**LAPORAN AKHIR**

**Penerapan K-Means Clustering dalam Mengelompokkan Siswa**

**Berdasarkan Gejala Psikologis SRQ-20**

Diajukan untuk memenuhi proyek akhir

Mata Kuliah Pembelajaran Mesin



**Disusun Oleh:**

1. **Muhammad Dyaka Faiz Yuhendra A11.2022.14712**
2. **Muhamad Aziz Zahrul Zamzami A11.2023.15156**
3. **Nathanael Christian Arifana A11.2023.15289**
4. **Noor Adekah Apriyana**  **A11.2022.14382**
5. **Anggita Alya Salsabila A11.2022.14401**

**A11.4409**

**FAKULTAS ILMU KOMPUTER**

**UNIVERSITAS DIAN NUSWANTORO**

**SEMARANG**

**2024/2025**

# DAFTAR ISI

[DAFTAR ISI 2](#_Toc202348888)

[DAFTAR GAMBAR 4](#_Toc202348889)

[DAFTAR TABEL 5](#_Toc202348890)

[BAB I PENDAHULUAN 6](#_Toc202348891)

[1.1. Latar Belakang 6](#_Toc202348892)

[1.2. Rumusan Masalah 6](#_Toc202348893)

[1.3. Tujuan Penelitian 6](#_Toc202348894)

[1.4. Batasan Penelitian 7](#_Toc202348895)

[1.5. Manfaat Penelitian 7](#_Toc202348896)

[1.5.1. Bagi Peneliti 7](#_Toc202348897)

[1.5.2. Bagi Akademi 7](#_Toc202348898)

[1.5.3. Bagi Objek Penelitian 7](#_Toc202348899)

[1.5.4. Bagi Masyarakat 8](#_Toc202348900)

[BAB II LANDASAN TEORI 9](#_Toc202348901)

[3.1. Tinjauan Studi 9](#_Toc202348902)

[2.2. Dasar Teori 11](#_Toc202348903)

[2.2.1. *Self Reporting Questionnaire* (SQR-20) 11](#_Toc202348904)

[2.2.2. Psikologi 11](#_Toc202348905)

[2.2.3. Klastering 11](#_Toc202348906)

[2.2.4. Algoritma K-Means 12](#_Toc202348907)

[2.2.5. *Principal Component Analysis* (PCA) 12](#_Toc202348908)

[BAB III METODOLOGI PENELITIAN 14](#_Toc202348909)

[3.1. Lingkungan Pengembangan 14](#_Toc202348910)

[3.2. Sumber dan Bentuk Data 14](#_Toc202348911)

[3.3. Data Understanding 14](#_Toc202348912)

[3.4. Data Preparation 14](#_Toc202348913)

[3.5. Modelling 15](#_Toc202348914)

[3.6. Evaluation 15](#_Toc202348915)

[3.7. Deployment 15](#_Toc202348916)

[BAB IV HASIL DAN PEMBAHASAN 16](#_Toc202348917)

[4.1. Hasil Klasterisasi 16](#_Toc202348918)

[4.2. Visualisasi Klaster 17](#_Toc202348919)

[4.3. Evaluasi Cutoff WHO dan Validasi Interna 17](#_Toc202348920)

[4.4. Implementasi Sistem Interaktif 18](#_Toc202348921)

[BAB V PENUTUP 21](#_Toc202348922)

[5.1. Kesimpulan 21](#_Toc202348923)

[5.2. Saran 21](#_Toc202348924)

[DAFTAR PUSTAKA 23](#_Toc202348925)

# DAFTAR GAMBAR

[Gambar 4. 1 Silhouette Score untuk Berbagai Nilai k 17](#_Toc202378743)

[Gambar 4. 2 Pemetaan Klaster Berdasarkan Nilai Rata-rata TOTAL\_YA 18](#_Toc202378744)

[Gambar 4. 3 Scatter plot PCA hasil klasterisasi untuk k = 2 18](#_Toc202378745)

[Gambar 4. 4 Menentukan Klaster Bergejala Bedasarkan Nilai Rata-rata 19](#_Toc202378746)

[Gambar 4. 5 Tampilan Dashboard Prediksi SRQ-20 20](#_Toc202378747)

[Gambar 4. 6 Tampilan Dashboard Prediksi SRQ-20 21](#_Toc202378748)

[Gambar 4. 7 Tampilan Dashboard Prediksi SRQ-20 21](#_Toc202378749)

# DAFTAR TABEL

[Tabel 2. 1 State of the Art 10](#_Toc202354894)

# BAB I PENDAHULUAN

## Latar Belakang

Masalah kesehatan mental telah menjadi perhatian global, terutama pada kelompok remaja dan mahasiswa yang rentan terhadap tekanan akademik, sosial, dan emosional.Masa remaja merupakan periode transisi dari anak-anak menuju dewasa, yang ditandai dengan perubahan fisik, kognitif, sosial, dan emosional yang kompleks. Ketidakseimbangan dalam proses ini dapat meningkatkan risiko gangguan psikologis seperti stres, depresi, dan kecemasan (Dara Safitri et al., 2024).

Salah satu instrumen yang banyak digunakan untuk mendeteksi gangguan psikologis adalah *Self Reporting Questionnaire* (SRQ-20), yang disusun oleh World Health Organization (WHO) dan telah divalidasi dalam berbagai konteks budaya, termasuk Indonesia (Fajriyah et al., 2023). SRQ-20 terdiri dari 20 item yang menyasar gejala somatik, emosional, dan perilaku, dan telah terbukti memiliki reliabilitas tinggi serta dapat digunakan dalam skrining awal gangguan mental pada populasi remaja maupun dewasa (Dara Safitri et al., 2024; Rachmat et al., 2022).

Namun, analisis manual terhadap hasil SRQ-20 masih memiliki keterbatasan dalam mengelompokkan responden berdasarkan tingkat risiko psikologis. Oleh karena itu, pendekatan berbasis pembelajaran mesin, seperti algoritma K-Means Clustering, menawarkan solusi otomatis dan objektif untuk mengelompokkan siswa berdasarkan pola jawaban mereka. Dengan memanfaatkan algoritma ini, konselor atau pihak terkait dapat lebih mudah melakukan deteksi dini, klasifikasi risiko, dan tindak lanjut yang tepat.

