

KLASTERISASI PROVINSI DI INDONESIA BERDASARKAN PERSENTASE PENGELUARAN MAKANAN DAN BUKAN MAKANAN SEBAGAI DASAR SEGMENTASI PASAR MENGGUNAKAN ALGORITMA K-MEANS

Alia Marlina¹, Arief Wibowo²

^{1,2}Universitas Budi Luhur, Jakarta Selatan, Indonesia

Email: 2531600670@student.budiluhur.ac.id, arief.wibowo@budiluhur.ac.id

Abstrak

Penelitian ini dilakukan dengan tujuan agar provinsi di Indonesia bisa dikelompokkan berdasarkan persentase pengeluaran makanan dan bukan makanan menggunakan algoritma K-Means sebagai dasar segmentasi pasar. Publikasi resmi Badan Pusat Statistik (BPS) merupakan sumber didapatkannya data sekunder yang digunakan sebagai sumber data dalam penelitian ini. Data itu berisi informasi terkait persentase pengeluaran per kapita sebulan makanan dan bukan makanan di daerah perkotaan dan perdesaan menurut provinsi, 2025. Penelitian dilakukan beberapa tahap mulai dari pengumpulan data, persiapan data, klusterisasi menggunakan algoritma K-Means, di lain sisi Davies Bouldin Index (DBI) dimanfaatkan untuk mengevaluasi hasil clustering dan interpretasi hasil pengelompokan. Empat cluster dengan karakteristik yang berbeda dihasilkan oleh hasil clustering yang digunakan dalam penelitian. Cluster 0 memuat 16 provinsi dengan pengeluaran makanan lebih tinggi. 15 provinsi dengan pola pengeluaran yang relatif seimbang termuat di dalam cluster 1. 2 provinsi dengan persentase pengeluaran makanan tertinggi termuat di dalam cluster 2. 5 provinsi dengan persentase pengeluaran bukan makanan tertinggi termuat di dalam cluster 3. Temuan penelitian memperlihatkan bahwa provinsi bisa dikelompokkan oleh algoritma K-Means berdasarkan karakteristik konsumsi masyarakat sehingga bisa dimanfaatkan sebagai dasar segmentasi pasar dan membantu penyusunan strategi pemasaran yang lebih tepat sasaran.

Kata Kunci: K-Means Clustering, Data Mining, Segmentasi Pasar, Sistem Informasi Manajemen, Pengeluaran Makanan, Bukan Makanan.

Abstract

This study aimed to group provinces in Indonesia based on their food and non-food expenditure percentages using the K-Means algorithm as a basis for market segmentation. Official publications from the Central Statistics Agency (BPS) served as the source of secondary data used in this study. The data contained information on the percentage of monthly per capita food and non-food expenditures in urban and rural areas by province, in 2025. The study involved several stages, including data collection, data preparation, and clustering using the K-Means algorithm. The Davies-Bouldin Index (DBI) was used to evaluate and interpret the clustering results. Four clusters with distinct characteristics were generated by the clustering results used in the study. Cluster 0 contains 16 provinces with higher food expenditures. Fifteen provinces with relatively balanced expenditure patterns are included in cluster 1. The two provinces with the highest percentage of food expenditure are included in cluster 2. The five provinces with the highest percentage of non-food expenditure are included in cluster 3. The research findings show that provinces can be grouped using the K-Means algorithm based on community consumption characteristics, thus serving as a basis for market segmentation and assisting in the development of more targeted marketing strategies.

Keywords: *K-Means Clustering, Data Mining, Market Segmentation, Management Information Systems, Food and Non-Food Expenditure.*

A. PENDAHULUAN

Perkembangan teknologi informasi sudah mendorong perusahaan dan organisasi menggunakan data sebagai dasar dalam mengambil keputusan strategis. Sanjaya *et al.*, (2026) memaparkan bahwasanya efektivitas pengambilan keputusan organisasi bisa ditingkatkan oleh peran dari sistem informasi manajemen. Data terintegrasi bisa dikelola dan informasi yang akurat, relevan, dan tepat waktu bisa diberikan oleh sistem informasi manajemen, yang membantu manajemen menganalisis masalah dan membuat pilihan alternatif yang selaras dengan tujuan organisasi. Informasi yang dihasilkan sistem informasi manajemen bisa dimanfaatkan di berbagai fungsi organisasi, termasuk mendukung kegiatan pemasaran dan menyusun strategi yang tepat sasaran. Telaumbanua *et al.*, (2025) memaparkan bahwasanya hubungan yang menguntungkan dengan pasar sasaran bisa dibangun serta tujuan perusahaan, memuat identifikasi kebutuhan konsumen, penyusunan strategi pemasaran, dan pelaksanaan program pemasaran bisa dicapai secara efektif saat kegiatan pemasaran bisa dianalisis, direncanakan, dilaksanakan, dan dikendalikan oleh proses yang disebut dengan manajemen pemasaran. Strategi pemasaran yang efektif bisa dibuat oleh manajemen pemasaran dengan memahami karakteristik konsumen.

Segmentasi pasar merupakan salah satu cara yang bisa diimplementasikan. Faza *et al.*, (2024) memaparkan bahwasanya segmentasi pasar adalah cara strategis dalam pemasaran yang melibatkan membagi pasar ke sejumlah kelompok khusus berdasarkan kesamaan karakteristik atau cara berbelanja dari para pembeli. Wijaya & Daulay, (2025) memaparkan bahwasanya karakteristik dan kebutuhan unik konsumen bisa dipahami oleh perusahaan dengan melakukan proses segmentasi dengan cermat, yang memungkinkan produk atau layanan bisa disesuaikan oleh mereka dengan preferensi pasar yang lebih terarah. Berdasarkan definisi itu bisa diambil kesimpulan bahwa proses di mana pasar dibagi ke dalam sejumlah kelompok konsumen yang memiliki karakteristik, kebutuhan, atau pola perilaku yang sama sehingga strategi pemasaran yang lebih sesuai dan efektif untuk masing-masing kelompok itu bisa dibuat oleh perusahaan disebut dengan segmentasi pasar.

Perbedaan karakteristik konsumen antar wilayah menyebabkan kebutuhan dan pola konsumsi masyarakat tidak bisa disamaratakan. Salah satu cara untuk memahami karakteristik itu dengan melihat pola pengeluaran masyarakat. Pengeluaran untuk makanan dan pengeluaran bukan makanan merupakan dua pembagian dari pengeluaran rumah tangga. Besarnya pengeluaran itu bisa menggambarkan tingkat kesejahteraan, preferensi konsumsi, dan perilaku ekonomi masyarakat di suatu wilayah. Masyarakat dengan pengeluaran makanan yang tinggi berfokus pada pemenuhan kebutuhan dasar, di lain sisi kebutuhan yang lebih beragam dan kesejahteraan yang lebih baik lazimnya dimiliki oleh masyarakat dengan pengeluaran bukan makanan yang tinggi.

