

Dynamic Segmentation Analysis for Expedition Services: Integrating K-Means dan Decision Tree

[Analisis Segmentasi Dinamis Untuk Layanan Ekspedisi: Mengintegrasikan K-Means dan Decision Tree]

Dwi Himatul Khoiriyah¹⁾, Rita Ambarwati Sukmono ^{*,2)}

¹⁾Program Studi Manajemen, Universitas Muhammadiyah Sidoarjo, Indonesia

²⁾ Program Studi Manajemen, Universitas Muhammadiyah Sidoarjo, Indonesia

*Email Penulis Korespondensi: ritaambarwati@umsida.ac.id

Abstract. *Technological developments have an impact on increasing the level of competition between companies in acquiring and retaining customers. With this competition, companies must maximise efforts to reach consumers and understand customer service needs so that the business can continue to survive and experience development. In this effort, The data processing was done using rapidminer with k-means clustering and decision tree methods. The research results show that k-means clustering achieved the lowest Davies Bouldin Index (DBI) accuracy results, namely -0,943 in cluster_8. In the research using the decision tree methods, accuracy results were obtained at 49.83%, with the good cluster being cluster_7. In this case, better accuracy values can be achieved by using the k-means clustering methods. This research can provide an illustration of the importance of utilizing the k-means and decision tree algorithm in classifying sales data as a tool for optimizing marketing and service efforts.*

Keywords – *K-Means Clustering; Decision Tree; Expedition Services; Marketplace*

Abstrak. *Perkembangan teknologi berdampak pada meningkatnya tingkat persaingan antar perusahaan dalam memperoleh dan mempertahankan pelanggan. Dengan adanya persaingan tersebut, perusahaan harus memaksimalkan upayanya untuk menjangkau konsumen dan memahami kebutuhan layanan pelanggan agar bisnis dapat terus bertahan dan mengalami perkembangan. Dalam upaya tersebut, pengolahan data dilakukan menggunakan rapidminer dengan metode k-means clustering dan decision tree. Hasil penelitian menunjukkan k-means clustering berhasil mendapatkan hasil akurasi Davies Bouldin Index (DBI) terendah yaitu -0,943 pada cluster_8. Pada hasil penelitian menggunakan metode decision tree diperoleh hasil akurasi sebesar 49,83% dengan cluster terbaik adalah cluster_7. Dalam hal ini, nilai akurasi yang lebih baik dicapai dengan menggunakan metode k-means clustering. Penelitian ini dapat memberikan gambaran tentang pentingnya pemanfaatan algoritma k-means dan decision tree dalam mengklasifikasikan data penjualan sebagai alat untuk mengoptimalkan upaya pemasaran dan layanan.*

Kata Kunci – *K-Means Clustering; Decision Tree; Layanan Ekspedisi; Marketplace*

I. PENDAHULUAN

Perkembangan teknologi mempengaruhi operasi ritel dan logistik, dimana konsumen kini dapat melakukan transaksi pembelian hanya melalui media online saja. Perusahaan yang dapat secara efektif memenuhi kebutuhan konsumen melalui media online akan membuat konsumen bersedia untuk bersikap loyal dan terus berinteraksi dengan perusahaan [1]. Perkembangan teknologi tersebut sejalan dengan meningkatnya persaingan di dunia bisnis. Persaingan membuat perusahaan harus bisa memaksimalkan kemampuan yang dimiliki dengan sebaik-baiknya agar dapat terus bertahan menghadapi persaingan [2]. Salah satu cara untuk memaksimalkan kemampuan bisnis adalah dengan memanfaatkan adanya media online seperti marketplace, kegiatan pemasaran ini memudahkan dalam upaya menjangkau konsumen, serta memberikan peluang untuk memaksimalkan keuntungan karena dapat mengurangi biaya pemasaran sehingga pelaku bisnis dapat memperluas pasar, memberikan kemudahan akses dan pelayanan pada konsumen [3][4]. Platform penjualan online dapat mengoptimalkan upaya penjualan dengan kemudahan penyebaran promosi melalui sistem elektronik atau serangkaian jaringan komputer, sehingga konsumen dapat melakukan transaksi hanya dengan melalui media online secara cepat [5].

Pelayanan konsumen yang optimal akan menghasilkan output perusahaan berupa keuntungan yang menjadi tujuan utama kegiatan bisnis. Dalam usaha untuk mempertahankan bisnis dan memperluasnya, perusahaan perlu melakukan serangkaian upaya yang berfokus pada pelanggan. Dalam hal ini perusahaan dapat memberikan pelayanan terbaik untuk memberikan pengalaman transaksi yang baik dengan perusahaan. Pengalaman transaksi yang buruk akan sangat berpengaruh pada loyalitas pelanggan, penyebab terjadinya pengalaman transaksi yang buruk yakni seperti ketidakmampuan perusahaan untuk memberikan solusi pada permasalahan yang dialami oleh pelanggan, ketidakmampuan perusahaan dalam mengidentifikasi permasalahan yang sering dialami oleh pelanggan yang dapat memberikan resiko hilangnya pelanggan yang dimiliki oleh perusahaan [2]

Perusahaan dapat memanfaatkan kemudahan dalam melakukan pemasaran saat ini untuk memperluas jangkauan konsumen dan terus berupaya mempertahankan pelanggan agar dapat terus bertahan menghadapi persaingan bisnis. Melalui marketplace, perusahaan dapat mendapatkan keuntungan berupa fleksibilitas dan efisiensi operasional, sehingga bisa memberikan pelayanan kemudahan interaksi berupa tanya jawab dengan konsumen dan mempermudah konsumen untuk mendapatkan akses informasi terkait produk dan pelayanan yang ada karena pemasaran melalui media online tidaklah rumit [6], [7], [8].

Perusahaan ritel pada umumnya akan melakukan dua kegiatan pemasaran, yakni secara offline dan online. Oleh karena itu, perlu dilakukan optimalisasi kegiatan pemasaran yang dimiliki perusahaan berupa pemetaan kelompok transaksi yang dilakukan oleh pelanggan berdasarkan frekuensi, rasio, dan keberagaman yang terdapat di dalamnya. Dalam hal ini, pengelola media penjualan dalam perusahaan akan sangat berdampak pada keberhasilan dan keberlangsungan bisnis. Untuk dapat mengoptimalkan hal tersebut, penelitian ini menggunakan metode data mining berupa pengelompokan akun marketplace dan jasa ekspedisi menggunakan metode k-means clustering, serta menggunakan metode decision tree guna menggambarkan secara sederhana hasil dari pengelompokan yang telah diperoleh.

