

Penerapan Kecerdasan Buatan Dengan Metode K-Means Dalam Penentuan Target Client Di Perusahaan CV.Project Master

Muhammad Sholahudin^{1*}, Herliyani Hasanah², Afu Ichsan Pradana³

¹Teknik Informatika

Universitas Duta Bangsa

^{1*} muhammadsholahudin96@gmail.com

²Teknik Informatika

Universitas Duta Bangsa

² herliyani_hasanah@udb.ac.id

³Teknik Informatika

Universitas Duta Bangsa

³ afu_ichsan@udb.ac.id

Abstrak— Kemajuan dalam teknologi informasi serta penerapan Kecerdasan Buatan (AI) mendorong organisasi untuk memaksimalkan data pelanggan sebagai landasan demi keputusan bisnis yang lebih efektif. Salah satu kendala yang dihadapi oleh CV. Project Master adalah proses identifikasi pelanggan yang masih tergolong manual, sehingga memakan waktu yang cukup lama dan kemungkinan adanya kesalahan selama proses analisis. Studi ini berupaya untuk mengevaluasi dan mengklasifikasikan informasi klien demi mendukung perusahaan dalam mengidentifikasi pelanggan secara lebih terstruktur, dan didukung oleh data, serta untuk mengembangkan prototipe sistem analisis pelanggan yang didukung oleh AI. Penelitian ini mengadopsi K-Means Clustering. Pendekatan ini memanfaatkan atribut data seperti jenis usaha, skala usaha, lama usaha, lokasi usaha, dan rating Google. Prosedur penelitian mencakup pengumpulan dan pra-pemrosesan data, transformasi, normalisasi, penentuan parameter kluster, implementasi algoritma K-Means untuk pengelompokan, evaluasi hasil klusterisasi, dan Prosedur penelitian mengikuti tahapan KDD, di mana hasil segmentasi diimplementasikan ke dalam sistem pada tahap Knowledge Presentation untuk mendukung pengambilan keputusan bisnis. Penelitian ini mengadopsi K-Means Clustering dengan memanfaatkan atribut rating Google, jumlah ulasan, dan status website. Prosedur penelitian mengikuti tahapan Knowledge Discovery in Databases yang meliputi selection, preprocessing, transformation, data mining, serta evaluation/interpretation. Hasil penelitian menunjukkan bahwa algoritma K-Means berhasil mengelompokkan 120 data calon client menjadi 3 cluster optimal berdasarkan evaluasi metode Elbow. Kualitas pengelompokan divalidasi dengan nilai Silhouette Score sebesar 0.2821, di mana proses mencapai kondisi konvergen pada iterasi ke-6. Implementasi sistem pada tahap knowledge presentation membuktikan bahwa segmentasi berbasis AI mampu meningkatkan efektivitas penentuan target client secara objektif dan terukur bagi CV. Project Master.

Kata kunci— Artificial Intelligence, K-Means Clustering, Target Client, Knowledge Discovery in Databases

Abstract— Advances in information technology and the application of Artificial Intelligence (AI) encourage organizations to maximize customer data as a basis for more effective business decisions. One of the obstacles faced by CV. Project Master is the customer identification process which is still relatively manual, so it takes a long time and the possibility of errors during the analysis process. This study seeks to evaluate and classify client information to support the company in identifying customers in a more structured, data-driven manner, and to develop a prototype of an AI-powered customer analysis system. This study adopts K-Means Clustering. This approach utilizes data attributes such as business type, business scale, business duration, business location, and Google ratings. The research procedure includes data collection and pre-processing, transformation, normalization, determination of cluster parameters, implementation of the K-Means algorithm for clustering, evaluation of clustering results, and The research procedure follows the KDD stages, where the segmentation results are implemented into the system at the Knowledge Presentation stage to support business decision making. This study adopts K-Means Clustering by utilizing Google rating attributes, number of reviews, and website status. The research procedure followed the stages of Knowledge Discovery in Databases which include selection, preprocessing, transformation, data mining, and evaluation/interpretation. The results showed that the K-Means algorithm successfully grouped 120 prospective client data into 3 optimal clusters based on the Elbow method evaluation. The quality of the clustering was validated with a Silhouette Score of 0.2821, where the process reached a convergent condition in the 6th iteration. The implementation of the system at the knowledge presentation stage proved that AI-based segmentation was able to increase the effectiveness of determining client targets objectively and measurably for CV. Project Master.