Indonesia memiliki berbagai kondisi sosial dan ekonomi yang berbeda di setiap provinsi. Perbedaan dalam penghasilan, jumlah penduduk, kondisi wilayah, dan pertumbuhan ekonomi membuat pola pengeluaran masyarakat berbeda di setiap provinsi. Variasi itu menjadi informasi yang penting bagi pemerintah dan pelaku usaha dalam membuat kebijakan serta strategi pemasaran yang disesuaikan dengan karakteristik setiap wilayah. Variasi itu perlu dipetakan agar bisa dimanfaatkan sebagai dasar segmentasi pasar. Namun jumlah data yang banyak sering kali membuat proses identifikasi pola dan karakteristik provinsi secara manual menjadi sulit, sehingga diperlukan metode yang bisa mengelompokkan data secara efektif.

Untuk mengatasi masalah itu, dibutuhkan metode untuk mengelompokkan provinsi berdasarkan pola pengeluaran masyarakat. Salah satu pendekatan yang bisa digunakan adalah teknik data mining. Khalish *et al.*, (2025) memaparkan bahwasanya informasi, pola, hubungan,

atau pengetahuan yang tersembunyi dari sekumpulan data yang sangat besar bisa ditemukan melalui proses yang disebut dengan data mining proses itu dilakukan dengan tujuan agar informasi yang membantu dalam pengambilan keputusan bisa dihasilkan. Ada metode clustering atau klasterisasi yang termuat dalam data mining, Khalish *et al.*, (2025) memaparkan bahwasanya data bisa dikelompokkan menjadi beberapa cluster berdasarkan tingkat kesamaan karakteristik yang dimiliki oleh data itu dengan mengimplementasikan metode penting dalam data mining yang disebut dengan teknik clustering. Karakteristik yang sama dimiliki oleh objek yang berada dalam satu kelompok, di lain sisi karakteristik yang berbeda dimiliki oleh objek pada kelompok yang berbeda.

Algoritma K-Means merupakan salah satu metode clustering yang banyak digunakan. Nazihah *et al.*, (2025) memaparkan bahwasanya data bisa dikelompokkan ke beberapa klaster berdasarkan tingkat kesamaan antar data itu dengan memanfaatkan metode klasterisasi non-hierarki yang disebut dengan K-Means clustering. Data dibagi oleh K-Means ke dalam beberapa klaster berdasarkan jarak terdekat ke pusat klaster (centroid). Sehubungan dengan hal itu, algoritma K-Means dinilai sesuai untuk mengelompokkan provinsi di Indonesia berdasarkan persentase pengeluaran makanan dan bukan makanan.

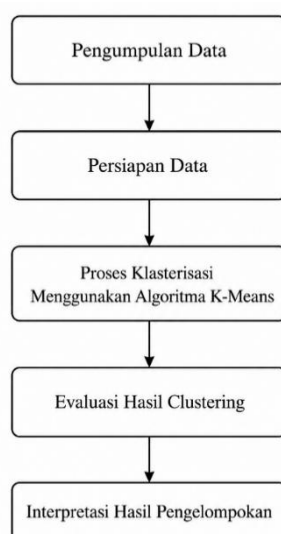
Penelitian terdahulu memperlihatkan bahwa berbagai pengelompokan data bisa dilakukan dengan memanfaatkan metode K-Means secara efektif. Penelitian yang dijalankan oleh Ardana *et al.*, (2024) yang berjudul Segmentasi Pelanggan Penjualan Online Menggunakan Metode K-Means Clustering, memperlihatkan bahwa cara berperilaku tiap-tiap pelanggan bisa diketahui oleh perusahaan dengan mengimplementasikan metode K-Means clustering dalam segmentasi pelanggan pada penjualan online, sehingga membantu dalam membuat strategi pemasaran yang lebih baik. Penelitian yang dijalankan oleh Purba & Handhayani, (2024) yang berjudul Perbandingan Algoritma K-Means, Affinity Clustering, Dan Minibatch K-Means Untuk Analisis Segmentasi Pasar, memperlihatkan bahwa K-Means clustering berjalan cukup baik sebab bisa mengelompokkan pelanggan dengan tingkat kesamaan yang cukup baik. Penelitian yang dijalankan oleh Rohman & Wibowo, (2024) yang berjudul Perbandingan Metode K-Medoids dan Metode K-Means Dalam Analisis Segmentasi Pelanggan Mall, memperlihatkan bahwa algoritma K-Means dipilih agar pelanggan bisa dikategorikan dengan lebih baik, sehingga strategi penjualan produk bisa dipahami secara lebih mudah oleh pengelola mall. Penelitian yang dijalankan oleh Dewi *et al.*, (2025) yang berjudul Komparasi Hasil Segmentasi Metode K-Means Dan Agglomerative Hierarchical Terhadap Provinsi di Indonesia Berdasarkan Profil Perjalanan Wisata Tahun 2024, memperlihatkan bahwa kelompok dengan anggota yang seimbang dan lebih bisa beradaptasi dengan cara distribusi data bisa dibentuk oleh metode K-Means, sehingga lebih efektif dalam mengenali pola aktivitas wisata secara jelas. Penelitian yang dijalankan oleh Iqbal *et al.*, (2025) yang berjudul Segmentasi Pelanggan Menggunakan K-Means Clustering Berdasarkan Data Kepribadian dan Pola Konsumsi, memperlihatkan bahwa penelitian menggunakan algoritma K-Means pada data kampanye pemasaran dan mengelompokkan pelanggan ke dalam empat segmen yang berbeda, sehingga strategi pemasaran yang lebih personal dan efektif bisa diimplementasikan.

Meskipun sudah banyak penelitian yang memanfaatkan algoritma K-Means di berbagai bidang, penelitian yang secara khusus mengelompokkan provinsi di Indonesia berdasarkan persentase pengeluaran makanan dan bukan makanan sebagai dasar segmentasi pasar masih terbatas. Penelitian ini dilakukan untuk melakukan klasterisasi provinsi di Indonesia berdasarkan persentase pengeluaran makanan dan bukan makanan menggunakan algoritma K-Means. Hasil klasterisasi diharapkan bisa menjadi dasar segmentasi pasar bagi pelaku usaha untuk menentukan strategi pemasaran yang tepat sasaran dan menjadi pertimbangan bagi pemerintah dalam menyusun kebijakan pembangunan ekonomi yang selaras dengan karakteristik masing-masing provinsi.