Clustering mampu membagi data berdasarkan dengan karakteristik dari data itu sendiri, kemudian menggabungkannya menjadi cluster-cluster serupa dan memisahkannya dengan cluster yang berbeda [9]. Dengan kata lain, clustering digunakan untuk mengelompokkan data berdasarkan kemiripan antara satu sama lain dan yang tidak memiliki kemiripan dengan objek cluster lain, guna dapat membagi informasi dari dataset yang ada berdasarkan kemiripannya [10][11]. Tujuan utama metode k-means adalah untuk mendapatkan jumlah cluster atau kelompok yang paling stabil berdasarkan hasil nilai terendah Davies Bouldin Indeks (DBI) [11][12].

Decision tree atau pohon keputusan menjadi salah satu alat yang digunakan dalam penelitian ini. Dimana decision tree dapat menghasilkan keputusan yang dirasa relevan dengan tujuan penelitian. Kami menggunakan metode decision tree guna mendapatkan informasi akhir dari hasil cluster yang ada. Decision tree merupakan pemodelan dengan tampilan sederhana dan dapat memberikan hasil yang baik untuk pengelolaan big data atau data dalam jumlah besar. Pemodelan ini akan menghasilkan output berupa klasifikasi instance berdasarkan nilai probabilitas data [13].

Pada studi terdahulu yang dilakukan oleh Indivar, dkk [11] mengenai pengelompokan objek serupa ke dalam cluster menghasilkan bahwa k-means berhasil membentuk cluster dari bid data e commerce. Menurut Yi Lei dan Xiandong [13], metode decision tree digunakan untuk mengevaluasi e commerce lintas batas, tingkat akurasi prediksi berada di atas 95%. Metode k-means dan decision tree digunakan dalam penelitian ini untuk menghindari kesalahan acak dan mendapatkan hasil yang optimal sesuai dengan tujuan penelitian yakni segmentasi guna optimalisasi akun marketplace serta dapat berdampak pada efisiensi dan efektifitas kinerja pegawai bidang pemasaran pada perusahaan ritel.

Rumusan masalah : Kolaborasi metode k-means dan decision tree yang diaplikasikan secara efektif untuk efisiensi layanan dan pemasaran.

Pertanyaan Penelitian : Bagaimana metode algoritma k-means mampu mengelompokkan data marketplace dan layanan ekspedisi. Bagaimana metode algoritma decision tree mampu mengelompokkan fokus marketplace dan layanan ekspedisi.

Tujuan Penelitian : Segmentasi transaksi guna optimalisasi akun amrketplace serta dapat berdampak pada efisiensi dan efektifitas kinerja pegawai bidang pemasaran dalam melakukan pelayanan pada perusahaan ritel.

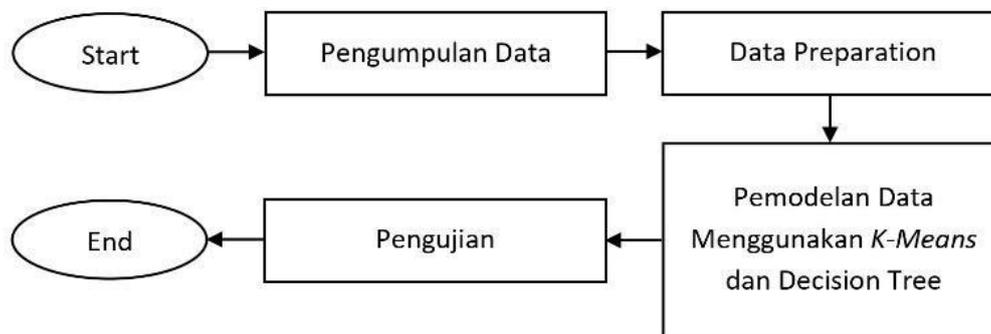
Kategori SDGs : Penelitian ini masuk dalam kategori delapan (8) dari tujuh belas (17) kategori *Sustainable Development Goals* (SDGs) <https://sdgs.un.org/goals>.

II. METODE

Penelitian ini menggunakan data sekunder rekam penjualan perusahaan ritel alat kesehatan yang kemudian data tersebut akan diolah menggunakan pemodelan yang terdapat dalam data mining, yakni k-means dan decision tree. Data mining merupakan suatu proses yang dilakukan untuk mendapatkan informasi yang berguna dari suatu data yang besar dan kompleks mencakup teknik algoritma, statistik, serta berfungsi untuk menemukan pola tersembunyi dalam suatu data dengan cara menganalisa data lalu disimpulkan menjadi informasi yang berguna [14][15]. Menurut Han dan Kamber, data mining terbagi menjadi dua kategori utama yakni [15]:

- Kategori prediktif, yakni memiliki guna sebagai alat prediksi nilai dari suatu atribut tertentu berdasarkan nilai atribut-atribut lain. Atribut yang diprediksi adalah target atau variabel tak bebas, dan atribut lain yang digunakan untuk membuat prediksi adalah berupa variabel bebas atau explanatory.
- Kategori deskriptif, bertujuan untuk menentukan pola-pola yang biasanya terdiri atas korelasi, trend, teritori, anomali, dan cluster. Pada kategori ini, data mining deskriptif seringkali harus memerlukan teknik post - preprocessing untuk validasi atau pengujian akurasi dan penjelasan hasil.

Dalam penelitian ini, data mining yang digunakan adalah kategori deskriptif. Dimana pada kategori tersebut, penelitian ini termasuk ke dalam pola pembentukan cluster. Adapun tahapan penelitian yang dilakukan tersaji dalam alur tahapan penelitian seperti Gambar 1. di bawah ini.



Gambar 1. Tahapan Alur Penelitian

a. Pengumpulan Data

Dataset yang diolah adalah dataset yang tersaji dalam bentuk laporan excel. Peneliti menggunakan data yang diperoleh dari rekam penjualan perusahaan ritel alat kesehatan selama periode awal tahun 2020 hingga periode awal tahun 2023 yakni dari bulan Januari sampai dengan tanggal 9 Juni. Dataset yang diolah berjumlah total sebanyak 15.274 data yang terdiri atas 15 atribut meliputi tanggal/hari kerja, no, nama customer, nomor telepon, asal kota, asal transaksi, nama barang, kuantitas, dropship, layanan ekspedisi, harga barang, ongkos kirim, asuransi, biaya tambahan, dan total harga.

b. Persiapan Data

Dataset yang ada dipersiapkan untuk kemudian dapat dimasukkan ke dalam mesin learning. Persiapan data yang dimaksud adalah dengan melakukan pemeriksaan terkait adanya data yang tidak sesuai secara penulisan, ejaan, serta adanya missing value yang akan membuat mesin learning tidak optimal dalam menghasilkan sebuah output. Hal pertama yang dilakukan adalah data cleaning untuk memastikan tidak ada data yang tidak konsisten secara penulisan dan ejaan, data transformation untuk mengubah bentuk data sesuai dengan input yang diinginkan oleh mesin learning namun tetap tidak merubah value data yang ada, serta melakukan data selection untuk memilih atribut yang akan diteliti pada mesin learning.