Keywords— Artificial Intelligence, K-Means Clustering, Target Client, Knowledge Discovery in Databases

I. PENDAHULUAN

Perkembangan teknologi informasi dan komunikasi yang pesat telah mendorong transformasi digital dalam berbagai sektor bisnis, termasuk perusahaan skala kecil dan menengah. Laporan dari Asosiasi Penyelenggara Jasa Internet Indonesia (APJII) menunjukkan bahwa

penetrasi internet di Indonesia terus mengalami peningkatan signifikan dalam beberapa tahun terakhir, yang berdampak langsung pada meningkatnya aktivitas transaksi dan interaksi pelanggan secara digital. Kondisi ini menyebabkan perusahaan memiliki volume data pelanggan yang semakin besar dan kompleks [1]. Dalam buku *Data Mining: Concepts and*

Techniques, data dalam jumlah besar memerlukan teknik analitik yang sistematis agar dapat menghasilkan informasi yang bernilai bagi pengambilan keputusan. Fenomena tersebut menunjukkan pentingnya pemanfaatan kecerdasan buatan dalam pengelolaan data pelanggan, khususnya dalam menentukan target client yang potensial secara objektif dan berbasis data [2]. Oleh karena itu, penelitian mengenai penerapan kecerdasan buatan dalam penentuan target client menjadi relevan untuk menjawab kebutuhan perusahaan terhadap strategi pemasaran yang lebih efektif dan efisien.

Laporan *The State of AI* menunjukkan bahwa penggunaan AI dalam sektor bisnis terus meningkat, terutama dalam analisis pelanggan, personalisasi pemasaran, serta pengambilan keputusan berbasis data. AI memungkinkan perusahaan untuk mengolah data dalam skala besar secara efisien serta mengidentifikasi pola perilaku pelanggan yang tidak mudah ditemukan metode konvensional [3].

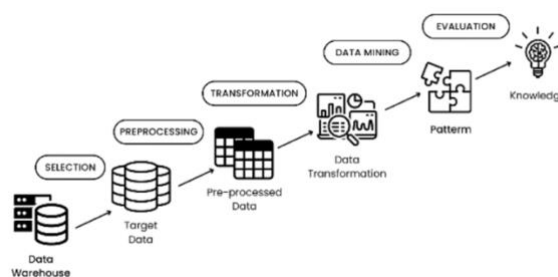
Pada konteks studi kasus di CV. Project Master, proses penentuan target client selama ini masih dilakukan melalui pendekatan analisis deskriptif sederhana berdasarkan data historis klien, seperti riwayat transaksi, frekuensi pemesanan, serta komunikasi langsung antara tim pemasaran dan klien. Data pelanggan dikelola menggunakan pencatatan berbasis spreadsheet dan evaluasi manual oleh tim marketing untuk menentukan prioritas klien yang dianggap potensial. Metode ini memiliki kelebihan, yaitu mudah diterapkan, tidak memerlukan infrastruktur komputasi yang kompleks, serta memungkinkan adanya pertimbangan pengalaman praktis dari tim pemasaran dalam mengambil keputusan. Namun demikian, pendekatan tersebut memiliki beberapa kelemahan, antara lain tingginya subjektivitas dalam menentukan prioritas klien, keterbatasan dalam mengolah data dalam jumlah besar, serta kesulitan dalam mengidentifikasi pola tersembunyi dari perilaku pelanggan secara sistematis [4].

II. METODOLOGI PENELITIAN

A. Tahapan Penelitian

Metodologi penelitian ini menerapkan kerangka kerja *Knowledge Discovery in Databases* (KDD) yang sistematis untuk mentransformasi data mentah menjadi informasi strategis bagi CV. Project Master

. Tahap awal dimulai dengan Data Selection, di mana 120 data calon klien dikumpulkan melalui teknik scraping pada Google Maps dengan fokus pada tiga atribut reputasi digital: rating Google, jumlah ulasan, dan status kepemilikan website



Gambar 31. Proses knowledge discovery in database (KDD)

Berkut beberapa tahapan KDD [5]:

1. Data Selection

Tahap pertama adalah Data Selection, di mana peneliti mengumpulkan dataset klien dari CV. Project Master dan memilih atribut yang relevan untuk proses segmentasi. Hal ini bertujuan untuk memfokuskan analisis pada variabel yang paling berpengaruh terhadap pengelompokan pelanggan [6].