## Rumusan Masalah

Setelah pemaparan pada latar belakang, maka diperoleh rumusan masalah:

1. Bagaimana cara mengelompokkan siswa berdasarkan hasil pengisian SRQ-20 menggunakan metode K-Means Clustering?
2. Bagaimana hasil visualisasi klaster dapat mendukung pemahaman kondisi psikologis siswa secara lebih efektif?

## Tujuan Penelitian

Dari rumusan masalah tersebut maka akan didapat tujuan dari penelitian ini, yaitu:

1. Mengimplementasikan algoritma K-Means Clustering untuk mengelompokkan siswa berdasarkan gejala psikologis dari SRQ-20.
2. Menyajikan hasil klaster dalam bentuk visualisasi yang informatif untuk mendukung pengambilan keputusan.

## Batasan Penelitian

Untuk menjaga fokus dan keterjangkauan ruang lingkup, penelitian ini dibatasi pada:

1. Penggunaan data hasil kuesioner SRQ-20 dari siswa dalam bentuk jawaban ya/tidak.
2. Penggunaan algoritma K-Means untuk metode pengelompokan.
3. Visualisasi dilakukan menggunakan teknik PCA (Principal Component Analysis).

## Manfaat Penelitian

### Bagi Peneliti

1. Menambah pengalaman dan pemahaman mengenai penerapan algoritma K-Means dalam bidang psikologi pendidikan.
2. Menjadi sarana eksplorasi dalam integrasi antara psikologi dan teknologi.

### Bagi Akademi

1. Memberikan referensi penelitian interdisipliner dalam pengolahan data gejala psikologis dengan pendekatan machine learning.
2. Menjadi dasar bagi penelitian lanjutan dengan algoritma yang lebih kompleks.

### Bagi Objek Penelitian

1. Memberikan gambaran yang lebih jelas mengenai kondisi psikologis pribadi berdasarkan hasil analisis data SRQ-20, sehingga siswa dapat memahami potensi risiko psikologis yang mungkin mereka alami.
2. Menjadi media refleksi diri untuk meningkatkan kesadaran akan pentingnya menjaga kesehatan mental, terutama di lingkungan pendidikan yang penuh tekanan akademik dan sosial.
3. Memberikan akses terhadap pendekatan teknologi yang informatif dan objektif sebagai alternatif dari metode tradisional yang biasanya bersifat subjektif atau memerlukan wawancara langsung dengan konselor.
4. Membuka peluang bagi siswa untuk mendapatkan intervensi atau pendampingan lebih awal, terutama jika mereka termasuk dalam klaster risiko sedang atau tinggi.
5. Mendorong terciptanya lingkungan belajar yang lebih suportif dan responsif, karena hasil klasifikasi dapat menjadi acuan bagi lembaga pendidikan dalam merancang program dukungan emosional dan mental yang lebih terarah.

### Bagi Masyarakat

1. Menyediakan *sarana pendeteksian dini risiko gangguan psikologis* yang berbasis teknologi, yang dapat diadaptasi oleh lembaga pendidikan, layanan kesehatan mental, maupun komunitas untuk meningkatkan kualitas layanan psiko edukatif.
2. Menumbuhkan *kesadaran kolektif akan pentingnya kesehatan mental*, terutama pada generasi muda, sehingga masyarakat lebih tanggap terhadap gejala psikologis ringan sebelum berkembang menjadi gangguan serius.
3. Mendorong partisipasi masyarakat dalam menciptakan *lingkungan sosial yang suportif dan inklusif*, terutama di sekolah, rumah, dan komunitas, sebagai upaya preventif terhadap tekanan mental berlebih.
4. Mengurangi stigma terhadap isu gangguan psikologis dengan menghadirkan pendekatan berbasis data dan analisis, sehingga masyarakat dapat memahami masalah mental secara lebih rasional dan terbuka.
5. Menjadi *model pengembangan sistem berbasis machine learning* yang bisa direplikasi dalam isu sosial lainnya seperti pendidikan karakter, monitoring kesejahteraan siswa, atau deteksi risiko perilaku menyimpang.

# BAB II LANDASAN TEORI

## Tinjauan Studi

Untuk mendukung penyusunan landasan teori, penulis mengacu pada beberapa studi sebelumnya yang memiliki keterkaitan dengan topik penelitian ini. Studi-studi tersebut digunakan sebagai referensi komparatif, khususnya yang membahas penerapan *Self Reporting Questionnaire* (SRQ-20) dalam deteksi gangguan psikologis serta penggunaan metode klastering seperti K-Means dalam pengelompokan risiko psikologis.

Salah satu penelitian dilakukan oleh (Dara Safitri et al., 2024) menguji validitas dan reliabilitas versi modifikasi SRQ-20 pada siswa SMA di Surakarta. Instrumen ini terbukti valid dengan nilai korelasi antar item sebesar r = 0,708–0,941 dan reliabel dengan Cronbach’s Alpha sebesar 0,981. Penelitian ini menunjukkan bahwa SRQ-20 dapat digunakan secara efektif untuk mengukur kondisi mental remaja di lingkungan pendidikan.

Penelitian lainnya oleh (Fajriyah et al., 2023)yang menganalisis kualitas psikometris SRQ-20 menggunakan model Rasch terhadap 165 responden berusia 12–44 tahun. Hasilnya menunjukkan bahwa seluruh item SRQ-20 memenuhi syarat item-model fit dengan reliabilitas tinggi (α = 0,823), menunjukkan bahwa SRQ-20 layak digunakan untuk skrining kesehatan mental di Indonesia.

