

Analisis Kluster Kriteria Gangguan Kecemasan Sosial Berdasarkan Fase Perawatannya

<http://dx.doi.org/10.28932/jutisi.v10i1.8400>

Riwayat Artikel

Received: 15 Februari 2024 | Final Revision: 17 Februari 2024 | Accepted: 18 Februari 2024

Creative Commons License 4.0 (CC BY – NC)



Panji Yudasetya Wiwaha^{✉#1}, Hapnes Toba^{#2}, Oscar Karnalim^{#3}

[#] Program Magister Ilmu Komputer, Universitas Kristen Maranatha
Jl. Prof. drg. Surya Sumantri, M.P.H. No. 65, Bandung, 40164, Indonesia

¹2279003@maranatha.ac.id

³oscar.karnalim@maranatha.ac.id

²hapnes.toba@maranatha.ac.id

[✉]Corresponding author: 2279003@maranatha.ac.id

Abstrak — Penelitian ini bertujuan untuk mengelompokkan data aktivitas pasien penderita gangguan kecemasan sosial di Perusahaan *Mental Healthcare* yang berlokasi di Belanda dan mengukur afinitas kluster terhadap fase perawatan teridentifikasi berdasarkan tingkat kemiripan densitas *features*-nya. Metode pengelompokan data dilakukan dengan cara sebagai berikut: 1) melakukan data *pre-processing*, 2) melakukan *exploratory data analysis*, dan 3) membangun model *clustering* menggunakan algoritma *k-Means*. Efektifitas metode pengelompokan data diukur berdasarkan afinitas kluster terhadap fase perawatan teridentifikasi. Bobot *features* di setiap kluster diobservasi untuk menemukan karakteristik unik yang menjadi pembeda fase perawatan. Model *k-Means* berhasil mengelompokkan data aktivitas pasien menjadi 10 cluster. Berdasarkan persentase rasio afinitas, 3 cluster dipilih karena cukup mirip dengan fase perawatan yang terjadi di perusahaan. Temuan dalam penelitian ini menunjukkan bahwa jumlah hari sejak pasien melakukan registrasi data, jumlah registrasi data yang terkait dengan gangguan kecemasan sosial dalam seminggu terakhir, polaritas emosi bertendensi negatif, nilai kuesioner, dan nilai rendah di salah satu indikator kuesioner merupakan *features* pembeda untuk setiap fase perawatan. Selain itu, *features* tersebut memiliki kesesuaian tingkat urgensi paling relevan dengan daftar prioritas terapis pada saat menangani pasien. Meskipun demikian, penelitian lebih lanjut dan komprehensif perlu dilakukan untuk memahami dampak dari *feature* pembeda di setiap kluster agar dapat diarahkan pada pembuatan model klasifikasi untuk membuat sistem rekomendasi pasien berdasarkan tingkat urgensi penanganannya.

Kata kunci— analisis kluster; afinitas kluster; fase perawatan; gangguan kecemasan sosial; *k-means*.

Clusters Analysis of Social Anxiety Disorder Criteria Based on Stages of the Treatment

Abstract — This study aims to cluster the activity dataset of patients who suffer from social anxiety disorder at a *Mental Healthcare Company* located in the Netherlands and measure the affinity of the cluster to the identified treatment phase based on the similarity of its feature density. The methodology of data clustering is carried out in the following ways: 1) data *pre-processing*, 2) *exploratory data analysis*, and 3) building a clustering model using the *k-Means* algorithm. The effectiveness of the data clustering method is measured based on the affinity of the cluster to the identified treatment phase. The features' weights in each cluster are observed to find unique characteristics that can differentiate each phase of treatment. The *k-Means* model succeeded in grouping patient activity data into 10 clusters. Based on the percentage affinity ratio, 3 clusters are chosen because they are pretty similar to the treatment phase that occurs in the company. The findings in this study show that the number of days since the patient made a registration, number of registrations

in the past week, polarity of the negative emotions, questionnaire scores, and low scores in any questionnaire indicators are dominant in each treatment phase. In addition, the urgency of those features matches the therapist's top priorities when treating their clients. Nonetheless, further and comprehensive research must be conducted to understand the impact of the dominant features in each cluster so the classification model for creating a list of recommended patients based on their urgency level of treatment can be built.

Keywords— cluster analysis; density affinity; k-means; social anxiety disorder; treatment phase.

I. PENDAHULUAN

Gangguan kecemasan sosial (SAD) atau dikenal juga dengan fobia sosial, umumnya ditandai dengan perasaan takut untuk melakukan atau menghindari interaksi sosial karena kekhawatiran akan penilaian orang lain [1]. SAD dianggap sebagai gangguan kesehatan mental yang sangat heterogen dan kemungkinan terbagi kedalam beberapa sub-kategori. Misalnya, gangguan kecemasan umum (*Generalized Anxiety Disorder*), gangguan kecemasan terbatas (*Non-Generalized Anxiety Disorder*), dan gangguan kepribadian menghindar (*Avoidant Personality Disorder*). Setiap sub-kategori dibedakan berdasarkan tingkat kemarahan, ketidakramahan, dan ketidakpercayaan individu terhadap orang lain. Individu SAD biasanya dianggap sebagai individu yang pemalu, penurut, dan perilaku sosialnya terhambat (*behaviourally inhibited*) [2] [3]. Berdasarkan data yang dimuat oleh *National Comorbidity Survey* (NCS) dan *National Comorbidity Survey Replication* (NCS-R), SAD merupakan jenis gangguan mental yang umum terjadi pada 16% dan 12,1% dari total populasi survey [4]. Tanpa penanganan yang baik SAD dapat menstimulus gangguan mental lainnya, seperti stres dan depresi. Di Eropa, stres paling banyak terjadi, yaitu 16% dari total populasi dan menelan biaya penyembuhan sebesar 74,4 miliar Euro setiap tahunnya [5]. Dalam penelitiannya, Safren et al. [6] menyebutkan bahwa kepuasan akan perubahan perilaku dan kualitas hidup penderita SAD dapat diukur setelah melakukan terapi kelompok perilaku-kognitif (CBGT) selama 12 minggu yang berdampak terhadap perubahan positif pada empat domain kepuasan hidup penderita, yaitu: pencapaian, fungsi sosial, tumbuh kembang individu, dan lingkungan.

Besarnya biaya yang dibutuhkan dalam penanganan pasien dengan gangguan mental di Eropa, mendorong dilakukannya berbagai riset penelitian menggunakan pendekatan *Machine Learning* (ML) untuk mendeteksi berbagai jenis gangguan kecemasan seperti gangguan kecemasan umum (GAD), gangguan stres pasca trauma (PTSD), panik berlebih (PD), agorafobia, ataupun SAD menggunakan metode klasifikasi *Bayesian networks*, *Regression*, *Hybrid methods*, *SVM*, *ANN*, *Ensemble method*, *Fuzzy systems*, dan *Decision tree* [5]. Meskipun tingkat akurasi pengklasifikasian ML bervariasi tergantung dengan jenis gangguan kecemasan dan metode yang digunakan, namun riset tersebut menunjukkan adanya indikasi positif bahwa beberapa proses diagnosis awal terkait gangguan kecemasan dapat dilakukan oleh mesin dan hasilnya dapat dijadikan sebagai bahan pendukung bagi terapis untuk melakukan diagnosa lebih lanjut dalam menentukan jenis terapi yang tepat dan lebih baik bagi pasien.

Biaya perawatan dan pemanfaatan AI adalah dua hal yang saat ini menjadi perhatian utama Perusahaan *Mental Healthcare* di Belanda agar perusahaan dapat memberikan layanan konsultasi jiwa secara daring (*online coaching*) dengan efektif, namun harganya tetap kompetitif. Salah satu kasus yang umum terjadi di perusahaan adalah terapis seringkali mengalami kesulitan untuk menemukan daftar pasien penderita SAD yang sedang menjalani perawatan dan harus diprioritaskan penanganannya melalui sistem aplikasi layanan konsultasi jiwa milik perusahaan. Hal inilah yang kemudian memotivasi Perusahaan *Mental Healthcare* untuk membangun suatu sistem berbasis AI dalam membuat daftar rekomendasi prioritas pasien yang harus ditangani berdasarkan pola aktivitasnya di fase perawatan tertentu.

Tujuan dilakukannya penelitian ini adalah mengelompokkan data aktivitas pasien penderita gangguan kecemasan sosial di Perusahaan *Mental Healthcare*, kemudian mengukur afinitas klaster yang terbentuk terhadap fase perawatan teridentifikasi (awal, tengah, dan akhir) berdasarkan tingkat kemiripan densitas *features*-nya. Hasil pengelompokan data dianalisis lebih lanjut agar dapat menghadirkan wawasan dasar mengenai karakteristik aktivitas pasien yang menjadi pembeda di setiap fasenya. Karakteristik pembeda inilah yang kemudian akan dikaji lebih dalam untuk menentukan tingkat urgensi pasien secara keseluruhan sehingga dapat dijadikan sebagai acuan pengembangan model klasifikasi untuk membangun sistem rekomendasi prioritas pasien penderita SAD berdasarkan aktivitasnya, sebagaimana penelitian-penelitian yang sudah dilakukan sebelumnya terkait proses mengelompokkan data individu dengan berbagai gangguan kecemasan berdasarkan tingkat urgensi penanganannya [7] [8] [9]. Data yang digunakan dalam penelitian merupakan data anonim pasien penderita SAD yang disediakan oleh Perusahaan *Mental Healthcare* dari Februari 2019 sampai dengan Oktober 2023. Data yang digunakan dalam penelitian sudah dipastikan tersaring dengan baik untuk menjaga kerahasiaan pengguna layanan sebagai bentuk kepatuhan pada regulasi umum perlindungan data (GDPR) yang diterapkan di Eropa [10].

