

# IMPLEMENTATION OF MACHINE LEARNING IN BUSINESS INTELLIGENCE FOR CUSTOMER SEGMENTATION AND LOYALTY AT PT. INTI GROUP

## IMPLEMENTASI MACHINE LEARNING PADA BISNIS INTELEGENSI UNTUK SEGMENTASI DAN LOYALITAS PELANGGAN PT. INTI GROUP

Galuh Pandu Siwi Ambarsari<sup>1</sup>, Ichsan Ibrahim<sup>2</sup>

<sup>1,2</sup>Teknik Informatika, STMIK Indonesia Mandiri Bandung

*e-mail: galuhpandusiwi@gmail.com<sup>1</sup>, ichsanibrahim@gmail.com<sup>2</sup>*

**Abstract-** *This study addresses the need for integrated data analytics and machine learning in PT Inti Group's BI dashboard by implementing an unsupervised K-Means clustering method on historical training data (January 2021–May 2025) extracted directly from a PostgreSQL database and analyzed using Python. The analysis process includes data preprocessing and feature engineering to create key variables: number of participants, training-type frequency, recency (days since the last training), and engagement duration. Cluster determination was evaluated using the Elbow method (4 clusters), Silhouette score (2 clusters), and Davies–Bouldin index (9 clusters). Based on business interpretation and the balance between cluster compactness and separation, four clusters were selected: Loyal & High-Value Customers, Inactive, Growing/Potential, and New/Sporadic. Customers who attended training more than ten times were classified as loyal. The segmentation results are visualized in a Power BI dashboard integrated directly with the data source, supporting rapid data-driven managerial decisions. This study demonstrates that integrating unsupervised learning with BI effectively enhances understanding of customer characteristics and serves as a basis for designing more targeted marketing strategies. A limitation of this study is that the data cover only up to May 2025.*

*Keywords: Machine Learning, Business Intelligence, Customer Segmentation, Customer Loyalty*

**Abstrak** - Penelitian ini mengisi kebutuhan integrasi analitik data dan machine learning pada dashboard BI PT Inti Group dengan mengimplementasikan metode unsupervised learning K-Means Clustering pada data historis pelatihan (Januari 2021–Mei 2025) yang diambil langsung dari database PostgreSQL dan dianalisis menggunakan Python. Proses analisis meliputi preprocessing dan feature engineering untuk membentuk variabel kunci: jumlah peserta, frekuensi jenis pelatihan, recency (jarak hari sejak pelatihan terakhir), dan durasi keterlibatan. Penentuan jumlah kluster dievaluasi melalui Metode Elbow (4 klaster), Silhouette Score (2 klaster), dan Davies-Bouldin Index (9 klaster). Berdasarkan interpretasi bisnis serta keseimbangan antara kekompakan dan pemisahan, dipilih empat klaster: Pelanggan Loyal & Bernilai Tinggi, Tidak Aktif, Berkembang/Potensial, dan Baru/Jarang. Kategori loyal ditetapkan bagi pelanggan dengan frekuensi pelatihan lebih dari sepuluh kali. Hasil segmentasi divisualisasikan dalam dashboard Power BI yang terintegrasi langsung dengan sumber data, sehingga mendukung pengambilan keputusan manajerial berbasis data secara cepat. Penelitian ini menunjukkan bahwa integrasi unsupervised learning dengan BI efektif meningkatkan pemahaman karakteristik pelanggan dan menjadi dasar perancangan strategi pemasaran yang lebih terarah. Keterbatasan penelitian mencakup cakupan data hingga Mei 2025. Penelitian ini memberikan kontribusi.

**Kata Kunci** : Machine Learning, Bisnis Inteligensi, Segmentasi Pelanggan, Loyalitas Pelanggan

## I. PENDAHULUAN

PT. Inti Group merupakan perusahaan penyelenggara pelatihan profesional yang telah memiliki ribuan data pelanggan dari berbagai sektor industri. Namun, data historis pelanggan tersebut selama ini belum dimanfaatkan secara optimal untuk mendukung strategi segmentasi dan retensi pelanggan. Proses pengelompokan pelanggan masih dilakukan secara manual dan berdasarkan intuisi, tanpa dasar kuantitatif yang terstruktur. Akibatnya, perusahaan kesulitan dalam merancang program pemasaran yang tepat sasaran dan dalam mengidentifikasi kelompok pelanggan yang bernilai tinggi maupun yang berisiko churn. Business Intelligence (BI) merupakan pendekatan sistematis untuk mengubah data mentah menjadi informasi yang bernilai guna mendukung pengambilan keputusan strategis. Dengan mengintegrasikan machine learning ke dalam kerangka BI, perusahaan dapat memperoleh wawasan yang lebih dalam mengenai perilaku pelanggan. Salah satu teknik yang umum digunakan adalah algoritma K-Means Clustering, yaitu metode unsupervised learning yang mampu mengelompokkan data berdasarkan kemiripan atribut secara otomatis[1],[2]. Penggunaan algoritma ini dalam segmentasi pelanggan telah menunjukkan efektivitasnya dalam berbagai sektor, termasuk e-commerce, perbankan, dan kini mulai diterapkan dalam industri pelatihan dan pengembangan SDM. Penggabungan metode ini dengan visualisasi dashboard seperti Power BI memberikan keunggulan dalam hal interpretasi hasil dan implementasi strategi bisnis secara cepat[3].

