

BAB II

LANDASAN TEORI

2.1 Kajian Teori

2.1.1 Segmentasi Pelanggan

A. Konsep Dasar Segmentasi

Segmentasi merupakan suatu konsep dasar dari strategi pemasaran, yang mana menurut Sula dalam Zuanda (2019) segmentasi pasar merupakan proses pemisahan pelanggan dalam sub-sub kelompok yang disesuaikan dengan jenis produk tertentu yang membutuhkan bauran pemasaran (*mix marketing*) tersendiri (Zuanda, 2019). Kotler dalam Mirantika (2022) mengemukakan definisi segmentasi pasar sebagai rangkaian proses guna meningkatkan ketepatan dalam memasarkan produk maupun layanan. Segmentasi pasar dilakukan dengan klasifikasi kelompok sesuai dengan kebutuhan, keinginan, daya beli, sikap pembeli, serta letak geografis pelanggan (Mirantika, 2022).

Hasil dari adanya segmentasi pasar dapat dijadikan acuan dalam menentukan strategi pemasaran apa yang cocok sesuai dengan sub-sub kelompok pelanggan, sehingga hal tersebut dapat menunjang peningkatan penjualan serta layanan produk dan output nya yakni mencapai tujuan perusahaan.

B. Fungsi Segmentasi Pasar

Segmentasi pasar memiliki beberapa fungsi untuk memudahkan proses pemasaran dalam perusahaan, antara lain adalah sebagai berikut :

- 1) Mengetahui segmentasi pasar terkecil untuk meningkatkan kualitas layanan maupun produk

- 2) Mengetahui sisi unik dari sub-sub kelompok terkecil, sehingga memudahkan strategi pemasaran dengan cara yang kreatif
- 3) Memudahkan alokasi sumber daya perusahaan agar tepat sasaran
- 4) Sebagai kunci utama untuk memenangkan persaingan antara kompetitor lain (Zuanda, 2019).

C. Faktor-faktor Segmentasi Pasar

Kotler dalam Mirantika (2022) mengemukakan variabel utama dalam segmentasi pasar yang dituangkan dalam “Marketing Manajemen” adalah sebagai berikut :

1) Segmentasi Psikografis

Segmentasi psikografis merupakan pengelompokan pelanggan berdasarkan karakteristik psikologis pelanggan yang meliputi aspek kepribadian, sikap, minat, maupun gaya hidup (*life style*) pelanggan. Seperti halnya, segmen pasar yang menerapkan gaya hidup hemat dengan tingkat produktivitas rendah memiliki sikap yang cenderung memilih kualitas bagus dengan biaya yang rendah. Berbeda dengan segmen pasar dengan pelanggan yang menerapkan gaya hidup konsumtif dan mewah akan berbeda dari segi minat maupun cara menyikapi suatu produk maupun layanan dari perusahaan.

2) Segmentasi Geografis

Segmentasi geografis merupakan pengelompokan pelanggan berdasarkan kondisi wilayah yang meliputi negara, cuaca, maupun lingkungan sekitar. Perbedaan letak geografis dari kelompok pelanggan tentunya dapat mempengaruhi adanya jumlah permintaan produk maupun layanan yang ditawarkan oleh

perusahaan dikarenakan kondisi lingkungan maupun cuaca yang berbeda. Selain itu, perbedaan geografis pelanggan tersebut juga mempengaruhi biaya operasional yang dikeluarkan oleh perusahaan.

3) Segmentasi Demografis

Segmentasi demografis merupakan pengelompokan yang umum digunakan dalam segmentasi pasar. Segmentasi ini dikelompokkan berdasarkan usia, jenis kelamin, jenis pekerjaan, penghasilan yang didapat, siklus kehidupan keluarga, agama, ras, maupun kelas sosial dari kelompok pelanggan. Segmentasi ini tentunya untuk membagi pelanggan ke dalam sub kelompok berdasarkan faktor demografis untuk diketahui kebutuhan serta keinginan apa yang sesuai dengan hal tersebut.