II. METODE PENELITIAN

A. Studi Literatur

Berbagai kajian pemodelan *Unsupervised ML* dengan metode *k-Means* telah dilakukan di bidang *mental healthcare* dan terbukti efektif dalam mengelompokkan profil individu dengan gejala fobia sosial, depresi, dan gangguan kecemasan umum berdasarkan kategori tertentu. Piqueras et al. [7] menggunakan model *k-Means* dan *Euclidean distance* untuk mengukur tingkat keparahan fobia sosial pada 971 profil remaja di Spanyol yang berpartisipasi dalam penelitian berdasarkan faktor demografis, dan ciri-ciri psikopatologi nya ke dalam lima klaster, yaitu: NO SP (tidak ada fobia sosial), SSP (fobia sosial spesifik), LGSP (fobia sosial umum ringan), MGSP (fobia sosial umum sedang), dan SGSP (fobia sosial umum berat). Dalam melakukan analisis keanggotaan klaster yang terbentuk, setiap informasi kuantitatif dianalisis dengan menggunakan metode *one-way analysis of variance* (ANOVA), sedangkan data dikotomi dianalisis dengan pengujian X². Yang et al. [8] menggunakan model *k-Means* untuk mengelompokkan 8063 profil siswa SMP dan SMA yang berpartisipasi dalam penelitian kedalam empat klaster yang merepresentasikan tingkat keparahan gejala depresi, yaitu: *None*, *Mild*, *Moderate*, dan *Severe*. Pengujian validitas klaster yang dihasilkan oleh model dilakukan dengan cara membandingkan setiap anggota klaster yang terbentuk dengan hasil dari metode pengklasifikasian tradisional berbasis norma (dan variasinya). Performa dan kinerja model juga dapat diukur dengan mempertimbangkan skor *accuracy*, *AUC*, *sensitivity*, serta asosiasi *feature* profil partisipan terhadap *feature* tingkat keparahan depresi yang sudah diketahui. Liu et al. [9] menggunakan model *k-Means* dan *Latent Profile Analysis* (LPA) untuk mengidentifikasi 9738 data profil siswa yang berpotensi memiliki gangguan kecemasan. Profil siswa disediakan dan dikelola oleh *Bureau of Education Yingshan*, China, dari bulan Juli hingga Agustus 2020. Hasil penelitiannya menunjukkan bahwa profil partisipan berhasil dikelompokkan menjadi tiga jenis klaster risiko, yaitu: berisiko rendah (*low-risk*), berisiko sedang (*mild-risk*), dan berisiko tinggi (*high-risk*). Metode LPA digunakan untuk mengklasifikasikan siswa ke dalam klaster yang paling mungkin berdasarkan profil kecemasannya dengan asumsi bahwa setiap siswa merupakan bagian dari salah satu *n* profil laten, yang jumlah atau ukurannya tidak diketahui secara apriori.

Pada penelitian-penelitian sebelumnya, algoritma *k-Means* digunakan untuk mengelompokkan data simtom gangguan kecemasan yang diderita oleh partisipan berdasarkan tingkat keparahannya, sedangkan dalam penelitian ini, *k-Means* digunakan untuk mengelompokkan data aktivitas pasien penderita gangguan kecemasan sosial berdasarkan fase perawatannya, sehingga dataset penelitian dapat dipastikan unik. Selain itu, *novelty* lain yang dihadirkan adalah metode untuk mengukur efektivitas pengelompokan data dilakukan dengan cara menghitung rasio afinitas klaster, dimana metode pengukuran ini terbilang cukup efektif dalam menguji efektivitas pengelompokan data yang setiap *feature*-nya memiliki sebaran data poin yang saling tumpang tindih (rapat) [11].

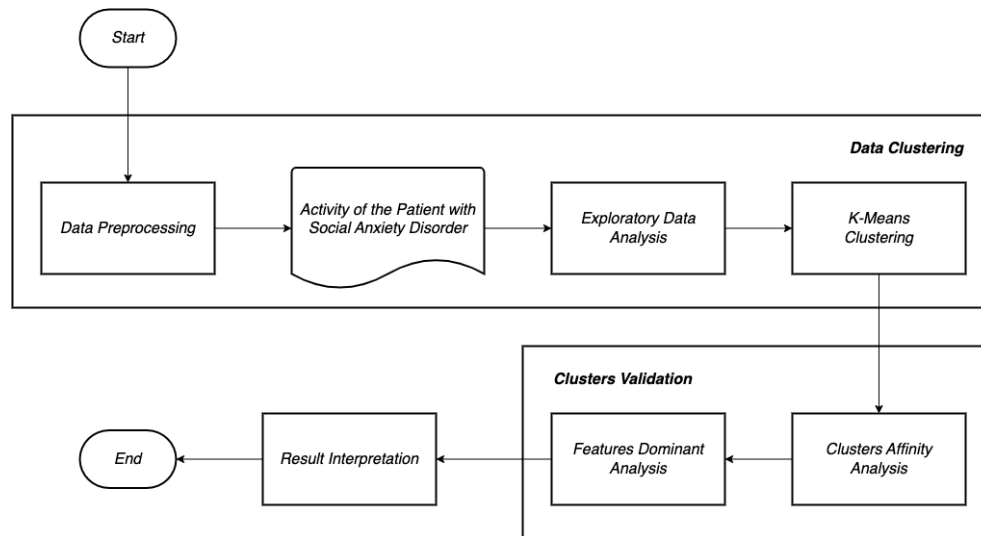
B. Hipotesis Penelitian

Penelitian ini dilakukan berdasarkan hipotesis-hipotesis sebagai berikut: 1) Klaster aktivitas pasien dapat dikelompokkan menjadi tiga klaster yang paling merepresentasikan fase awal, tengah, dan akhir perawatan. 2) Pola komunikasi pasien berpotensi untuk menjadi *feature* pembeda dalam setiap fase perawatan, dengan asumsi: pada fase awal perawatan, komunikasi pasien dengan terapis kemungkinan intens terjadi melalui panggilan audio/video. Sedangkan menjelang akhir perawatan komunikasi lebih banyak dilakukan melalui pesan singkat. 3) Pola aktivitas pasien dalam melakukan registrasi data berpotensi juga untuk menjadi *feature* pembeda di setiap fase perawatan, dengan asumsi: frekuensi dan jumlah registrasi data yang dilakukan pasien terbilang rendah di fase awal perawatan, mengalami peningkatan di pertengahan perawatan, dan mengalami penurunan menjelang fase akhir perawatan. 4) Pola penurunan gejala gangguan kecemasan sosial yang ditandai dengan perubahan emosi pasien dalam dua minggu terakhir berpotensi untuk menjadi *feature* pembeda di setiap fase perawatan, dengan asumsi: jumlah registrasi data bertendensi negatif di fase awal perawatan akan mengalami peningkatan sampai dengan fase tengah perawatan, kemudian mengalami penurunan menjelang fase akhir perawatan. Sebaliknya, data bertendensi positif akan mengalami peningkatan menjelang fase akhir perawatan. 5) Pola penyelesaian tugas yang harus dilakukan pasien dalam dua minggu terakhir berpotensi juga untuk menjadi *feature* pembeda di setiap fase perawatan, dengan asumsi: tugas-tugas yang diberikan terapis atau atas inisiatif sendiri semakin banyak yang terselesaikan menjelang fase akhir perawatan. 6) Pola skor tanggapan yang dikirim pasien terkait relevansi perawatan terhadap keluhan pasien dalam bentuk kuesioner akan semakin membaik menjelang fase akhir perawatan.

C. Prosedur Penelitian

Penelitian dilakukan dengan mengikuti tahapan desain penelitian pada diagram alir yang dapat dilihat di Gambar 1, yaitu proses untuk mengelompokkan data (*data clustering*) dimulai dengan menjalankan data preprocessing untuk mengolah data anonim pasien menjadi dokumen data aktivitas pasien penderita gangguan kecemasan sosial, kemudian melakukan proses *exploratory data analysis*, dan membuat model untuk mengelompokkan data aktivitas pasien menggunakan metode *k-Means*. Setelah proses pengelompokan data berhasil dilakukan, proses kemudian dilanjutkan dengan menguji efektivitas pengelompokan data (*clusters validation*), dimulai dengan melakukan proses analisis afinitas klaster terhadap fase perawatan

teridentifikasi dan melakukan proses analisis *features* dominan di setiap kluster. Proses kemudian dilanjutkan dengan menginterpretasikan kelompok data terpilih untuk menentukan karakteristik kluster dan mengkaji dampak *features* dominan terhadap urgensi penanganan pasien. Setelah proses interpretasi hasil kluster selesai dilakukan, maka rangkaian proses penelitian pun berakhir.



Gambar 1. Diagram alir penelitian.

D. Pengelompokan Data

Secara garis besar, pengelompokan data aktivitas pasien penderita gangguan kecemasan sosial menggunakan *k-Means* dilakukan setelah melalui dua sub pemrosesan data, yaitu data *preprocessing* dan *exploratory data analysis*.

1) *Data Preprocessing*: Tujuan utama dari tahap ini adalah untuk mengakuisisi data anonim pasien dari layanan penyedia data yang dikelola oleh Perusahaan *Mental Healthcare*, kemudian mengkonversi data anonim tersebut menjadi *dataset* aktivitas pasien penderita gangguan kecemasan sosial. Proses akuisisi dan transformasi data ini dikembangkan dalam bentuk *service* sehingga dapat dijalankan dalam kontainer *Docker* secara periodik sampai dengan Oktober 2023. Pengembangan *service* dilakukan dengan menggunakan bahasa pemrograman Python 3.9, bersifat *open source*, disimpan dalam *repository Github*, dapat diakses secara publik, dan arsitekturnya dapat dilihat pada Gambar 2. Modul utama seperti *Loader* dan *Transformator*, telah dilengkapi dengan pengujian unit, fungsional, dan pengujian *blackbox* untuk memastikan *dataset* aktivitas pasien dihasilkan dengan benar.

2) *Exploratory Data Analysis*: Tujuan utama dari tahap ini adalah untuk melihat pola sebaran data aktivitas pasien penderita gangguan kecemasan sosial terhadap kemunculan kasus per-harinya, melakukan standarisasi data menggunakan *Standard Scaler* untuk mengubah data ke dalam bentuk *z-score* [11] agar data terdistribusi secara normal, serta menentukan jumlah kluster ideal berdasarkan metode *Elbow* yang memiliki ketebalan dan rata-rata siluet kluster paling baik. Metode *Elbow* mengacu pada nilai *sum of squared error* (SSE), yaitu metode paling sederhana dan umum digunakan untuk melakukan pemisahan berdasarkan kemiripan objek di dalam suatu kluster [12] [13]. SSE didefinisikan pada persamaan (1) dimana C_k merupakan merupakan kelompok objek dalam kluster k dan μ_k merupakan nilai rata-rata dari kluster k .

$$SSE = \sum_{k=1}^K \sum_{\forall x_i \in C_k} \|x_i - \mu_k\|^2 \quad (1)$$

Sedangkan ketebalan dan rata-rata siluet kluster terbaik ditentukan berdasarkan nilai rata-rata jarak intra-kluster dan rata-rata jarak terpendek inter-kluster. Nilai siluet memiliki rentang -1 sampai dengan $+1$, semakin nilainya mendekati $+1$, maka performa algoritma pengelompokan data dianggap semakin baik [14]. Metode pengukuran nilai siluet didefinisikan pada persamaan (2) dimana a merupakan nilai rata-rata jarak intra-kluster dan b merupakan nilai rata-rata jarak terpendek inter-kluster [15].