Tujuan penelitian ini adalah untuk menganalisis karakteristik provinsi-provinsi di Indonesia berdasarkan persentase pengeluaran untuk makanan dan bukan makanan sebagai gambaran pola konsumsi masyarakat. Selanjutnya, penelitian ini bertujuan menerapkan algoritma K-Means dalam mengklasterisasi provinsi berdasarkan kesamaan karakteristik persentase pengeluaran tersebut sehingga diperoleh kelompok-kelompok provinsi yang memiliki pola konsumsi yang relatif homogen. Selain itu, penelitian ini juga bertujuan mengkaji pemanfaatan hasil klasterisasi sebagai dasar segmentasi pasar, sehingga dapat memberikan informasi yang bermanfaat bagi pelaku usaha, pemasar, maupun pembuat kebijakan dalam menyusun strategi pemasaran, menentukan target pasar, serta merancang kebijakan ekonomi yang lebih tepat sasaran sesuai dengan karakteristik masing-masing kelompok provinsi.

B. METODE

Penelitian kuantitatif adalah pendekatan yang diimplementasikan oleh penelitian ini dengan memanfaatkan metode data mining menggunakan algoritma K-Means clustering, yang bertujuan untuk mengelompokkan provinsi di Indonesia ke dalam beberapa kelompok berdasarkan persentase pengeluaran masyarakat untuk makanan dan bukan makanan. Objek dalam penelitian ini adalah seluruh provinsi di Indonesia yang dianalisis berdasarkan persentase pengeluaran masyarakat untuk makanan dan bukan makanan. Publikasi resmi Badan Pusat Statistik (BPS) merupakan sumber didapatkannya data sekunder yang dijadikan sebagai sumber data penelitian ini. Data itu berisi informasi terkait persentase pengeluaran per kapita sebulan makanan dan bukan makanan di daerah perkotaan dan perdesaan menurut provinsi, 2025.



Gambar 1 Tahap Penelitian

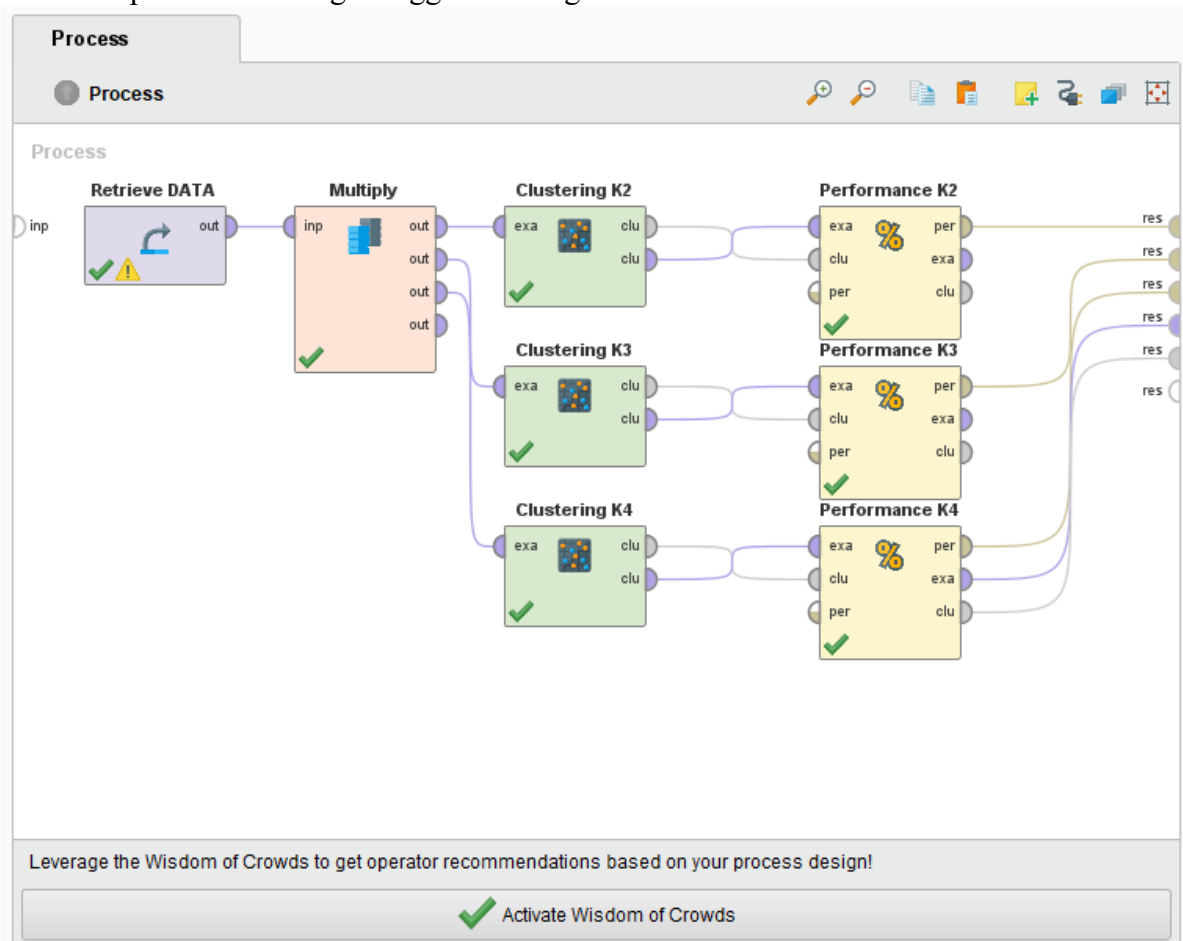
Penelitian ini dilakukan beberapa tahap, mulai dari pengumpulan data, persiapan data, proses klasterisasi menggunakan algoritma K-Means, evaluasi hasil clustering, hingga interpretasi hasil pengelompokan. Aplikasi RapidMiner perangkat lunak yang menyediakan berbagai fitur data mining dimanfaatkan untuk melakukan seluruh proses analisis yang bisa mempermudah proses pengolahan data. Tahap pengumpulan data, yang selanjutnya tahap persiapan data atau preprocessing melakukan pemeriksaan data memastikan tidak ada data kosong, data ganda, atau kesalahan input. Algoritma K-Means dengan aplikasi RapidMiner Studio dimanfaatkan dalam proses clustering. Diawali dengan menentukan jumlah klaster, sistem menentukan centroid awal, kemudian data ke klaster terdekat dikelompokkan, sesudah itu centroid diperbarui dan proses diulang hingga centroid stabil sehingga menghasilkan klaster yang terbaik. Davies Bouldin Index (DBI) dimanfaatkan dalam tahap evaluasi hasil clustering.