Pada tahapan data transformation, atribut data awal diubah menjadi bentuk yang lebih seragam dan sesuai dengan syarat masukan mesin learning rapidminer. Pada tahapan ini, atribut asal kota diubah menjadi provinsi 4 5 secara manual menggunakan microsoft excel. Setelah data dimasukkan ke dalam mesin learning, data kemudian ditransformasi ke bentuk numerical. Atribut yang dimasukkan ke dalam mesin learning adalah atribut provinsi, asal transaksi, dan layanan ekspedisi seperti terdapat dalam Tabel 1 dan Tabel 2.

Tabel 1. Dataset Atribut

Provinsi	Asal Transaksi	Ekspedisi
Jawa Timur	Shopee nusantarapersadasurabaya	Shopee Express
Jawa Barat	Tokopedia Deda Medica	SiCepat Reguler
Kalimantan Barat	Shopee Mastha Medika	J&T Economy
Jawa Barat	Tokopedia Mastha Medica Jakarta	SiCepat Reguler
Jawa Timur	Tokopedia Mastha Medica (Surabaya)	JNE Oke
Jawa Barat	Shopee Mastha Medika	SiCepat Reguler
Kalimantan Tengah	Tokopedia Mastha Medica Jakarta	JNE Reguler
Jawa Barat	Shopee Mastha	Shopee Express
Bali	Shopee Mastha Medika	J&T Economy
Jakarta	Tokopedia Mastha Medica Jakarta	SiCepat Reguler
Surabaya	Tokopedia Fanmed Sby	GoSend

Tabel 2. Keterangan Atribut

Atribut	Keterangan
<i>Provinsi</i>	Informasi tentang wilayah asal konsumen yang membeli produk
<i>Asal Transaksi</i>	Informasi asal akun transaksi yang dipilih konsumen untuk membeli produk
<i>Layanan Ekspedisi</i>	Informasi tentang layanan pengiriman atau distributor yang dipilih oleh konsumen untuk mengirimkan pembelian produk

c. Pemodelan Data

Pemodelan data yang dilakukan menggunakan dua metode pemodelan data mining yakni k-means clustering dan decision tree. Pada pemodelan ini peneliti menggunakan aplikasi rapidminer karena hasil output yang mudah dapat secara sederhana ditampilkan dan mudah untuk dipahami.

- K-Means

K-means merupakan teknik pembelajaran yang digunakan untuk mengelompokkan kumpulan data ke dalam kelompok-kelompok terpisah sehingga tiap kelompok dapat dibagi sesuai dengan karakteristik yang serupa [16]. Untuk bisa dilakukan pengelompokkan dengan hasil akurat, percobaan memvariasikan jumlah cluster perlu dilakukan. Dalam hal ini dilakukan percobaan sebanyak $k=10$ untuk melihat cluster mana yang memiliki nilai Davies Bouldin Index (DBI) terbaik, yakni yang memiliki nilai paling rendah dibandingkan dengan nilai cluster lainnya.

Hasil cluster menggunakan metode K-Means sangat dipengaruhi oleh nilai k atau jumlah k yang ditentukan [17]. Percobaan sebanyak $k=10$ yakni dengan mengambil sampel nilai $k=2$ sampai dengan nilai $k=10$ untuk melihat cluster yang memiliki nilai paling rendah dibandingkan dengan nilai cluster lainnya.

Algoritma ini dinilai sebagai salah satu pemodelan cluster paling kuat dan populer [16]. Kekurangan k-means yakni hanya bisa menerima input berbentuk numeric data, dan sensitif terhadap pemilihan titik awal centroid [11][14]. Untuk itu perlu dilakukan transformasi data dan percobaan dengan memvariasikan jumlah k untuk mendapatkan nilai terbaik. Untuk validasi metode dapat dilakukan dengan melakukan pengukuran terhadap klasterisasi dan iterasi dengan cara sebagai berikut:

$$V_{ij} = \frac{1}{N_i} \sum_{k=0}^{N_i} X_{kj}$$

$$D = \sqrt{(X_i - S_i)^2 + (y_i - t_i)^2} \quad [18]$$

- Decision Tree

Decision tree adalah salah satu algoritma yang digunakan untuk pengambilan keputusan. Decision tree mengubah kriteria menjadi bentuk node yang saling terhubung hingga membentuk struktur seperti pohon. Setiap pohon memiliki cabang yang mewakili atribut yang harus dipenuhi untuk mencapai cabang berikutnya hingga tidak ada cabang lagi pada struktur decision tree yang terbentuk. Algoritma decision tree memproses keputusan dengan memformat ulang data tabular menjadi model pohon yang menghasilkan aturan dan disederhanakan. Algoritma ini merepresentasikan secara sederhana metode klasifikasi pada banyak kelas dengan cara memberi tanda dengan nama atribut pada simpul internal dan simpul akar yang terbentuk, serta memberi label nilai atribut pada rusuk-rusuknya dan menandai dengan kelas-kelas yang berbeda pada simpul daunnya [19].

Decision tree atau pohon keputusan merupakan pemodelan yang digunakan untuk mengklasifikasikan data berdasarkan input data berupa distribusi probabilitas dengan tampilan sederhana yang dapat memberikan hasil yang baik untuk data dalam jumlah besar. Pemodelan ini menghasilkan output berupa klasifikasi instance berdasarkan nilai probabilitas dari data [13]. Decision tree dipilih sebagai pemodelan kedua karena klasifikasi menggunakan decision tree mampu menangani data dalam jumlah besar berdasarkan set pelatihan dan label dari data [20]. Pemodelan dengan metode ini dapat menyederhanakan gambaran dari hasil perhitungan jarak dan penentuan cluster yang telah dilakukan dengan metode k-means.

d. Pengujian

Pengujian yang dilakukan adalah untuk melihat nilai akurasi dari hasil pengolahan data dengan menggunakan metode k-means dan decision tree. Dalam proses ini, pengukuran kualitas hasil cluster dilakukan dengan menggunakan fitur cluster distance performance untuk mendapatkan cluster yang memiliki nilai Davies Bouldin Index (DBI) terendah/terbaik.