Selain itu, (Prasetio et al., 2022)juga melakukan analisis faktor terhadap SRQ-20 versi Indonesia dengan metode *Confirmatory Factor Analysis* (CFA) dan *Exploratory Factor Analysis* (EFA). Hasilnya menunjukkan bahwa SRQ-20 dalam konteks Indonesia terbagi ke dalam lima faktor utama: Energi, Kognitif, Depresi, Fisiologis, dan Kecemasan. Penelitian ini memperkuat bahwa SRQ-20 adalah instrumen yang valid untuk mengukur gangguan mental baik dalam bentuk satu dimensi umum maupun dalam lima dimensi khusus.

Berbagai penelitian tersebut menunjukkan bahwa SRQ-20 merupakan alat skrining yang valid dan reliabel, serta dapat dikombinasikan dengan pendekatan *machine learning* seperti K-Means Clustering untuk meningkatkan efektivitas deteksi gangguan psikologis berbasis data.

Tabel 2. 1 State of the Art

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **No** | **Judul** | **Tahun** | **Metode** | **Hasil** |
| 1. | Analisis Validitas Self Reporting Questionnaire (SQR) Terhadap Kesehatan Mental Pada Remaja (Dara Safitri et al., 2024) | 2024 | Pearson Correlation & Cronbach Alpha | Hasil uji validitas menunjukkan korelasi antar item SRQ-20 dalam rentang r = 0,708–0,941, dan nilai reliabilitas Cronbach Alpha sebesar 0,981. Ini membuktikan SRQ-20 valid dan sangat reliabel untuk siswa SMA. |
| 2. | Psychometric Properties Analysis of Self-Reported Questionnaire (SRQ)-20 Instrument with Rasch Model (Fajriyah et al., 2023) | 2023 | Rasch Model | SRQ-20 menunjukkan validitas item yang tinggi dan reliabilitas keseluruhan sebesar α = 0,823. Semua item sesuai dengan model Rasch, menunjukkan kesesuaian antara teori dan data responden usia 12–44 tahun. |
| 3. | Psychometric Properties of Self-Report Questionnaire-20 (SRQ-20) Indonesian Version (Prasetio et al., 2022) | 2022 | CFA & EFA | Analisis faktor eksploratori dan konfirmatori mengidentifikasi lima faktor utama: Energi, Kognitif, Depresi, Fisiologis, dan Kecemasan. Ini memperkuat struktur multidimensi SRQ-20 di populasi Indonesia. |

## Dasar Teori

### *Self Reporting Questionnaire* (SRQ-20)

*Self Reporting Questionnaire* (SRQ-20) adalah alat skrining gangguan mental non-psikotik yang dikembangkan oleh World Health Organization (WHO). Instrumen ini terdiri dari 20 pertanyaan dengan format ya/tidak yang bertujuan mendeteksi gejala depresi, kecemasan, stres, dan gangguan somatik dalam kurun waktu 30 hari terakhir. Beberapa penelitian menyatakan bahwa SRQ-20 memiliki struktur faktor yang terdiri dari energi, kognitif, depresi, fisiologis, dan kecemasan (Fajriyah et al., 2023; Prasetio et al., 2022). Selain itu, SRQ-20 juga terbukti valid dan reliabel dalam konteks remaja (Dara Safitri et al., 2024).

### Psikologi

Psikologi adalah studi ilmiah tentang perilaku dan proses mental manusia. Psikologi mempelajari bagaimana individu berpikir, merasakan, dan berperilaku dalam berbagai situasi. Dalam konteks pendidikan, kesehatan mental siswa sangat berperan dalam pencapaian akademik dan perkembangan sosial. Masalah psikologis seperti stres akademik dan kecemasan sosial dapat berdampak negatif terhadap performa siswa. Oleh karena itu, pemantauan kondisi psikologis siswa melalui pendekatan kuantitatif sangat diperlukan (Dara Safitri et al., 2024).

### Klastering

Klastering adalah proses pengelompokan objek yang mirip ke dalam kelompok yang terpisah, atau lebih khusus lagi, membagi kumpulan data menjadi subset sehingga informasi di setiap subset memiliki arti yang berguna. Sebuah lokasi dalam sebuah klaster terdiri dari sekumpulan objek yang mirip satu sama lain tetapi berbeda dengan objek di klaster lain. Klastering dapat berperan penting dalam kehidupan sehari-hari, karena tidak terlepas dari banyaknya informasi yang menghasilkan informasi yang memenuhi kebutuhan hidup. Salah satu cara terpenting untuk mengolah data adalah mengkategorikan atau mengelompokkannya ke dalam kategori atau klaster. Ada banyak aplikasi di berbagai bidang yang mendukung klastering. Misalnya data mining, pengenalan pola, analisis citra dan bioinformatika adalah contoh metode pengumpulan data yang digunakan untuk analisis statistic data (Indraputra & Fitriana, 2020).

### Algoritma K-Means

Algoritma K-Means merupakan salah satu metode *unsupervised learning* yang digunakan secara luas dalam pengelompokan data berdasarkan kesamaan karakteristik. Proses kerja algoritma ini melibatkan pembentukan sejumlah klaster berdasarkan nilai centroid awal yang ditentukan secara acak. Setiap data akan dihitung jaraknya terhadap semua centroid, dan selanjutnya diklasifikasikan ke dalam klaster dengan jarak terdekat. Setelah proses klasifikasi awal, posisi centroid diperbarui berdasarkan rata-rata dari seluruh anggota klaster. Proses ini akan terus berulang hingga tidak terjadi lagi perubahan signifikan, atau hingga mencapai konvergensi.

Keunggulan K-Means adalah kecepatannya dalam memproses data dalam jumlah besar dan kemampuannya menghasilkan klaster yang representatif apabila data memiliki pola tersembunyi yang jelas. Namun demikian, kelemahan dari metode ini terletak pada pemilihan jumlah klaster (*k*) yang harus ditentukan terlebih dahulu dan sensitivitas terhadap nilai inisialisasi centroid. Dalam penelitian (Yunita Sari et al., n.d.), algoritma K-Means digunakan untuk mengelompokkan mahasiswa berdasarkan kesejahteraan dan tekanan psikologis menjadi tiga klaster utama, yaitu klaster “tertekan”, “netral”, dan “bahagia”. Hasil evaluasi menggunakan Silhouette Coefficient menunjukkan struktur klaster yang tergolong sedang hingga kuat, dengan skor antara 0,39 hingga 0,60, yang menunjukkan efektivitas metode ini dalam mengidentifikasi risiko psikologis berbasis data survei mahasiswa.

