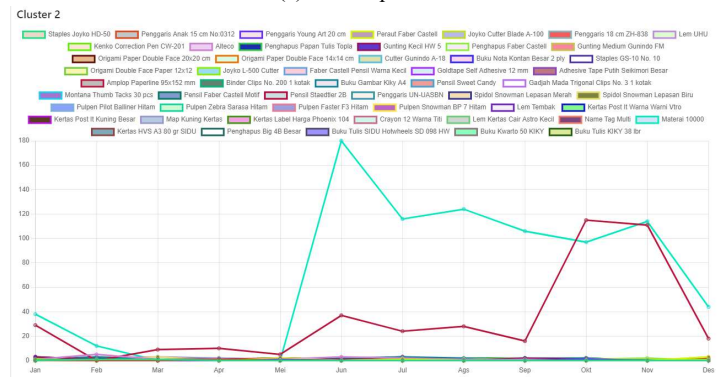
Hasil ditunjukkan pada data 2021 di cluster pertama berjumlah 62 data. Volume penjualan naik dan turun sepanjang tahun dalam

rentang hingga 12 penjualan. Pada cluster kedua terdapat 52 data dengan volume penjualan dibawah 10.

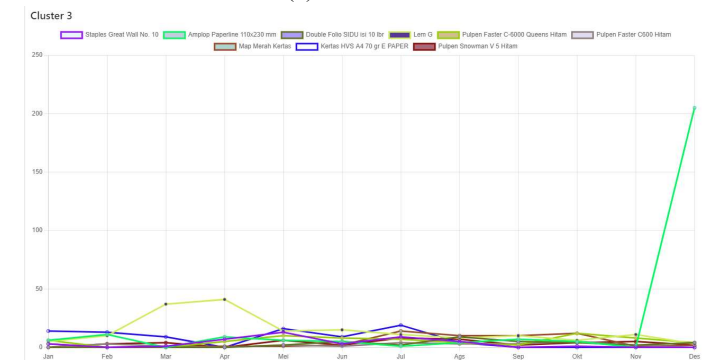
Sedangkan pada cluster ketiga terdapat 9 data dengan pola volume penjualan cenderung dibawah 20 penjualan.



(a) Cluster pertama



(b) Cluster kedua



(c) Cluster ketiga

Gambar 3 Hasil pengelompokan data tahun 2021 kategori alat tulis

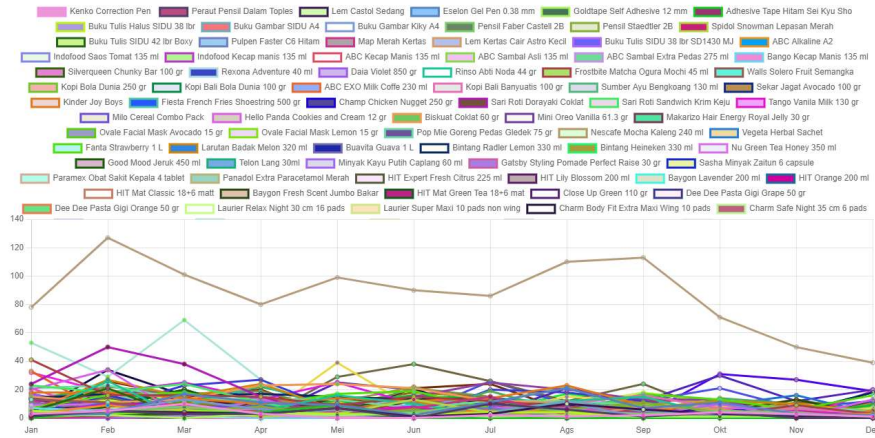
3.3. Pengelompokan data semua barang

Pengelompokan data tahun 2020 berjumlah 3668 dilakukan dengan mempartisipasinya menjadi 3 cluster. Hasil yang didapat menunjukkan bahwa cluster pertama berjumlah 244 data. Gambar 4(a) menunjukkan kenaikan penjualan pada awal tahun, kemudian menurun di pertengahan tahun, lalu naik dan kembali menurun saat akhir tahun. Hasil cluster kedua ini berjumlah 524 data. Gambar 4(b) menunjukkan jumlah penjualan yang terhitung stabil di awal hingga pertengahan tahun, namun beberapa item menurun saat menuju akhir tahun. Hasil cluster ketiga ini berjumlah 2900 data. Gambar 4(c) menunjukkan di awal tahun penjualan terlihat rendah, hanya ada beberapa item yang

mengalami kenaikan namun menurun lagi. Saat di pertengahan hingga akhir tahun, beberapa item menaik lalu kembali turun.