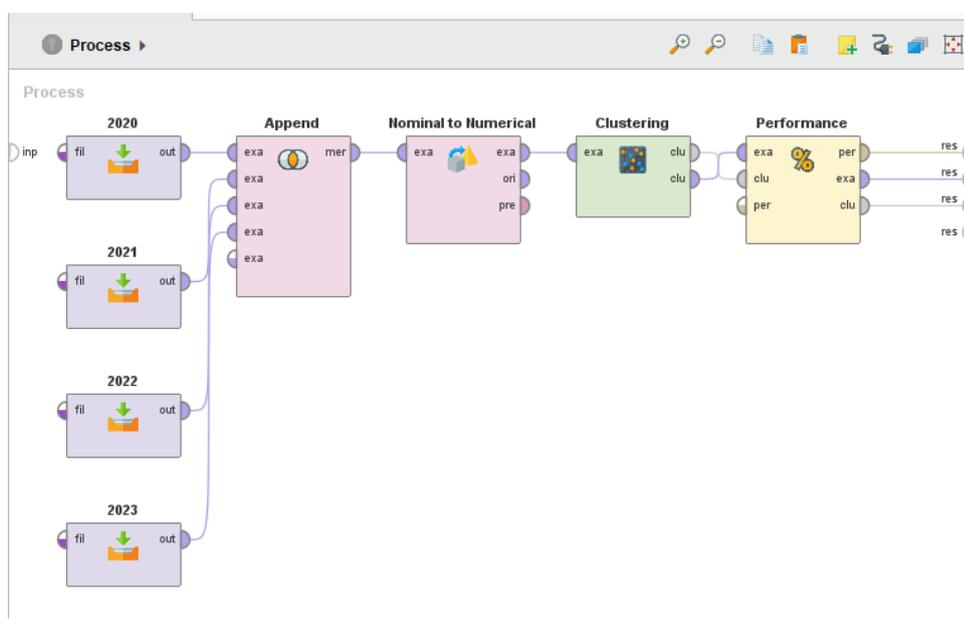
Setelah mendapatkan cluster terbaik, dilakukan pemodelan kembali menggunakan metode decision tree dengan melakukan testing pada data hasil cluster. Peneliti menggunakan pengukuran performance classification untuk mendapatkan nilai akurasi tingkat keberhasilan sistem dalam membentuk klasifikasi terhadap input data yang telah diperoleh sebelumnya.

III. HASIL DAN PEMBAHASAN

Dalam penelitian ini, dataset yang diolah adalah sebanyak 15.274 total data. Aplikasi rapidminer digunakan untuk melakukan pemodelan data dengan metode k-means dan decision tree. Untuk melakukan pemodelan, harus dilakukan data preparation pada dataset yang akan diolah melalui serangkaian proses preprocessing dengan memperbaiki ejaan, penulisan, menghilangkan missing value, dan merubah bentuk data ke dalam bentuk yang diterima dan sesuai dengan pemodelan metode dan tujuan penelitian.

A. K-Means

Pemodelan pertama yang dilakukan adalah dengan menggunakan metode k-means yang akan divariasikan sebanyak jumlah clusternya untuk mendapatkan nilai Davies Bouldin Index (DBI) terendah/terbaik. Pada tahap ini, dilakukan data transformation pada atribut asal kota menjadi provinsi, dan keseluruhan data telah ditransformasi menjadi bentuk numeric. Atribut yang dipilih sebagai label adalah ekspedisi dengan id asal transaksi.



Gambar 2. Pemodelan Metode K-Means

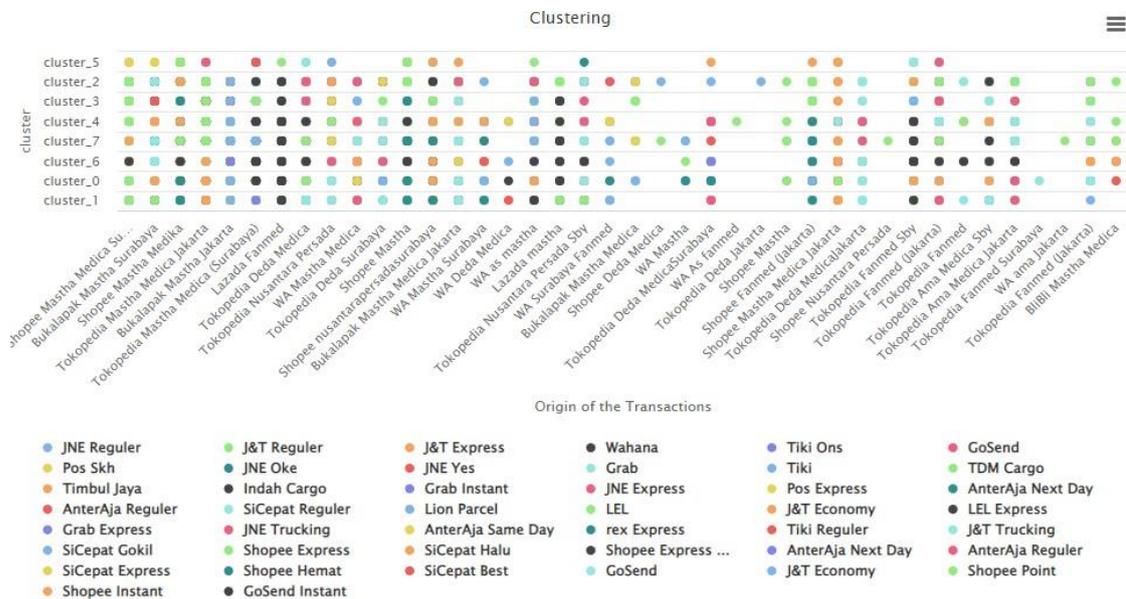
Pada gambar 2. tersaji alur pemrosesan data pada rapidminer untuk dapat dilakukan pengolahan menggunakan metode k-means. Dataset secara terpisah ditempatkan sesuai urutan tahun untuk memudahkan dalam mengidentifikasi adanya data yang masih mengalami kecacatan. Data kemudian disatukan dan ditransformasi ke dalam bentuk numerical agar memenuhi syarat input data metode k-means. Performance digunakan untuk

mengetahui nilai Davies Bouldin Index (DBI), dimana peneliti menggunakan penilaian cluster distance performance sebagai tolak ukur nilai cluster yang diuji.