2. Preprocessing

Tahap ini dilakukan untuk membersihkan data dari nilai yang hilang (*missing values*) atau data yang tidak konsisten. Tahap ini bertujuan untuk melakukan proses pra-pemrosesan pada data Pelanggan agar dapat mengurangi kata-kata yang tidak diperlukan atau

tidak relevan, sehingga mempermudah proses pengolahan pada tahap selanjutnya. Data juga diperiksa konsistensinya dan disesuaikan agar layak untuk pemrosesan lebih lanjut. Tahapan data *preprocessing* yang dilakukan yaitu, data cleaning, normalization [7].

3. Transformation

Pada tahap ini, data melalui proses transformasi untuk menyesuaikan format atribut agar sesuai dengan kebutuhan algoritma yang akan digunakan. Pada penelitian ini proses *klasterisasi* dilakukan dengan memanfaatkan representasi fitur *Min-Max Scaling* untuk menyamakan skala antar atribut, terutama antara rating (skala 1-5) dan jumlah ulasan (skala ratusan/ribuan. Transformasi ini bertujuan agar tidak ada atribut yang mendominasi proses perhitungan jarak pada algoritma *K-Means* [8].

4. Data Mining

Merupakan inti dari penelitian dengan mengimplementasikan algoritma *K-Means Clustering*. Proses dimulai dengan penentuan jumlah cluster optimal menggunakan Metode *Elbow*, dilanjutkan dengan perhitungan jarak *Euclidean Distance* hingga mencapai kondisi konvergen atau stabil. Hasil dari tahap ini adalah segmentasi calon client ke dalam tiga kelompok potensi [9].

5. Evaluation

Tahap terakhir adalah mengevaluasi kualitas cluster menggunakan *Silhouette Score* dan memvisualisasikan persebaran data melalui *Principal Component Analysis*. Sebagai bentuk presentasi pengetahuan Knowledge Presentation, hasil analisis ini diintegrasikan ke dalam sebuah prototipe sistem berbasis web

untuk membantu CV. Project Master menginterpretasikan data secara praktis dalam pengambilan keputusan bisnis [10].

6. Knowledge Presentation

Sistem website yang dikembangkan bertindak sebagai antarmuka cerdas yang menyederhanakan kompleksitas algoritma *K-Means* bagi pengguna. Pengguna cukup mengunggah data mentah klien, dan sistem secara otomatis menjalankan preprocessing dan clustering di *backend* [11].

III. HASIL DAN PEMBAHASAN

A. Deskripsi Dataset Penelitian

Tahap awal penelitian dilakukan dengan mengumpulkan data klien CV. Project Master yang diperoleh dari proses *scrapping* di google maps. Data tersebut digunakan sebagai dasar dalam proses segmentasi target client menggunakan metode *K-Means Clustering*.

Tabel 1. Contoh Table Dataset

No	t	ngan
	Google	ating Usaha
	Ulasan	Jumlah Ulasan Mengenai usaha
e		Keppemilikan Website Usaha tersebut

3.1.Data Selection

Tahap awal dalam penelitian ini adalah Data Selection, yaitu proses pemilihan dan seleksi data mentah yang akan digunakan dalam prosedur pengelompokan Data penelitian ini dikumpulkan melalui teknik *scrapping* pada Google Maps yang menghasilkan

sebanyak 120 data calon klien CV. Project Master Dari sekumpulan data tersebut, dipilih tiga atribut kunci yang merepresentasikan reputasi dan karakteristik digital klien, yaitu Rating Google, Jumlah Ulasan, dan status kepemilikan Website, sebagaimana dirinci dalam Tabel 1. Penyeleksian fitur-fitur ini bersifat krusial untuk memastikan bahwa model *clustering* yang dibangun memiliki kualitas dan relevansi yang tinggi dalam menentukan target klien potensial

3.2. Preprocessing

Data yang telah dipilih sebelumnya dilakukan proses pengecekan terhadap keberadaan data yang hilang (*missing value*).

a. Data Cleaning

Proses pertama yang dilakukan adalah penghapusan data duplikat. Data duplikat umumnya muncul karena suatu usaha tercatat lebih dari satu kali pada hasil scraping. Keberadaan data duplikat dapat memengaruhi hasil clustering karena menyebabkan suatu usaha memiliki bobot yang lebih besar dibandingkan data lainnya. Oleh karena itu, seluruh data yang teridentifikasi sebagai duplikat dihapus sehingga setiap usaha hanya direpresentasikan oleh satu data.