$$S = \frac{b - a}{\max(a, b)} \quad (2)$$

3) *K-Means Clustering*: Setelah jumlah n kluster paling optimal telah diketahui dan *dataset* aktivitas pasien sudah distandarisasi, maka pengelompokan data menggunakan metode *k-Means* akan dijalankan. Algoritma *k-Means* pertama kali diperkenalkan oleh J.B. MacQueen pada tahun 1967 untuk mengelompokkan data-data numerik, dimana setiap kelompok data memiliki titik pusat yang disebut dengan *mean* [16]. Metode ini akan mengklasifikasikan kelompok data ke dalam kluster k yang berbeda secara iteratif sampai konvergen ke titik lokal minimum [17]. Algoritma ini biasanya digunakan untuk *data mining* dan pengenalan pola [18]. Tahap awal algoritma *k-Means* adalah menentukan titik pusat kluster secara acak dengan jumlah kluster k yang konstan. Selanjutnya, algoritma akan bekerja dengan cara mengalokasikan setiap data poin diluar radius titik pusat ke kluster C terdekat. Jarak setiap data poin umumnya diukur menggunakan formula *Euclidean distance*. Persamaan (3) merupakan formula *Euclidean distance* antara vektor $p = [p_1, p_2, \dots, p_n]$ dan $q = [q_1, q_2, \dots, q_n]$:

$$d_{Euclid}(p, q) = \sqrt{\sum_{i=1}^n (p_i - q_i)^2} \quad (3)$$

Jika setiap data poin telah tergabung dalam kelompok kluster (C_1, C_2, \dots, C_k), maka seluruh proses tahap awal dapat dikatakan telah selesai. Selanjutnya, nilai rata-rata untuk setiap kluster (C_1, C_2, \dots, C_k), akan dihitung ulang dan keanggotaan klasternya akan terus diubah sampai dengan nilai fungsi objektifnya (*inertia*) menjadi minimum atau keanggotaan setiap kluster menjadi tetap. Nilai *inertia* merupakan jumlah rata-rata setiap jarak intra-kluster antara kluster k_i dengan titik *centroid* (C_i) [17] [19], dengan persamaan (4) sebagai berikut ini:

$$I = \sum_{i=1}^k \sum_{x \in C_i} d_{Euclid}(x, (C_i)) \quad (4)$$

dimana (C_i) merupakan *centroid* dari kluster C_i , dan $d_{euc}(x, (C_i))$ merupakan *Euclidean distance* antara x dengan (C_i).

E. Pengujian Kluster

Pengujian kluster dilakukan dengan cara mengukur kedekatan kluster aktivitas pasien terhadap fase perawatan yang teridentifikasi Perusahaan *Mental Healthcare*, berdasarkan rasio kerapatan *feature* di dalam kluster k terhadap kerapatan *feature* pada fase perawatan teridentifikasi. Metode pengujian ini disebut sebagai afinitas kluster [20], dimana formula tingkat kemiripan (rasio afinitas) kluster didefinisikan dengan persamaan (5) berikut ini:

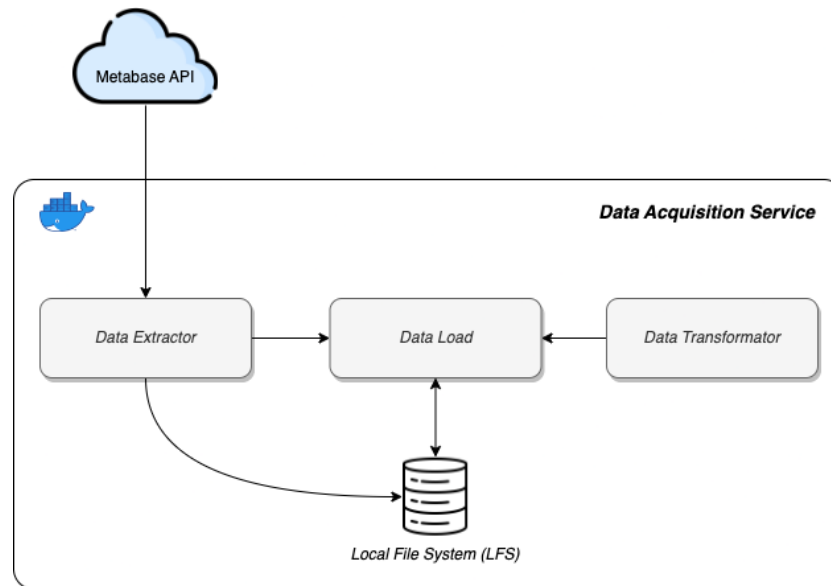
$$A_{ratio} = \frac{\text{count}(F_{\text{within cluster}})}{\text{count}(F'_{\text{within cluster}})} \times 100\% \quad (5)$$

A_{ratio} merupakan rasio kemiripan (rasio afinitas) kluster terhadap fase perawatan teridentifikasi, $F_{\text{within cluster}}$ merupakan jumlah *feature* di kluster k , dan $F'_{\text{within cluster}}$ adalah jumlah *feature* teridentifikasi pada fase perawatan tertentu.

Dominansi *features* di dalam kluster k ditentukan berdasarkan besaran nilai bobotnya, yaitu rasio dari rata-rata *features* di dalam kluster terhadap rata-rata globalnya. Nilai bobot yang dihasilkan menunjukkan seberapa unik suatu *feature* di dalam kluster, didefinisikan dengan persamaan (6) sebagai berikut ini:

$$W_{F_{\text{within cluster}}} = \frac{\text{avg}(F_{\text{within cluster}})}{\text{avg}(F_{\text{global}})} \quad (6)$$

dimana $W_{F_{\text{within cluster}}}$ merupakan nilai bobot *feature* dalam kluster k , $\text{avg}(F_{\text{within cluster}})$ merupakan nilai rata-rata *feature* di dalam kluster k , dan $\text{avg}(F_{\text{global}})$ merupakan nilai rata-rata *feature* secara global.



Gambar 2. Arsitektur *data acquisition service*. *Data Extractor* merupakan modul yang bertugas untuk mengekstraksi data anonim pasien dari penyedia data—*Metabase* melalui *REST API* dan menyimpannya ke *local file system*, dimana hak aksesnya dibatasi dan diberikan langsung oleh perusahaan. *Data Load* merupakan modul yang bertugas untuk membaca data anonim milik pasien dari *local file system*, mengkonversi data mentah menjadi *dataset* aktivitas pasien penderita gangguan kecemasan sosial, menyimpan data hasil konversi kembali ke *local file system*, dan hanya dijalankan setelah modul *Data Extractor* selesai dieksekusi. *Data Transformator* merupakan modul yang bertugas untuk menyediakan fungsi-fungsi transformasi data yang digunakan oleh modul *Data Load* untuk mengkonversi data anonim pasien menjadi *dataset* aktivitas pasien penderita gangguan kecemasan sosial.

III. HASIL DAN PEMBAHASAN

A. Pengelompokan Data

1) *Data Preprocessing*: *Data acquisition service* yang dikembangkan dan dijalankan dalam *Docker container*, berhasil mengakuisisi data anonim klien dari layanan penyedia data Perusahaan *Mental Healthcare* dari Februari 2019 sampai dengan Oktober 2023. Data anonim yang dimaksud meliputi: 111 data anonim pasien, 3.845 data komunikasi, 3.886 data pelacakan gangguan kecemasan sosial, 4.024 data riwayat catatan harian, 21.368 data pemberitahuan, 15.032 data penyelesaian aktivitas terjadwal, 301 data kuesioner terkait relevansi sesi perawatan, 901 data riwayat sesi perawatan, dan 487 data riwayat registrasi hal-hal yang dipikirkan oleh pasien (*thought record*). Data anonim tersebut kemudian dikonversi menjadi *dataset* aktivitas pasien penderita gangguan kecemasan sosial, dengan *volume dataset* yang dihasilkan sampai dengan Oktober 2023 berjumlah 2.316. Adapun sampel 5 data teratas dan 5 data terbawah dari 2.316 baris data aktivitas pasien dapat dilihat pada Gambar 3 dan deskripsi setiap kolomnya dapat dilihat pada TABEL 1.

TABEL 1
DESKRIPSI KOLOM DATASET AKTIVITAS PASIEN PENDERITA GANGGUAN KECEMASAN SOSIAL

Kolom	Deskripsi	Tipe Data
<i>case_id</i>	ID terenkripsi yang merujuk pada kasus pasien dengan gangguan kecemasan sosial	Teks
<i>case_created_at</i>	Tanggal kemunculan kasus	Teks
<i>client_id</i>	ID terenkripsi (<i>anonymous</i>) yang merujuk pada pasien dengan gangguan kecemasan sosial	Teks
<i>a_by_call</i>	Jumlah hari sejak terakhir pasien melakukan komunikasi melalui panggilan audio/video dengan terapis	Numerik
<i>a_by_chat</i>	Jumlah hari sejak terakhir pasien melakukan komunikasi melalui pesan singkat dengan terapis	Numerik
<i>b</i>	Jumlah hari sejak terakhir pasien melakukan registrasi data	Numerik
<i>c</i>	Jumlah registrasi data yang berkaitan dengan gangguan kecemasan sosial dalam 7 hari terakhir	Numerik
<i>d</i>	Polaritas registrasi data bertendensi negatif yang terjadi dari 0 – 7 hari terakhir dengan 7 – 14 hari sebelumnya	Numerik-Kategorikal 0: <i>Decrease</i> 1: <i>Stable</i> 2: <i>Small Increase</i> 3: <i>Big Increase</i>

Kolom	Deskripsi	Type Data
<i>e</i>	Polaritas perubahan registrasi data bertendensi positif yang terjadi dari 0 – 7 hari terakhir dengan 7 – 14 hari sebelumnya	Numerik-Kategorikal 0: <i>Increase</i> 1: <i>Stable</i> 2: <i>Decrease</i>
<i>f__is_scheduled</i>	Bernilai 1 jika terdapat aktivitas terjadwal dalam 7 hari terakhir, dan 0 jika tidak ada aktivitas yang dijadwalkan	Numerik-Kategorikal 0: <i>Unplanned</i> 1: <i>Planned</i>
<i>f__completion_status</i>	Status penyelesaian aktivitas terjadwal dalam 7 hari terakhir	Numerik-Kategorikal 0: <i>None</i> 1: <i>Incomplete</i> 2: <i>Some Complete</i> 3: <i>Complete</i>
<i>g__is_reminder_activated</i>	Bernilai 1 jika pengingat untuk melakukan registrasi data <i>thought record</i> dalam 7 hari terakhir diaktifkan, dan 0 jika pengingat tidak diaktifkan	Numerik-Kategorikal 0: <i>Unreminded</i> 1: <i>Reminded</i>
<i>g__is_completed</i>	Bernilai 1 jika pasien minimal melakukan registrasi data <i>thought record</i> satu kali dalam 7 hari terakhir, dan 0 jika tidak terdapat registrasi data	Numerik-Kategorikal 0: <i>Incomplete</i> 1: <i>Complete</i>
<i>h</i>	Perubahan skor kuesioner sesi perawatan terakhir dan kuesioner sebelumnya	Numerik-Kategorikal 1: <i>Large Decrease</i> 2: <i>Stable</i> 3: <i>Large Increase</i>
<i>h__low_score</i>	Bernilai 1 jika terdapat indikasi nilai rendah di salah satu indikator penilaian pada kuesioner sesi perawatan terakhir, dan 0 jika tidak terdapat nilai rendah	Numerik-Kategorikal 0: <i>False</i> 1: <i>True</i>
<i>i__is_reminder_activated</i>	Bernilai 1 jika pengingat untuk melakukan registrasi data catatan harian dalam 7 hari terakhir diaktifkan, dan 0 jika pengingat tidak diaktifkan	Numerik-Kategorikal 0: <i>Unreminded</i> 1: <i>Reminded</i>
<i>i__is_completed</i>	Bernilai 1 jika pasien minimal melakukan registrasi data catatan harian satu kali dalam 7 hari terakhir, dan 0 jika tidak terdapat registrasi data	Numerik-Kategorikal 0: <i>Incomplete</i> 1: <i>Complete</i>

```

CRITERIA_DATASET = load_criteria('valid_criteria')
CRITERIA_DATASET

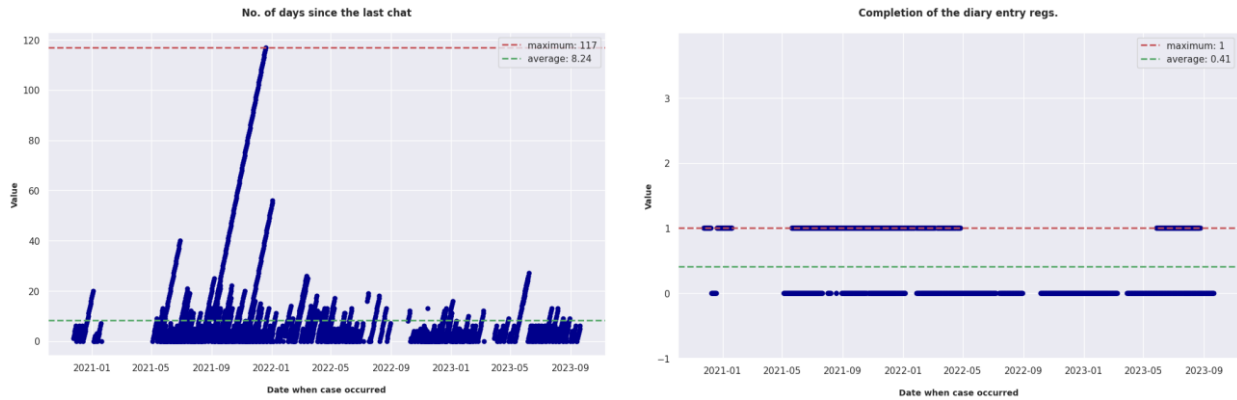
```