4) Segmentasi Perilaku

Segmentasi perilaku merupakan pembagian segmen pasar berdasarkan pengetahuan, minat, serta respon para pelanggan terhadap suatu produk maupun layanan terlepas dari faktor-faktor demografi, geografis, maupun psikologis.

Dari variabel-variabel segmentasi pasar tersebut, penelitian ini hanya menggunakan variabel segmentasi perilaku yang meliputi pengelompokan pelanggan berdasarkan pengetahuan, minat, serta respon para pelanggan terhadap suatu produk maupun layanan terlepas dari faktor-faktor demografi, geografis, maupun psikologis untuk dijadikan data penelitian. Alasan lain dari pemilihan variabel segmentasi perilaku yaitu, data perilaku seperti lama berlangganan, penggunaan layanan, dan tagihan sangat relevan dan terdapat keterkaitan dengan kebutuhan serta keinginan konsumen, utamanya dalam memilih produk layanan *digital service*.

D. Karakteristik Segmentasi Pasar

Segmentasi pasar haruslah memenuhi beberapa karakteristik agar segmentasi pasar dapat bermanfaat serta tepat dalam menentukan strategi pemasaran. Berikut beberapa karakteristik segmentasi pasar adalah sebagai berikut :

- 1) Terukur (*Measurable*), yakni dalam segmentasi pasar perlunya variabel yang dapat diukur, seperti daya beli ataupun profil segmen. Proses pengukuran dilakukan melalui metode penghitungan jarak (*distance metrics*), seperti *Euclidean* atau *Manhattan distance*, untuk mengidentifikasi seberapa dekat atau seberapa mirip data pelanggan satu sama lain. Meskipun ada variabel yang sulit diukur, metode ini dapat mengelompokkan data dengan pendekatan berbasis jarak atau kesamaan (*similarity*).
- 2) Substansial, artinya segmen harus cukup besar dan menguntungkan untuk dilayani. Setelah proses pengelompokan, perlu dilakukan analisis untuk memastikan bahwa jumlah anggota dalam setiap kluster cukup besar dan bernilai strategis. Kluster yang terlalu kecil mungkin tidak relevan untuk dijadikan target pasar. Oleh karena itu, proses pemotongan dendogram perlu diperhatikan agar menghasilkan jumlah segmen yang substansial.
- 3) Dapat diakses (*Accessible*) artinya dapat dijangkau. hasil klasterisasi perlu diterjemahkan ke dalam variabel-variabel yang mudah dipahami dan dapat digunakan dalam strategi pemasaran. Misalnya, jika kluster terbentuk dari atribut lokasi geografis, perilaku pembelian, atau preferensi produk, maka

informasi tersebut harus dapat diterapkan secara operasional oleh tim pemasaran.

- 4) Dapat berdiferensiasi (*Differentiable*). dalam metode *Hierarchical Clustering*, segmen-segmen yang terbentuk harus berbeda secara signifikan satu sama lain. Segmen yang terbentuk dari proses klasterisasi seharusnya memiliki karakteristik unik yang dapat dibedakan dari klaster lainnya. Perbedaan ini dapat dilihat dari centroid atau ciri khas klaster, yang memungkinkan tim pemasaran untuk mengembangkan strategi yang berbeda untuk setiap kelompok. Jika dua segmen memiliki pola yang terlalu mirip, maka perlu dilakukan pengelompokan ulang atau pengaturan jumlah klaster yang lebih optimal.
- 5) Dapat ditindaklanjuti (*Actionable*). Hasil dari *Hierarchical Clustering* harus memberikan wawasan yang dapat ditindaklanjuti. Program atau strategi pemasaran dapat disusun berdasarkan pola yang teridentifikasi dari klaster. Misalnya, jika satu klaster terdiri dari pelanggan dengan preferensi tinggi terhadap diskon, maka perusahaan dapat membuat program promosi khusus untuk klaster tersebut. Oleh karena itu, segmen yang terbentuk melalui metode *Hierarchical Clustering* harus memberikan informasi yang berguna dan dapat diterapkan dalam pengambilan keputusan pemasaran.