Tabel 3. Hasil DBI Metode K-Means

Cluster	Davies Bouldin Index
cluster_2	-0,856
cluster_3	-0,842
cluster_4	-0,806
cluster_5	-0,778
cluster_6	-0,718
cluster_7	-0,888
cluster_8	-0,943
cluster_9	-0,939
cluster_10	-0,938

Peneliti melakukan uji coba sebanyak $k=10$ untuk menemukan nilai Davies Bouldin Index (DBI) terendah sebagai cluster terbaik yang dibentuk oleh metode k-means. Pengujian dilakukan secara random dari $k=2$ sampai dengan $k=10$ seperti yang telah tersaji pada Tabel 3. Pada tabel tersebut diantara 9 cluster yang diujikan, cluster_8 menunjukkan nilai terendah dibandingkan dengan cluster lainnya, yakni sebesar -0,943, dimana angka terendah tersebut yang menunjukkan bahwa cluster tersebut adalah cluster terbaik diantara cluster lain yang juga diujikan sebelumnya.



Gambar 3. Visualisasi Hasil Cluster

Setelah menemukan nilai Davies Bouldin Index terbaik, peneliti melakukan visualisasi atas hasil cluster_8 seperti tersaji pada Gambar 3. dan didapatkan bahwa terdapat cluster yang tidak terdapat beberapa akun dan layanan ekspedisi yang biasa dipilih oleh konsumen, serta terdapat cluster yang memiliki hampir semua akun dan layanan ekspedisi. Cluster_8 memiliki 8 kelompok dengan total items, jumlah akun, dan jumlah layanan ekspedisi yang berbeda-beda seperti pada Gambar 4 dan Tabel 4.

Cluster Model

```

Cluster 0: 2544 items
Cluster 1: 890 items
Cluster 2: 2631 items
Cluster 3: 994 items
Cluster 4: 2894 items
Cluster 5: 92 items
Cluster 6: 2857 items
Cluster 7: 2368 items
Total number of items: 15270

```

Gambar 4. Total Items K-Means

Tabel 4. Hasil Temuan Cluster Pada K-Means

Variasi Cluster	Jumlah Akun	Jumlah Ekspedisi
cluster_0	34 akun	30 ekspedisi
cluster_1	30 akun	23 ekspedisi
cluster_2	34 akun	26 ekspedisi
cluster_3	26 akun	24 ekspedisi
cluster_4	33 akun	30 ekspedisi
cluster_5	18 akun	16 ekspedisi
cluster_6	32 akun	31 ekspedisi
cluster_7	35 akun	30 ekspedisi

Hasil dari pengolahan data yang dilakukan menggunakan metode k-means ditemukan hasil analisis yakni sebagai berikut:

1. Cluster_0, cluster_2, cluster_4, cluster_6, dan cluster_7 adalah kelompok yang memiliki lebih dari 2000 total transaksi dengan kemiripan kebiasaan pelanggan dalam pemilihan akun asal transaksi beserta dengan layanan ekspedisi yang dipilih dalam melakukan pembelian, yakni terdapat setidaknya ≥ 30 jumlah akun asal transaksi dan ≥ 25 layanan ekspedisi yang dipilih. Kelompok ini merupakan kelompok berisikan hal umum atau mayoritas yang biasa dilakukan oleh konsumen. Semakin umum, semakin banyak pilihan yang dipilih, dan semakin banyak rasio yang terdapat di dalam kelompok ini menjadi dasar untuk adanya pengawasan dan pengendalian lebih karena banyaknya konsentrasi yang harus dibagi dan diatur.
2. Cluster_1, cluster_3, dan cluster_5 merupakan kelompok pembelian dengan jumlah transaksi kemiripan < 1000 . Kelompok ini merupakan kelompok konsumen yang memiliki kebiasaan pembelian yang lebih sempit jika dibandingkan dengan kelompok pertama yakni terdapat ≤ 30 akun asal transaksi dan memiliki tidak lebih dari 25 layanan pengiriman yang disediakan oleh perusahaan, dan terdapat lebih sedikit akun asal transaksi yang dipilih oleh konsumen untuk melakukan pembelian. Kelompok ini merupakan kelompok yang memiliki rasio lebih sedikit dalam segala aspek dibandingkan dengan kelompok pertama.

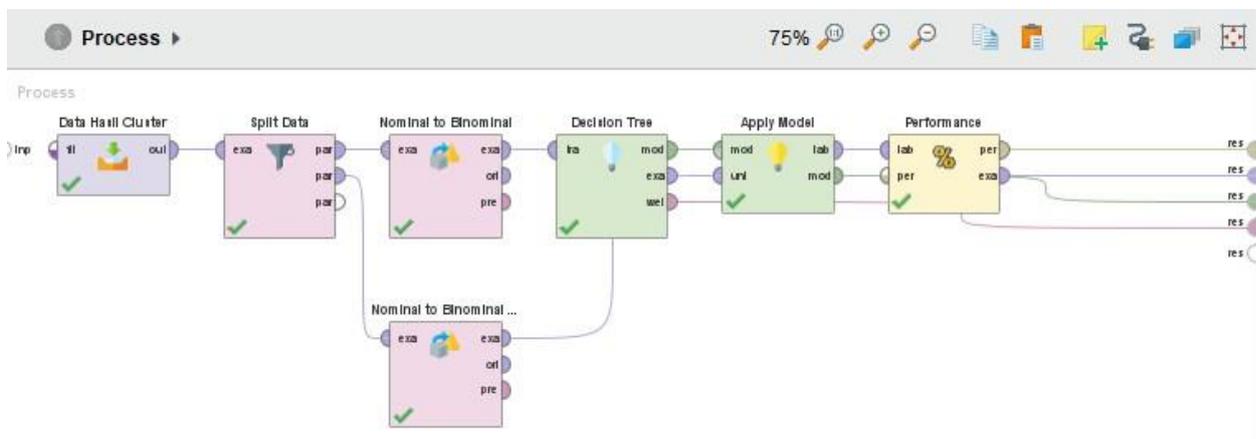
B. Decision Tree

Hasil pembentukan cluster oleh metode k-means kemudian diekspor ke dalam bentuk excel untuk kemudian dapat dimodelkan kembali ke dalam pemodelan metode decision tree. Pemodelan ini dilakukan dengan aplikasi yang sama yakni rapidminer dengan menambahkan satu atribut baru yakni answer, seperti pada Gambar 5. di bawah ini. Hal ini dilakukan guna menyederhanakan bentuk hasil cluster yang telah diperoleh.