	case_id	case_created_at	client_id	a_by_call	...	h	h_low_score	i_is_reminder_activated	i_is_completed
0	3bdc547ce0c2d8956bcf32c3170bab	2021-06-23	8c582f39ed7af97a900c9720c3d3c79c	0	...	2	1	0	1
1	e196a655d611dfb7d39f6e29b73982a4	2021-06-22	8c582f39ed7af97a900c9720c3d3c79c	6	...	2	1	0	1
2	6a6442992c949043891d389687c509f7	2021-06-21	8c582f39ed7af97a900c9720c3d3c79c	5	...	2	1	0	1
3	962700491b1342925450ed9d5e7cbe59	2021-06-20	8c582f39ed7af97a900c9720c3d3c79c	4	...	2	1	0	1
4	7c10dbf4d1a0e52770f7b059e8dab8d9	2021-06-19	8c582f39ed7af97a900c9720c3d3c79c	3	...	2	1	0	1
...
2311	1a7f1bcbda5762ac52b32fe810537930	2023-09-18	2b0f8e152dd4b04e83ea5b2fb5eda989	4	...	2	0	0	0
2312	db5f101a4ca7fa81b3795478f36e1997	2023-09-17	2b0f8e152dd4b04e83ea5b2fb5eda989	3	...	2	0	0	0
2313	f55225964b199361addf96036e71f5e1	2023-09-16	2b0f8e152dd4b04e83ea5b2fb5eda989	2	...	2	0	0	0
2314	551f7bb9e545e27a25af2167dbb226d	2023-09-15	2b0f8e152dd4b04e83ea5b2fb5eda989	1	...	2	0	0	0
2315	854e0d00d76798e4549ebc6e6a738c43	2023-09-20	2b0f8e152dd4b04e83ea5b2fb5eda989	0	...	2	0	0	0

2316 rows x 17 columns

Gambar 3. Sampel data aktivitas pasien penderita gangguan kecemasan sosial.

2) *Exploratory Data Analysis*: Temuan dari hasil visualisasi sebaran nilai *dataset* aktivitas pasien penderita gangguan kecemasan sosial terhadap tanggal kemunculan kasus per-harinya menunjukkan adanya rentang nilai aktivitas pasien yang berbeda signifikan antara satu dengan yang lainnya, misalnya nilai jumlah hari sejak terakhir kontak melalui pesan singkat (*a_by_chat*) dengan nilai status penyelesaian registrasi catatan harian dalam seminggu terakhir (*i__is_completed*) dengan rentang nilai masing-masing 0 – 117 dan 0 – 1 sebagaimana yang dapat dilihat pada Gambar 4. Adapun hasil standarisasi nilai *dataset* aktivitas pasien ke dalam bentuk *z-score* menggunakan *Standard Scaler* dapat dilihat pada Gambar 5.



Gambar 4. Perbandingan sebaran nilai jumlah hari sejak terakhir kontak lewat pesan singkat (kiri) dengan nilai status penyelesaian registrasi data catatan harian (kanan), dimana sumbu x menjelaskan tanggal kemunculan kasus per-harinya dan sumbu y menjelaskan nilai dari aktivitas pasien. Gambar kiri menegaskan bahwa nilai maksimum untuk jumlah hari sejak terakhir kontak bernilai 117 sedangkan gambar di sebelah kanannya menunjukkan bahwa nilai maksimum untuk status penyelesaian registrasi data catatan harian bernilai 1.

```
features = [f[F_CODE] for f in FEATURES]
SCALED_CRITERIA[features].round(decimals=3)
```

	a_by_call	a_by_chat	b	c	...	h	h_low_score	i_is_reminder_activated	i_is_completed
0	-1.092	-0.528	-0.703	0.605	...	-0.335	2.800	-0.246	1.201
1	0.118	-0.528	-0.703	0.372	...	-0.335	2.800	-0.246	1.201
2	-0.084	-0.272	-0.703	0.140	...	-0.335	2.800	-0.246	1.201
3	-0.285	-0.336	-0.703	-0.093	...	-0.335	2.800	-0.246	1.201
4	-0.487	-0.400	-0.350	-0.558	...	-0.335	2.800	-0.246	1.201
...
2311	-0.285	-0.272	3.536	-0.558	...	-0.335	-0.357	-0.246	-0.832
2312	-0.487	-0.336	3.359	-0.558	...	-0.335	-0.357	-0.246	-0.832
2313	-0.689	-0.400	3.182	-0.558	...	-0.335	-0.357	-0.246	-0.832
2314	-0.890	-0.464	3.006	-0.558	...	-0.335	-0.357	-0.246	-0.832
2315	-1.092	-0.144	3.889	-0.558	...	-0.335	-0.357	-0.246	-0.832

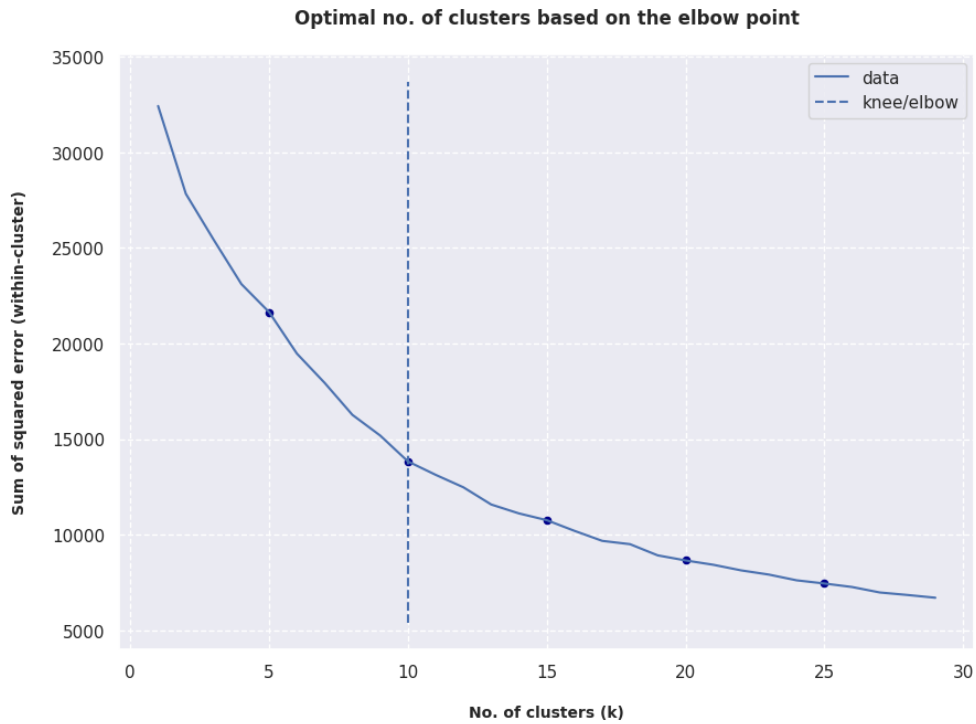
2316 rows x 14 columns

Gambar 5. Sampel data aktivitas pasien penderita gangguan kecemasan sosial setelah distandarisasi.

Berdasarkan kurva distribusi nilai SSE untuk setiap kluster di Gambar 6, maka jumlah n kluster ideal yang disarankan metode *Elbow* adalah 10 yang ditegaskan dengan ketebalan rata-rata siluet pada Gambar 7. Meskipun data pada TABEL 2 menunjukkan bahwa nilai rata-rata terbaik terbentuk pada saat jumlah n kluster = 14, namun pemilihan jumlah kluster tersebut bukanlah pilihan terbaik karena kluster dengan label 9 memiliki nilai siluet dibawah rata-rata (< 0.246) sebagaimana yang ditunjukkan oleh Gambar 9. Sedangkan Gambar 8, menunjukkan bahwa jumlah n kluster = 12 bukan pula pilihan jumlah terbaik, karena kluster dengan label 7 nilai siluetnya hampir berada di bawah rata-rata global (≤ 0.244).