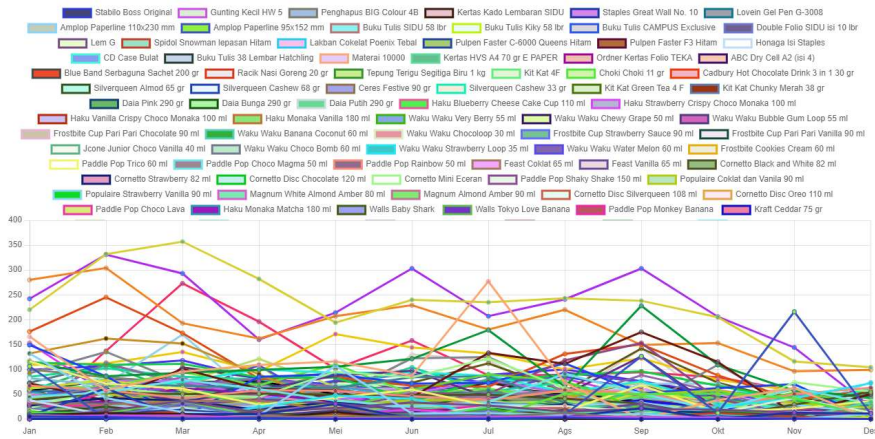
Pada data tahun 2021 berjumlah 3109 data menunjukkan pola 3 cluster berbeda. Pada cluster pertama berisi sejumlah 417 data. Gambar 5(a) menunjukkan penjualan yang cenderung menaik dari awal tahun yang sempat turun kemudian perlahan naik kembali saat berjalan dari pertengahan hingga akhir tahun. Hasil cluster kedua ini berjumlah 2403 data. Gambar 5(b) menunjukkan penjualan yang stabil di awal tahun. Menuju pertengahan tahun mulai terjadi kenaikan. Lalu di akhir tahun sempat turun namun kembali naik. Hasil cluster ketiga ini berjumlah 289 data. Gambar 5(c) menunjukkan penjualan yang menurun saat awal tahun, namun terlihat stabil hingga akhir tahun.