### *Principal Component Analysis* (PCA)

Principal Component Analysis (PCA) merupakan metode statistik yang digunakan untuk mereduksi dimensi data dengan cara mengubah variabel-variabel berkorelasi menjadi sekumpulan komponen utama yang saling tidak berkorelasi. Metode ini pertama kali diperkenalkan oleh Karl Pearson pada tahun 1901 dan dikembangkan lebih lanjut oleh Harold Hotelling. PCA bertujuan menyederhanakan struktur data kompleks tanpa kehilangan banyak informasi penting, sehingga memudahkan analisis dan visualisasi.

Tahapan utama PCA meliputi normalisasi data, pembentukan matriks kovarians, perhitungan nilai dan vektor eigen, pemilihan komponen utama berdasarkan nilai eigen terbesar, dan proyeksi data ke ruang berdimensi lebih rendah. Komponen yang dipilih menggambarkan arah variansi terbesar dalam data, sehingga informasi utama tetap terjaga.

Dalam penelitian ini, PCA digunakan untuk memvisualisasikan hasil klastering dari kuesioner SRQ-20, yang terdiri dari 20 item. Melalui proyeksi ke dua dimensi, pola distribusi antar siswa dalam klaster dapat divisualisasikan secara lebih jelas dan intuitif. PCA terbukti efektif mendukung proses klasifikasi berbasis machine learning dan telah banyak diterapkan di berbagai bidang seperti kesehatan, pendidikan, dan keuangan (Alaudin Shalih et al., 2025).

# BAB III METODOLOGI PENELITIAN

## Lingkungan Pengembangan

Penelitian ini dijalankan di platform Google Colaboratory (Colab) yang memungkinkan eksekusi kode Python secara daring tanpa instalasi lokal. Colab terintegrasi dengan Google Drive, sehingga memudahkan penyimpanan dan akses dataset, model .pkl, serta hasil analisis secara kolaboratif (Nazar, 2024). Dengan dukungan pustaka seperti pandas, scikit-learn, matplotlib, seaborn, dan joblib, Colab sangat andal untuk melakukan *preprocessing*, pelatihan model K‑Means, serta penyimpanan dan pemuatan kembali model secara efisien. Selain itu, framework Streamlit digunakan untuk membuat antarmuka web interaktif yang memungkinkan pengguna menginput data SRQ‑20 dan melihat hasil klasterisasi secara langsung. Aplikasi ini kemudian di-*deploy* menggunakan pyngrok sebagai terowongan dari Colab ke alamat publik, sebuah metode yang banyak dipraktikkan dan terbukti efektif dalam berbagai riset berbasis Streamlit di Indonesia, seperti pada sistem visualisasi penghitungan kerumunan menggunakan Streamlit (Prihandoko et al., 2025).

### 3.2. Sumber dan Bentuk Data

Penelitian ini menggunakan data primer dari kuesioner SRQ‑20 yang diisi oleh siswa melalui formulir. SRQ‑20 merupakan alat skrining kesehatan mental yang telah divalidasi dalam konteks Indonesia dan teruji reliabel untuk mendeteksi gejala gangguan emosional dan psikosomatik pada remaja (Dara Safitri et al., 2024). Kuesioner terdiri dari 20 pertanyaan dengan respons biner (0 = tidak, 1 = ya), sesuai standar WHO. Instrumen ini digunakan untuk memperoleh skor TOTAL\_YA sebagai indikator jumlah gejala yang dialami responden. Skor tersebut kemudian dipakai sebagai variabel utama dalam proses klasterisasi menggunakan K‑Means.

### 3.3. Data Understanding

Setelah data dikumpulkan, proses *data understanding* dimulai dengan melakukan eksplorasi awal, melihat lima baris pertama dataset dan memeriksa tipe data tiap kolom untuk memastikan kesesuaian format (Anggraini et al., 2022). Selanjutnya, dilakukan pengecekan nilai kosong (*missing values*), nilai unik, dan baris duplikat untuk memastikan kualitas data yang baik sebelum analisis lanjutan. Kolom-kolom dari Q1 hingga Q20 kemudian diidentifikasi sebagai variabel utama untuk pemrosesan lebih lanjut dan eventualisasi pembentukan kolom TOTAL\_YA, yang mewakili jumlah jawaban “ya” dari setiap responden. Tahapan ini penting memastikan data bersih dan siap digunakan dalam pemodelan klasterisasi.

### 3.4. Data Preparation

Pada tahap data preparation, data SRQ-20 dari kolom Q1 hingga Q20 terlebih dahulu dikonversi menjadi format numerik menggunakan fungsi pd.to\_numeric(errors='coerce'), sehingga nilai non-numerik dapat diidentifikasi sebagai NaN. Selanjutnya, *missing value* diisi menggunakan nilai modus pada masing-masing kolom agar tidak mengganggu distribusi data asli. Seluruh nilai selain 0 kemudian dikonversi menjadi 1 untuk mempertahankan skala jawaban biner (0 dan 1). Setelah itu, ditambahkan kolom baru bernama TOTAL\_YA, yaitu jumlah total jawaban "ya" dari masing-masing responden. Tahapan ini bertujuan untuk menghasilkan data yang bersih, terstandar, dan siap digunakan dalam proses klasterisasi menggunakan algoritma K-Means. Pendekatan serupa dilakukan dalam penelitian pengelompokan mahasiswa berdasarkan data psikologis menggunakan metode K-Means, di mana proses pembersihan dan transformasi data dilakukan sebelum model diterapkan (Solang & Nugroho, 2023).