2.1.2 Preprocessing Data

Preprocessing data bertujuan untuk mempersiapkan data agar siap untuk dianalisis dengan algoritma *Hierarchical Divisive Clustering*. Proses ini melibatkan beberapa langkah berikut:

1. Pembersihan Data (*data cleaning*): Menghapus data yang hilang, duplikasi, atau kesalahan entri.
2. Transformasi Data: Mengonversi data menjadi format yang dapat digunakan oleh algoritma. Misalnya, mengonversi data kategorik menjadi data numerik menggunakan teknik *label encoding* atau *one-hot encoding*.
3. Penghapusan *Outlier*: Menghapus data yang tidak normal atau terlalu berbeda dari pola umum dalam dataset.
4. Pengisian Data yang Hilang: Mengisi nilai yang hilang dengan metode tertentu, seperti *mean imputation* atau *median imputation*.
5. Pengecekan Konsistensi: Memastikan bahwa data memiliki format, satuan, dan skala yang konsisten.

2.1.3 Pemilihan Fitur

Pemilihan Fitur adalah proses seleksi variabel-variabel yang relevan dan signifikan dalam pembentukan kluster. Tidak semua variabel diperlukan dalam analisis klusterisasi. Oleh karena itu, dilakukan seleksi untuk mengidentifikasi fitur-fitur yang memiliki kontribusi besar terhadap pemisahan kluster (Irawan, 2019). Dalam metode *Hierarchical Divisive Clustering*, pemilihan fitur penting untuk mengurangi kompleksitas data dan meningkatkan keakuratan klusterisasi. Pemilihan fitur dapat menggunakan metode seperti *Principal Component Analysis* (PCA) (2.1), analisis korelasi antar variabel (2.2), atau metode berbasis pengurangan dimensi. Fitur yang sering digunakan dalam pengelompokan pelanggan IndiHome dapat mencakup variabel demografis, perilaku penggunaan

internet, durasi berlangganan, tingkat kepuasan, jumlah keluhan, serta jenis paket layanan yang dipilih.

$$PC_i = \sum (loading_{i,j} * X_j)$$

$$r = \frac{\sum((X_i - \bar{X})(Y_i - \bar{Y}))}{(\sqrt{\sum(X_i - \bar{X})^2} * \sqrt{\sum(Y_i - \bar{Y})^2})}$$

Keterangan :

PC_i = nilai komponen utama ke-i untuk suatu observasi

loading_{i,j} = loading (bobot) dari variabel ke-j pada komponen utama ke-i

X_j = nilai variabel ke-j untuk observasi tersebut

R = koefisien korelasi

X_i, Y_i = nilai observasi untuk variabel X dan Y

\bar{X} , \bar{Y} = rata-rata variabel X dan Y

2.1.4 Normalisasi Data

Normalisasi data adalah proses penyesuaian skala nilai dari data yang ada agar memiliki rentang nilai yang sama. Normalisasi diperlukan karena algoritma *Hierarchical Divisive Clustering* menggunakan jarak *Euclidean* atau metrik jarak lainnya untuk mengelompokkan data. Jika data memiliki skala yang berbeda-beda, maka jarak antara data bisa menjadi bias. Teknik normalisasi yang umum digunakan adalah *Min-Max Normalization* atau *Z-score Normalization*, di mana data disesuaikan agar berada dalam rentang [0,1] atau memiliki distribusi normal dengan nilai rata-rata nol dan standar deviasi satu (Ulansari, 2019). Berikut adalah persamaan dari *Min-Max Normalization* (2.3) atau *Z-score Normalization* (2.4) :