TABEL 2
DISTRIBUSI NILAI RATA-RATA SILHOUETTE

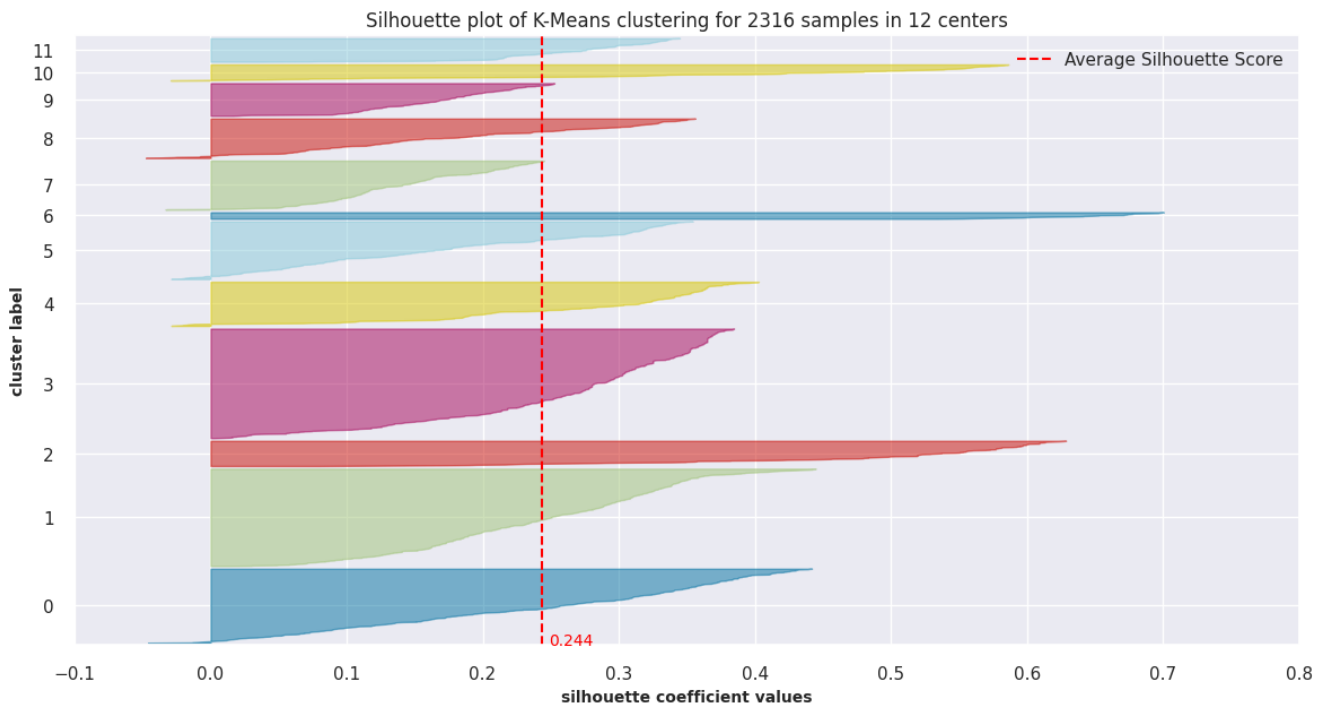
Kluster	Rata-rata
6	0.197
7	0.218
8	0.214
9	0.229
10	0.240
11	0.234
12	0.244
13	0.243
14	0.246



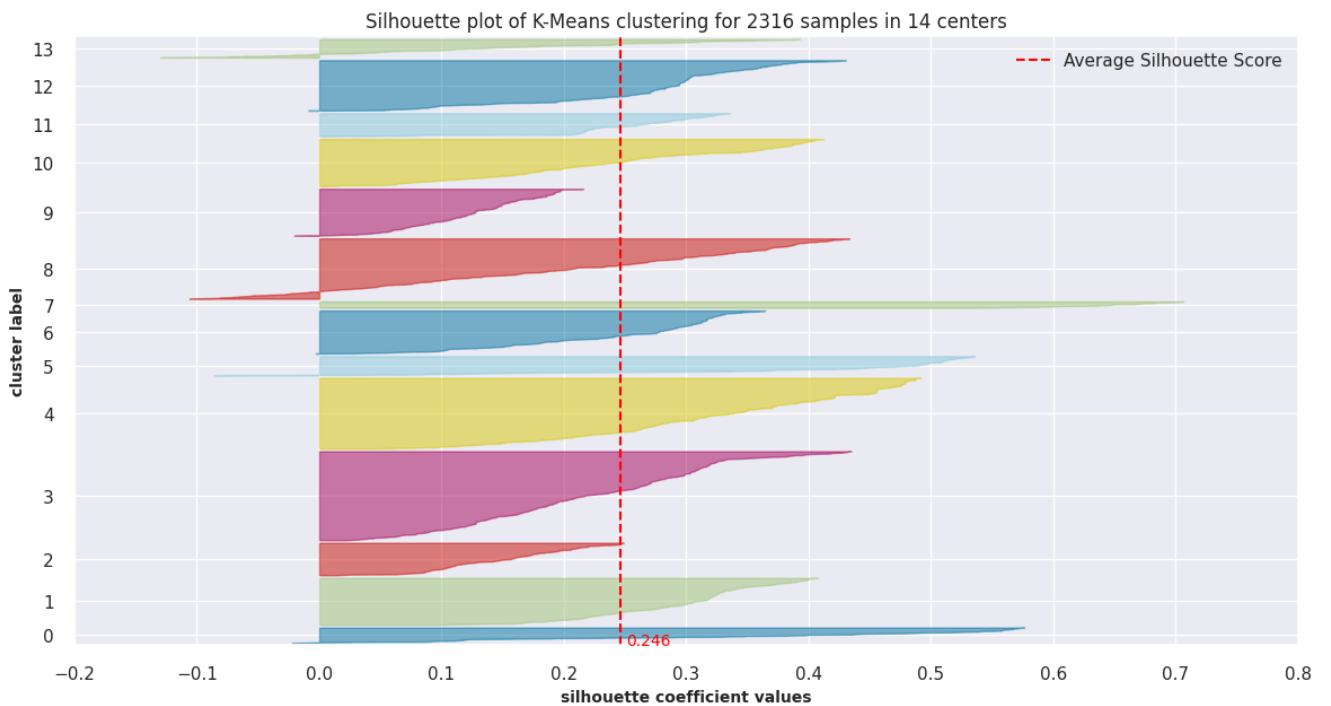
Gambar 6. Plot jumlah kluster ideal berdasarkan Metode *Elbow*. Garis vertikal putus-putus menunjukkan bahwa posisi *elbow* yang ideal terbentuk pada saat jumlah kluster bernilai 10.



Gambar 7. Plot ketebalan siluet dan distribusi nilai rata-rata siluet untuk setiap kluster ketika jumlah kluster bernilai 10. Garis vertikal putus-putus menunjukkan bahwa rata-rata siluet secara global bernilai 0.240 dan menegaskan bahwa jumlah kluster tersebut merupakan pilihan paling optimal karena tidak ditemukan satupun kluster yang nilai siluet-nya dibawah rata-rata.

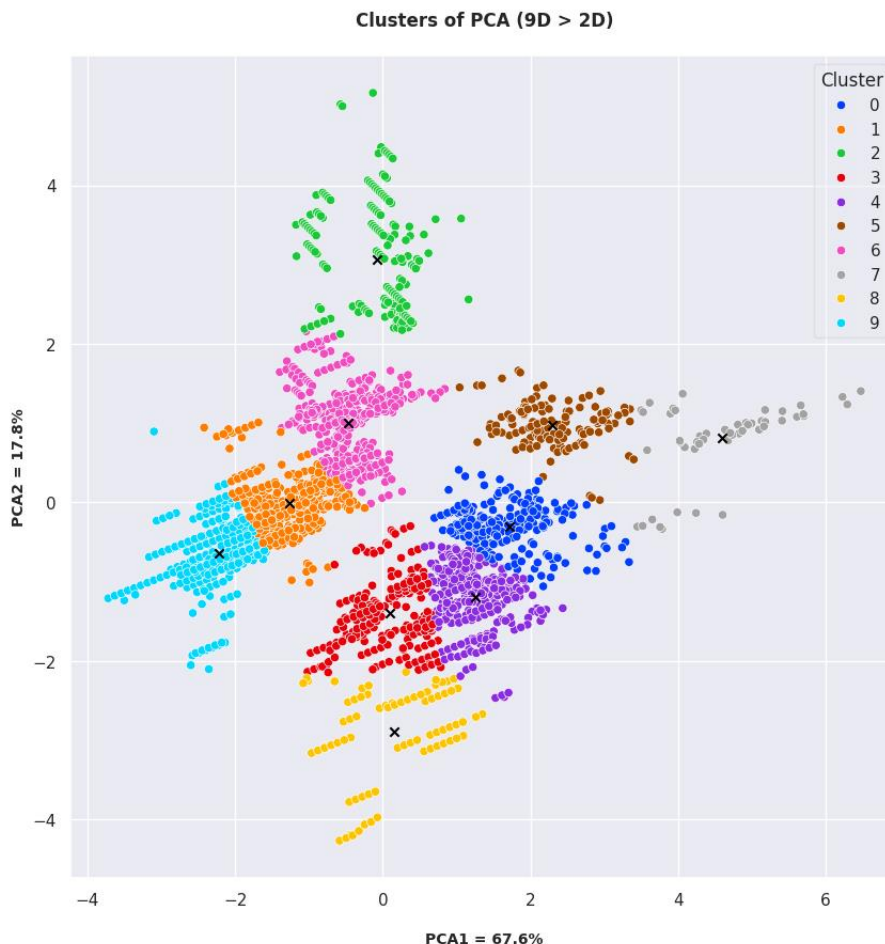


Gambar 8. Plot ketebalan siluet dan distribusi nilai rata-rata siluet untuk setiap kluster ketika jumlah kluster bernilai 12. Garis vertikal putus-putus menunjukkan bahwa rata-rata nilai siluet secara global bernilai 0.244 dan menegaskan bahwa jumlah kluster tersebut bukan pilihan paling optimal karena kluster dengan label 7 nilai siluet-nya hampir berada dibawah rata-rata global.



Gambar 9. Plot ketebalan siluet dan distribusi nilai rata-rata siluet untuk setiap kluster ketika jumlah kluster bernilai 14. Garis vertikal putus-putus menunjukkan bahwa rata-rata nilai siluet secara global bernilai 0.246 dan menegaskan bahwa jumlah kluster tersebut bukan pilihan paling optimal karena kluster dengan label 9 nilai siluet-nya berada dibawah rata-rata global.

3) *K-Means Clustering*: Pola 10 kluster data yang dihasilkan oleh metode *k-Means* dapat dilihat pada Gambar 10, dimana jumlah dimensi *dataset* aktivitas pasien direduksi dari 14 menjadi 9 menggunakan teknik *Principal Component Analysis* [21] dengan tetap mempertahankan nilai kumulatif rasio variansi sebesar 83.3%.



Gambar 10. Hasil pengelompokan data menggunakan *k-Means* dengan 9 komponen utama yang ditampilkan pada 2 komponen utama, yaitu PCA1 dan PCA2 dengan nilai kumulatif rasio variansi masing-masing sebesar 67.6% dan 17.8%.

B. Pengujian Kluster

1) *Afinitas Kluster*: Setiap fase perawatan yang teridentifikasi Perusahaan *Mental Healthcare* di dalam *dataset* aktivitas pasien dikodekan sebagai kolom 'p', dimana nilai 0 merepresentasikan fase awal perawatan, nilai 1 untuk fase tengah perawatan, dan nilai 2 untuk fase akhir perawatan, sebagaimana yang dapat dilihat pada Gambar 11. Adapun hasil kalkulasi rasio jumlah *features* di dalam kluster terhadap jumlah *features* pada fase perawatan teridentifikasi, direpresentasikan sebagai tabel *confusion matrix* di TABEL 3.