Permasalahan utama dalam penelitian ini adalah belum adanya sistem otomatis yang mengintegrasikan segmentasi pelanggan berbasis machine learning ke dalam dashboard BI di lingkungan PT. Inti Group. Belum tersedia acuan berbasis data untuk mengklasifikasikan loyalitas pelanggan maupun per segmen. Selain itu, belum pernah dilakukan evaluasi empiris mengenai efektivitas metode clustering yang paling sesuai terhadap karakteristik data pelanggan PT. Inti Group. Penelitian ini bertujuan untuk mengimplementasikan metode unsupervised machine learning dalam sistem Business Intelligence guna mengelompokkan pelanggan dan menganalisis loyalitas mereka terhadap program pelatihan. Fokus utama mencakup preprocessing data historis, pembentukan fitur segmentasi, pemodelan klaster, penilaian loyalitas, serta visualisasi hasil dalam dashboard Power BI terintegrasi. Dengan pendekatan ini, diharapkan perusahaan dapat merancang strategi pemasaran dan retensi pelanggan secara lebih presisi dan berbasis data.

## II. SIGNIFIKANSI STUDI

### A. Studi Literatur

Penelitian ini berakar pada berbagai literatur yang mendukung penerapan unsupervised machine learning dalam sistem Business Intelligence (BI) untuk segmentasi dan analisis loyalitas pelanggan. Business Intelligence didefinisikan sebagai integrasi proses, arsitektur, dan teknologi untuk mengubah data menjadi informasi bermakna yang mendukung pengambilan keputusan strategis[4]. Visualisasi interaktif berbasis data juga semakin penting dalam mempercepat keputusan manajerial[5]. Segmentasi pelanggan, adalah strategi pengelompokan berdasarkan kesamaan karakteristik perilaku untuk meningkatkan efektivitas pemasaran. Dalam pendekatan machine learning, metode unsupervised seperti K-Means Clustering menjadi populer karena kemampuannya untuk memetakan perilaku tanpa memerlukan label data historis[6]. Loyalitas pelanggan umumnya diukur dari konsistensi pembelian dan frekuensi keterlibatan[7].

## B. Gap Penelitian

Hingga saat ini, belum ditemukan implementasi sistematis yang mengintegrasikan model segmentasi pelanggan berbasis machine learning ke dalam dashboard Business Intelligence di industri pelatihan, khususnya di lingkungan PT. Inti Group. Sebagian besar penelitian serupa masih didominasi oleh studi di sektor e-commerce dan finansial[8],[9], sementara sektor pelatihan belum banyak dieksplorasi.

Oleh karena itu, penelitian ini berkontribusi dengan:

1. Menyediakan sistem segmentasi otomatis berbasis data aktual pelanggan PT. Inti Group, yang sebelumnya hanya dikelompokkan secara manual.
2. Mengintegrasikan algoritma K-Means Clustering ke dalam pipeline BI secara langsung, dari PostgreSQL hingga visualisasi Power BI.
3. Memberikan studi kasus empirik pada industri pelatihan, memperluas aplikasi unsupervised learning dalam domain non-komersial.

## C. Data Penelitian

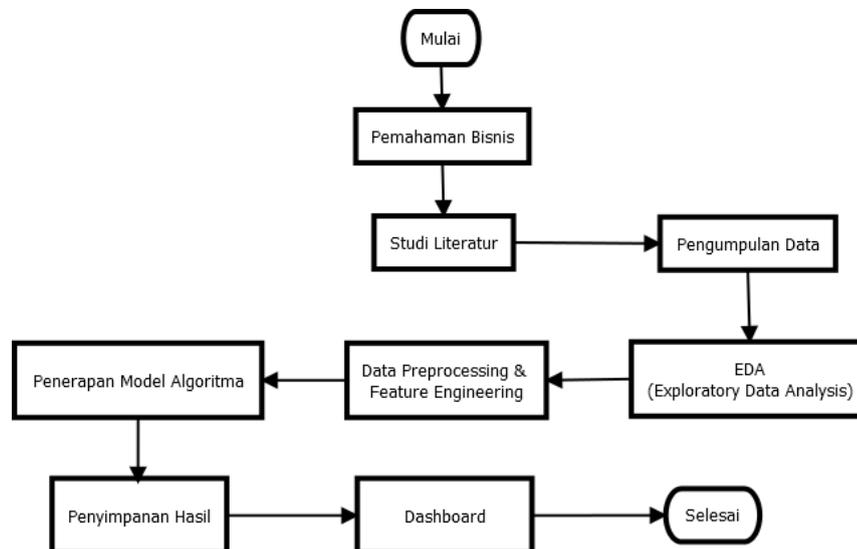
Penelitian ini dilakukan dengan memanfaatkan data historis yang diperoleh dari sistem internal PT. Inti Group, yang disimpan dalam RDBMS PostgreSQL. Dataset terdiri atas informasi peserta pelatihan, jenis program, frekuensi pelatihan, dan waktu keterlibatan, dengan cakupan periode Januari 2021 hingga Mei 2025. Total entri pelanggan korporat mencakup lebih dari 500 perusahaan unik. Data diambil langsung menggunakan bahasa Python dan pustaka *psycopg2*, kemudian diproses dalam format *pandas*. DataFrame untuk analisis eksploratif, pembentukan fitur, serta pemodelan machine learning [16].