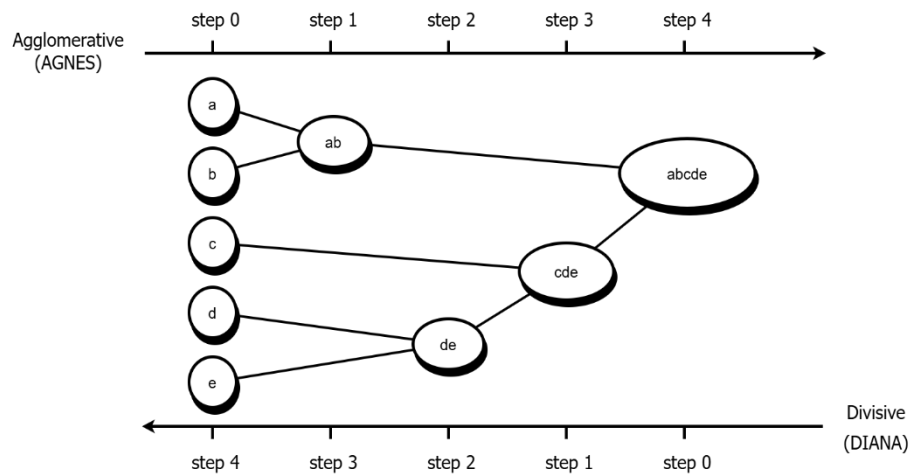
$$X_{norm} = (X - X_{min}) / (X_{max} - X_{min}) \dots\dots\dots(2.3)$$

$$X_{norm} = (X - mean) / std\dots\dots\dots(2.4)$$

Normalisasi ini penting terutama ketika variabel memiliki skala pengukuran yang berbeda, seperti penghasilan pelanggan, durasi langganan layanan, dan tingkat penggunaan layanan internet.

2.1.5 Metode Clustering

Metode clustering adalah teknik pembagian data ke dalam kelompok-kelompok yang memiliki kemiripan satu sama lain. Menurut Tan (2006) ada dua jenis utama metode clustering, yaitu *clustering partisi* dan *clustering hierarki* (Mirantika, 2022). Dalam clustering hierarki, data dikelompokkan berdasarkan tingkat kemiripannya. Data yang sangat mirip akan berada dalam satu kelompok kecil, sedangkan data yang kurang mirip akan berada dalam kelompok yang lebih besar. Hasil dari pengelompokan ini adalah sebuah diagram berbentuk pohon yang menunjukkan hubungan hirarkis antara kelompok-kelompok data. Clustering hirarki dibagi lagi menjadi dua metode, yaitu *agglomerative clustering* (dari bawah ke atas) dan *divisive clustering* (dari atas ke bawah). Pada penelitian ini, metode yang digunakan adalah *Divisive Hierarchical Clustering*, di mana seluruh data dimasukkan ke dalam satu klaster besar, kemudian dibagi secara bertahap menjadi sub-klaster yang lebih kecil. Metode ini efektif dalam pengelompokan data pelanggan yang jumlahnya besar dan memiliki karakteristik yang berbeda-beda.



Gambar 2.1 Proses *Hierarchical Clustering*

(Sumber : Budhi dkk., 2019)

2.1.1 Metode *Hierarchical Clustering*

Metode *Hierarchical Clustering* adalah metode pengelompokan data yang membangun hirarki dari data tersebut. Proses pembentukan kluster dapat dilakukan dengan dua cara:

1. *Agglomerative Clustering* : Dimulai dari setiap data sebagai kluster tunggal, kemudian kluster-kluster kecil tersebut digabungkan secara berulang hingga membentuk satu kluster besar.
2. *Divisive Clustering*: Dimulai dari satu kluster besar yang mencakup semua data, kemudian kluster tersebut dibagi menjadi kluster-kluster yang lebih kecil secara bertahap (Suyono dkk., 2024).

Penelitian ini berfokus pada metode *Divisive Clustering*, karena metode ini lebih efektif dalam mengidentifikasi kluster yang tidak jelas sejak awal. Metode ini menggunakan pendekatan pemisahan *top-down*, yang dimulai dari satu kluster besar, kemudian membaginya secara bertahap hingga membentuk sub-kluster yang

lebih kecil. Hasil dari proses clustering ini divisualisasikan menggunakan dendogram yang menggambarkan pemisahan hierarkis antar kluster