Tahap ini sangat penting karena kualitas data akan memengaruhi hasil analisis yang dihasilkan sistem.

Tabel 2. Tabel Hasil Data Cleaning

Pelanggan	Rating	Jumlah Ulasan	Website
1	4.5	456	1
2	4.5	55	0
3	4.7	96	0
4	4.2	108	0

	4.8	68	0
5			

b. Normalization

Normalisasi dilakukan karena setiap atribut memiliki rentang nilai yang berbeda. Sebagai contoh, atribut rating Google memiliki rentang nilai antara 1 hingga 5, sedangkan jumlah ulasan dapat mencapai ratusan bahkan ribuan. Perbedaan skala tersebut dapat menyebabkan atribut dengan nilai yang lebih besar mendominasi proses perhitungan jarak pada algoritma K-Means.

3.3. Transformation

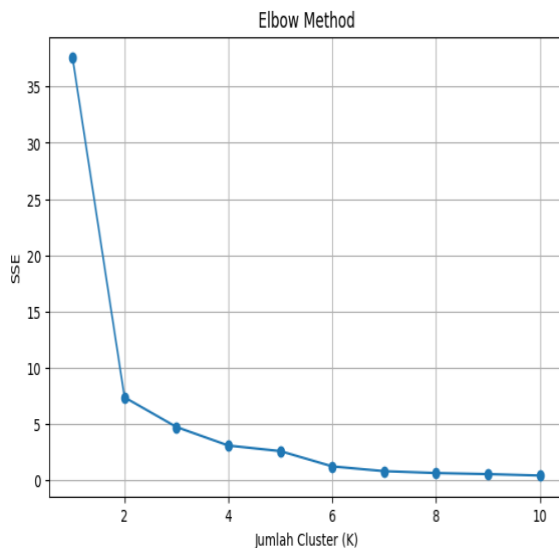
Tahap transformasi data ini berfokus pada konversi data kategorikal menjadi format numerik serta normalisasi nilai atribut menggunakan metode *Min-Max Scaling*. Proses ini mutlak diperlukan karena algoritma *K-Means* beroperasi secara matematis dengan menghitung jarak dalam ruang vektor multidimensi, sehingga seluruh data harus memiliki skala yang seragam untuk mencegah dominasi satu atribut terhadap atribut lainnya. Peneliti melakukan penskalaan pada fitur 'rating' dan 'jumlah_ulasan' ke rentang 0 hingga 1 menggunakan *MinMaxScaler* dari pustaka *Scikit-learn*. Transformasi ini memastikan data siap diproses secara optimal oleh model machine learning untuk menghasilkan segmentasi ke dalam 3 kluster target klien yang terstruktur dan objektif.

3.4. Data Mining

a. Pengujian *Elbow Method*

Pada grafik Elbow, titik yang membentuk pola menyerupai siku (*elbow point*) menunjukkan jumlah cluster yang paling sesuai untuk digunakan. Titik tersebut menandakan bahwa penambahan *cluster* setelah titik tersebut tidak memberikan

penurunan nilai SSE yang signifikan. Berdasarkan hasil pengujian menggunakan metode *Elbow*, diperoleh nilai K yang digunakan sebagai jumlah cluster pada proses K-Means. Nilai tersebut kemudian menjadi dasar dalam proses segmentasi target client pada penelitian ini.



Gambar 32. Grafik Elbow Method

Berdasarkan grafik *Elbow Method*, titik siku terlihat pada nilai K = 3 sehingga jumlah *cluster* yang digunakan dalam penelitian ini adalah 3 cluster.

b. Penentuan *Centroid* Awal

Centroid merupakan titik pusat yang digunakan sebagai acuan dalam proses pengelompokan data. Pada penelitian ini, *centroid* awal dipilih secara acak dari data yang terdapat dalam dataset sesuai dengan jumlah *cluster* yang telah ditentukan sebelumnya.