Cluster 1



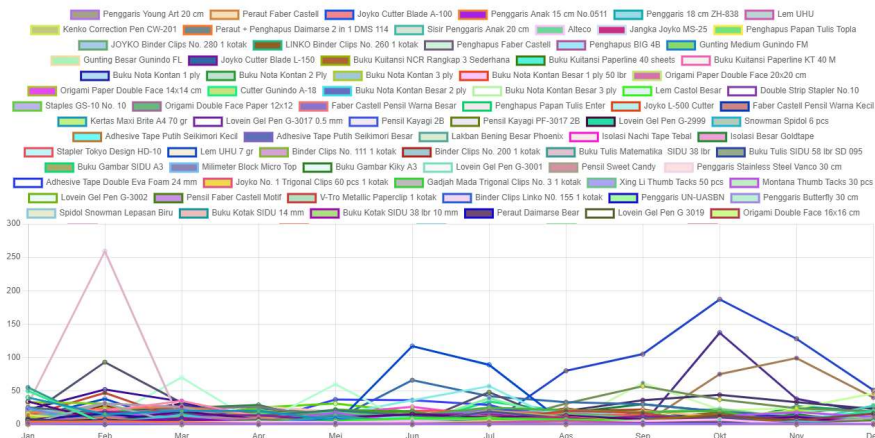
(a) Cluster pertama

Cluster 2



(b) Cluster kedua

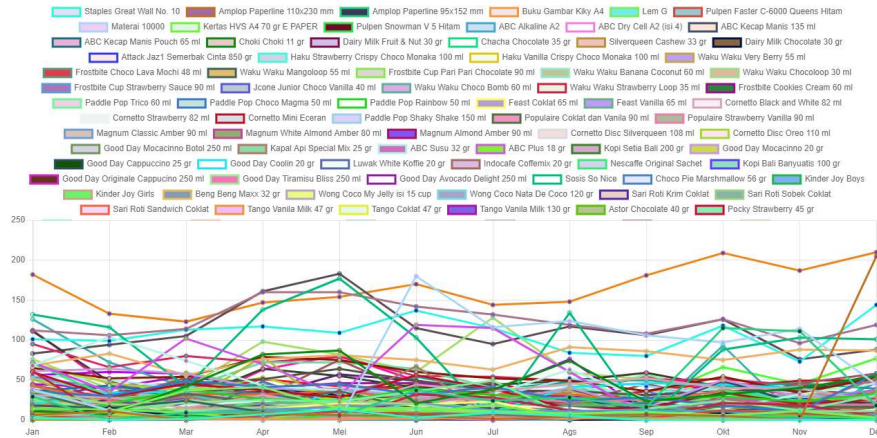
Cluster 3



(c) Cluster ketiga

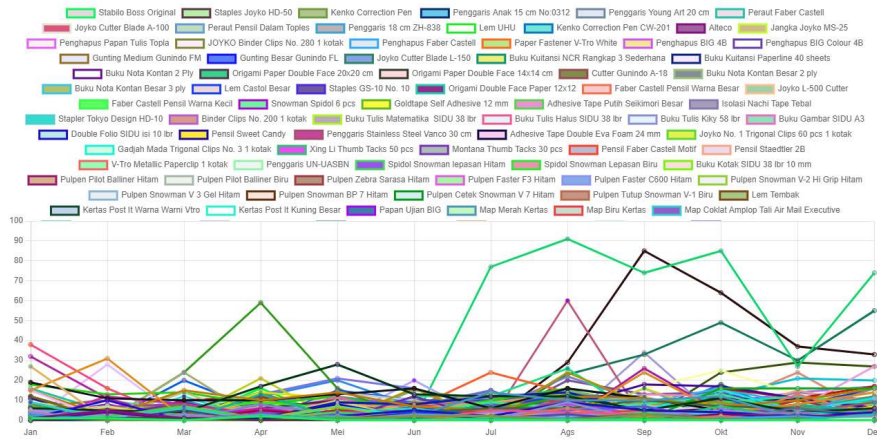
Gambar 4 Hasil pengelompokan data trend penjualan tahun 2020

Cluster 1



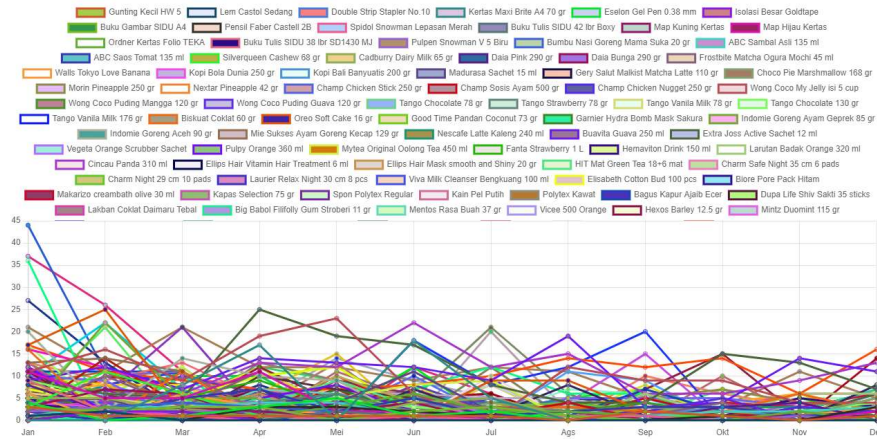
(a) Cluster pertama

Cluster 2



(b) Cluster kedua

Cluster 3



(c) Cluster ketiga

Gambar 5 Hasil pengelompokan data trend penjualan tahun 2021

4. PEMBAHASAN

Dari penelitian yang telah dilakukan sebelumnya dapat dilihat bahwa pengelompokan trend penjualan menggunakan metode *K-Means* dengan perhitungan jarak DTW menunjukkan hasil yang sudah baik dengan kelompok yang sesuai karakter alamnya. Perbedaan antara cluster satu dengan yang lain juga terlihat. Ada