2.1.2 Algoritma *Divisive Hierarchical Clustering*

Algoritma *Divisive Hierarchical Clustering* dimulai dengan seluruh data dimasukkan ke dalam satu kluster besar, kemudian kluster tersebut dipecah menjadi dua sub-kluster. Proses ini terus berlanjut hingga semua data dikelompokkan ke dalam sub-kluster terkecil yang diinginkan. Langkah-langkah algoritma ini adalah sebagai berikut:

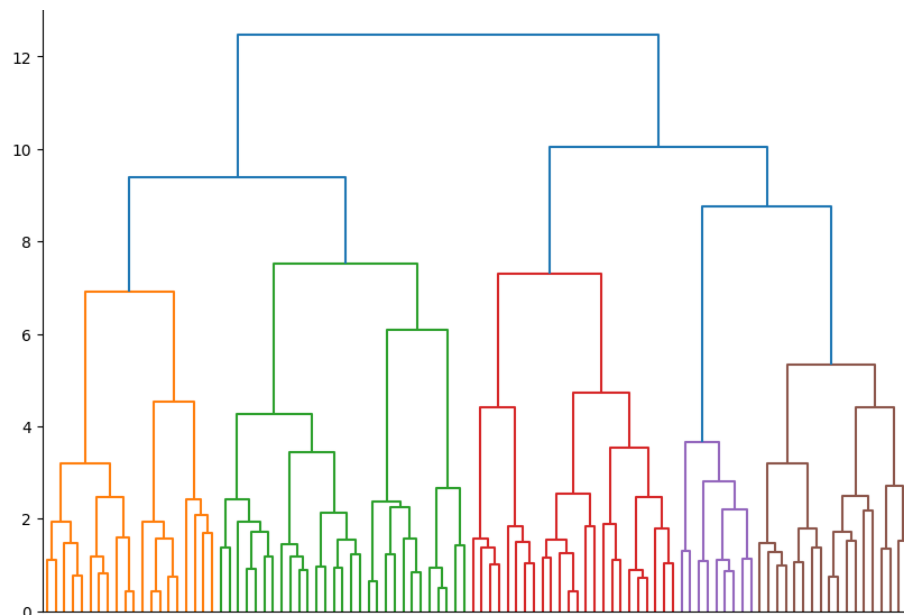
1. Inisialisasi Kluster Awal: Memasukkan seluruh data ke dalam satu kluster besar.
2. Penghitungan Jarak dan Pemisahan Kluster: Menghitung jarak antara titik-titik data dalam kluster menggunakan matrik jarak, seperti *Euclidean distance* (2.5), *Manhattan distance* (2.6), atau metrik jarak lainnya.

$$d(x, y) = \sqrt{\sum (x_i - y_i)^2} \dots\dots\dots(2.5)$$

$$d(x, y) = \sum |x_i - y_i| \dots\dots\dots(2.6)$$

3. Pembagian Kluster: Memisahkan kluster besar menjadi dua kluster kecil berdasarkan perbedaan jarak terbesar. Data yang memiliki jarak terbesar dari rata-rata kluster dipisahkan ke kluster baru.
4. Iterasi: Proses pembagian terus berlangsung hingga jumlah kluster yang diinginkan tercapai atau hingga kriteria tertentu terpenuhi, seperti ukuran kluster minimal atau batas jarak tertentu.

5. Dendrogram: Hasil dari algoritma *Divisive Clustering* dapat divisualisasikan menggunakan dendrogram, yang menggambarkan hubungan hirarki antar kluster.



Gambar 2.2 Hasil Dendrogram Algoritma Divisive Hierarchical

(Sumber : Diolah Peneliti, 2024)

Keunggulan dari algoritma *divisive* ini adalah kemampuannya untuk menemukan kluster dengan pola yang tidak terdefinisi sejak awal. Prosesnya lebih efisien jika dibandingkan dengan *agglomerative clustering* ketika jumlah data yang diproses sangat besar. Proses pengelompokan pelanggan IndiHome dapat menggunakan variabel-variabel seperti tingkat penggunaan internet, lokasi pelanggan, jumlah keluhan, dan frekuensi pembayaran tagihan. Algoritma ini juga memungkinkan perusahaan untuk mengidentifikasi pelanggan yang memiliki karakteristik serupa dan mengembangkan strategi pemasaran yang lebih spesifik berdasarkan kluster-kluster tersebut.