## D. Metode Penelitian dan Pemilihan Model

1. *Exploratory Data Analysis (EDA)*, Tahap EDA bertujuan memahami struktur data[10], mendeteksi pola, dan mengidentifikasi anomali. Hasil eksplorasi ini menjadi dasar pembentukan fitur-fitur seperti recency, frequency, variety, dan durasi keterlibatan.
2. *Data Preprocessing dan Feature Engineering*, Data dibersihkan, dikonversi menjadi numerik, dan diformulasikan ulang ke dalam format agregat per pelanggan[11]. Fitur-fitur tersebut dinormalisasi untuk menghindari dominasi variabel tertentu saat dilakukan clustering.
3. *Pemilihan dan Evaluasi Model Machine Learning*, Penelitian ini menggunakan algoritma K-Means Clustering[12] karena:
  - Cocok untuk data numerik dan tidak berlabel.
  - Hasil kluster mudah diinterpretasikan secara bisnis dan visual.
  - Tidak menandai pelanggan bernilai tinggi sebagai outlier, seperti yang terjadi pada DBSCAN. Pada DBSCAN mengklasifikasikan kelompok pelanggan bernilai tinggi (loyal) sebagai outlier, sehingga menyulitkan penargetan bisnis.
  - Lebih efisien secara komputasi dan skalabel dibanding Hierarchical Clustering. Dengan Hierarchical Clustering bisa lebih lambat untuk data yang semakin besar.
  - Evaluasi model dilakukan menggunakan tiga metrik: Elbow (Inertia), Silhouette Score, dan Davies–Bouldin Index. Dari hasil uji, konfigurasi 4 kluster (Metode Elbow) dipilih karena menghasilkan pemisahan terbaik yang juga bermakna secara strategis dan operasional[13].

#### 4. Penyimpanan dan Visualisasi

Hasil segmentasi disimpan kembali dalam PostgreSQL dan divisualisasikan ke dalam dashboard Power BI interaktif. Pendekatan ini mendukung pengambilan keputusan real-time oleh manajemen PT. Inti Group.



Gambar 1. Proses Tahapan Penelitian

Proyek dimulai dengan penetapan tujuan segmentasi dan definisi loyalitas, lalu studi literatur untuk memilih metode RFM dan K-Means. Data pelatihan (Jan 2021–Mei 2025) diekstrak dari PostgreSQL dan dieksplorasi untuk mengidentifikasi outlier dan pola awal. Setelah preprocessing dan feature engineering (jumlah peserta, frekuensi, recency, durasi), K-Means dijalankan dan dievaluasi (Elbow, Silhouette, Davies–Bouldin), lalu dipilih empat kluster berdasarkan keseimbangan kekompakan dan pemisahan. Hasil segmentation dan metrik evaluasi disimpan ke data warehouse dan divisualisasikan di dashboard Power BI terintegrasi, mendukung keputusan pemasaran berbasis data secara real time.

### III. HASIL DAN PEMBAHASAN

#### A. Pengumpulan Data

Pengumpulan data dalam penelitian ini dilakukan secara langsung dari sistem database PostgreSQL milik PT. Inti Group yang menyimpan histori pelaksanaan pelatihan peserta dari berbagai perusahaan klien. Data yang diambil merupakan data primer dari sistem manajemen pelatihan perusahaan secara operasional.

Data ditarik menggunakan bahasa pemrograman Python dengan bantuan library `psycopg2`, yang memungkinkan koneksi langsung ke sistem basis data relasional. Proses pengambilan data dilakukan sebagai berikut:

- 1) Koneksi Database

Dibuat koneksi aman menggunakan parameter seperti host, user, password, port, dan nama database.

- 2) Eksekusi Query SQL

Perintah SQL (SELECT \* FROM data\_pelatihan\_training) dijalankan untuk mengekstrak semua catatan pelatihan yang tersedia.

3) Pemuatan ke DataFrame

Data yang diperoleh dari database langsung dimuat ke dalam objek pandas.DataFrame untuk selanjutnya dianalisis di lingkungan Python (Google Colaboratory).

4) Integrasi Langsung ke Analisis

Seluruh proses preprocessing dan segmentasi dilakukan langsung dari hasil query tanpa menyimpan atau mengekspor ke file eksternal.

Data mentah yang dikumpulkan terdiri dari beberapa atribut penting:

- nama\_peserta: Nama lengkap peserta pelatihan
- perusahaan\_pengirim: Instansi atau perusahaan asal peserta
- jenis\_pelatihan: Nama kategori atau topik pelatihan
- tanggal\_pelatihan: Tanggal pelaksanaan pelatihan
- lokasi\_pelatihan: Tempat penyelenggaraan pelatihan

## B. Data Preprocessing dan Feature Engineering

Sebelum dilakukan Data Preprocessing dan Feature Engineering pada dataset, dilakukan dulu EDA yang bertujuan untuk memahami struktur data, mengidentifikasi pola, mendeteksi anomali, serta merumuskan strategi preprocessing dan pemodelan yang tepat[14]. Setelah dilakukan EDA, selanjutnya dilakukan Data Preprocessing dan Feature Engineering yang bertujuan untuk mempersiapkan data agar sesuai untuk proses modeling. Langkah-langkah yang dilakukan meliputi:

1) Pembersihan Data

- Menghapus Duplikasi  
Baris yang sama persis dihapus agar tidak memengaruhi hasil analisis.
- Penanganan Missing Values  
Nilai kosong pada kolom penting dihapus.
- Normalisasi Format Tanggal  
Seluruh data tanggal dikonversi ke format datetime.