yang di awal tahun memiliki penjualan yang tinggi kemudian menurun saat akhir tahun, ada yang stabil setiap bulannya, ada yang rendah di awal tahun tapi perlahan naik hingga akhir tahun, dan berbagai macam trend lainnya. Walaupun ada beberapa data penjualan yang sedikit berbeda pada satu cluster. Bisa terjadi karena batas jumlah cluster yang ditentukan.

4.1. Analisa hasil pengelompokan data kategori alat tulis

Eksperimen yang dilakukan pada data kategori alat tulis tahun 2020 berjumlah 164 data menunjukkan hasil yang beragam. Pada cluster pertama terdapat 6 data dengan pola yang peningkatan volume penjualan pada rentang waktu tertentu yaitu bulan Juli dan Nopember. Ada dua barang yang mempunyai pola berbeda yaitu Amplop Paperline 95x152 mm dan Materai 10000. Kedua barang ini memiliki volume penjualan yang sangat tinggi dibanding 4 barang lainnya. Volume penjualan Materai sangat tinggi di bulan Juli, sedangkan Amplop sangat tinggi dibulan Nopember. Empat data lainnya tidak memiliki volume yang menonjol dibanding 2 barang tersebut. Cluster kedua dengan jumlah 124 data menunjukkan pola dimana volume penjualan barang mempunyai pola yang mirip hampir di semua barang. Kecuali hanya satu barang yaitu DVD-R120 min/4.7 GB yang memiliki volume sangat tinggi diawal tahun yaitu 50 penjualan. Data penjualan pada barang lainnya juga naik turun namun semuanya memiliki pola yang sama. Cluster ketiga berisi 34 data juga memiliki pola naik turun, namun ada satu barang, yaitu Lem G, dengan volume penjualan yang sangat tinggi sepanjang tahun. Barang tersebut tergolong sangat laris dibandingkan dengan barang lainnya. Meskipun juga naik turun penjualannya namun volumenya paling tinggi. Ini menjadi perhatian yang harus dilakukan oleh manajemen agar stok penjualan juga disediakan sepanjang tahun.

Hasil pengelompokan data kategori alat tulis tahun 2021 menunjukkan pola cluster yang berbeda dibanding tahun sebelumnya. Pada cluster pertama, dari 62 barang yang bergabung, ada Kertas Maxi Brite A4 70 gr yang memiliki volume penjualan dibulan April dan Juni dengan volume hingga 18 penjualan, sedangkan barang lainnya mayoritas dibawah 10 penjualan. Cluster 2 dari 53 barang yang bergabung, hanya Mterai 10000 yang volumenya sangat tinggi hingga 180 penjualan mulai Juni hingga akhir tahun, sedangkan barang lainnya dibawah 10 penjualan. Cluster ketiga pola penjualan barang hampir sama semua pada 9 barang yang bergaung. Hanya Amplop Paperline 110x230 mm yang lebih tinggi hingga 200 penjualan, itupun hanya di akhir tahun saja.

4.2. Analisa hasil pengelompokan data semua kategori

Hasil pengelompokan data tahun 2020, sejumlah 3668 data, pada cluster pertama menunjukkan pola penjualan naik turun di volume sekitar 20 penjualan. Namun dari semua 244 barang di cluster ini, ada satu barang yang memiliki pola penjualan paling banyak dibanding barang lainnya yaitu Sampoerna Mild 16. Volume penjualannya 80 hingga 120 sepanjang Januari hingga Oktober, volume ini jauh dibanding data lainnya di cluster pertama. Lain hasil yang didapatkan di cluster kedua, dari 524 data yang tergabung, ada empat barang yang memiliki volume penjualan yang tinggi yaitu Teh Gelas Original 180 ml, Aqua Botol 1500 ml, Teh Pucuk Harum 350 ml, Aqua 600 ml. Barang lainnya memiliki volume dibawah 150 penjualan. Cluster ketiga memiliki pola yang tidak terlalu besar fluktuasi penjualannya. Semua barang mencapai volume penjualan dibawah 50 penjualan. Hanya ada satu barang yang memiliki penjualan tinggi mulai Agustus hingga akhir tahun yaitu Aqua Galon MAMART.