2.1.3 Python

Python juga digunakan secara luas dalam pembelajaran mesin dan kecerdasan buatan. Library seperti *Scikit-learn*, *TensorFlow*, dan *PyTorch* menyediakan alat yang kuat untuk pengembangan dan implementasi model pembelajaran mesin dan kecerdasan buatan (Géron, 2019).

2.2 Kajian Empiris

Tabel 2.1 Kajian Empiris

Nama Peneliti	Tahun	Judul	Jurnal	Persamaan dan Perbedaan
Yudi Irawan	2019	Penerapan Data Mining untuk Evaluasi Data Penjualan Menggunakan Metode Clustering dan Algoritma Hirarki Divisive	JTIULM	Persamaan penelitian tersebut yakni menggunakan metode algoritma <i>Hierarchical Divisive</i> dalam pengelompokan kluster Perbedaannya yaitu dalam fokus permasalahan yaitu pengelompokan data mining pelanggan loyal terhadap perusahaan Media World Pekanbaru
Tiris Sudrartono	2019	Pengaruh Segmentasi Pasar Terhadap Tingkat Penjualan Produk Fashion Usaha Mikro Kecil	Coopetition : Jurnal Ilmiah Manajemen	Persamaan dari penelitian ini dari topik yang diangkat yaitu segmentasi pasar Perbedaannya yakni objek penelitian serta metode penelitian yang digunakan yaitu jenis deskriptif
Ilham Adnan Kasoqi, dkk	2021	Pengelompokan Desa atau Kelurahan di Kutai Kartanegara Menggunakan Algoritma Divisive Analysis	Jurnal Statistika	Persamaan penelitian tersebut yakni sama-sama menggunakan metode algoritma <i>Hierarchical Divisive</i> Perbedaan dengan penelitian ini yaitu perbedaan topik dan fokus penelitian, dimana penelitian ini bertujuan untuk mengelompokkan desa guna penyaluran bantuan secara tepat.
Aulia Nur Rahmawati dan Luky susilowati	2021	Analisis Kualitas Pelayanan dan Persepsi Harga Terhadap	Jurnal Sosial Ekonomi dan Politik	Persamaan penelitian dari studi kasus yang diangkat yakni pelanggan IndiHome

Nama Peneliti	Tahun	Judul	Jurnal	Persamaan dan Perbedaan
		Kepuasan Pelanggan (Studi Kasus Pada Pelanggan IndiHome di STO Injoko Witel Surabaya Selatan)		Perbedaannya yaitu dari fokus permasalahan, metode yang digunakan, serta perbedaan lokasi
Aan Rohanah, dkk	2021	Perbandingan Naïve Bayes dan Support Vector Machine untuk Klasifikasi Ulasan Pelanggan IndiHome	Jurnal STRING (Satuan Tulisan Riset dan Inovasi Teknologi)	Persamaannya dari topik penelitian yang diangkat yaitu pelanggan IndiHome Perbedaannya yaitu dari segi fokus penelitian dan metode yang digunakan yaitu algoritma <i>Naïve Bayes Classifier</i> dan <i>Support Vector Machine Kernel Linear</i>
Aastha Gupta, Himanshu Sharma, Anas Akhtar	2021	A Comparative Analysis of K-Means and Hierarchical Clustering	EPRA International Journal of Multidisciplinary Research (IJMR)-Peer Reviewed Journal	Persamaan dari penelitian ini yakni terdapat pada metode yang digunakan, yakni menggunakan clustering. Perbedaan penelitian ini terdapat pada fokus penelitian, permasalahan, serta objek penelitian
Pranav Shetty, Suraj Singh	2021	Hierarchical Clustering: A Survey	International Journal of Applied Research	Persamaan penelitian ini yakni membahas tentang bagaimana cara kerja algoritma clustering Namun, perbedaan yang mencolok yakni fokus permasalahan yang diangkat oleh peneliti
Mochammad Ainul Yaqin, dkk	2022	Segmentasi Pelanggan Agen <i>Beauty Local Brand</i> Menggunakan Algoritma <i>K-Modes</i> Berbasis <i>Python</i>	Jurnal NJCA	Persamaannya yakni sama-sama membahas topik segmentasi pelanggan Perbedaannya dari topik yang diangkat yaitu <i>Beauty Local Brand</i> dan teknik pengelompokan menggunakan Algoritma <i>K-Modes</i> Berbasis <i>Python</i>
Musthofa Galih Pradana, dkk	2023	Optimalisasi Segmentasi Pelanggan Menggunakan <i>Hierarchical Clustering</i>	JTI : Jurnal Teknik Informasi	Persamaannya yakni penerapan metode <i>Hierarchical Clustering</i> pada suatu objek