2) Pembuatan Fitur (Feature Engineering)

- jumlah\_peserta: Jumlah total peserta dari masing-masing perusahaan.
- jenis\_pelatihan\_unik: Jumlah jenis pelatihan berbeda yang diikuti.
- pelatihan\_terakhir dan pelatihan\_pertama: Digunakan untuk menghitung recency dan lama langganan.
- recency\_days: Jumlah hari sejak pelatihan terakhir hingga tanggal dilakukan analisa
- lama\_langganan: Selisih tahun antara pelatihan pertama dan terakhir.

Fitur-fitur ini digunakan sebagai input dalam proses K-Means Clustering yang akan dilakukan pada tahap berikutnya.

### B. Penerapan Model Algoritma

**DBSCAN** adalah algoritma berbasis kepadatan yang mengelompokkan titik-titik padat dan menandai titik di area berdensitas rendah sebagai outlier (segmen -1). Berikut Interpretasi Segmen DBSCAN dari analisa :

- Segmen -1 (Outlier/Noise):

Pada plot, titik-titik ini adalah yang tidak termasuk dalam kluster padat manapun. Mereka tersebar di sekitar plot atau berada di area yang berdensitas rendah. Berdasarkan analisis deskriptif, meskipun outlier, titik-titik ini memiliki rata-rata jumlah\_peserta, jenis\_pelatihan\_unik, dan lama\_langganan yang cukup tinggi dibandingkan dengan sebagian besar kluster lainnya.

- Segmen 0 (DBSCAN):

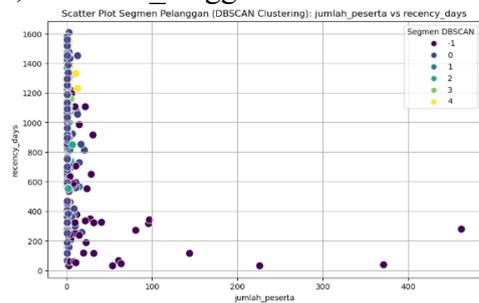
Pada plot, ini adalah salah satu kluster padat yang ditemukan. Dapat dilihat kumpulan titik-titik dengan warna yang sama. Berdasarkan analisis deskriptif, segmen ini dicirikan oleh jumlah\_peserta sangat rendah, jenis\_pelatihan\_unik rendah (hanya 1), recency\_days sangat tinggi, dan lama\_langganan sangat rendah. Ini adalah kluster utama untuk Pelanggan Tidak Aktif dengan Interaksi Singkat.

- Segmen 1, 2, 3, 4 (DBSCAN):

Pada plot, ini adalah kluster padat lainnya yang ditemukan oleh DBSCAN. Masing-masing akan muncul sebagai kumpulan titik-titik dengan warna yang berbeda. Berdasarkan analisis deskriptif, kluster-kluster ini cenderung memecah kelompok pelanggan yang kurang aktif atau jarang berinteraksi menjadi sub-kelompok yang lebih spesifik berdasarkan nuansa dalam jumlah\_peserta, jenis\_pelatihan\_unik, recency\_days, dan lama\_langganan.

segment_dbscan	jumlah_peserta	jenis_pelatihan_unik	recency_days	lama_langganan	loyal_customer
-1	40.830189	2.924528	374.188879	844.924528	0.566038
0	1.792453	1.000000	896.803774	21.912736	0.033019
1	4.272727	2.000000	721.545455	704.181818	0.090909
2	2.705882	2.000000	881.529412	234.294118	0.000000
3	3.800000	3.000000	1134.800000	230.400000	0.000000
4	5.714288	2.000000	1379.571429	53.428571	0.285714

Gambar 2. Hasil Analisa DBSCAN

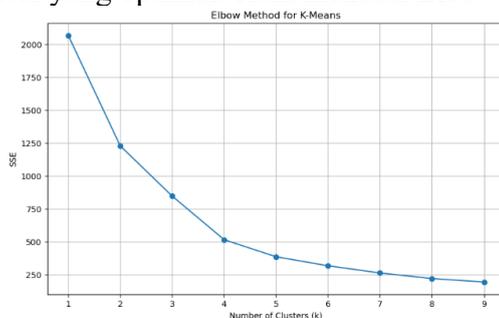


Gambar 3. Scatter Plot (DBSCAN Clustering)

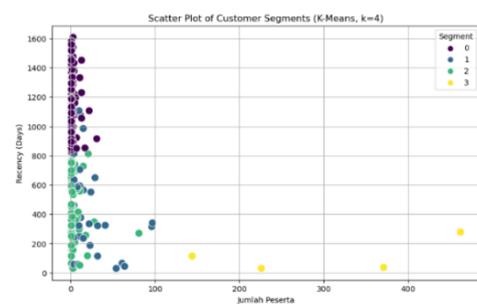
DBSCAN berhasil membedakan outlier pelanggan super-aktif (kluster -1) dari kelompok pelanggan reguler dan jarang, serta mengidentifikasi segmen baru/potensial (kluster 4) yang memiliki waktu “baru” sejak pendaftaran dan frekuensi pelatihan menengah. Visualisasi ini membantu tim pemasaran menentukan strategi retensi bagi kluster aktif tua (2 & 3), re-engagement untuk kluster sangat jarang (0), dan upsell untuk kluster potensial (4).