Hasil pengelompokan data tahun 2021 menunjukkan pola yang berbeda. Data berjumlah 3109 terbagi menjadi 3 cluster dimana cluster pertama menunjukkan volume penjualan mayoritas hingga 80 penjualan, sedangkan sebaian kecil lainnya hingga 200 penjualan. Cluster kedua volume penjualannya mayoritas 20 penjualan. Sedangkan cluster ketiga volume penjualan berfluktuasi hingga 25 penjualan sepanjang tahun.

Secara umum, hasil pengelompokan menunjukkan bahwa kelompok trend data penjualan terjadi cluster mayoritas dengan jumlah yang paling banyak dibanding cluster lainnya. Data tahun 2020 dan 2021 menunjukkan fluktuasi yang sering sepanjang tahun, sedangkan data tahun 2021 ada sedikit perbedaan dimana pada awal tahun memiliki volume penjualan relative tinggi kemudian berfluktuasi sepanjang tahun. Hasil-hasil tersebut menunjukkan bahwa penggunaan DTW dapat melakukan partisi sesuai dengan karakter periodisasi data. Ini sejalan dan didukung oleh penelitian sebelumnya [20], [21] yang juga membuktikan bahwa pengelompokan data periodik dengan jarak DTW memberikan hasil dan kualitas cluster yang baik pula. Semua kelompok yang dihasilkan dalam eksperimen penelitian ini menunjukkan hasil cluster dengan data-data yang serupa ditiap cluster dan berbeda di cluster yang lain. Hasil ini juga didukung oleh validitas cluster dengan DBI yang menunjukkan bahwa nilai indeksnya -1.215. Nilai indeks ini mendekati nol yang berarti bahwa kualitas cluster yang dicapai sudah baik.

5. KESIMPULAN

Berdasarkan dari penelitian yang telah dilakukan, maka dapat disimpulkan bahwa pengelompokan data periodik dapat diselesaikan menggunakan K-Means dengan metric jarak *Dynamic Time Warping* (DTW). Tahap-tahap dalam proses pengelompokan menggunakan metode *k-means*, dimulai dari pemilihan beberapa *centroid* secara acak sesuai jumlah *cluster* yang diinginkan, kemudian menghitung jarak antara data selain *centroid* dengan masing-masing *centroid* yang ada menggunakan *dynamic time warping*, lalu dikelompokkan sesuai jarak yang terdekat, kemudian menghitung nilai fungsi objektif, apabila nilai fungsi objektif. Ketika nilai fungsi objektif kurang dari sama dengan 0 maka proses pengelompokan selesai. Hasil eksperimen pengelompokan *trend* penjualan menggunakan metode *K-Means* dan jarak DTW menunjukkan hasil yang baik dan sesuai dengan karakter alami data dimana data terbagi menjadi 3 cluster dengan nilai validitas DBI -1.215. Walau ada beberapa item yang berbeda *trend* penjualannya dalam satu *cluster*.

Saran penelitian berikutnya yang dapat dilakukan adalah penyisihan data anomali trend penjualan yang masih ada, dimana data anomali dapat menyebabkan distorsi hasil pengelompokan. Masalah ini dapat diantisipasi nantinya dengan deteksi anomali sebelum penerapan metode pengelompokan.