Tabel 3. Tabel Hasil Penentuan Centroid Awal

Cluster	Rating	Jumlah Ulasan	Website
C0	1	0.008574	1
C1	1	0.010102	1
C2	0	0.001822	0

c. Perhitungan Jarak Data Ke *centroid*

Setelah *centroid* awal ditentukan, proses selanjutnya adalah menghitung jarak antara setiap data dengan seluruh *centroid* yang tersedia. Perhitungan jarak dilakukan menggunakan metode *Euclidean Distance* untuk mengetahui tingkat kedekatan masing-masing data terhadap *centroid*.

Tabel 4. Tabel Perhitungan Jarak Ke Centroid

Pelanggan	Jarak C0	Jarak C1	Jarak C2
1	0.41668	0.41667	1.03083
2	1.08336	1.08337	0.25
3	1.03079	1.0308	0.08334
4	1.20186	1.20187	0.5
5	1.01382	1.01383	0

d. Hasil Iterasi *K-Means*

Proses clustering berakhir ketika kondisi *konvergen* tercapai, yaitu ketika tidak terdapat perubahan signifikan pada posisi *centroid* maupun anggota cluster pada *iterasi* berikutnya. Proses clustering menggunakan algoritma *K-Means* ini mencapai kondisi konvergen dan resmi terhenti pada *iterasi* ke-6.

Tabel 5. Tabel Hasil Dari Iterasi Ke 6

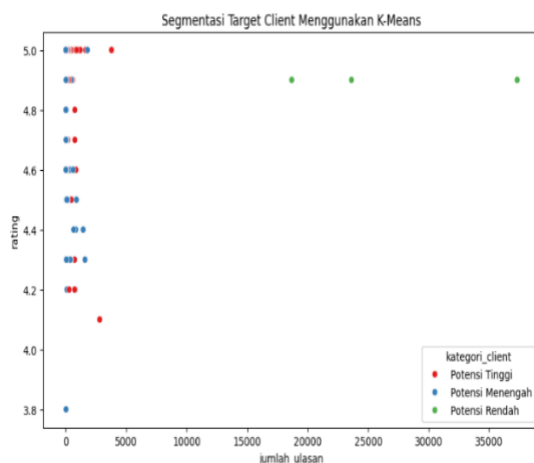
Pelanggan	Jarak C0	Jarak C1	Jarak C2
1	0.349209	0.142241	1.014473
2	1.05926	1.011624	0.170694
3	1.016557	1.022386	0.005219

4	1.165809	1.045843	0.420646
5	1.004949	1.038	0.079473

e. Hasil Pembentukan *Cluster*

Tabel 6. Tabel Hasil Pembentukan Cluster

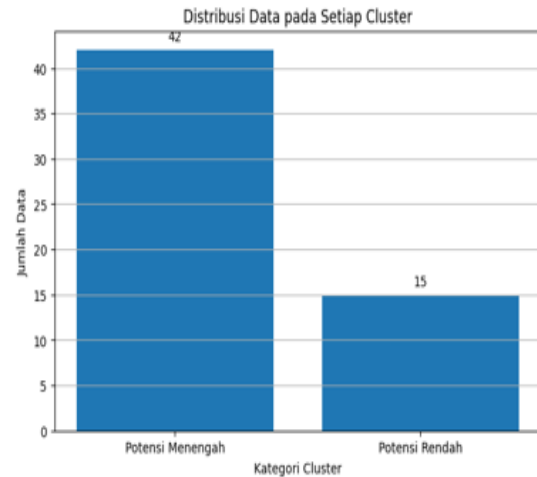
ster	Jumlah Data
ster 0	62
ster 1	54
ster 2	3



Gambar 33. Visualisasi Hasil Pembentukan Cluster

f. Grafik Distribusi *Cluster*

Apabila jumlah data pada suatu *cluster* lebih besar dibandingkan *cluster* lainnya, maka dapat disimpulkan bahwa karakteristik usaha pada kelompok tersebut lebih banyak ditemukan dalam dataset penelitian. Sebaliknya, cluster dengan jumlah anggota yang lebih sedikit menunjukkan kelompok usaha dengan karakteristik yang lebih spesifik.