Menurut Sinaga & Sipayung, (2025) kualitas klaster bisa diukur dengan Davies Bouldin Index, yang dihitung dengan menghitung rasio antara penyebaran di dalam klaster dan pemisahan antar klaster. Klasterisasi yang optimal tercermin dari nilai DBI yang paling rendah. Tahap interpretasi hasil klaster, setiap klaster dianalisis berdasarkan nilai centroid untuk mengetahui karakteristik pengeluaran masyarakat pada masing-masing kelompok provinsi. Hasil pengelompokan digunakan sebagai dasar segmentasi pasar sehingga bisa membantu perusahaan maupun pemerintah dalam menyusun strategi dan kebijakan yang lebih tepat sasaran.

C. HASIL DAN PEMBAHASAN

1. Pengolahan Data

Penelitian ini menggunakan data persentase pengeluaran per kapita sebulan makanan dan bukan makanan di daerah perkotaan dan perdesaan menurut provinsi, 2025. Data kemudian diolah menggunakan algoritma K-Means di aplikasi RapidMiner untuk mengelompokkan berbagai provinsi yang memiliki karakteristik pengeluaran yang serupa. Proses clustering dilakukan dengan dua atribut yaitu persentase pengeluaran makanan dan persentase pengeluaran bukan makanan. Sebelum melakukan proses clustering, data diperiksa terlebih dahulu agar bisa dipastikan semua data bisa digunakan dalam proses analisis. Hasil pemeriksaan memperlihatkan bahwa semua data yang digunakan sudah lengkap dan tidak ada nilai yang kosong atau data ganda. Selanjutnya, data diimpor ke dalam RapidMiner dan dilakukan proses clustering menggunakan algoritma K-Means.



Gambar 2 Proses Klasterisasi Menggunakan RapidMiner

Gambar 2 memperlihatkan bahwa proses dimulai dengan penggunaan operator Retrieve untuk mengambil data yang akan dianalisis. Operator Multiply untuk menduplikasi data

sehingga bisa dilakukan pengujian beberapa cluster, K2, K3, dan K4. Data diproses menggunakan algoritma K-Means untuk membentuk kelompok berdasarkan tingkat kemiripan karakteristik pengeluaran. Hasil clustering masing-masing jumlah cluster dievaluasi menggunakan operator Performance clustering dengan tujuan agar nilai Davies-Bouldin Index (DBI) bisa didapatkan.

2. Proses Klasterisasi Menggunakan Algoritma K-Means

Proses klasterisasi bisa dilakukan dengan memanfaatkan algoritma K-Means sesudah data diimpor ke RapidMiner. Dalam penelitian ini, sistem menentukan centroid awal sebagai pusat sementara setiap klaster untuk nilai K2, K3, dan K4, kemudian data dikelompokkan ke klaster terdekat, sesudah itu centroid diperbarui dan proses diulang hingga centroid stabil sehingga menghasilkan klaster yang terbaik.

Untuk menentukan jumlah klaster yang terbaik, metode Davies Bouldin Index (DBI) kemudian diimplementasikan dalam mengevaluasi setiap hasil clustering. Nilai DBI yang paling rendah memperlihatkan klasterisasi yang optimal. Berdasarkan pada temuan dari pengolahan data menggunakan RapidMiner diperoleh nilai DBI sebagai berikut:

Tabel 1 Nilai Davies Bouldin Index

Jumlah Klaster (K)	Nilai DBI
K2	0.689
K3	0.504
K4	0.420

Sumber: Data diolah penulis

Berdasarkan pada temuan uji Davies Bouldin Index (DBI) yang sudah dilakukan pada beberapa jumlah cluster, maka diperoleh nilai Davies Bouldin Index (DBI) sebagai berikut: K2 = 0.689, K3 = 0.504, K4 = 0.420. Berdasarkan nilai itu jumlah cluster terbaik adalah K4 sebab menghasilkan nilai DBI paling rendah dibandingkan cluster yang lain.

3. Hasil Klasterisasi Menggunakan Algoritma K-Means

Hasil pengolahan data menggunakan algoritma K-Means dengan jumlah klaster empat atau K4 diperoleh empat kelompok provinsi dengan karakteristik pengeluaran makanan dan bukan makanan yang berbeda, sebagai berikut:

Cluster Model

```
Cluster 0: 16 items
Cluster 1: 15 items
Cluster 2: 2 items
Cluster 3: 5 items
Total number of items: 38
```

Gambar 3. Cluster Model

Hasil clustering memperlihatkan bahwa ada 16 provinsi pada cluster 0, 15 provinsi pada cluster 1, 2 provinsi pada cluster 2, dan 5 provinsi pada cluster 3, yang memuat: Cluster 0 memuat 16 provinsi, yakni Aceh, Sumatera Utara, Sumatera Barat, Riau, Jambi, Sumatera Selatan, Lampung, Kepulauan Bangka Belitung, Nusa Tenggara Barat, Nusa Tenggara Timur, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Sulawesi Tengah, Sulawesi Barat, dan Papua Selatan. Cluster 1 memuat 15 provinsi, yakni Bengkulu, Jawa Barat, Jawa Tengah, Jawa Timur, Banten, Kalimantan Utara, Sulawesi Utara, Sulawesi Selatan, Sulawesi Tenggara, Gorontalo, Maluku, Maluku Utara, Papua Barat, Papua Barat Daya, dan Papua. Cluster 2