TABEL 3
PERSENTASE AFINITAS KLUSTER TERHADAP FASE PERAWATAN TERIDENTIFIKASI

% dist.	c0	c1	c2	c3	c4	c5	c6	c7	c8	c9
awal perawatan	11.83	20.51	6.11	5.72	9.47	1.58	22.09	0.00	4.54	18.15
tengah perawatan	9.84	19.78	9.38	9.19	14.11	6.22	15.78	2.69	3.16	9.84
akhir perawatan	10.93	14.75	6.28	13.11	13.11	7.38	18.99	3.01	2.46	9.97
all	10.62	18.35	7.69	9.67	12.78	5.57	18.18	2.20	3.24	11.70

Data persentase afinitas kluster di TABEL 3, menunjukkan bahwa karakteristik fase awal dan akhir perawatan cenderung direpresentasikan oleh kluster 'c6' dengan persentase afinitas tertinggi masing-masing sebesar 22.09% dan 18.99%. Sedangkan karakteristik fase tengah perawatan cenderung direpresentasikan oleh kluster 'c1' dengan persentase afinitas

tertinggi sebesar 19.78%. Temuan ini menunjukkan bahwa fase awal dan akhir perawatan cenderung memiliki *features* dominan yang serupa. Sehingga, apabila dilihat berdasarkan urutan afinitas klaster di TABEL 4, klaster dengan *features* dominan yang memiliki kemungkinan paling besar untuk menjadi pembeda di fase perawatan awal, tengah, dan akhir terdapat di urutan ketiga ['c9', 'c4', 'c3'] (ditandai warna hijau) dan di urutan ke delapan ['c8', 'c5', 'c2'] (ditandai warna biru).

TABEL 4
POLA AFINITAS KLASTER DI SETIAP FASE PERAWATAN

Fase Perawatan	Urutan Afinitas Klaster
awal	c6 – c1 – c9 – c0 – c4 – c2 – c3 – c8 – c5 – c7
tengah	c1 – c6 – c4 – c9 – c0 – c2 – c3 – c5 – c8 – c7
akhir	c6 – c1 – c3 – c4 – c0 – c9 – c5 – c2 – c7 – c8

2) *Features Dominan*: Apabila diambil 5 *feature* paling dominan yang dianggap dapat membedakan fase awal, tengah, dan akhir perawatan pada klaster 'c9', 'c4', dan 'c5', maka data bobot *features* di TABEL 5 dapat menegaskan bahwa jumlah hari sejak pasien melakukan registrasi data (*b*), tanggapan dengan skor rendah di salah satu indikator kuesioner relevansi sesi perawatan (*h_low_score*), status aktivasi pengingat untuk melakukan registrasi data *thought record* (*g_is_reminder_activated*), jumlah hari sejak terakhir kontak melalui panggilan audio/video (*a_by_call*), dan status aktivasi pengingat untuk melakukan registrasi data catatan harian (*i_is_reminder_activated*) adalah *features* paling unik di fase awal perawatan. Sedangkan data bobot *features* di TABEL 6, menegaskan bahwa status penyelesaian aktivitas terjadwal (*f_completion_status*), adanya aktivitas terjadwal minimal satu dalam seminggu terakhir (*f_is_scheduled*), jumlah registrasi data yang berkaitan dengan gangguan kecemasan sosial (*c*), perubahan registrasi data bertendensi negatif dalam seminggu terakhir dibandingkan dengan seminggu sebelumnya (*d*), serta tanggapan pasien terhadap relevansi perawatan (*h*) merupakan *features* paling unik di fase tengah perawatan. Kemudian, data bobot *features* di TABEL 7 menegaskan bahwa adanya aktivitas terjadwal minimal satu dalam seminggu terakhir (*f_is_scheduled*), jumlah hari sejak pasien melakukan registrasi data (*b*), status penyelesaian aktivitas terjadwal (*f_completion_status*), status aktivasi pengingat untuk melakukan registrasi data catatan harian (*i_is_reminder_activated*), dan jumlah hari sejak terakhir kontak melalui panggilan audio/video (*a_by_call*) merupakan *features* paling unik di fase akhir perawatan.

TABEL 5
BOBOT FEATURES PADA KLASTER C9

Feature	Bobot
<i>b</i>	3.61
<i>h_low_score</i>	2.06
<i>g_is_reminder_activated</i>	1.58
<i>a_by_call</i>	1.53
<i>i_is_reminder_activated</i>	1.36
<i>e</i>	1.06
<i>h</i>	1.03
<i>a_by_chat</i>	0.83
<i>d</i>	0.74
<i>c</i>	0.03
<i>i_is_completed</i>	0.01
<i>f_is_scheduled</i>	0.00
<i>f_completion_status</i>	0.00
<i>g_is_completed</i>	0.00

TABEL 6
BOBOT FEATURES PADA KLASTER C4

Feature	Bobot
<i>f_completion_status</i>	2.49
<i>f_is_scheduled</i>	2.28
<i>c</i>	1.27
<i>d</i>	1.17
<i>h</i>	1.02
<i>e</i>	0.99
<i>a_by_call</i>	0.80
<i>a_by_chat</i>	0.50
<i>b</i>	0.41
<i>i_is_completed</i>	0.40
<i>i_is_reminder_activated</i>	0.36
<i>g_is_reminder_activated</i>	0.00
<i>g_is_completed</i>	0.00
<i>h_low_score</i>	0.00

TABEL 7
BOBOT FEATURES PADA KLASTER C3

Feature	Bobot
<i>f_is_scheduled</i>	2.21
<i>b</i>	1.63
<i>f_completion_status</i>	1.56
<i>i_is_reminder_activated</i>	1.33
<i>a_by_call</i>	1.18
<i>e</i>	1.03
<i>h</i>	1.01
<i>a_by_chat</i>	0.60
<i>d</i>	0.56
<i>h_low_score</i>	0.36
<i>c</i>	0.27
<i>i_is_completed</i>	0.14
<i>g_is_completed</i>	0.06
<i>g_is_reminder_activated</i>	0.00

```

focused_features = ['p', 'case_id', 'case_created_at'] + [f[F_CODE] for f in FEATURES]
IDENTIFIED_CRITERIA = load_criteria_with_correct_labels()
IDENTIFIED_CRITERIA[focused_features]

```

	p	case_id	case_created_at	a__by_call	...	h	h__low_score	i__reminder_priority	i__completion_priority
0	0	3bdcd547ce0c2d8956bc32c3170bab	2021-06-23	0	...	2	3	3	3
1	0	e196a655d611dfb7d39f6e29b73982a4	2021-06-22	6	...	2	3	3	3
2	0	6a6442992c949043891d389687c509f7	2021-06-21	5	...	2	3	3	3
3	0	962700491b1342925450ed9d5e7cbe59	2021-06-20	4	...	2	3	3	3
4	0	7c10db4d1a0e52770f7b059e8dab8d9	2021-06-19	3	...	2	3	3	3
...
2311	1	1a711bcda5762ac52b32fe810537930	2023-09-18	4	...	2	0	3	1
2312	1	db5f101a4ca7fa81b3795478f36e1997	2023-09-17	3	...	2	0	3	1
2313	1	f55225964b199361addf96036e71f5e1	2023-09-16	2	...	2	0	3	1
2314	1	55117fbb9e545e27a25af2167dbb226d	2023-09-15	1	...	2	0	3	1
2315	2	854e0d00d76798e4549ebcc6a738c43	2023-09-20	0	...	2	0	3	1

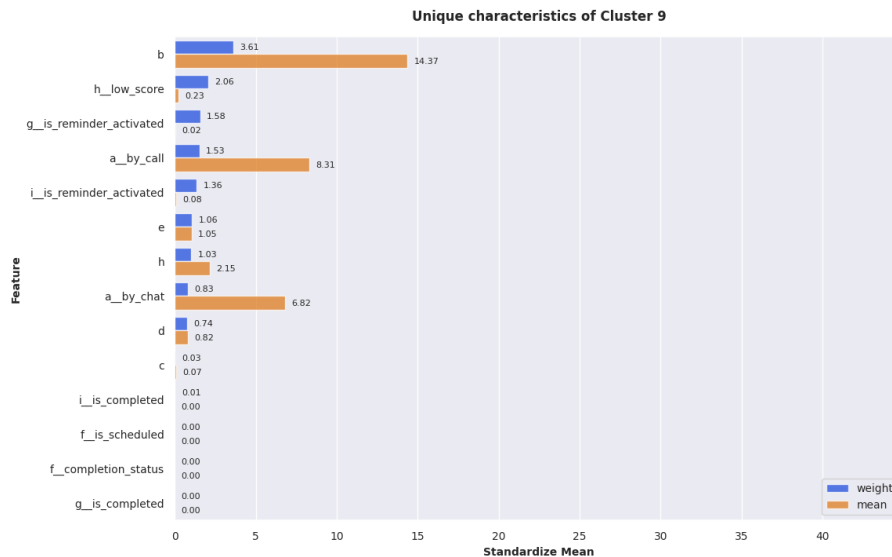
2316 rows x 17 columns

Gambar 11. Sampel data aktivitas pasien teridentifikasi Perusahaan *Mental Healthcare*.

C. Pembahasan

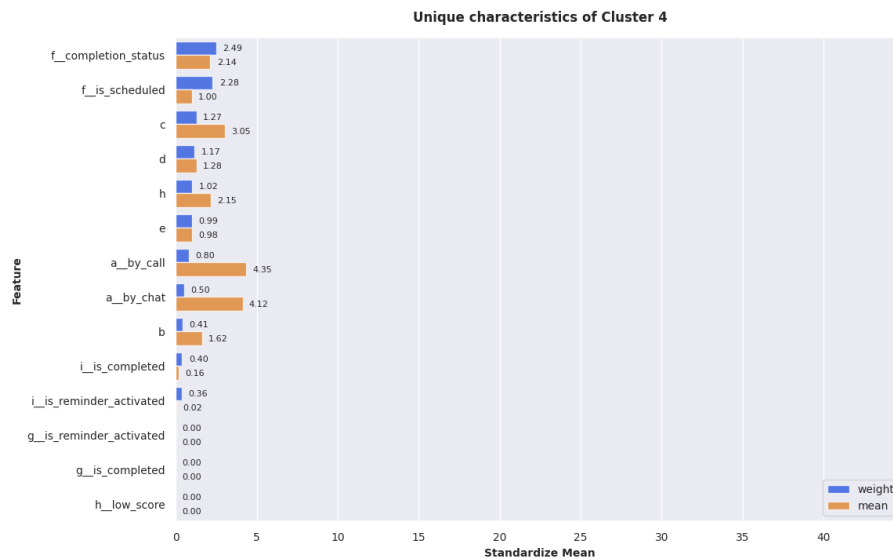
1) *Seleksi Kluster*: Hasil penelitian menunjukkan bahwa tingkat kemiripan tertinggi kluster yang merepresentasikan fase perawatan teridentifikasi oleh Perusahaan *Mental Healthcare* tidak cukup dominan / unik, dimana kluster 'c6' memiliki 22.09 % karakteristik fase awal perawatan, 15.78% karakteristik fase tengah perawatan, dan 18.99% karakteristik fase akhir perawatan. Sedangkan 'c1' memiliki 20.51% karakteristik awal perawatan, 19.78% karakteristik fase tengah perawatan, dan 14.75% karakteristik fase akhir perawatan. Hal ini jelas mengindikasikan bahwa kluster 'c6' dan 'c1' saling beririsan satu sama lain di setiap fase perawatan. Oleh karena itu, urutan afinitas kluster menjadi penting untuk dijadikan bahan pertimbangan dalam melakukan seleksi kluster, sehingga karakteristik yang benar-benar unik di setiap fase perawatan dapat ditemukan. Berdasarkan TABEL 4, diketahui bahwa kluster yang berpotensi memiliki karakteristik dominan untuk fase perawatan awal, tengah, dan akhir terdapat di urutan ketiga yaitu kluster 'c9', 'c4', dan 'c3' dengan persentase kemiripan terhadap fase awal, tengah, dan akhir perawatan teridentifikasi masing-masing sebesar 18.5%, 14.11%, dan 13.11%. Sedangkan kluster 'c8', 'c5', dan 'c2' di urutan kedelapan memiliki persentase kemiripan terhadap fase awal, tengah, dan akhir masing-masing sebesar 4.54%, 6.22%, dan 6.28%. Dari kedua opsi tersebut, urutan kluster yang dipilih adalah urutan ketiga, yaitu 'c9', 'c4', dan 'c3' karena memiliki tingkat persentase kemiripan dengan jumlah populasi yang lebih baik dari opsi kedua.