Asal Transaksi	Ekspedisi	cluster	Answer
Shopee Mastha Medica Surabaya	JNE Reguler	cluster_1	No
Bukalapak Mastha Surabaya	J&T Reguler	cluster_0	Yes
Shopee Mastha Medica Surabaya	J&T Express	cluster_6	Yes
Shopee Mastha Medika	JNE Reguler	cluster_6	Yes
Tokopedia Mastha Medica Jakarta	Wahana	cluster_7	Yes
Tokopedia Mastha Medica Jakarta	Tiki Ons	cluster_7	Yes
Shopee Mastha Medika	Tiki Ons	cluster_4	Yes
Tokopedia Mastha Medica Jakarta	Tiki Ons	cluster_4	Yes
Shopee Mastha Medika	JNE Reguler	cluster_4	Yes
Bukalapak Mastha Jakarta	J&T Reguler	cluster_7	Yes
Bukalapak Mastha Jakarta	J&T Reguler	cluster_0	Yes
Tokopedia Mastha Medica (Surabaya)	J&T Reguler	cluster_6	Yes
Tokopedia Mastha Medica (Surabaya)	JNE Reguler	cluster_6	Yes
Tokopedia Mastha Medica (Surabaya)	GoSend	cluster_6	Yes
Shopee Mastha Medika	J&T Express	cluster_3	No
Shopee Mastha Medika	JNE Reguler	cluster_7	Yes
Shopee Mastha Medika	J&T Express	cluster_2	Yes
Tokopedia Mastha Medica Jakarta	JNE Reguler	cluster_7	Yes
Shopee Mastha Medika	JNE Reguler	cluster_1	No

Gambar 5. Jawaban Cluster

Atribut answer dibuat menggunakan aplikasi microsoft excel dengan rumus vlookup, yakni dengan ketentuan cluster yang membutuhkan pengawasan lebih karena memiliki total items >2000 dengan jumlah ≥ 30 memiliki jawaban Yes. Sedangkan jawaban No diperuntukan untuk cluster yang tidak membutuhkan pengawasan lebih rendah dari cluster yang memiliki jawaban Yes, dimana pada cluster tersebut terdapat lebih sedikit total items yakni <1000 dengan jumlah akun ≤ 30 . Setelah dilakukan penambahan atribut answer menggunakan rumus vlookup pada microsoft excel, dataset tersebut kemudian dimasukkan ke dalam mesin learning rapidminer dengan alur seperti pada Gambar 6.



Gambar 6. Pemodelan Metode Decision Tree

Peneliti menetapkan atribut cluster sebagai label dan menggunakan split data dengan rasio pembagian 7:3, dimana 70% adalah data training dan 30% adalah data testing. Data kemudian ditransformasi ke dalam bentuk binominal agar dapat diterima oleh fitur metode decision tree tanpa merubah value dalam data. Decision tree yang digunakan membandingkan antara criteria information_gain untuk mendapatkan hasil akurasi seleksi fitur tertinggi. Keseluruhan jenis akun yang terdapat dalam atribut asal transaksi dimasukkan ke dalam maximal depth yakni dengan total sebanyak 35. Dalam proses ini, peneliti tidak menggunakan apply pruning dan apply prepruning untuk mendapatkan hasil yang maksimal dan memiliki hasil analisa yang lebih jelas dan menyeluruh.

Hal yang selanjutnya dilakukan yakni melakukan pengujian dan pengukuran tingkat akurasi yang didapat dari menggunakan criteria information_gain pada pemodelan data hasil cluster seperti pada Gambar 7 dan Gambar 8.

accuracy: 49.83%

	true cluster_1	true cluster_0	true cluster_6	true cluster_7	true cluster_4	true cluster_3	true cluster_2	true cluster_5	class precision
pred. cluster_1	169	0	0	0	0	89	0	20	60.79%
pred. cluster_0	0	204	23	114	103	0	70	0	39.69%
pred. cluster_6	0	57	501	33	79	0	21	0	72.50%
pred. cluster_7	0	81	20	141	24	0	109	0	37.60%
pred. cluster_4	0	306	275	175	616	0	146	0	40.58%
pred. cluster_3	97	0	0	0	0	208	0	8	66.45%
pred. cluster_2	0	115	38	247	46	0	443	0	49.83%
pred. cluster_5	1	0	0	0	0	1	0	0	0.00%
class recall	63.30%	26.74%	58.46%	19.86%	70.97%	69.80%	56.15%	0.00%	

Gambar 7. Decision Tree Accuracy View Tabel

PerformanceVector

PerformanceVector:

accuracy: 49.83%

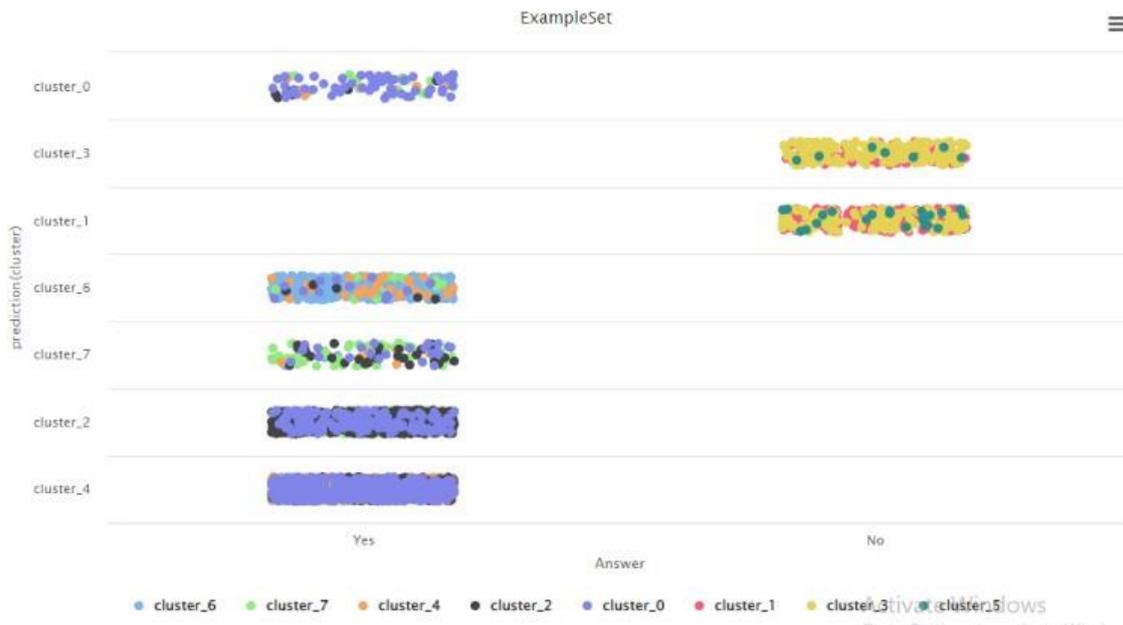
ConfusionMatrix:

True:	cluster_1	cluster_0	cluster_6	cluster_7	cluster_4	cluster_3	cluster_2	cluster_5
cluster_1:	169	0	0	0	89	0	20	0
cluster_0:	0	204	23	114	103	0	70	0
cluster_6:	0	57	501	33	79	0	21	0
cluster_7:	0	81	20	141	24	0	109	0
cluster_4:	0	306	275	175	616	0	146	0
cluster_3:	97	0	0	0	0	208	0	8
cluster_2:	0	115	38	247	46	0	443	0
cluster_5:	1	0	0	0	0	1	0	0

Gambar 8. Confusion Matrix

Information_Gain adalah criteria yang umum digunakan pada klasifikasi menggunakan metode decision tree [21]. Pada pemodelan ini menggunakan metode decision tree dengan kriteria information_gain dan maximal depth 35 didapatkan hasil akurasi sebesar 49.83%, confusion matrix perbandingan total True Positive (TP) sebanyak 2.282 dan False Positive (FP) sebanyak 2.296 seperti terdapat dalam Gambar 7 dan Gambar 8.

Setelah dilakukan pemodelan dengan menggunakan metode decision tree di atas, diketahui nilai akurasi menggunakan metode decision tree adalah sebesar 49.83%. Nilai tersebut belum bisa dikatakan baik karena nilai yang dihasilkan kurang dari 50% dan belum bisa dikatakan bahwa metode decision tree berhasil membentuk kelas jika data memiliki banyak atribut dan fitur di dalamnya. Hasil akurasi pada pemodelan decision tree sangat dipengaruhi oleh kedalaman tree, yakni jarak terpanjang yang diukur dari akar hingga daun tree [22]. Selain itu, banyaknya data, tingginya dimensi data, dan konsistensi data yang ada pada dataset juga sangat berpengaruh pada hasil akurasi dari sebuah data mining [23]. Berdasarkan hasil confusion matrix pada Gambar 9, satu-satunya cluster yang tidak memiliki True Positive (TP) value dan memiliki False Positive (FP) value adalah cluster_5.



Gambar 9. Visualisasi Scatter/Bubble

Pada Gambar 9. metode decision tree menunjukkan bahwa untuk answer No memiliki kesimpulan bahwa hanya terdapat 2 cluster, yakni cluster_3 dengan rasio terbanyak dan disusul oleh cluster_1, sedang kan True Positif cluster_5 terletak terbagi pada cluster_3 dan cluster_1. Pada cluster lain yang menunjukkan answer Yes di dalamnya, terdapat urutan cluster_4 dengan rasio terbanyak, disusul cluster_6, cluster_2, cluster_0, dan cluster_7. Berdasarkan hasil yang telah diperoleh melalui metode decision tree, hanya terdapat 7 cluster dari hasil prediksi metode decision tree dari dataset cluster oleh k-means. Total items yang memiliki nilai True Positive (TP) value yang dihasilkan oleh decision tree tersaji secara singkat pada tabel 5 di bawah ini.

Tabel 5. Total Items True Value (TP) Decision Tree

<i>Cluster</i>	<i>Total Items True Positive(TP)</i>
<i>cluster_0</i>	204
<i>cluster_1</i>	169
<i>cluster_2</i>	443
<i>cluster_3</i>	208
<i>cluster_4</i>	616
<i>cluster_5</i>	0
<i>cluster_6</i>	501
<i>cluster_7</i>	141

Hasil pengolahan data dengan menggunakan metode decision tree yang tersaji dalam Gambar 8. Dan Tabel 7. ditemukan hasil analisis sebagai berikut:

1. Cluster berlabelkan Yes yang mencakup cluster_0, cluster_2, cluster_4, cluster_7, dan cluster_6 memiliki sebaran yang merata pada label yang sama. Dimana cluster label yes memiliki jumlah true positive yang tinggi dan keberadaannya saling tersebar diantara cluster dengan label yang sama.
2. Cluster berlabelkan No yang mencakup cluster_1 dan cluster_3 memiliki jumlah true positive yang tinggi dan keberadaannya tersebar merata pada cluster yang sama-sama memiliki label No.
3. Cluster_5 yang merupakan hasil dari metode k-means dihilangkan atau dihapuskan karena tidak memiliki nilai true positive dan hanya terdapat dua sebaran false positive di dalam cluster. Sehingga cluster yang dianggap berhasil hanya ada 7 cluster

IV. SIMPULAN

Berdasarkan hasil pengolahan dan uji coba yang dilakukan, diperoleh kesimpulan bahwa metode k-means menghasilkan pola kelompok terbaik yakni 8 cluster dengan nilai Davies Bouldin Index (DBI) sebesar -0,943. K-means berhasil membentuk kelompok label Yes yakni, cluster_4, lalu cluster_6, cluster_2, cluster_0, cluster_7, serta kelompok dengan label No yang mencakup cluster_3, cluster_1, dan cluster_5. Sedangkan analisis menggunakan metode decision tree menunjukkan bahwa berdasarkan nilai akurasi, metode decision tree belum dapat menghasilkan output yang maksimal untuk pengelolaan dari data besar yang memiliki fitur serta atribut yang banyak dan beragam. Hal ini dibuktikan dengan nilai akurasi yang rendah yakni sebesar 49.83%. Decision tree menunjukkan hasil 7 cluster, yakni label Yes secara urut berdasarkan banyaknya items berada di cluster_4, cluster_2, cluster_6, cluster_7, dan cluster_0, serta label No berada di cluster_3 dan cluster_1. Label Yes dan No menggambarkan kebutuhan pengawasan agar kegiatan operasional pemasaran dapat berjalan lancar. Berdasarkan hasil akurasi kedua metode, dapat disimpulkan bahwa k-means memiliki tingkat akurasi yang lebih tinggi dibandingkan dengan metode decision tree meskipun hasil atau output keduanya hampir berada di jumlah pembagian pengawasan yang sama.

UCAPAN TERIMA KASIH

Peneliti mengucapkan terima kasih kepada Allah SWT. yang telah melimpahkan rahmat-Nya sehingga tugas akhir ini dapat selesai sebagaimana mestinya. Peneliti juga mengucapkan banyak terima kasih atas doa dan dukungan dari ayah, almarhumah ibu, kakak, serta teman-teman saya. Tak lupa pula ucapan terima kasih kepada dosen pembimbing saya, ibu Rita Ambarwati Sukmono dan dosen penguji sekalian yang telah banyak membantu serta membimbing saya dalam menyelesaikan artikel ini sehingga layak untuk dipublikasikan. Terima kasih pula kepada instansi universitas yang telah membantu melancarkan proses penyelesaian serangkaian penemuan tugas akhir ini.