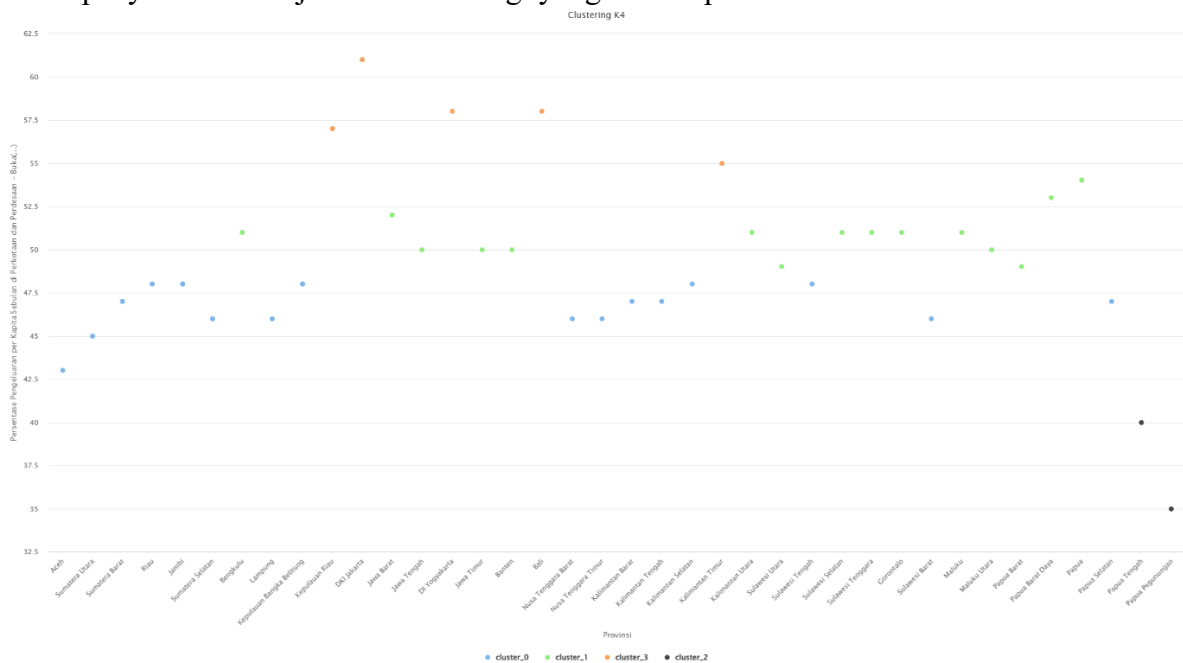
memuat 2 provinsi, yakni Papua Tengah dan Papua Pegunungan. Cluster 3 memuat 5 provinsi, yakni Kepulauan Riau, DKI Jakarta, Daerah Istimewa Yogyakarta, Bali, dan Kalimantan Timur.

Tabel 2 Hasil Pengelompokan Cluster

Cluster	Jumlah Anggota	Kategori Cluster	Karakteristik Cluster
Cluster 0	16 Provinsi	Pengeluaran Makanan Tinggi	Didominasi oleh pengeluaran untuk kebutuhan makanan dibandingkan kebutuhan bukan makanan.
Cluster 1	15 Provinsi	Pengeluaran Seimbang	Pengeluaran makanan dan bukan makanan relatif seimbang.
Cluster 2	2 Provinsi	Pengeluaran Makanan Sangat Tinggi	Dominasi pengeluaran makanan yang paling tinggi dibandingkan cluster lain.
Cluster 3	5 Provinsi	Pengeluaran Bukan Makanan Tinggi	Didominasi oleh pengeluaran untuk kebutuhan bukan makanan dibandingkan kebutuhan makanan.

Sumber: Data diolah penulis

Hasil pengelompokan memperlihatkan ada perbedaan pola konsumsi antar provinsi di Indonesia, setiap cluster memiliki karakteristik pengeluaran yang berbeda dan menjadi dasar dalam penyusunan kebijakan atau strategi yang lebih tepat sasaran.



Gambar 4 Scatter Plot Hasil Clustering

Berdasarkan gambar 4 memperlihatkan hasil visualisasi pengelompokan 38 provinsi di Indonesia menggunakan algoritma K-Means dengan empat cluster. Setiap warna memperlihatkan kelompok provinsi yang memiliki karakteristik pengeluaran makanan dan bukan makanan yang serupa. Visualisasi itu memperlihatkan bahwa ada perbedaan pola pengeluaran antar provinsi sehingga terbentuk empat kelompok yang berbeda.

4. Analisis Nilai Centroid

Nilai centroid digunakan untuk memperlihatkan karakteristik rata-rata setiap cluster. Berdasarkan pada temuan pengolahan data menggunakan RapidMiner diperoleh nilai centroid sebagai berikut:

Tabel 3 Nilai Centroid

Atribut	Cluster 0	Cluster 1	Cluster 2	Cluster 3
Persentase pengeluaran per kapita sebulan untuk makanan	53.375	49.133	62.500	42.200
Persentase pengeluaran per kapita sebulan untuk bukan makanan	46.625	50.867	37.500	57.800
Kategori	Pengeluaran Makanan Tinggi	Pengeluaran Seimbang	Pengeluaran Makanan Sangat Tinggi	Pengeluaran Bukan Makanan Tinggi

Sumber: Data diolah penulis

Berdasarkan pada tabel 3, cluster 0 memiliki nilai centroid pengeluaran makanan sebanyak 53.375 dan pengeluaran bukan makanan sebanyak 46.625. Nilai itu memperlihatkan bahwa provinsi dalam cluster ini umumnya menghabiskan lebih banyak pengeluaran untuk kebutuhan makanan daripada kebutuhan bukan makanan. Cluster 1 memiliki nilai centroid pengeluaran makanan sebanyak 49.133 dan pengeluaran bukan makanan sebanyak 50.867. Nilai itu memperlihatkan bahwa pengeluaran masyarakat di berbagai provinsi dalam cluster itu seimbang antara kebutuhan untuk makanan dan kebutuhan bukan makanan. Cluster 2 memiliki nilai centroid pengeluaran makanan sebanyak 62.500 dan pengeluaran bukan makanan sebanyak 37.500. Nilai itu memperlihatkan bahwa tingkat pengeluaran untuk makanan yang paling tinggi dan tingkat pengeluaran untuk bukan makanan yang paling rendah di antara semua cluster. Hal ini memperlihatkan provinsi di cluster ini masih sangat fokus pada pemenuhan kebutuhan pangan. Cluster 3 memiliki nilai centroid pengeluaran makanan sebanyak 42.200 dan pengeluaran bukan makanan sebanyak 57.800. Nilai itu memperlihatkan bahwa pengeluaran bukan makanan dalam cluster ini lebih tinggi daripada pengeluaran makanan, cluster ini memiliki nilai pengeluaran bukan makanan tertinggi di antara cluster lain. Hal ini memperlihatkan bahwa provinsi dalam cluster ini umumnya mengalokasikan pengeluaran lebih besar untuk kebutuhan bukan makanan, seperti perumahan, pendidikan, kesehatan, transportasi, dan kebutuhan lainnya.