Nama Peneliti	Tahun	Judul	Jurnal	Persamaan dan Perbedaan
				Perbedaannya yakni terletak pada objek penelitian serta fokus permasalahan yang diangkat.

Berdasarkan dari penelitian terdahulu yang relevan dengan penelitian yang diangkat peneliti, terdapat beberapa pembahasan dari kajian terdahulu tersebut adalah sebagai berikut :

Penelitian yang dikaji oleh Yudi Irawan (2019) yang berjudul “ Penerapan Data Mining untuk Evaluasi Data Penjualan Menggunakan Metode *Clustering* dan Algoritma *Hirarki Divisive*” membahas tentang pengelompokan data mining dari pelanggan loyal terhadap perusahaan Media World Pekanbaru. Penelitian ini bertujuan untuk menerapkan strategi bisnis serta kebijakan yang tepat untuk penjualan produk dari perusahaan, seperti produk mesin digital printing. Persamaannya dari topik penelitian yang dibahas yaitu pengelompokan menggunakan Algoritma *Hirarki Divisive*. Perbedaan dari penelitian yang dikaji peneliti yaitu pertama fokus penelitian yang mengangkat pengelompokan data mining pelanggan. Kedua, yaitu objek penelitian berfokus pada perusahaan Media World Pekanbaru.

Penelitian kedua yang dikaji oleh Tiris Sudrartono (2019) yaitu mengenai bagaimana pengaruh segmentasi pasar terhadap UMKM Bumi Orange khususnya terhadap produk *fashion*. Penelitian ini dilatarbelakangi karena UMKM di Cinunuk Bandung belum menyentuh angka maksimal dalam penjualan produk. Penelitian ini menggunakan metode deskriptif dengan analisis regresi linear sederhana. Persamaan penelitian ini yakni membahas tentang segmentasi pasar. Namun,

perbedaan dari penelitian ini yaitu fokus permasalahan , metode yang digunakan, dan jenis penelitiannya.

Selanjutnya, penelitian yang dilakukan pada 2021 oleh Ilham Adnan, dkk mengkaji tentang pengelompokan warga desa atau kelurahan di Kutai Kartanegara menggunakan *Hierarchical Divisive Analysis*. Penelitian ini berfokus pada pengelompokan warga untuk memudahkan pemerintah daerah untuk menyalurkan bantuan desa. Penelitian ini memiliki persamaan metode penelitian, namun yang membedakan dari penelitian yang dikaji peneliti yaitu terhadap fokus permasalahan dan lokasi yang digunakan.

Studi lanjutan yang diteliti oleh Rahmawati dan Susilowati (2021) yang menganalisis bagaimana pengaruh kualitas layanan perusahaan internet dan persepsi harga terhadap kepuasan pelanggan IndiHome. Penelitian ini menggunakan metode deskriptif dengan teknik purposive sampling. Penelitian ini memiliki persamaan yakni objek yang diteliti sama-sama mengangkat pelanggan IndiHome. Perbedaan yang signifikan yakni dari fokus penelitian, jenis penelitian yang digunakan, metode, serta lokasi penelitian.