### 3.5. Modelling

Pada tahap modelling, algoritma *K‑Means Clustering* digunakan untuk mengelompokkan responden berdasarkan skor gejala psikologis dari hasil SRQ‑20. Pertama, dilakukan pemilihan jumlah klaster (k) dengan metode Elbow untuk menemukan jumlah klaster paling representatif. Selanjutnya model dilatih menggunakan pustaka scikit‑learn. Hasil percobaan untuk k antara 2 hingga 5 dianalisis menggunakan metrik *Silhouette Score*, dan didapati bahwa nilai tertinggi diperoleh pada k = 2, sehingga dipilih sebagai model utama. Pendekatan serupa juga digunakan dalam penelitian implementasi K‑Means untuk strategi promosi kampus, yang menunjukkan efektivitas model dalam klasifikasi dan segmentasi data survei (Latifah et al., 2024).

### 3.6. Evaluation

Pada tahap *evaluation*, kualitas klaster dari model K‑Means dievaluasi menggunakan *Silhouette Score*, yang mengukur sejauh mana objek berada dalam klaster yang sama (cohesion) dibandingkan klaster lainnya (separation). Untuk menetapkan jumlah klaster optimal, diuji nilai k dari 2 hingga 5, dan didapati bahwa nilai *Silhouette Score* tertinggi diperoleh pada k = 2 menunjukkan pemisahan klaster yang baik untuk dataset SRQ‑20. Metode serupa juga diterapkan dalam penelitian lokal, seperti pengelompokan wilayah stunting dan evaluasi jumlah klaster optimal menggunakan *Silhouette* dan *Elbow*, yang memberikan hasil akurasi cluster sebesar 0,37 (Amalia & Arianto, 2024; Firman Ashari et al., 2023) .

### 3.7. Deployment

Data akhir dari hasil klasterisasi dipadukan dengan informasi demografis (nomor responden, usia, jenis kelamin) dalam sebuah DataFrame baru. Kemudian, dilakukan reduksi dimensi menggunakan PCA dan label klaster ditambahkan ke DataFrame. Dataset ini disimpan sebagai file CSV untuk digunakan dalam aplikasi Streamlit. Aplikasi ini memungkinkan pengguna menjawab 20 pertanyaan SRQ-20, lalu langsung mendapatkan prediksi klaster, interpretasi berdasarkan cutoff WHO, dan visualisasi kondisi psikologis yang terindikasi.

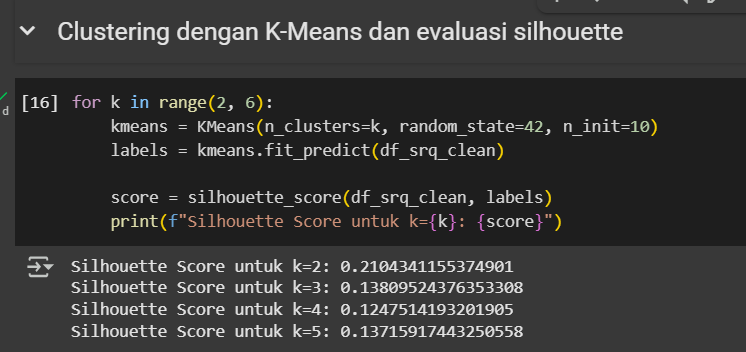
Aplikasi Streamlit dijalankan langsung dari Google Colaboratory (Colab) dan diakses secara publik menggunakan pyngrok, setelah pustaka dipasang dan authtoken dikonfigurasi. Dengan demikian, seluruh proses mulai dari input, analisis, hingga penyampaian hasil kepada pengguna telah terintegrasi dalam satu sistem berbasis web yang responsif dan interaktif. Pendekatan ini sejalan dengan tren pengembangan aplikasi pendeteksi dini kesehatan mental berbasis web di Indonesia, seperti aplikasi *MentalCare* yang dirancang sebagai aplikasi skrining kesehatan mental remaja (Hertini et al., 2024).

# BAB IV HASIL DAN PEMBAHASAN

## Hasil Klasterisasi

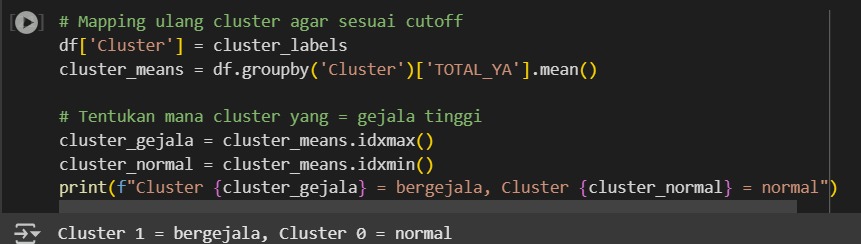
Hasil klasterisasi menggunakan algoritma K-Means dengan nilai k = 2 menunjukkan bahwa data siswa dapat dikelompokkan ke dalam dua klaster yang mencerminkan perbedaan tingkat risiko psikologis. Klaster pertama (Klaster 0) umumnya memiliki skor TOTAL\_YA yang rendah, yang diasumsikan mewakili kelompok siswa dengan kondisi psikologis relatif stabil. Sementara itu, Klaster kedua (Klaster 1) menunjukkan skor yang tinggi, mengindikasikan adanya indikasi gejala psikologis lebih besar. Rata-rata skor pada tiap klaster diperoleh sebagai dasar identifikasi klaster bergejala dan normal. Penentuan ini sejalan dengan pendekatan berbasis data dalam skrining psikologis yang telah digunakan dalam penelitian sebelumnya (Solang & Nugroho, 2023).

Rata-rata skor TOTAL\_YA dari masing-masing klaster dihitung untuk memperkuat interpretasi. Klaster dengan rata-rata tertinggi secara konseptual ditetapkan sebagai Kelompok Bergejala, sedangkan yang terendah sebagai Kelompok Normal. Analisis ini dilakukan secara data-driven dan menjadi landasan awal dalam pemetaan risiko psikologis.



Gambar 4. 1 Silhouette Score untuk Berbagai Nilai k

Dari Gambar 4.1, evaluasi menunjukkan bahwa nilai k=2 memberikan Silhouette Score tertinggi (0.21), menandakan pemisahan klaster yang paling optimal dibanding opsi k lainnya.