Gambar 34. Visualisasi Grafik Distribusi Cluster

g. Analisis karakteristik setiap cluster

1. *Cluster 0* Target Potensial Tinggi Dengan Karakteristik: Rating tinggi, Lama Usaha Tinggi, Skala Usaha Menengah, Memiliki Website Aktif.
2. *Cluster 1* Target Potensial menengah dengan Karakteristi: Rating cukup baik, Lama Usaha Sedang, Skala Usaha Kecil.
3. *Cluster 2* Target Potensial Rendah dengan Karakterisitik: Rating Rendah, Usaha baru berkembang, Aktivitas Digital masih terbatas.

3.5. Evaluation

a. Evaluasi *Silhouette Score*

Nilai *Silhouette Score* berkisar antara -1 hingga 1. Nilai yang mendekati 1 menunjukkan bahwa objek tersebut dikelompokkan dengan sangat baik dan memiliki jarak yang jauh dari cluster tetangga. Sebaliknya, nilai yang mendekati 0 menunjukkan bahwa objek berada di perbatasan antar cluster, dan nilai negatif menunjukkan kemungkinan kesalahan penempatan cluster.

Tabel 7. Table Silhouette Score

Cluster	Silhouette Score
2	0.2638
3	0.2821
4	0.2633
5	0.3067
6	0.3255
7	0.3429
8	0.3608
9	0.3810
10	0.3765

Hasil PCA menunjukkan persebaran data berdasarkan hasil *clustering* yang telah dilakukan. Setiap warna merepresentasikan cluster yang berbeda sehingga memudahkan proses interpretasi terhadap pola pengelompokan data.

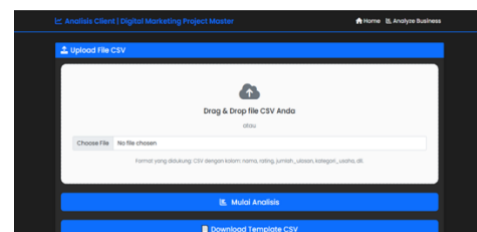
3.6. Knowledge Presentation

Implementasi Prototype sistem

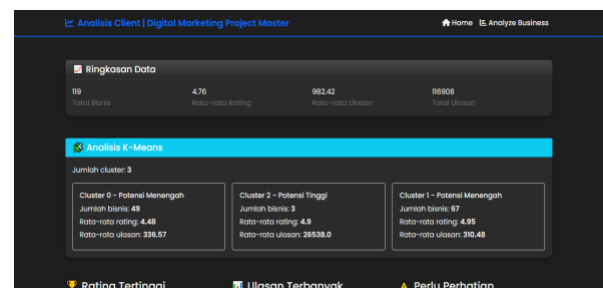
1. Halaman Utama

**Gambar 36. Halaman Utama**

2. Halaman Upload Data Client

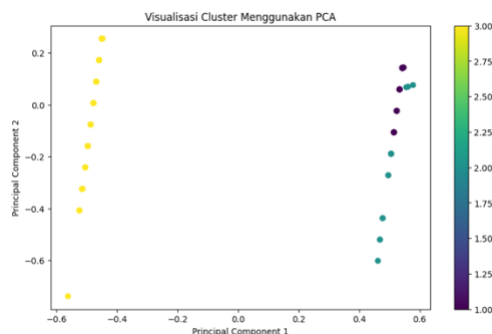
**Gambar 37. Halaman Upload Data Client**

3. Halaman Hasil Segmentasi

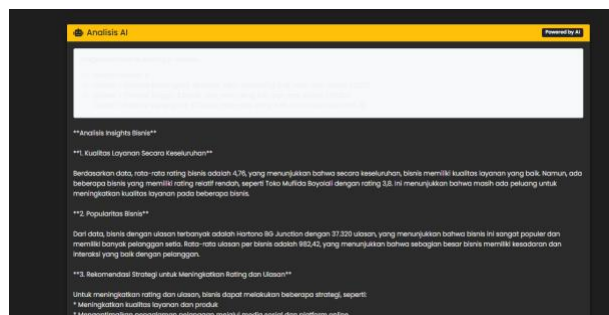
**Gambar 38. Halaman Hasil Segmentasi**

b. Persebaran Data Melalui PCA

Karena proses clustering dilakukan menggunakan lebih dari dua atribut, maka diperlukan teknik reduksi dimensi untuk mempermudah visualisasi data. Pada penelitian ini digunakan *Principal Component Analysis* (PCA) untuk mereduksi atribut menjadi dua komponen utama tanpa menghilangkan sebagian besar informasi yang terkandung dalam data [12].