**K-Means** adalah algoritma partisional yang membagi data menjadi kluster, di mana setiap titik data termasuk dalam kluster dengan pusat (centroid) terdekat[15]. Penentuan jumlah kluster (k) yang optimal sangat penting dalam K-Means. Pada kasus ini menggunakan tiga metode untuk mengevaluasi jumlah kluster: Metode Elbow, Silhouette Score, dan Davies-Bouldin Index.

**Metode Elbow** melihat plot nilai SSE (jumlah kuadrat jarak setiap titik ke pusat klasternya) sebagai fungsi dari jumlah kluster k. Penurunan SSE akan signifikan pada awalnya seiring penambahan kluster, tetapi pada titik tertentu, penambahan kluster tidak lagi memberikan penurunan SSE yang substansial, membentuk "siku" atau "elbow" pada grafik dimana titik siku ini dianggap sebagai jumlah kluster yang optimal. Rekomendasi Elbow k = 4.



Gambar 4. Grafik Elbow



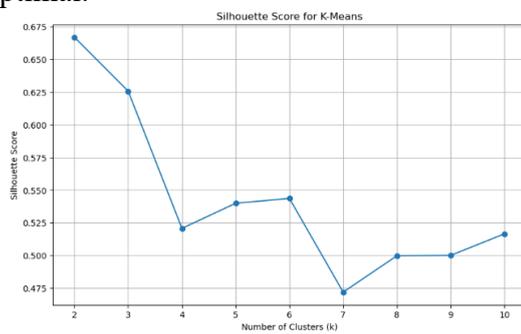
Gambar 5. Scatter Plot (K-Means k=4)

segment	jumlah_peserta	jenis_pelatihan_unik	recency_days	lama_langganan	loyal_customer	
0	0.0	1.959707	1.102564	1198.485201	21.809524	0.032967
1	1.0	16.808511	2.893617	464.617021	912.127860	0.446809
2	2.0	2.839378	1.072539	458.181347	68.005181	0.087358
3	3.0	300.750000	5.000000	114.750000	1344.500000	1.000000

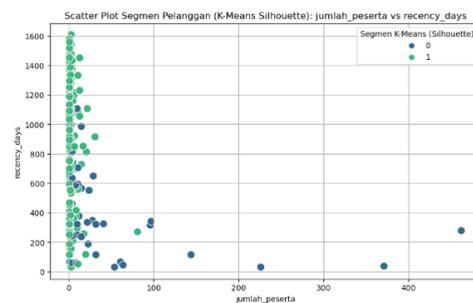
Gambar 6. Hasil Analisa K-Means (k=4)

Berdasarkan grafik Elbow Method yang menunjukkan titik belok pada k=4, analisis K-Means kemudian menghasilkan empat kluster pelanggan dengan profil berbeda: Segmen 0 (inaktif) hanya mengikuti 1–2 pelatihan, recency ~1.200 hari, langganan singkat, dan loyalitas hampir nol; Segmen 1 (loyal & bernilai tinggi) mengikuti rata-rata 17 pelatihan, recency ~465 hari, berlangganan ~912 hari, dengan skor loyalitas ~0,45; Segmen 2 (baru/potensial rendah) berpartisipasi 2–3 kali, recency ~458 hari, langganan ~68 hari, loyalitas rendah; dan Segmen 3 (super-customer/outlier) ekstrem dengan >300 pelatihan, recency ~115 hari, langganan ~1.345 hari, serta loyalitas sempurna. Scatter plot “jumlah\_peserta vs. recency\_days” mengonfirmasi pemisahan ini: kluster super-aktif terpecah di kanan bawah, kluster loyal di tengah bawah, kluster potensial di kiri bawah, dan kluster inaktif di kiri atas.

**Metode Silhouette** mengukur seberapa mirip setiap titik data dengan klusternya sendiri dibandingkan dengan kluster terdekat lainnya. Skor berkisar dari -1 (kluster yang buruk) hingga +1 (kluster yang sangat baik), dengan skor 0 menunjukkan tumpang tindih antar kluster. Rata-rata Silhouette Score dihitung untuk setiap jumlah kluster k, dan nilai k dengan skor rata-rata tertinggi dianggap optimal.