Temuan selanjutnya yang dikaji oleh Aan Rohanah, dkk (2021) yang berjudul “Perbandingan *Naïve Bayes* dan *Support Vector Machine* untuk Klasifikasi Ulasan Pelanggan IndiHome”. Penelitian ini mengkaji tentang bagaimana ulasan pelanggan IndiHome terhadap tingkat layanan perusahaan. Data yang diambil yakni dari aplikasi *Twitter* yang kemudian diolah dan diklasifikasi dengan dua metode yakni menggunakan algoritma *Naïve Bayes* dan *Support Vector Machine*. Penelitian ini memiliki kesamaan yakni terletak pada objek penelitiannya.

Perbedaan yang signifikan yaitu terletak pada fokus permasalahan serta metode yang digunakan.

Penelitian internasional yang juga dilakukan oleh Aastha Gupta, Himanshu Sharma, Anas Akhtar pada 2021 dimana penelitian ini berfokus pada bagaimana perbedaan antara algoritma K-Means dan Clustering dengan menggunakan data mining WEKA. Persamaan penelitian ini yakni sama-sama meneliti tentang algoritma clustering, namun perbedaan yang cukup signifikan yakni pada fokus permasalahan yang diangkat.

Selanjutnya yaitu penelitian yang dilakukan pada tahun 2021 yang dikaji oleh Pranav Shetty dan Suraj Singh yang berfokus pada bagaimana peneliti membedakan antara metode partisi dengan metode clustering. Tujuan penelitian ini yakni mengetahui bagaimana implementasi dari algoritma pengelompokan hierarki sehingga dapat memahami cara kerja dari algoritma tersebut dan menghasilkan pendekatan dan inovasi baru untuk perbaikan. Perbedaan yang mencolok yakni fokus permasalahan yang diangkat serta objek yang digunakan.

Mochammad Ainul Yaqin dkk pada 2022 juga telah mengkaji tentang segmentasi pelanggan terhadap agen *Beauty Local Brand* dengan menggunakan metode K-Modes berbasis Python. Penelitian ini berfokus pada bagaimana agen *Beauty Local Brand* dapat mengatasi permasalahan strategi pemasarannya yakni dengan mengelompokkan karakteristik pelanggan. Persamaannya yakni sama-sama mengangkat topik segmentasi pasar, namun yang membedakan yakni objek penelitian dan metode yang digunakan.

Penelitian selanjutnya yang berjudul “Optimalisasi Segmentasi Pelanggan Menggunakan *Hierarchical Clustering*” yang diteliti oleh Perdana, dkk pada 2023 yakni mengkaji tentang bagaimana proses metode *Hierarchical Clustering* pada data sehingga dapat dijadikan sebuah referensi dalam menentukan strategi pemasaran terhadap perusahaan. Data yang digunakan yaitu data pengunjung *mall* yang dikumpulkan melalui kuesioner online. Persamaan dari penelitian ini yaitu penerapan dari metode *Hierarchical Clustering*, namun perbedaannya yakni terletak pada objek penelitian dan fokus permasalahan, dimana penelitian yang dikaji ini lebih kepada proses penerapannya dengan objek yang umum.

Dari penelitian terdahulu yang telah dipaparkan diatas, terdapat satu referensi yang paling mendekati dengan penelitian yang akan diteliti, yakni penelitian yang berjudul “Segmentasi Pelanggan Agen *Beauty Local Brand* Menggunakan Algoritma *K-Modes* Berbasis *Python*” yang dikaji oleh Mochammad Ainul Yaqin dkk pada 2022. Penelitian ini sama-sama mengkaji tentang bagaimana agen tersebut mengelompokkan pelanggan berdasarkan karakteristik tertentu. Algoritma yang digunakan yaitu *K-Modes* dimana metode ini merupakan jenis data yang bersifat kategorikal dengan mengukur kemiripan kategori, sedangkan algoritma *Hierarchical Clustering* memiliki jenis data numerik maupun kategorikal, seperti pengelompokan data berdasarkan ukuran jarak pada data numerik dan ukuran ketidakmiripan pada data kategorikal.