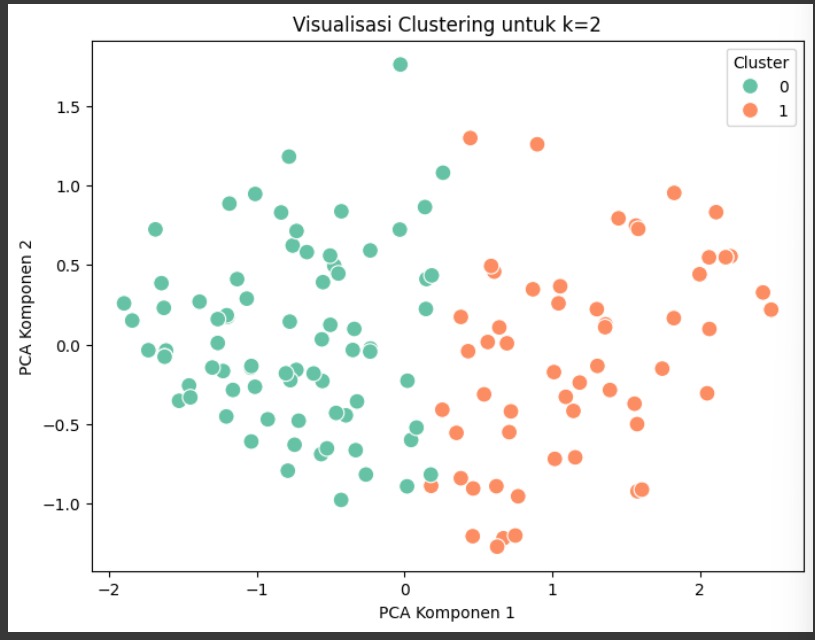


Gambar 4. 2 Pemetaan Klaster Berdasarkan Nilai Rata-rata TOTAL\_YA

Dari Gambar 4.2 menunjukkan identifikasi klaster dengan gejala tinggi berdasarkan pengelompokan rata-rata skor TOTAL\_YA. Klaster 1 dipetakan sebagai kelompok bergejala, dan Klaster 0 sebagai kelompok normal.

## Visualisasi Klaster

Untuk memperjelas struktur hasil klasterisasi, dilakukan reduksi dimensi menggunakan *Principal Component Analysis* (PCA). Proyeksi dua dimensi hasil PCA divisualisasikan dalam bentuk scatter plot, dengan tiap titik mewakili satu siswa dan diberi warna berdasarkan klaster yang terbentuk. Visualisasi ini memberikan pemahaman visual yang intuitif mengenai sebaran kelompok siswa dan efektivitas model K-Means dalam memisahkan kelompok berdasarkan pola respon SRQ-20. Penggunaan PCA dalam visualisasi hasil klasterisasi juga didukung oleh penelitian lokal yang menunjukkan efektivitas pendekatan ini dalam memperjelas pola klasifikasi data kesehatan mental (Alaudin Shalih et al., 2025).

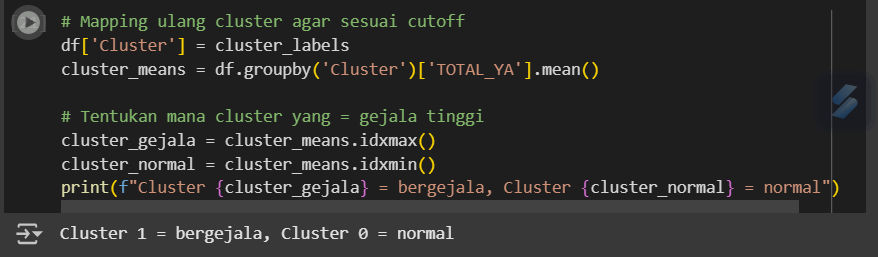


Gambar 4. 3 Scatter plot PCA hasil klasterisasi untuk k = 2

Dari Gambar 4.3, menampilkan distribusi dua klaster hasil K-Means berdasarkan proyeksi PCA. Titik berwarna menunjukkan pemisahan antara klaster Normal dan Bergejala.

## Evaluasi Cutoff WHO dan Validasi Internal

Sebagai bentuk validasi tambahan, dilakukan perbandingan hasil klasterisasi dengan cutoff klinis dari WHO, yaitu skor TOTAL\_YA ≥ 6 sebagai batas awal indikasi gangguan psikologis. Untuk mempermudah interpretasi, dilakukan analisis terhadap rata-rata skor TOTAL\_YA di masing-masing klaster. Hasil analisis menunjukkan bahwa Klaster 1 memiliki nilai rata-rata TOTAL\_YA yang lebih tinggi dibandingkan Klaster 0, sehingga diidentifikasi sebagai Klaster Bergejala, sedangkan Klaster 0 sebagai Klaster Normal.



Gambar 4. 4 Menentukan Klaster Bergejala Bedasarkan Nilai Rata-rata

Dari Gambar 4.4, menunjukkan bahwa proses unsupervised learning yang dilakukan melalui K-Means Clustering secara statistik mampu mengelompokkan data siswa sesuai dengan kecenderungan klinis. Artinya, klaster dengan skor gejala psikologis tinggi memang sesuai dengan cutoff WHO. Dengan demikian, dapat disimpulkan bahwa hasil klasterisasi memiliki konsistensi terhadap standar skrining WHO, dan menunjukkan potensi sebagai alat bantu skrining awal.