Perbedaan karakteristik pengeluaran setiap cluster memperlihatkan bahwa provinsi di Indonesia dikelompokkan ke beberapa segmen pasar yang berbeda berdasarkan pola konsumsi masyarakat. Provinsi dengan cluster dominasi pengeluaran makanan bisa menjadi target pasar bagi perusahaan yang bergerak di sektor pangan, kebutuhan pokok. Provinsi dengan cluster dominasi pengeluaran bukan makanan bisa menjadi target pasar untuk produk dan jasa bukan makanan, sehubungan dengan hal itu hasil segmentasi menggunakan algoritma K-Means bisa membantu perusahaan maupun pemerintah untuk memahami karakteristik pasar pada setiap kelompok provinsi.

5. Analisis Karakteristik Cluster

Empat kelompok provinsi dengan karakteristik pengeluaran untuk makanan dan bukan makanan yang berbeda dihasilkan oleh hasil klasterisasi menggunakan algoritma K-Means. Perbedaan karakteristik itu bisa membantu memahami pola konsumsi masyarakat di setiap wilayah. Bagi manajemen pemasaran informasi itu penting sebab bisa digunakan sebagai dasar segmentasi pasar berdasarkan perilaku konsumsi masyarakat. Dengan adanya segmentasi itu,

perusahaan bisa menyusun strategi pemasaran yang lebih efektif selaras dengan karakteristik konsumen di wilayah tertentu.

Cluster 0 memiliki nilai centroid sebanyak 53.375 untuk pengeluaran makanan dan 46.625 pengeluaran bukan makanan. Cluster 0 memuat Aceh, Sumatera Utara, Sumatera Barat, Riau, Jambi, Sumatera Selatan, Lampung, Kepulauan Bangka Belitung, Nusa Tenggara Barat, Nusa Tenggara Timur, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Sulawesi Tengah, Sulawesi Barat, dan Papua Selatan. Nilai centroid memperlihatkan bahwa masyarakat umumnya pengeluaran untuk kebutuhan makanan. Berdasarkan segmentasi pasar, cluster ini dikategorikan sebagai pasar yang fokus pada kebutuhan dasar. Perusahaan di sektor kebutuhan pokok memiliki peluang besar untuk berkembang di wilayah dalam cluster ini. Strategi pemasaran yang menekankan kualitas produk dengan harga yang kompetitif lebih disukai oleh masyarakat.

Cluster 1 memiliki nilai centroid sebanyak 49.133 untuk pengeluaran makanan dan 50.867 untuk pengeluaran bukan makanan. Cluster 1 memuat Bengkulu, Jawa Barat, Jawa Tengah, Jawa Timur, Banten, Kalimantan Utara, Sulawesi Utara, Sulawesi Selatan, Sulawesi Tenggara, Gorontalo, Maluku, Maluku Utara, Papua Barat, Papua Barat Daya, dan Papua. Cluster ini seimbang antara pengeluaran makanan dan bukan makanan. Segmentasi pasar cluster ini dikategorikan sebagai pasar berkembang. Wilayah dalam kelompok memiliki peluang cukup besar untuk berbagai jenis produk, produk kebutuhan pokok atau produk sekunder. Perusahaan bisa mengimplementasikan strategi pemasaran yang variatif dengan menyesuaikan dengan karakteristik konsumen. Kualitas layanan dan promosi bisa ditingkatkan oleh strategi pemasaran yang bisa diimplementasikan, yakni diferensiasi produk. Serta menggunakan pemasaran digital yang berpotensi lebih efektif.

Cluster 2 memiliki nilai centroid tertinggi untuk pengeluaran makanan sebanyak 62.500 dan pengeluaran bukan makanan sebanyak 37.500, yang memuat dua provinsi yaitu Papua Tengah dan Papua Pegunungan. Kebutuhan pokok fokus utama dalam pengeluaran masyarakat. Dalam pemasaran cluster ini termasuk pasar kebutuhan primer, perusahaan yang ingin menjangkau wilayah itu sebaiknya menggunakan strategi pemasaran yang fokus pada harga yang terjangkau, distribusi yang mudah, dan ketersediaan produk.

Cluster 3 memiliki nilai centroid sebanyak 42.200 pengeluaran makanan dan 57.800 pengeluaran bukan makanan, yang memuat provinsi Kepulauan Riau, DKI Jakarta, Daerah Istimewa Yogyakarta, Bali, dan Kalimantan Timur. Cluster ini persentase pengeluaran bukan makanan lebih tinggi daripada cluster lain. Hal itu memperlihatkan masyarakat memiliki pengeluaran yang besar untuk kebutuhan seperti pendidikan, kesehatan, transportasi, teknologi, dan lain-lain. Dalam segmentasi pasar cluster ini dikategorikan sebagai pasar modern. Wilayah cluster ini cocok untuk memasarkan produk premium, strategi yang tepat yaitu pemasaran yang fokus pada nilai, diferensiasi produk, pembangunan citra merek, dan pemasaran digital yang terintegrasi.

Hasil segmentasi itu berguna bagi perusahaan untuk menentukan target pasar, menyusun strategi promosi, menetapkan harga, dan mengembangkan produk yang selaras dengan karakteristik konsumen di setiap wilayah. Sehubungan dengan hal itu, pengelompokan provinsi bisa dihasilkan berdasarkan pola pengeluaran dan gambaran terkait potensi pasar pada tiap-tiap wilayah di Indonesia bisa diberikan oleh implementasi algoritma K-Means dalam penelitian ini.

6. Pembahasan Hasil Penelitian

Hasil penelitian memperlihatkan bahwa algoritma K-Means bisa mengelompokkan 38 provinsi di Indonesia ke dalam empat cluster berdasarkan persentase pengeluaran makanan dan bukan makanan. Pengelompokan ini memperlihatkan perbedaan pola konsumsi masyarakat antar provinsi, yang bisa membantu untuk memahami karakteristik pasar di setiap

wilayah. Perbedaan terlihat dari pengeluaran masyarakat untuk memenuhi kebutuhan makanan maupun kebutuhan bukan makanan.

Hasil clustering memperlihatkan bahwa pengeluaran makanan yang dimiliki oleh kelompok provinsi cluster 0 masih melampaui pengeluaran bukan makanan, masyarakat masih memprioritaskan kebutuhan pokok sebagai sumber pengeluaran utama. Cluster 1 memiliki pengeluaran yang relatif seimbang antara makanan dan bukan makanan, masyarakat mulai mengalokasikan pendapatan untuk hal lain selain konsumsi makanan. Cluster 2 memiliki persentase pengeluaran makanan tertinggi, masyarakat masih fokus pada kebutuhan pokok. Cluster 3 memiliki pengeluaran bukan makanan tertinggi, masyarakat pada cluster ini memiliki pola konsumsi lebih beragam.