2) *Interpretasi Karakteristik Kluster*: Gambar 12 menunjukkan bahwa karakteristik fase awal perawatan dapat diinterpretasikan sebagai berikut: 1) *Pola Komunikasi*; Komunikasi pasien tidak terjadi secara intens baik melalui panggilan audio/video (*a__by_call*) ataupun pesan singkat (*a__by_chat*), dimana rata-rata jumlah hari sejak terakhir kontak untuk masing-masing jenis komunikasi adalah 8 dan 7 hari. 2) *Pola Registrasi Data*; Jumlah hari sejak terakhir pasien melakukan registrasi data (*b*) menjadi *feature* paling dominan dalam kluster 'c9', dengan nilai rata-rata sebesar 14 hari. Sedangkan jumlah registrasi data yang spesifik terkait dengan gangguan kecemasan sosial (*c*) hampir tidak ada, ditegaskan dengan nilai rata-rata *feature* sebesar 0.07. 3) *Pola Penurunan Gejala Gangguan Kecemasan Sosial*; Perubahan emosi negatif pasien (*d*) dalam seminggu terakhir dibandingkan dengan seminggu sebelumnya cenderung stabil, ditegaskan dengan nilai rata-rata *feature* sebesar 0.82. Begitupun dengan pola perubahan emosi positif pasien (*e*) dalam seminggu terakhir dibandingkan dengan seminggu sebelumnya, cenderung stabil. Hal ini ditegaskan dengan nilai rata-rata *feature* sebesar 1.05. 4) *Pola Penyelesaian Tugas*; Pasien rata-rata tidak melakukan registrasi data sekalipun dalam seminggu terakhir, baik untuk catatan harian (*i__is_completed*), ataupun thought record (*g__is_completed*). Meskipun pengingat untuk melakukan registrasi data keduanya, diaktifkan oleh sebagian kecil pasien yang ditegaskan dengan nilai rata-rata pengingat *thought record* (*g__is_reminder_activated*) dan catatan harian (*i__is_reminder_activated*) masing-masing sebesar 0.02 dan 0.08, dengan bobot tertinggi masing-masing sebesar 1.58 dan 1.36. 5) *Pola Tanggapan Relevansi Perawatan*; Beberapa pasien memberikan tanggapan dengan skor rendah pada salah satu indikator penilaian kuesioner (*h__low_score*), ditegaskan dengan nilai rata-rata sebesar 0.23, dan menjadi *feature* kedua paling dominan pada kluster 'c9' dengan nilai bobot 2.06. Tanggapan yang diberikan oleh sebagian kecil pasien terhadap relevansi perawatan cenderung mengalami peningkatan (*h*), yang ditegaskan dengan nilai rata-rata *feature* sebesar 2.15 dengan bobot 1.03.



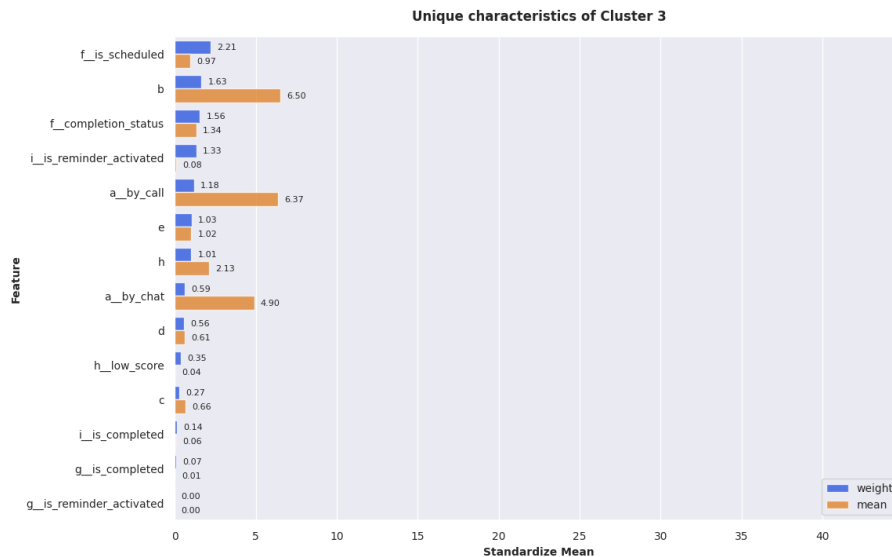
Gambar 12. Karakteristik fase awal perawatan yang direpresentasikan oleh kluster 'c9'.

Gambar 13 menunjukkan bahwa karakteristik fase tengah perawatan dapat diinterpretasikan sebagai berikut: 1) *Pola Komunikasi*; Komunikasi pasien mulai sering dilakukan baik melalui panggilan audio/video (*a_by_call*) ataupun pesan singkat (*a_by_chat*), dimana rata-rata jumlah hari sejak terakhir kontak untuk keduanya adalah 4 hari. 2) *Pola Registrasi Data*; Pasien cukup intens melakukan registrasi data yang ditegaskan dengan jumlah hari sejak terakhir pasien melakukan registrasi data (*b*) adalah 2 hari. Sedangkan rata-rata jumlah registrasi data yang spesifik terkait dengan gangguan kecemasan sosial (*c*) adalah 3. 3) *Pola Penurunan Gejala Gangguan Kecemasan Sosial*; Perubahan emosi negatif pasien (*d*) dalam seminggu terakhir dibandingkan dengan seminggu sebelumnya memiliki kecenderungan mengalami sedikit peningkatan, ditegaskan dengan nilai rata-rata *feature* sebesar 1.28. Sedangkan pola perubahan emosi positif pasien (*e*) dalam seminggu terakhir dibandingkan dengan seminggu sebelumnya, belum sepenuhnya menuju stabil, yang ditegaskan dengan nilai rata-rata *feature* sebesar 0.98. 4) *Pola Penyelesaian Tugas*; Setiap pasien umumnya memiliki satu aktivitas terjadwal (*f_is_scheduled*) dan mampu menyelesaikan minimal satu aktivitas terjadwal (*f_completion_status*) dalam seminggu terakhir, ditegaskan dengan nilai rata-rata untuk masing-masing *feature* sebesar 1.00 dan 2.1, dimana keduanya merupakan *feature* paling dominan dalam kluster 'c4'. Kemudian, sebagian kecil pasien mulai memiliki kecenderungan untuk melakukan registrasi data catatan harian (*i_is_completed*) minimal satu dalam seminggu terakhir, ditegaskan dengan nilai rata-rata *feature* sebesar 0.16. Meskipun pengingat untuk melakukan registrasi data catatan harian (*i_is_reminder_activated*) umumnya tidak diaktifkan oleh pasien, ditegaskan dengan nilai rata-rata *feature* sebesar 0.02. 5) *Pola Tanggapan Relevansi Perawatan*; Tanggapan dengan skor rendah pada salah satu indikator penilaian kuesioner (*h_low_score*) tidak diberikan pasien sama sekali, ditegaskan dengan nilai rata-rata *feature* dan bobotnya masing-masing bernilai 0. Sedangkan, tanggapan yang diberikan sebagian kecil pasien terhadap relevansi perawatan cenderung mengalami peningkatan (*h*), yang ditegaskan dengan nilai rata-rata *feature* sebesar 2.15 dengan bobot 1.02.



Gambar 13. Karakteristik fase tengah perawatan yang direpresentasikan oleh kluster 'c4'.

Gambar 14 menunjukkan bahwa karakteristik fase akhir perawatan dapat diinterpretasikan sebagai berikut: 1) *Pola Komunikasi*; Komunikasi pasien mulai jarang dilakukan baik melalui panggilan audio/video (*a_by_call*) ataupun pesan singkat (*a_by_chat*), ditegaskan dengan nilai rata-rata jumlah hari untuk keduanya masing-masing berjumlah 6 dan 5 hari. 2) *Pola Registrasi Data*; Jumlah hari sejak terakhir pasien melakukan registrasi data (*b*) menjadi *feature* kedua paling dominan dalam kluster 'c3', dengan nilai rata-rata 7 hari dan ditegaskan dengan bobot *feature* sebesar 1.63. Sedangkan rata-rata jumlah registrasi data yang spesifik terkait dengan gangguan kecemasan sosial (*c*) adalah 1. 3) *Pola Penurunan Gejala Gangguan Kecemasan Sosial*; Perubahan emosi negatif pasien (*d*) dalam seminggu terakhir dibandingkan dengan seminggu sebelumnya memiliki kecenderungan mengalami penurunan, ditegaskan dengan nilai rata-rata *feature* sebesar 0.61. Sedangkan pola perubahan emosi positif pasien (*e*) dalam seminggu terakhir dibandingkan dengan seminggu sebelumnya adalah stabil, yang ditegaskan dengan nilai rata-rata *feature* sebesar 1.02. 4) *Pola Penyelesaian Tugas*; Setiap pasien umumnya memiliki satu aktivitas terjadwal (*f_is_scheduled*), namun dalam seminggu terakhir, sebagian besar pasien tidak menyelesaikannya (*f_completion_status*), ditegaskan dengan nilai rata-rata untuk masing-masing *feature* sebesar 0.97 dan 1.34, dimana keduanya merupakan *feature* paling dominan dalam kluster 'c3'. Selain itu, dalam seminggu terakhir pasien rata-rata tidak lagi melakukan registrasi catatan harian (*i_is_completed*), ataupun *thought record* (*g_is_completed*). 5) *Pola Tanggapan Relevansi Perawatan*; Tanggapan dengan skor rendah pada salah satu indikator penilaian kuesioner (*h_low_score*) hampir tidak diberikan pasien sama sekali, ditegaskan dengan nilai rata-rata *feature* dan bobotnya masing-masing bernilai 0.04 dan 0.35. Sedangkan, tanggapan yang diberikan sebagian kecil pasien terhadap relevansi perawatan cenderung mengalami peningkatan (*h*), yang ditegaskan dengan nilai rata-rata *feature* sebesar 2.13 dengan bobot 1.01.