**Gambar 35. Visualisasi PCA**

4. Halaman hasil analisis oleh AI



Gambar 39. Tampilan Hasil analisis oleh AI

B. Pembahasan Hasil Penelitian

Hasil pengelompokan menunjukkan bahwa metode *K-Means* mampu mengidentifikasi perbedaan karakteristik calon client secara lebih objektif dibandingkan proses seleksi manual yang sebelumnya digunakan oleh perusahaan. Adanya cluster dengan kategori potensi tinggi memungkinkan perusahaan untuk memfokuskan sumber daya pemasaran pada segmen usaha yang memiliki peluang konversi lebih besar.

Selain itu, hasil clustering memperlihatkan bahwa atribut rating Google dan jumlah ulasan memiliki pengaruh yang cukup signifikan dalam membedakan tingkat potensi suatu usaha. Usaha dengan reputasi digital yang baik cenderung berada pada cluster potensi tinggi karena menunjukkan tingkat kepercayaan pelanggan yang lebih besar serta peluang pengembangan bisnis yang lebih baik [13].

Temuan ini sejalan dengan konsep segmentasi pelanggan berbasis data yang menyatakan bahwa pengelompokan berdasarkan karakteristik tertentu dapat membantu perusahaan dalam menentukan strategi pemasaran yang lebih efektif. Dengan adanya sistem segmentasi berbasis *K-Means*, CV. Project Master dapat melakukan pendekatan pemasaran yang lebih terarah, mengurangi waktu pencarian calon client, serta meningkatkan efisiensi proses pengambilan keputusan.

Secara keseluruhan, penerapan metode *K-Means* tidak hanya menghasilkan kelompok data client, tetapi juga memberikan informasi strategis yang dapat digunakan sebagai dasar dalam

menentukan prioritas target pemasaran dan pengembangan layanan perusahaan.

IV. KESIMPULAN

Penelitian ini berhasil menerapkan metode *K-Means Clustering* sebagai pendekatan berbasis kecerdasan buatan untuk membantu proses penentuan target client pada CV. Project Master. Melalui tahapan pengumpulan data, preprocessing, dan proses clustering, data calon client dapat dikelompokkan berdasarkan karakteristik usaha yang dimiliki sehingga menghasilkan segmentasi yang lebih terstruktur dibandingkan metode seleksi manual.

Hasil penelitian menunjukkan bahwa metode *K-Means* mampu mengidentifikasi kelompok client dengan tingkat potensi yang berbeda dari total 120 data calon client yang diolah. Berdasarkan pengujian menggunakan metode *Elbow*, telah ditentukan bahwa jumlah *cluster* paling optimal adalah $K=3$, yang membagi data ke dalam tiga kategori: Potensi Tinggi (*Cluster 0*), Potensi Menengah (*Cluster 1*), dan Potensi Rendah (*Cluster 2*). Kualitas pengelompokan ini divalidasi dengan nilai *Silhouette Score* sebesar 0.2821, yang menunjukkan bahwa struktur cluster yang terbentuk memiliki pemisahan yang cukup signifikan. Secara teknis, algoritma *K-Means* mencapai kondisi stabil atau konvergen pada iterasi ke-6, yang menandakan hasil pengelompokan telah optimal. Dengan hasil ini, perusahaan dapat menentukan prioritas target pemasaran secara lebih objektif dan efisien dibandingkan pendekatan manual. Informasi strategis yang dihasilkan dari proses *clustering* ini dapat digunakan sebagai dasar dalam penyusunan strategi pemasaran yang lebih tepat sasaran serta mendukung pengambilan keputusan berbasis data.

Dengan demikian, tujuan penelitian untuk membangun sistem segmentasi target client menggunakan metode *K-Means* telah tercapai. Penerapan sistem ini diharapkan dapat meningkatkan efektivitas proses pencarian calon

client dan membantu perusahaan dalam mengoptimalkan sumber daya pemasaran yang dimiliki.