Tingkat konsumsi masyarakat tidak sama di setiap wilayah, seperti yang ditunjukkan oleh perbedaan karakteristik antar cluster. Hasil itu memperlihatkan beragamnya kondisi sosial ekonomi masyarakat Indonesia. Semakin tinggi pengeluaran bukan makanan, maka semakin besar kemungkinan masyarakat bisa memenuhi kebutuhan sekunder maupun tersier sesudah kebutuhan dasar terpenuhi. Sebaliknya wilayah dengan pengeluaran makanan tinggi memperlihatkan sebagian besar pendapatan digunakan untuk kebutuhan pokok.

Penelitian ini memperlihatkan bahwa penggunaan data mining bisa mengubah data statistik menjadi informasi yang bermanfaat segmentasi wilayah yang mudah dipahami serta dimanfaatkan sebagai dasar pengambilan keputusan. Dengan informasi hasil clustering, memudahkan pengambil keputusan dalam memahami pola pengeluaran antar provinsi, sehubungan dengan hal itu sistem informasi manajemen berperan dalam mengubah data menjadi informasi yang mendukung pengambilan keputusan lebih efektif dan strategis.

Temuan penelitian ini bisa membantu perusahaan untuk memahami karakteristik konsumen berdasarkan wilayah. Salah satu tujuan pemasaran adalah mengetahui kebutuhan dan perilaku konsumen agar strategi yang diimplementasikan selaras dengan kondisi pasar. Dengan menggunakan hasil clustering perusahaan bisa mengetahui wilayah yang lebih potensial untuk produk kebutuhan pokok dan wilayah yang memiliki peluang lebih besar untuk produk sekunder dan tersier.

Hasil penelitian memperlihatkan bahwa pengeluaran makanan dan bukan makanan bisa digunakan untuk membedakan karakteristik pasar antar wilayah. Segmentasi yang terbentuk menghasilkan empat kelompok pasar dengan karakteristik yang berbeda. Kelompok pertama pasar yang berorientasi pada kebutuhan dasar, kelompok kedua pasar yang berkembang, kelompok ketiga pasar dengan kebutuhan dasar yang sangat dominan, dan kelompok keempat pasar yang lebih modern dengan konsumsi bukan makanan yang lebih tinggi. Perusahaan bisa memanfaatkan segmentasi itu untuk membuat strategi pemasaran yang lebih spesifik. Untuk memenuhi kebutuhan masing-masing segmen perusahaan bisa memilih produk yang tepat, menetapkan harga yang sesuai, memilih media promosi yang efektif, dan menyusun strategi distribusi yang sesuai. Sehubungan dengan hal itu, sumber daya pemasaran bisa digunakan dengan efektif dan peluang keberhasilan pemasaran meningkat.

Temuan penelitian ini selaras dengan penelitian yang dijalankan oleh Ardana *et al.*, (2024) yang berjudul Segmentasi Pelanggan Penjualan Online Menggunakan Metode K-Means Clustering, memperlihatkan bahwa dengan metode K-Means clustering dalam segmentasi pelanggan pada penjualan online, perusahaan bisa mengetahui cara berperilaku masing-masing pelanggan, sehingga strategi pemasaran yang lebih baik bisa dibuat. Pada penelitian ini, K-Means berhasil mengelompokkan provinsi di Indonesia berdasarkan persentase pengeluaran makanan dan bukan makanan sehingga menghasilkan segmentasi yang bisa digunakan untuk memahami karakteristik pasar pada masing-masing kelompok provinsi. Temuan penelitian ini juga selaras dengan penelitian yang dijalankan oleh Dewi *et al.*, (2025) yang berjudul Komparasi Hasil Segmentasi Metode K-Means Dan Agglomerative Hierarchical Terhadap Provinsi di Indonesia Berdasarkan Profil Perjalanan Wisata Tahun 2024,

memperlihatkan bahwa kelompok dengan anggota yang seimbang dan lebih bisa beradaptasi dengan cara distribusi data bisa dibentuk oleh implementasi metode K-Means, sehingga lebih efektif dalam mengenali pola aktivitas wisata secara jelas. Penelitian ini mengelompokkan provinsi berdasarkan pola pengeluaran masyarakat, kedua penelitian memperlihatkan bahwa algoritma K-Means bisa menghasilkan kelompok yang jelas dan mudah diinterpretasikan untuk mendukung pengambilan keputusan.

Temuan utama penelitian ini adalah bahwa pola pengeluaran makanan dan bukan makanan bisa membedakan karakteristik konsumsi masyarakat di berbagai provinsi Indonesia. Provinsi dikelompokkan menjadi ada empat cluster yang berbeda berdasarkan tingkat konsumsi dan prioritas pengeluaran dengan memanfaatkan algoritma K-Means. Hasil ini memperlihatkan bahwa pola pengeluaran masyarakat bisa digunakan untuk menentukan karakteristik pasar suatu wilayah.

Implikasi manajerial penelitian ini bisa dirasakan oleh perusahaan maupun pemerintah. Bagi perusahaan bisa menggunakan hasil clustering sebagai dasar untuk menentukan target pasar, membuat strategi pemasaran, mengembangkan produk, dan membuat program promosi yang selaras dengan karakteristik konsumen di setiap wilayah, dan perusahaan bisa mengambil keputusan yang lebih tepat berdasarkan data dan kondisi pasar yang sebenarnya. Bagi pemerintah hasil penelitian bisa menjadi pertimbangan saat membuat kebijakan yang meningkatkan kesejahteraan masyarakat dan pembangunan ekonomi. Informasi pola pengeluaran pada setiap cluster bisa digunakan untuk memahami kondisi konsumsi masyarakat dan menentukan program pembangunan yang selaras dengan karakteristik masing-masing wilayah. Penelitian ini memperlihatkan bahwa pengolahan data pengeluaran masyarakat bisa dilakukan dengan memanfaatkan algoritma K-Means, sehingga informasi yang bermanfaat bagi sistem informasi manajemen bisa dihasilkan, kegiatan manajemen pemasaran bisa didukung, dan proses segmentasi pasar yang lebih objektif bisa dilakukan.