Gambar 7. Grafik Silhouette



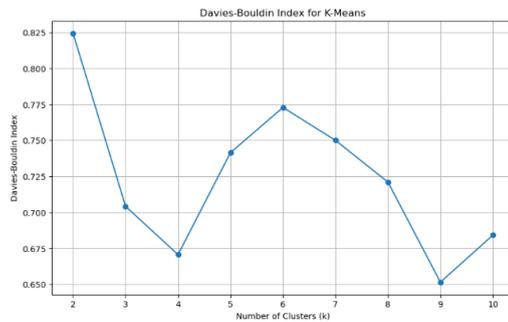
Gambar 8. Scatter Plot (K-Means k=2)

segment_kmeans_silhouette	jumlah_peserta	jenis_pelatihan_unik	recency_days	lama_langganan	loyal_customer
0	40.320000	3.080000	416.580000	949.040000	0.520000
1	2.289807	1.092077	893.098501	42.558887	0.044968

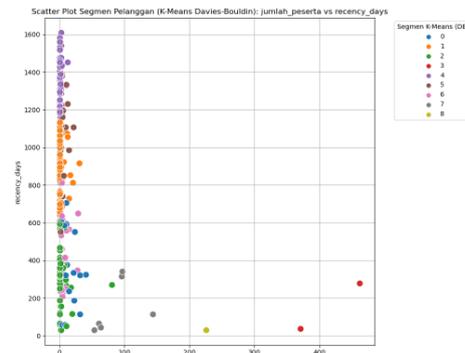
Gambar 9. Hasil Analisa K-Means (k=2)

Berdasarkan Grafik Silhouette (Gambar 7) yang menunjukkan skor tertinggi (~0,67) pada k=2, analisis K-Means dengan dua kluster (Gambar 9) membagi pelanggan menjadi: Segmen 0 yang sangat bernilai (rata-rata ~40 kali ikut pelatihan, recency ~416 hari, lama langganan ~949 hari, loyalitas ~0,52) dan Segmen 1 yang jarang/baru (rata-rata ~2 kali ikut, recency ~893 hari, langganan ~43 hari, loyalitas ~0,045); scatter plot (Gambar 8) memperjelas pemisahan ini dengan titik segmen 0 tersebar di kanan (jumlah peserta tinggi, recency menengah) dan titik segmen 1 berkumpul di kiri atas (jumlah peserta rendah, recency tinggi), menunjukkan bahwa k=2 memberikan klusterisasi paling compact dan jelas menurut metrik silhouette.

**Davies-Bouldin Index** mengukur rasio rata-rata antara dispersi dalam kluster (seberapa tersebar titik-titik di dalam kluster) dan pemisahan antar kluster. Nilai Davies-Bouldin Index yang lebih rendah menunjukkan clustering yang lebih baik, di mana kluster lebih padat (dispersi rendah) dan terpisah dengan baik (pemisahan tinggi). Nilai k dengan indeks terendah sering dianggap optimal. Davies-Bouldin Index mencapai nilai terendah pada  $k = 9$ . Rekomendasi Davies-Bouldin Index:  $k = 9$



Gambar 10. Grafik Davies-Bouldin



Gambar 11. Scatter Plot (K-Means k=9)

segment_kmeans_dbSCAN	jumlah_peserta	jenis_pelatihan_unik	recency_days	lama_langganan	loyal_customer
0	14.368421	3.368421	369.631579	970.157895	0.578947
1	1.939759	1.000000	910.210943	16.174699	0.036145
2	2.788618	1.024390	340.975610	40.243902	0.065041
3	416.500000	5.000000	157.500000	1279.000000	1.000000
4	1.510949	1.000000	1364.642336	4.328467	0.014599
5	4.843750	2.218750	1056.156250	204.625000	0.156250
6	6.483671	1.806452	511.000000	706.129032	0.193548
7	86.000000	5.166667	151.500000	1382.500000	1.000000
8	226.000000	3.000000	30.000000	1375.000000	1.000000

Gambar 12. Hasil Analisa K-Means (k=9)

Berdasarkan Grafik Davies–Bouldin (Gambar 10), nilai indeks mencapai titik terendah pada  $k = 9$ , sehingga K-Means dengan sembilan kluster (Gambar 12) dipilih sebagai konfigurasi optimal; kluster tersebut mencakup misalnya Segmen 3 (outlier super-user dengan rata-rata ~416 peserta, loyalitas = 1) dan Segmen 6 (high-value user dengan ~86 peserta, loyalitas = 1), Segmen 0 (mid-tier dengan ~14 peserta, loyalitas  $\approx 0,58$ ), serta beberapa kluster dengan partisipasi sangat rendah (Segmen 1, 2, 4, 7, 8: <3 peserta, recency tinggi) hingga kluster menengah (Segmen 5 dengan ~5 peserta, loyalitas 0,16); scatter plot (Gambar 11) memvisualisasikan pemisahan ini—titik kluster super-aktif terpecah di kanan bawah (jumlah peserta tinggi, recency rendah) sedangkan kluster inaktif terkonsentrasi di kiri atas (jumlah rendah, recency tinggi)—menegaskan bahwa  $k = 9$  memberikan segmentasi yang paling compact dan terpisah dengan baik untuk mendukung strategi pemasaran dan retensi pelanggan.

Dari metode K-Means (K=4) berikut karakteristik masing-masing segmen pelanggan :

1. Segmen 0 (Tidak Aktif)

- Karakteristik Utama: Partisipasi sangat rendah, recency sangat tinggi (lama tidak aktif), durasi langganan singkat.
- Konsekuensi Bisnis Kritis: Segmen ini mewakili pelanggan yang berisiko tinggi churn atau sudah tidak aktif. Mereka menyumbang sedikit nilai saat ini.
- Implikasi Strategi: Membutuhkan strategi reaktivasi yang agresif. Kampanye pemasaran ulang dengan penawaran menarik (diskon, pelatihan gratis/demo), komunikasi yang menyoroti nilai pelatihan terbaru, atau survei untuk memahami alasan

ketidakaktifan. Risiko investasi pemasaran tinggi karena potensi respons rendah.