## Implementasi Sistem Interaktif

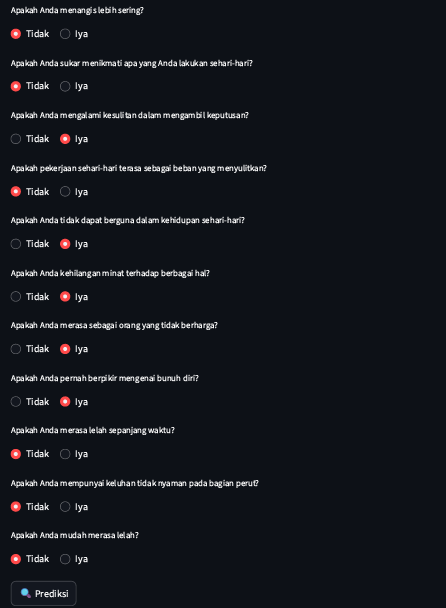
Seluruh hasil klasterisasi diimplementasikan ke dalam sistem berbasis Streamlit, yang memungkinkan pengguna untuk:

* Mengisi 20 pertanyaan SRQ-20 secara interaktif,
* Melihat skor TOTAL\_YA secara langsung,
* Mendapatkan prediksi klaster berdasarkan model K-Means,
* Melihat interpretasi hasil berdasarkan cutoff WHO.

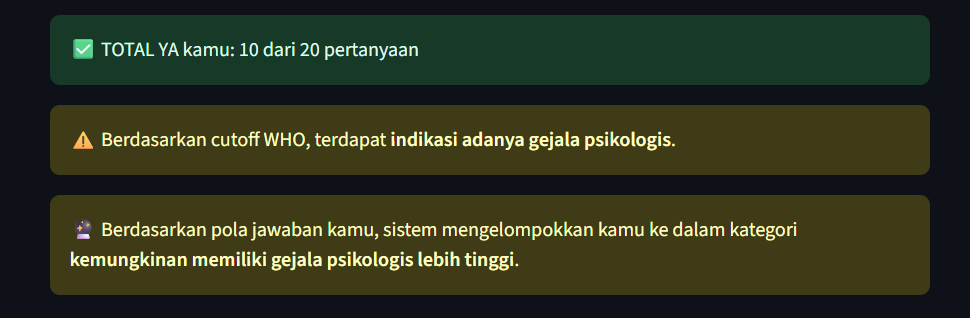
Dashboard ini dirancang agar mudah digunakan oleh guru, konselor, atau responden secara mandiri. Implementasi dilakukan di lingkungan Google Colab dan diakses secara publik melalui layanan ngrok, memungkinkan pengujian lapangan dan demonstrasi interaktif secara real-time.



Gambar 4. 5 Tampilan Dashboard Prediksi SRQ-20



Gambar 4. 6 Tampilan Dashboard Prediksi SRQ-20



Gambar 4. 7 Tampilan Dashboard Prediksi SRQ-20

Dari Gambar 4.5, 4.6, 4.7, menampilkan 20 pertanyaan SRQ-20 dengan pilihan jawaban. Setelah responden menyelesaikan pengisian, sistem memberikan skor TOTAL\_YA serta prediksi klaster dengan interpretasi berbasis cutoff WHO.

1. **Interpretasi dan Implikasi Hasil**

Hasil klasterisasi menunjukkan bahwa pendekatan unsupervised seperti K-Means mampu mengidentifikasi kelompok risiko secara efisien hanya dari pola jawaban kuesioner. Klaster dengan skor tinggi menunjukkan adanya konsentrasi responden yang menjawab “ya” pada banyak item SRQ-20, yang secara teoritis berkorelasi dengan potensi gangguan psikologis seperti depresi ringan, gangguan kecemasan, atau kelelahan mental (Dara Safitri et al., 2024; Rachmat et al., 2022).

Implikasi dari hasil ini adalah bahwa sistem dapat digunakan sebagai alat screening awal untuk membantu intervensi sebelum dilakukan diagnosis klinis yang lebih mendalam (Solang & Nugroho, 2023). Sistem ini juga membuka peluang untuk digunakan dalam skala besar, dengan modifikasi lebih lanjut terhadap fitur demografis seperti usia, jenis kelamin, atau riwayat psikososial, sehingga rekomendasi yang diberikan menjadi lebih personal dan kontekstual (Hertini et al., 2024).

# BAB V PENUTUP

## Kesimpulan

Penelitian ini berhasil mengimplementasikan algoritma K-Means Clustering dalam proses pengelompokan siswa berdasarkan hasil pengisian kuesioner SRQ-20. Hasil klasterisasi menunjukkan bahwa siswa dapat dibagi ke dalam tiga kategori berdasarkan tingkat risiko psikologis, yaitu rendah, sedang, dan tinggi.

Pendekatan berbasis data ini memungkinkan analisis hasil kuesioner dilakukan secara lebih objektif, efisien, dan akurat, dibandingkan dengan interpretasi manual. Selain itu, visualisasi hasil klasterisasi memberikan wawasan yang signifikan dan dapat digunakan sebagai dasar pengambilan keputusan dalam layanan konseling yang terarah dan berbasis data.

## Saran

1. Untuk meningkatkan generalisasi dan validitas hasil, disarankan agar penelitian selanjutnya menggunakan dataset yang lebih besar dan mencakup siswa dari berbagai sekolah dengan latar belakang yang beragam. Hal ini bertujuan untuk memperoleh representasi yang lebih luas terhadap kondisi psikologis siswa secara keseluruhan.
2. Selain itu, penting untuk melakukan perbandingan algoritma, dengan menguji efektivitas metode K-Means terhadap teknik klasterisasi lain seperti DBSCAN dan Agglomerative Clustering. Perbandingan ini dapat memberikan gambaran yang lebih komprehensif terkait akurasi dan keandalan pengelompokan berdasarkan karakteristik data.
3. Agar hasil interpretasi klasterisasi lebih akurat dan aplikatif, keterlibatan konselor atau psikolog profesional dalam proses analisis sangat dianjurkan. Validasi oleh ahli dapat memastikan bahwa pengelompokan yang dihasilkan sesuai dengan kondisi psikologis nyata yang dialami siswa.
4. Selanjutnya, pengembangan sistem dalam bentuk aplikasi berbasis web direkomendasikan agar hasil analisis lebih mudah diakses dan digunakan oleh pihak sekolah, khususnya oleh tenaga pendidik dan konselor.
5. Sebagai bentuk penguatan intervensi, sistem juga dapat dilengkapi dengan fitur pelaporan dan pemantauan yang memungkinkan pelacakan perkembangan siswa dari waktu ke waktu, misalnya melalui visualisasi grafik dinamika risiko psikologis atau rekam jejak tindak lanjut konseling.