D. KESIMPULAN

Berdasarkan pada temuan penelitian, 38 provinsi di Indonesia bisa dikelompokkan oleh algoritma K-Means clustering berdasarkan persentase pengeluaran makanan dan bukan makanan menjadi empat cluster yang memiliki karakteristik yang berbeda. Cluster 0 memperlihatkan provinsi pengeluaran makanan lebih tinggi dibandingkan pengeluaran bukan makanan. Cluster 1 memiliki pengeluaran makanan dan bukan makanan yang relatif seimbang. Cluster 2 persentase pengeluaran makanan tertinggi. Cluster 3 memiliki persentase pengeluaran bukan makanan tertinggi. Hasil pengelompokan itu memperlihatkan pola pengeluaran masyarakat bisa digunakan sebagai dasar segmentasi pasar sebab menggambarkan perbedaan karakteristik konsumen antar wilayah. Target pasar bisa ditentukan dan strategi pemasaran yang selaras dengan karakteristik tiap-tiap segmen bisa disusun oleh perusahaan dengan adanya kontribusi dari informasi yang dihasilkan. Penelitian ini membuktikan bahwa penerapan algoritma K-Means bisa mengubah data statistik menjadi informasi yang bisa dimanfaatkan untuk mendukung proses pengambilan keputusan.

Sejumlah keterbatasan dimiliki oleh penelitian ini: Pertama objek penelitian hanya memuat 38 provinsi di Indonesia, sehingga hasil segmentasi yang diperoleh masih pada tingkat provinsi dan belum bisa menggambarkan karakteristik konsumsi masyarakat lebih rinci pada tingkat kabupaten/kota. Kedua, hanya algoritma K-Means yang digunakan oleh penelitian ini sebagai metode clustering, meskipun data bisa dikelompokkan oleh metode ini berdasarkan tingkat kemiripan karakteristik secara efektif, penelitian ini belum menbandingkan hasilnya dengan algoritma clustering lainnya, seperti k-medoids, hierarchical clustering, dan fuzzy c-means. Ketiga, data yang digunakan hanya berasal dari satu periode pengamatan, sehingga hasil yang diperoleh hanya menggambarkan kondisi pola pengeluaran masyarakat pada periode itu. Sehubungan dengan hal itu, objek penelitian yang lebih rinci hingga tingkat

kabupaten/kota dilakukan bisa digunakan oleh penelitian selanjutnya dengan melakukan perbandingan beberapa metode clustering untuk memperoleh model terbaik, dan menggunakan data dari beberapa periode waktu agar bisa menghasilkan segmentasi yang lebih komprehensif dan bisa menggambarkan perubahan pola konsumsi masyarakat lebih akurat.

DAFTAR PUSTAKA

- Badan Pusat Statistik. (2025). *Survei Sosial Ekonomi Nasional (Susenas)*. Badan Pusat Statistik.
- Faza, A. H., Mutmainah, M., Kusumadini, L. R., Hidayat, R., & Ikaningtyas, M. (2024). Analisis Segmentasi Pasar Dalam Perencanaan Bisnis Industri Ritel. *Jurnal Ilmiah Multidisiplin*, 1(4), 40-49.
- Hafidz Ardana, C., Khoyum, A. A. A. A., & Faisal, M. (2024). Segmentasi pelanggan penjualan online menggunakan Metode K-Means Clustering. *JISKA (Jurnal Informatika Sunan Kalijaga)*, 9(1), 1-9.
- Iqbal, I., Hidayat, N., Gevano, D. P., & Ilahi, A. P. R. (2025). Segmentasi Pelanggan Menggunakan K-Means Clustering Berdasarkan Data Kepribadian dan Pola Konsumsi. *Jurnal Teknik Informatika (Jutif)*, 6(5), 3914–3924. <https://doi.org/10.52436/1.jutif.2025.6.5.5140>
- Khalish, F., Piranti, N. M., & Martadireja, O. (2025). Implementasi data mining menggunakan teknik clustering dengan metode K-Means. *JIIP-Jurnal Ilmiah Ilmu Pendidikan*, 8(5), 5392-5397.
- Nazihah, F., Danniswara, A., & Wibowo, A. (2025). Analisis Segmentasi Pelanggan Dengan Algoritma K-Means Pada Data Penjualan. *Jurnal Algoritma*, 22(2). <https://doi.org/10.33364/algoritma/v.22-2.2489>
- Ni Luh Ayu Nariswari Dewi, Azizah Zalfa Assyadida, Steffany Marcellia Witanto, Muhammad Nasrudin, & Kartika Maulida Hindrayani. (2025). Komparasi Hasil Segmentasi Metode K-Means dan Agglomerative Hierarchical Terhadap Provinsi di Indonesia Berdasarkan Profil Perjalanan Wisata Tahun 2024. *Statmat: Jurnal Statistika dan Matematika*, 7(3). <https://doi.org/10.32493/sm.v7i3.49999>
- Purba, A. C., & Handhayani, T. (2024). Perbandingan Algoritma K-Means, Affinity Clustering, Dan Minibatch K-Means Untuk Analisis Segmentasi Pasar. *Jurnal Ilmiah Komputer Dan Informatika*, 13(1), 54-63.
- Rohman, N., & Wibowo, A. (2024). Perbandingan Metode K-Medoids dan Metode K-Means Dalam Analisis Segmentasi Pelanggan Mall. *SINTECH (Science and Information Technology) Journal*, 7(1), 49-58.
- Sanjaya, R., Saputra, F. R., Ghiffari, M. H., Kurniawan, T., Nursodiq, A., & Informatika, T. (2026). Peran Sistem Informasi Manajemen Dalam Meningkatkan Efektivitas Pengambilan Keputusan Organisasi.
- Sinaga, C., & Sipayung, S. P. (2025). Implementasi K-Means Clustering Dengan Davies-Bouldin Index Evaluation Terhadap Pengelompokan Provinsi di Indonesia Berdasarkan Indikator Sosioekonomi. *Kakifikom (Kumpulan Artikel Karya Ilmiah Fakultas Ilmu Komputer)*, 109–117.
- Telaumbanua, M. D., Siahaan, G. O., Sinaga, A. Y., Pinasih, Y., & Lase, D. (2025). Meningkatkan Efektivitas Manajemen Pemasaran Dalam Konteks Persaingan Global. *Jurnal Penelitian Ilmiah Multidisipliner*, 2(01), 1596–1606.
- Wijaya, D. D. R., & Daulay, S. R. (2025). Strategi Pemasaran Berbasis Segmentasi: Menentukan, Menganalisis, Dan Meramalkan Pasar Sasaran Untuk Peningkatan Kinerja Bisnis. *Jurnal Ilmiah Ekonomi dan Manajemen*, 3(2), 464–478.