## 2. Segmen 1 (Berkembang/Potensial)

- Karakteristik Utama: Partisipasi sedang (jumlah peserta/variasi), recency sedang, durasi langganan lama, loyalitas moderat.
- Konsekuensi Bisnis Kritis: Ini adalah segmen berharga dengan potensi pertumbuhan dan retensi karena pelanggan sudah menunjukkan komitmen.
- Implikasi Strategi: Fokus pada pengembangan hubungan dan peningkatan nilai pelanggan (upsell/cross-sell), penawaran pelatihan lanjutan, program loyalitas, komunikasi personalisasi, dan dukungan premium, serta menjaga kepuasan untuk memastikan retensi jangka panjang.

## 3. Segmen 2 (Baru/Jarang)

- Karakteristik Utama: Partisipasi rendah, recency sedang (relatif baru berinteraksi), durasi langganan singkat.
- Konsekuensi Bisnis Kritis: Segmen ini adalah titik awal bagi pelanggan baru. Risiko churn dini tinggi jika tidak dikonversi.
- Implikasi Strategi: Fokus pada akuisisi dan konversi menjadi pelanggan berulang, kampanye onboarding, penawaran paket pelatihan awal, komunikasi rutin tentang jadwal dan manfaat pelatihan, serta pengumpulan feedback awal.

## 4. Segmen 3 (Loyal & Bernilai Tinggi / VIP)

- Karakteristik Utama: Partisipasi sangat tinggi (banyak peserta/variasi), recency rendah (aktif baru-baru ini), durasi langganan sangat lama, 100% loyal.
- Konsekuensi Bisnis Kritis: Segmen ini adalah aset paling berharga perusahaan. Kontribusi pendapatan dan potensi advocacy sangat tinggi. Jumlah perusahaan yang loyal sangat kecil (4 perusahaan) sehingga sangat penting untuk dijaga.
- Implikasi Strategi: Membutuhkan manajemen hubungan prioritas tinggi (Key Account Management), layanan yang sangat personalisasi, pelatihan yang custom, akses eksklusif, dan memastikan kepuasan maksimal untuk menjaga loyalitas jangka panjang.

Dalam menentukan jumlah kluster optimal untuk model K-Means, evaluasi dilakukan menggunakan Metode Elbow, Silhouette Score, dan Davies-Bouldin Index. Meskipun terdapat keterbatasan inheren pada algoritma K-Means dalam menangani data dengan ukuran kluster yang tidak seimbang dan metrik validasi internal memberikan rekomendasi yang berbeda ( $k=2$  dari Silhouette,  $k=9$  dari Davies-Bouldin Index), pada penelitian ini memilih memakai 4 kluster ( $k=4$ ) karena keputusan ini mencerminkan keseimbangan optimal antara bukti statistik, terutama dari Metode Elbow, dan yang terpenting, relevansi serta kemudahan interpretasi segmen dari perspektif bisnis operasional yang secara langsung dapat diterjemahkan menjadi strategi pemasaran, penjualan, dan retensi yang spesifik dan dapat ditindaklanjuti.

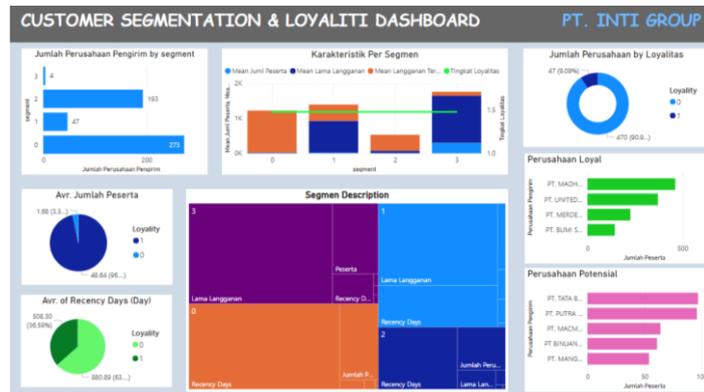
### C. Penyimpanan hasil

Setelah proses segmentasi selesai dilakukan, seluruh hasil analisis disimpan kembali ke dalam sistem basis data PostgreSQL agar dapat diakses oleh tim manajemen atau diintegrasikan dengan sistem informasi lain yang digunakan perusahaan. Penyimpanan dilakukan melalui koneksi langsung dari Python ke database.

## D. Pembuatan Dashboard BI

Hasil segmentasi dan analisis loyalitas divisualisasikan dalam bentuk dashboard interaktif di Power BI. Dashboard ini menjadi alat bantu manajerial untuk:

- Menyusun strategi penawaran berbeda antar segmen
- Mengidentifikasi pelanggan yang harus direaktivasi
- Mengalokasikan anggaran pemasaran secara lebih tepat sasaran



Gambar 13. Tampilan Dashboard Power BI

Dashboard “Customer Segmentation & Loyalty” untuk PT. Inti Group yang menampilkan beberapa visualisasi utama bar chart jumlah perusahaan per segmen, radar chart karakteristik per segmen, ring pie untuk proporsi loyal vs tidak loyal, treemap deskripsi segmen, dan bar chart vertikal untuk daftar perusahaan loyal dan potensial.