Gambar 14. Karakteristik fase akhir perawatan yang direpresentasikan oleh kluster 'c3'.

3) *Interpretasi Features Dominan di Setiap Kluster*: Apabila pembahasan interpretasi karakteristik kluster disederhanakan berdasarkan *features* dominannya, maka *feature* spesifik yang menjadi pembeda di setiap fase perawatan adalah (b), (c), (d), (h), dan (h_low_score), dimana kelima *feature* tersebut memiliki tingkat keselarasan paling tinggi dengan proses perawatan yang berlangsung pada setiap fase di Perusahaan *Mental Healthcare*, khususnya (h) dan (h_low_score), dimana setiap terapis di perusahaan wajib untuk segera melakukan *follow up* kepada pasien yang memberikan skor rendah terhadap salah satu indikator kuesioner relevansi perawatan. Artinya, *feature* (h) dan (h_low_score) memiliki tingkat urgensi paling tinggi bagi terapis untuk memastikan semua pasien mendapatkan penanganan yang sesuai dengan keluhan yang dideritanya di sepanjang fase perawatan. Sehingga, apabila setiap *feature* paling dominan di setiap kluster diurutkan berdasarkan tingkat urgensinya, maka akan didapatkan daftar urutan tingkat urgensi *feature* sebagai berikut: 1) Tanggapan dengan skor rendah di salah satu indikator penilaian kuesioner (h_low_score), 2) Tanggapan keseluruhan terhadap relevansi perawatan (h), 3) Jumlah hari sejak terakhir pasien melakukan registrasi data (b), 4) Jumlah registrasi data yang berkaitan dengan gangguan kecemasan sosial dalam seminggu terakhir (c), dan 5) Perubahan emosi negatif pasien dalam seminggu terakhir dibandingkan dengan seminggu sebelumnya (d). Namun apabila merujuk pada Gambar 12 terkait fase awal perawatan dan Gambar 14 untuk fase akhir perawatan, nilai rata-rata *feature* (h_low_score) tidak secara tegas didefinisikan dengan nilai 1 di fase awal atau 0 di fase akhir, meskipun dampaknya berpengaruh besar terhadap cara kerja terapis pada saat melakukan sesi terapi. Oleh karena itu, besar kemungkinan terdapat *feature* lain yang 'tersembunyi' dan tidak dapat dikelompokkan oleh mesin sebagai *feature* dominan dengan dampak signifikan yang menentukan tingkat urgensi perawatan pasien.

IV. SIMPULAN

Secara garis besar kesimpulan yang didapatkan adalah data aktivitas pasien dengan gangguan kecemasan sosial berhasil dikelompokkan menjadi 10 kluster, dimana 3 diantaranya dianggap memiliki karakteristik paling merepresentasikan fase perawatan yang terjadi di Perusahaan *Mental Healthcare*, yaitu 'c9' untuk fase awal perawatan, 'c4' untuk fase tengah perawatan, dan 'c3' untuk fase akhir perawatan dengan afinitas kluster masing-masing sebesar 18.5%, 14.11%, 13.11%. Setiap fase perawatan dapat dibedakan berdasarkan pola aktivitas pasien dalam melakukan registrasi data, pola penurunan gejala gangguan kecemasan sosial yang ditandai dengan perubahan emosi pasien dalam dua minggu terakhir, dan pola skor tanggapan yang dikirim pasien terkait relevansi perawatan. Sedangkan pola komunikasi pasien dan pola penyelesaian tugas yang harus dilakukan pasien dalam seminggu terakhir, cenderung tidak dapat dijadikan acuan untuk membedakan fase perawatan pasien karena fase awal dan akhir perawatan memiliki karakteristik yang mirip pada kedua pola tersebut. *Feature* paling dominan dalam pola aktivitas registrasi data yang dilakukan pasien adalah jumlah hari sejak terakhir pasien melakukan registrasi data dan jumlah registrasi data terkait gangguan kecemasan sosial dalam seminggu terakhir. Pada pola penurunan gangguan kecemasan sosial, *feature* paling dominan adalah perubahan emosi negatif pasien dalam seminggu terakhir dibandingkan dengan seminggu sebelumnya. Sedangkan pada pola skor tanggapan kuesioner, *feature* paling dominannya adalah tanggapan keseluruhan terhadap relevansi perawatan dan tanggapan dengan skor rendah di salah satu indikator penilaian kuesioner. Urutan tingkat urgensi *features* yang ditemukan dapat dijadikan sebagai wawasan bagi Perusahaan *Mental Healthcare* dalam menentukan rentang nilai yang mendefinisikan level urgensi pasien secara keseluruhan (misalnya urgen, sedang, atau tidak urgen) dari penerapan aturan-aturan yang terbentuk pada pola pembobotan *feature* di setiap kluster.

Untuk penelitian selanjutnya, akan lebih baik lagi apabila pendekatan *Latent Profile Analysis* (LPA) dilakukan dengan melibatkan tanggapan dari terapis dalam mengkaji karakteristik setiap *feature* dari dua kluster dengan persentase afinitas tertinggi, yaitu 'c6' dan 'c1', karena kedua kluster tersebut saling beririsan satu sama lain di setiap fase perawatan teridentifikasi (awal, tengah, dan akhir). Selain itu, pada saat membangun sistem rekomendasi penanganan pasien berdasarkan tingkat urgensinya, eksplorasi korelasi dengan memasukan efek "delay" (*delayed correlation*) dinilai akan sangat menjanjikan sebelum membuat model klasifikasi, melihat fakta bahwa hasil terapi terhadap pasien tidak akan terlihat secara langsung, namun akan terlihat di sesi selanjutnya.

DAFTAR PUSTAKA

- [1] F. Schneier and G. Goldmark, Social anxiety disorder. [Online] Available: https://www.researchgate.net/publication/283836688_Social_Anxiety_Disorder.
- [2] M. S. O'Toole, M. Arendt, H. N. Fentz, E. Hougaard and N. K. Rosenberg, "Cluster A personality pathology in social anxiety disorder: a comparison with panic disorder," *Nordic journal of psychiatry*, vol. 68, no. 7, pp. 460-463, 2014.
- [3] T. L. Rodebaugh, R. M. Holaway and R. G. Heimberg, "The treatment of social anxiety disorder," *Clinical psychology review*, vol. 24, no. 7, pp. 883-908, 2004.
- [4] D. J. Stein, C. C. W. Lim, A. M. Roest, P. de Jonge, S. Aguilar-Gaxiola, A. Al-Hamzawi, J. Alonso, C. Benjet, E. J. Bromet, R. Bruffaerts, G. de Girolamo, S. Florescu, O. Gureje and J. M. Haro, "The cross-national epidemiology of social anxiety disorder: Data from the World Mental Health Survey Initiative," *BMC Medicine*, vol. 15, 2017.
- [5] E. Pintelas, T. Kotsilieris, I. Livieris and P. Pintelas, "A review of machine learning prediction methods for anxiety disorders," p. 8–15, 2018.
- [6] W. Eng, M. E. Coles, R. G. Heimberg and S. A. Safren, "Domains of life satisfaction in social anxiety disorder: relation to symptoms and response to cognitive-behavioral therapy," *Journal of Anxiety Disorders*, vol. 19, pp. 143-156, 2005.
- [7] J. A. Piqueras, J. Olivares and J. A. López-Pina, "A new proposal for the subtypes of social phobia in a sample of Spanish adolescents," *Journal of Anxiety Disorders*, vol. 22, no. 1, pp. 67-77, 2008.
- [8] Yang, Zhenkai and Chen, Chuansheng and Li, Hanwen and Yao, Li and Zhao and Xiaojie, "Unsupervised Classifications of Depression Levels Based on Machine Learning Algorithms Perform Well as Compared to Traditional Norm-Based Classifications," *Frontiers in Psychiatry*, vol. 11, p. 45, 2020.
- [9] Liu, Fang and Yang, Dan and Liu, Yueguang and Zhang, Qin and Chen, Shiyu and Li, Wanxia and Ren, Jidong and Tian, Xiaobin and Wang and Xin, "Use of latent profile analysis and k-means clustering to identify student anxiety profiles," *BMC Psychiatry*, vol. 22, 2022.
- [10] I. Consulting, "General Data Protection Regulation," [Online]. Available: <https://gdpr-info.eu/>. [Accessed 3 June 2023].
- [11] H. Toba, E. Wianto, M. Malinda, A. W. Al Halim and C.-H. Chen, "Cluster Dominance Analysis of Strength Training Motion Characteristics," in *IEEE 12th Global Conference on Consumer Electronics (GCCE)*, 2023.
- [12] H. Karakala, A. K. Gogineni, V. Thadi, S. Tokala, M. K. Enduri and S. Anamalamudi, "Performance Evaluation of Machine Learning and Neural Network Algorithms for Wine Quality Prediction," in *International Conference on Computing and Networking Technology (ICCNT)*, 2023.
- [13] L. Rokach and O. Maimon, "Clustering Method," *Data Mining and Knowledge Discovery Handbook*, pp. 321-352, 2005.
- [14] R. G. Dias, S. Vieira, W. H. L. Pinaya and A. Mechelli, "Clustering Analysis," *Academic Press*, pp. 227-247, 2020.
- [15] A. P. Bello, E. Barbour, M. C. Gonzalez, S. Yilmaz, M. K. Patel and D. Parra, "How Does the Electricity Demand Profile Impact Attractiveness of PV-Coupled Battery System Combining Applications?," *Energies*, vol. 13, p. 4038, 2020.
- [16] A. Starczewski and A. Krzyżak, "Performance Evaluation of the Silhouette Index," in *International Conference on Artificial Intelligence and Soft Computing*, Switzerland, 2015.
- [17] G. Gan, C. Ma and J. Wu, *Data Clustering: Theory, Algorithms, and Applications*, Society for Industrial and Applied Mathematics, 2007.
- [18] S. Na, L. Xumin and G. Yong, "Research on k-means Clustering Algorithm: An Improved k-means Clustering Algorithm," 2010 Third International Symposium on Intelligent Information Technology and Security Informatics, Jian, China, 2010.
- [19] Y. Li and H. Wu, "A Clustering Method Based on K-Means Algorithm," *Physics Procedia*, vol. 25, pp. 1104-1109, 2012.
- [20] G. Bonaccorso, *Hands-On Unsupervised Machine Learning with Python*, Birmingham: Packt Publishing Ltd., 2019.
- [21] A. Maćkiewicz and W. Ratajczak, "Principal components analysis (PCA)," *Computers & Geosciences*, vol. 19, pp. 303-342, 1993.