REFERENSI

- [1] S. Hanifah, F. Akbar, and R. P. Santi, "Implementasi Business Intelligence dan Prediksi Menggunakan Regresi Linear pada Data Penjualan dan Breakage di PT

<https://doi.org/10.25077/TEKNOSI.v10i2.2024.100-109>

- XYZ,” *Jurnal Nasional Teknologi dan Sistem Informasi*, vol. 8, no. 3, pp. 144–152, Dec. 2022, doi: [10.25077/TEKNOSI.V8I3.2022.144-152](https://doi.org/10.25077/TEKNOSI.V8I3.2022.144-152).
- [2] S. Kasus, B. M. Al-Fahmi, E. Rahmawati, and T. Sagirani, “Penerapan K-Means Clustering Pada Pariwisata Kabupaten Bojonegoro Untuk Mendukung Keputusan Strategi Pemasaran,” *Jurnal Nasional Teknologi dan Sistem Informasi*, vol. 9, no. 2, pp. 141–149, Aug. 2023, doi: [10.25077/TEKNOSI.V9I2.2023.141-149](https://doi.org/10.25077/TEKNOSI.V9I2.2023.141-149).
- [3] R. E. van Ruitenbeek, G. M. Koole, and S. Bhulai, “A hierarchical agglomerative clustering for product sales forecasting,” *Decision Analytics Journal*, vol. 8, p. 100318, Sep. 2023, doi: [10.1016/J.DAJOUR.2023.100318](https://doi.org/10.1016/J.DAJOUR.2023.100318).
- [4] S. Lee, S. Ko, A. H. Roudsari, and W. Lee, “A deep learning model for predicting the number of stores and average sales in commercial district,” *Data & Knowledge Engineering*, vol. 150, p. 102277, Mar. 2024, doi: [10.1016/J.DATAK.2024.102277](https://doi.org/10.1016/J.DATAK.2024.102277).
- [5] G. Gustientiedina, M. H. Adiya, and Y. Desnelita, “Penerapan Algoritma K-Means Untuk Clustering Data Obat-Obatan,” *Jurnal Nasional Teknologi dan Sistem Informasi*, vol. 5, no. 1, pp. 17–24, 2019, doi: [10.25077/teknosi.v5i1.2019.17-24](https://doi.org/10.25077/teknosi.v5i1.2019.17-24).
- [6] W. Afifi, D. R. Nastiti, and Q. Aini, “Clustering K-Means Pada Data Ekspor (Studi Kasus: Pt. Gaikindo),” *Simetris: Jurnal Teknik Mesin, Elektro dan Ilmu Komputer*, vol. 11, no. 1, pp. 45–50, 2020, doi: [10.24176/simet.v11i1.3568](https://doi.org/10.24176/simet.v11i1.3568).
- [7] A. Supriyatna, I. Carolina, S. Janti, and ..., “Clustering Koridor Transjakarta Berdasarkan Jumlah Penumpang Dengan Algoritma K-Means,” *J-SAKTI (Jurnal Sains ...)*, vol. 4, no. September, pp. 682–693, 2020.
- [8] R. A. Indraputra and R. Fitriana, “K-Means Clustering Data COVID-19,” *Jurnal Teknik Industri*, vol. 10, no. 3, p. 3, 2020.
- [9] A. Bahauddin, A. Fatmawati, and F. Permata Sari, “Analisis Clustering Provinsi Di Indonesia Berdasarkan Tingkat Kemiskinan Menggunakan Algoritma K-Means,” *Jurnal Manajemen Informatika dan Sistem Informasi*, vol. 4, no. 1, p. 1, 2021, doi: [10.36595/misi.v4i1.216](https://doi.org/10.36595/misi.v4i1.216).
- [10] N. Purba, P. Poningsih, and H. S. Tambunan, “Penerapan Algoritma K-Means Clustering Pada Penyebaran Penyakit Infeksi Saluran Pernapasan Akut (ISPA) di Provinsi Riau,” *Journal of Information System Research (JOSH)*, vol. 2, no. 3, pp. 220–226, 2021.
- [11] D. G. Ramadhan, I. Prihatini, and F. Liantoni, “Analisis Clustering Pengelompokan Penjualan Paket Data Menggunakan Metode K-Means,” *Ultimatics: Jurnal Teknik Informatika*, vol. 13, no. 1, pp. 33–38, 2021, doi: [10.31937/ti.v13i1.1981](https://doi.org/10.31937/ti.v13i1.1981).