V. UCAPAN TERIMA KASIH

Penulis menyampaikan rasa hormat dan terima kasih yang sebesar-besarnya kepada Bapak Parmin dan Ibu Sumarni, selaku orang tua tercinta, yang tiada henti memberikan doa restu, kasih sayang, serta dukungan moril maupun materiil yang menjadi kekuatan utama penulis dalam menyelesaikan studi dan penelitian ini. Ungkapan terima kasih yang tulus juga penulis tujukan kepada, Andini Natasya, atas segala bentuk dukungan moral maupun jasmani, serta kesediaannya meluangkan waktu untuk membantu penulis dalam proses penyusunan dan penyempurnaan artikel penelitian ini hingga selesai. Semoga segala kebaikan dan bantuan yang telah diberikan mendapatkan balasan yang terbaik.

VI. REFERENSI

- [1] N. A. A. Nasution, "Peran Asosiasi Penyelenggara Jasa Internet Indonesia (APJII) Sumatera Utara Dalam Mensosialisasikan Keamanan Berinternet Kepada Masyarakat Menurut Perspektif Komunikasi Islam," *Al-Balagh: Jurnal Komunikasi Islam*, vol. 8, no. 1, pp. 24–35, 2024.
- [2] J. Han, J. Pei, and H. Tong, *Data mining: concepts and techniques*. Morgan kaufmann, 2022.
- [3] A. Singla, A. Sukharevsky, L. Yee, M. Chui, and B. Hall, "The state of AI," *How Organizations are Rewiring to Capture Value*. Publisher: McKinsey, 2025.
- [4] J. Salminen, M. Mustak, M. Sufyan, and B. J. Jansen, "How can algorithms help in segmenting users and customers? A systematic review and research agenda for algorithmic customer segmentation," *Journal of Marketing Analytics*, vol. 11, no. 4, pp. 677–692, 2023.
- [5] S. O. Rezende, R. B. T. Oliveira, L. C. M. Felix, and C. A. J. Rocha, "Visualization for knowledge discovery in database," *WIT Transactions on Information and Communication Technologies*, vol. 22, 2025.
- [6] H. Ivison, M. Zhang, F. Brahman, P. W. Koh, and P. Dasigi, "Large-scale data selection for instruction tuning," *arXiv preprint arXiv:2503.01807*, 2025.
- [7] C. Yan, "A review on spectral data preprocessing techniques for machine learning and quantitative analysis," *iScience*, vol. 28, no. 7, 2025.
- [8] K. Paraskevas and T. Christos, "Data preprocessing and feature engineering for data mining: Techniques, tools, and best practices," *AI*, vol. 6, no. 10, p. 257, 2025.
- [9] E. López-Meneses, P. C. Mellado-Moreno, C. Gallardo Herrerías, and N. Pelicano-Piris, "Educational data mining and predictive modeling in the age of artificial intelligence: An in-depth analysis of research dynamics," *Computers*, vol. 14, no. 2, p. 68, 2025.
- [10] N. A. Shidqi, H. Firmansyah, W. Asriyani, and R. P. Tulodo, "Analisis Segmentasi Pelanggan Berbasis Data Transaksi Menggunakan Metode K-Means," *Journal of Multidisciplinary Inquiry in Science, Technology and Educational Research*, vol. 3, no. 1, pp. 965–976, 2026.
- [11] S. N. Lathifah and Z. F. Azzahra, "AI-Driven Customers Segmentation Using K-Means Clustering," *G-Tech: Jurnal Teknologi Terapan*, vol. 9, no. 1, pp. 320–329, 2025.
- [12] C. Lumbantoruan and D. A. Megawaty, "Penerapan Metode K-Means Clustering dan Principal Component Analysis (PCA) untuk Pengelompokan Provinsi di Indonesia Berdasarkan Indikator Pendidikan," *Journal of Artificial Intelligence and Technology Information (JAITI)*, vol. 4, no. 2, pp. 120–131, 2026.
- [13] M. Alves Gomes and T. Meisen, "A review on customer segmentation methods for personalized customer targeting in e-commerce use cases," *Information Systems and e-Business Management*, vol. 21, no. 3, pp. 527–570, 2023.