REFERENSI

- [1] L. T. T. Tran, "Managing the effectiveness of e-commerce platforms in a pandemic," *J. Retail. Consum. Serv.*, vol. 58, no. September 2020, p. 102287, 2021, doi: 10.1016/j.jretconser.2020.102287.
- [2] J. Teknologi, E. Febrianty, L. Awalina, and W. I. Rahayu, "Optimalisasi Strategi Pemasaran dengan Segmentasi Pelanggan Menggunakan Penerapan K-Means Clustering pada Transaksi Online Retail Optimizing Marketing Strategies with Customer Segmentation Using K-Means Clustering on Online Retail Transactions," vol. 13, no. September, pp. 122–137, 2023, doi: 10.34010/jati.v13i2.
- [3] I. P. Artaya and T. Purworismiardi, "Efektifitas Marketplace Dalam Meningkatkan Konsentrasi," *Ekon. Dan Bisnis, Univ. Narotama Surabaya*, no. April, pp. 1–10, 2019, doi: 10.13140/RG.2.2.10157.95206.
- [4] W. Novita Sari., Achmad Hizazi., "Effect of Good Corporate Governance and Leverage on Profitability-Mediated Tax Avoidance (Study on Mining Companies listed on the Indonesia Stock Exchange 2016 – 2019)," *Int. J. Acad. Res. Account. Financ. Manag. Sci.*, vol. 11, no. 2, pp. 202–221, 2021, doi: 10.6007/IJARAFMS.
- [5] P. N. I. Sari, "Pengaruh Brand ambassador, kepercayaan dan resiko terhadap keputusan pembelian di e-commerce Shopee oleh mahasiswa di Pekanbaru," *J. Chem. Inf. Model.*, vol. 53, no. 9, pp. 1689–1699, 2020.
- [6] Ismai, "E-commerce dorong perekonomian Indonesia, selama pandemi covid 19 sebagai entrepreneur," *J. Manaj. dan Bisnis Prodi Kewirausahaan*, vol. 2, no. 2, pp. 111–124, 2020.
- [7] B. Algifari and A. Ariesta, "Penerapan E-Commerce Untuk Meningkatkan Penjualan Sepatu Pada Toko Garasi Spokat," pp. 99–105.
- [8] V. No and Z. Kedah, "Startupreneur Bisnis Digital (SABDA) Use of E-Commerce in The World of Business," vol. 2, no. 1, pp. 51–60, 2023.
- [9] B. Zhang, L. Wang, and Y. Li, "Precision Marketing Method of E-Commerce Platform Based on Clustering Algorithm," *Complexity*, vol. 2021, 2021, doi: 10.1155/2021/5538677.
- [10] F. A. Dewa and M. T. Jatipaningrum, "SEGMENTASI E-COMMERCE DENGAN CLUSTER K-MEANS DAN FUZZY C-MEANS (Studi Kasus : Media Sosial di Indonesia yang diunduh di Play Store)," vol. 4, no. 1, pp. 53–67, 2019.
- [11] I. Shaik, S. S. Nittela, T. Hiwarkar, and S. Nalla, "K-means Clustering Algorithm Based on E-Commerce Big Data," no. September, 2023, doi: 10.35940/ijitee.K2121.0981119.
- [12] E. Muningsih, I. Maryani, and V. R. Handayani, "Penerapan Metode K-Means dan Optimasi Jumlah Cluster dengan Index Davies Bouldin untuk Clustering Propinsi Berdasarkan Potensi Desa," vol. 9, no. 1, pp. 95–100, 2021.
- [13] Y. Lei and X. Qiu, "Research on the Evaluation of Overseas Strategic Climate Based on Decision Tree and Adaptive Boosting Classification Models," vol. 12, no. December, pp. 1–10, 2021, doi: 10.3389/fpsyg.2021.803989.
- [14] M. R. Nahjan, N. Heryana, A. Voutama, F. I. Komputer, U. S. Karawang, and R. Miner, "IMPLEMENTASI RAPIDMINER DENGAN METODE CLUSTERING K-MEANS UNTUK ANALISA PENJUALAN PADA TOKO OJ CELL," vol. 7, no. 1, pp. 101–104, 2023.
- [15] G. Indrawan, G. R. Dantes, P. Studi, I. Komputer, P. Pascasarjana, and U. P. Ganesha, "DATA MINING REKOMENDASI CALON MAHASISWA TECHNIQUE FOR OTHERS REFERENCE BY SIMILARITY TO IDEAL," no. 1, pp. 11–21, 2019.
- [16] M. Ahmed, R. Seraj, S. Mohammed, and S. Islam, "The k-means Algorithm : A Comprehensive Survey and Performance Evaluation," pp. 1–12, 2020, doi: 10.3390/electronics9081295.
- [17] B. H. Prakoso, E. Rachmawati, D. Rachmatta, and P. Mudiono, "Klasterisasi Puskesmas dengan K-Means Berdasarkan Data Kualitas Kesehatan Keluarga dan Gizi Masyarakat," vol. 14, no. April, pp. 60–68, 2023.
- [18] N. Suwaryo, A. Rahman, D. Marini, U. Atmaja, and A. Basri, "Klasterisasi Stok Produk Retail Untuk Menentukan Pergerakan Kebutuhan Konsumen Dengan Algoritma K-Means," vol. 4, no. 2, pp. 306–312, 2023.

- [19] M. Rizal *et al.*, “ALGORITMA DECISION TREE UNTUK ANALISIS SENTIMEN PUBLIC TERHADAP MARKETPLACE DI,” vol. 05, no. 01, pp. 18–25, 2023.
- [20] B. T. Jijo and A. M. Abdulazeez, “Classification Based on Decision Tree Algorithm for Machine Learning,” vol. 02, no. 01, pp. 20–28, 2021, doi: 10.38094/jastt20165.
- [21] A. Setyawan and D. N. Fauzi, “IMPLEMENTASI FUNGSI DISPERSION RATIO PADA PROSES SPLITING ATRIBUT ALGORITMA DECISION TREE,” vol. 2, no. 2, pp. 86–91, 2022, doi: 10.5281/zenodo.7782439.
- [22] V. M. Member, P. Casari, and S. Member, “A Novel Hyperparameter-free Approach to Decision Tree Construction that Avoids Overfitting by Design arXiv : 1906 . 01246v1 [cs . LG] 4 Jun 2019”.
- [23] K. Halim, D. E. Herwindiati, T. Sutrisno, T. Informatika, and U. Tarumanagara, “PENERAPAN METODE DECISION TREE UNTUK,” pp. 1–5.

Conflict of Interest Statement:

The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.