## IV. KESIMPULAN

Penelitian ini berhasil mengembangkan dan mengimplementasikan sistem segmentasi pelanggan berbasis K-Means Clustering yang terintegrasi langsung ke dalam dashboard Business Intelligence PT Inti Group, guna mendukung pengambilan keputusan data-driven secara real time. Berdasarkan evaluasi menggunakan Elbow Method, Silhouette Score, dan Davies–Bouldin Index, konfigurasi **empat klaster** ( $k=4$ ) dipilih sebagai yang paling optimal, karena memadukan kekompakan internal, pemisahan antar klaster, serta relevansi terhadap kebutuhan strategi bisnis. Keempat segmen yang diidentifikasi beserta proporsinya adalah:

1. **Tidak Aktif (36,6%)**, Aktivitas rendah, recency tinggi, durasi keterlibatan singkat; berisiko churn tinggi perlu program reaktivasi agresif
2. **Berkembang/Potensial (20,3%)**, Partisipasi menengah, berlangganan jangka panjang; berpotensi ditingkatkan melalui upsell, bundling, dan loyalty program.
3. **Baru/Jarang (40,2%)**, Pelanggan baru dengan keterlibatan terbatas; memerlukan program onboarding, edukasi produk, dan penawaran awal agar terikat lebih kuat.
4. **Loyal & Bernilai Tinggi (2,9%)**, Riwayat pelatihan intensif, recency terkini, dan skor loyalitas 100%.

Dengan dasar segmentasi ini, manajemen dapat merancang intervensi dan alokasi sumber daya secara tertarget: re-engagement untuk segmen inaktif, penguatan relasi pada segmen potensial,

onboarding efektif bagi segmen baru, serta penghargaan eksklusif bagi segmen loyal. Integrasi hasil clustering ke dalam dashboard Power BI mempercepat monitoring dinamika pelanggan dan evaluasi efektivitas strategi secara berkelanjutan. Kontribusi penelitian terletak pada demonstrasi aplikatif machine learning dalam industri pelatihan, di mana metode K-Means terbukti lebih sesuai dibanding DBSCAN atau hierarchical clustering untuk data PT Inti Group. Keterbatasan studi ini meliputi cakupan data hanya hingga Mei 2025 dan penggunaan variabel numerik tanpa atribut demografis/psikografis.

## REFERENSI

- [1] Han J, Pei J, Kamber M. *Data Mining: Concepts and Techniques*. 3rd ed. Cambridge: Morgan Kaufmann; 2016.
- [2] McKinney W. *Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython*. 2nd ed. Sebastopol: O'Reilly Media; 2018.
- [3] Kaufman L, Rousseeuw PJ. *Finding Groups in Data: An Introduction to Cluster Analysis*. New York: Wiley; 2019.
- [4] Dursun D, Ozkan S. A decision support framework for customer retention management using machine learning algorithms. *Decision Support Systems*. 2017;96:67–82.
- [5] Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*. 2015;12:2825–2830.
- [6] Zhang Y, Bell D. Clustering analysis for customer segmentation in e-commerce using K-Means. *Procedia Computer Science*. 2016;91:1015–1022.
- [7] Lee C, Kumar R. Integrating Business Intelligence and Machine Learning for Effective Customer Insights. *Journal of Business Analytics and Data Science*. 2023;5(2):55–72.
- [8] Tan L, Wu H, Zhang M. Customer segmentation using hybrid K-Means and DBSCAN for targeted marketing. *IEEE Access*. 2022;10:45234–45246.
- [9] Zhao F, Liu J, Chen Q. Evaluating clustering algorithms for customer segmentation in service-based industries. *International Journal of Data Science and Analytics*. 2023;11(4):233–248.
- [10] Provost F, Fawcett T. *Data Science for Business: What You Need to Know About Data Mining and Data-Analytic Thinking*. 2nd ed. Sebastopol: O'Reilly Media; 2016.
- [11] Hartigan JA, Wong MA. Algorithm AS 136: A K-Means clustering algorithm. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*. 2017;28(1):100–108.
- [12] Kimball R, Ross M, Thornthwaite W. *The Data Warehouse Lifecycle Toolkit*. 2nd ed. Indianapolis: Wiley; 2017.
- [13] Sathe S, Aggarwal C. Customer profiling and churn prediction using unsupervised clustering. *Proceedings of the ACM Symposium on Applied Computing*. 2021;1:742–747.
- [14] Verma A, Singh R. Business intelligence for customer segmentation in training industries: A practical approach. *International Journal of Business Analytics*. 2022;9(1):88–101.
- [15] Microsoft Corp. Power BI Documentation. Microsoft Learn; 2020. Available from: <https://learn.microsoft.com/en-us/power-bi/>
- [16] Tedyyana, Agus, Osman Ghazali, and Onno W. Purbo. "Machine learning for network defense: automated DDoS detection with telegram notification integration." *Indonesian Journal of Electrical Engineering and Computer Science* 34.2 (2024): 1102