# DAFTAR PUSTAKA

Alaudin Shalih, F., Akbar Ramadhan, R., & Syalaisa, N. (2025). *Tinjauan Komprehensif tentang Aplikasi dan Perkembangan Principal Component Analysis (PCA) A B S T R A K INFORMASI ARTIKEL*. https://ejournal.upi.edu/index.php/JEM

Amalia, M. F., & Arianto, D. B. (2024). Implementasi Algoritma K-Means Clustering Dalam Klasterisasi Kabupaten/Kota Provinsi Jawa Barat Berdasarkan Faktor Pemicu Stunting Pada Balita. *SIMKOM*, *9*(1), 36–46. https://doi.org/10.51717/simkom.v9i1.356

Anggraini, F. D. P., Aprianti, A., Setyawati, V. A. V., & Hartanto, A. A. (2022). Pembelajaran Statistika Menggunakan Software SPSS untuk Uji Validitas dan Reliabilitas. *Jurnal Basicedu*, *6*(4), 6491–6504. https://doi.org/10.31004/basicedu.v6i4.3206

Dara Safitri, D., Widodo, A., Studi Keperawatan, P., Ilmu Kesehatan, F., & Muhammadiyah Surakarta, U. (2024). ANALISIS VALIDITAS SELF REPORTING QUESTIONNAIRE (SRQ) TERHADAP KESEHATAN MENTAL PADA REMAJA. *Jurnal Kesehatan Tambussai*, *5*(1).

Fajriyah, N., Departemen, P., Sosial, I. K., Dakwah, F., Komunikasi, D., Sunan, U., & Yogyakarta, K. (2023). Psychometric Properties Analysis of Self-Reported Questionnaire (SRQ)-20 Instrument with Rasch Model. *Jurnal Psikologi Dan Kesehatan Mental (JISAN)*. https://doi.org/10.20473/jpkm.v8i22023.148

Firman Ashari, I., Dwi Nugroho, E., Baraku, R., Yanda, I. N., & Liwardana, R. (2023). Analysis of Elbow, Silhouette, Davies-Bouldin, Calinski-Harabasz, and Rand-Index Evaluation on K-Means Algorithm for Classifying Flood-Affected Areas in Jakarta. In *Journal of Applied Informatics and Computing (JAIC)* (Vol. 7, Issue 1). http://jurnal.polibatam.ac.id/index.php/JAIC

Hertini, R., Rohita, T., Wibowo, D. A., Muflihah, M., Oktavia, S. R., & Lesmana, I. (2024). Penerapan Aplikasi Mentari (Mental Health Remaja Indonesia) Assessment Pendeteksi Dini Kesehatan Mental Emosional Dan Perilaku Pada Remaja. *Kolaborasi: Jurnal Pengabdian Masyarakat*, *4*(4), 227–234. https://doi.org/10.56359/kolaborasi.v4i4.380

Indraputra, R. A., & Fitriana, R. (2020). K-Means Clustering Data COVID-19. *Jurnal Teknik Industri*.

Latifah, U. W., Bahri, S., & Satriandhini, M. (2024). Implementasi Algoritma K-Means Clustering untuk Strategi Promosi Kampus IBISA. *JIKO (Jurnal Informatika Dan Komputer)*, *8*(2), 292. https://doi.org/10.26798/jiko.v8i2.1307

Nazar, R. (2024). Implementasi Pemrograman Python Menggunakan Google Colab. *Jurnal Informatika Dan Komputer*, *15*, 50–56.

Prasetio, C. E., Triwahyuni, A., & Prathama, A. G. (2022). Psychometric Properties of Self-Report Questionnaire-20 (SRQ-20) Indonesian Version. *Jurnal Psikologi*, *49*(1), 69. https://doi.org/10.22146/jpsi.69782

Prihandoko, Agustina Rumapea, S., & Hanif Pratama, A. (2025). Integrasi Algoritma YoloV8 dan Streamlit Untuk Visualisasi Real-Time danAkurat dalam Perhitungan Kerumunan di Kawasan Stasiun Bekasi. *METHOMIKA : Jurnal Manajemen Informatika Dan Komputerisasi Akuntansi*, *9*(1). https://doi.org/10.46880/jmika.Vol9No1.pp179-187

Rachmat, M., Wati, F., Hasnaeni, H., Manti, S., Hasliani, A., Hasnita, H., Marbun, U., Bellani, E., Virani, D., Ikhsan, Muh., Asrianti, T., Hidayat, M., & Anwar, S. A. (2022). Detection of Mental Emotional Disorder Symptoms using SRQ-20 in Pregnant Women: A Case Example from South Sulawesi, Indonesia. *Jurnal Ilmiah Kesehatan (JIKA)*, *4*(1), 106–115. https://doi.org/10.36590/jika.v4i1.231

Solang, T., & Nugroho, A. (2023). ANALISIS KESEHATAN MENTAL MAHASISWA UNIVERSITAS KRISTEN SATYA WACANA MENGGUNAKAN METODE CLUSTERING ALGORITMA K-MEANS. *Jurnal TEKINKOM*, *6*(1). https://doi.org/10.37600/tekinkom.v6i1.641

Yunita Sari, F., Sukma Kuntari, M., Ari Yati, W., Khaulasari, H., & Hafiyusholeh, M. (n.d.). Jurnal Nasional Teknologi dan Sistem Informasi Attribution-ShareAlike 4.0 International Some rights reserved Artikel Penelitian Implementasi K-Means Clustering Melalui Pemanfaatan Sampling Kombinasi Pada Pengelompokan Pola Kesehatan Mental Mahasiswa Sains dan Teknologi. *Jurnal Teknologi Dan Sistem Informasi*, *11*. https://doi.org/10.25077/TEKNOSI.v11i1.2025.9-16