- [12] M. I. Rizki, T. A. Taqqiyuddin, and J. J. Cerelia, “K-Medoids Clustering dengan Jarak Dynamic Time Warping dalam Mengelompokkan Provinsi di Indonesia Berdasarkan Kasus Aktif Covid-19,” *Prosiding Seminar Nasional Matematika*, vol. 4, pp. 685–692, 2021.
- [13] S. Putriana, E. Erna, and D. Andreswari, “Clustering Data Titik Gempa Dengan,” vol. 9, no. 1, 2021.
- [14] M. B. Johra, “Soft Clustering Dengan Algoritma Fuzzy K-Means (Studi Kasus : Pengelompokan Desa Di Kota Tidore Kepulauan),” *BAREKENG: Jurnal Ilmu Matematika dan Terapan*, vol. 15, no. 2, pp. 385–392, 2021, doi: [10.30598/barekengvol15iss2pp385-392](https://doi.org/10.30598/barekengvol15iss2pp385-392).
- [15] P. S. Kumar, S. K. Laha, and L. A. Kumaraswamidhas, “Assessment of rolling element bearing degradation based on Dynamic Time Warping, kernel ridge regression and support vector regression,” *Applied Acoustics*, vol. 208, p. 109389, Jun. 2023, doi: [10.1016/J.APACOUST.2023.109389](https://doi.org/10.1016/J.APACOUST.2023.109389).
- [16] M. Kloska, G. Grmanova, and V. Rozinajova, “Expert enhanced dynamic time warping based anomaly detection,” *Expert Systems with Applications*, vol. 225, p. 120030, Sep. 2023, doi: [10.1016/J.ESWA.2023.120030](https://doi.org/10.1016/J.ESWA.2023.120030).
- [17] H. Ding, Z. Yuan, J. Yin, X. Shi, and M. Shi, “Evaluating ecosystem stability based on the dynamic time warping algorithm: A case study in the Minjiang river Basin, China,” *Ecological Indicators*, vol. 154, p. 110501, Oct. 2023, doi: [10.1016/J.ECOLIND.2023.110501](https://doi.org/10.1016/J.ECOLIND.2023.110501).
- [18] Z. Liu, T. Bai, C. Peng, H. Bao, and J. Lu, “A method of scaling distortion assessment based on dynamic time warping,” *Annals of Nuclear Energy*, vol. 177, p. 109304, Nov. 2022, doi: [10.1016/J.ANUCENE.2022.109304](https://doi.org/10.1016/J.ANUCENE.2022.109304).
- [19] Y. Ida, E. Fujita, and T. Hirose, “Classification of volcano-seismic events using waveforms in the method of k-means clustering and dynamic time warping,” *Journal of Volcanology and Geothermal Research*, vol. 429, p. 107616, Sep. 2022, doi: [10.1016/J.JVOLGEORES.2022.107616](https://doi.org/10.1016/J.JVOLGEORES.2022.107616).
- [20] M. A. Zen *et al.*, “Aplikasi Pengelompokan Data Runtun Waktu dengan Algoritma K-Medoids,” *Inferensi*, vol. 6, no. 2, pp. 117–123, Sep. 2023, doi: [10.12962/J27213862.V6I2.15864](https://doi.org/10.12962/J27213862.V6I2.15864).
- [21] H. Almiatus Soleha, W. Pura Nurmawanti, U. Hidayaturrohman, R. Haiban Hirzi, A. Septiani, and J. Statistika, “Penerapan Clustering Time Series pada Pengelompokan Provinsi di Indonesia (Studi Kasus : Nilai Ekspor Non Migas di Indonesia Tahun 2016-2020),” *J Statistika: Jurnal Ilmiah Teori dan Aplikasi Statistika*, vol. 15, no. 2, pp. 286–291, Dec. 2022, doi: [10.36456/JSTAT.VOL15.NO2.A5550](https://doi.org/10.36456/JSTAT.VOL15.NO2.A5550).