**PROPOSAL PENELITIAN**

**PEMBELAJARAN MESIN**

****

**Penerapan K-Means Clustering dalam Mengelompokkan Siswa Berdasarkan Gejala Psikologis SRQ-20**

**KETUA:**

Muhammad Dyaka Faiz Yuhendra A11.2022.14712

**ANGGOTA:**

Muhamad Aziz Zahrul Zamzami A11.2023.15156

Nathanael Christian Arifana A11.2023.15289

Noor Adekah Apriyana A11.2022.14382

Anggita Alya Salsabila A11.2022.14401

**UNIVERSITAS DIAN NUSWANTORO SEMARANG**

**2025**

| **A. JUDUL**  *Tuliskan judul usulan project maksimal 20 kata* |
| --- |
| Penerapan K-Means Clustering dalam Mengelompokkan Siswa Berdasarkan Gejala Psikologis SRQ-20 |
| **B. RINGKASAN**  *Isian ringkasan project tidak lebih dari 300 kata yang berisi urgensi, tujuan, metode, dan luaran yang ditargetkan* |
| Kesehatan mental siswa merupakan aspek penting yang mempengaruhi proses belajar. Namun, sering kali kondisi psikologis siswa tidak terdeteksi dengan baik. Kuesioner Self Reporting Questionnaire (SRQ-20) adalah alat ukur yang digunakan WHO untuk mendeteksi gejala psikologis seperti kecemasan, depresi, dan gangguan lainnya. Dalam penelitian ini, digunakan algoritma K-Means Clustering untuk mengelompokkan siswa berdasarkan kemiripan jawaban mereka terhadap 20 item SRQ-20. Tujuan utama dari penelitian ini adalah memberikan pendekatan otomatis dan berbasis data untuk mengidentifikasi kelompok risiko siswa berdasarkan gejala psikologis.  Dataset berupa jawaban biner dari SRQ-20 diproses dan dianalisis. Proses dimulai dari pra-pemrosesan data, eksplorasi jumlah cluster optimal dengan metode Elbow, pemodelan dengan K-Means Clustering, hingga visualisasi klaster menggunakan PCA. Luaran yang diharapkan adalah klasifikasi kelompok siswa dalam klaster risiko tinggi, sedang, dan rendah, serta visualisasi yang mudah dipahami bagi pihak sekolah atau konselor. |
| **C. KATA KUNCI**  *Isian 5 kata kunci* |
| k-means, clustering, srq-20, gejala psikologis, analisis kuesioner,siswa |
| **D. PENDAHULUAN**  *Pendahuluan project tidak lebih dari 1000 kata yang memuat, latar belakang, rumusan permasalahan yang akan diteliti, pendekatan pemecahan masalah, state-of-the-art dan kebaruan, peta jalan (road map) project setidaknya 5 tahun. Sitasi disusun dan ditulis berdasarkan sistem nomor sesuai dengan urutan pengutipan (Vancouver).* |
| **Latar Belakang:** Kesehatan mental remaja, khususnya siswa sekolah menengah, menjadi perhatian penting dalam dunia pendidikan. perubahan hormonal, tekanan akademik, dan dinamika sosial sering kali membuat siswa rentan terhadap gangguan psikologis seperti stres, kecemasan hingga depresi. namun, keterbatasan jumlah tenaga psikolog/konselor di sekolah serta kecenderungan siswa untuk tidak menyampaikan keluhan psikologis membuat banyak kasus luput dari deteksi dini.  Salah satu alat skrining yang cukup populer untuk mengukur kondisi psikologis adalah **Self Reporting Questionnaire (SRQ-20)** yang di kembangkan oleh WHO. Kuesioner ini terdiri dari 20 pertanyaan yang mencakup berbagai gejala psikologis seperti kesulitan tidur, mudah lelah, kehilangan minat, hingga pemikiran bunuh diri. di banyan institusi kesehatan, SRQ-20 telah digunakan secara luas untuk skrining awal gangguan mental non-psikotik.  Namun, tantangan muncul ketika data SRQ-20 dikumpulkan dari populasi besar seperti siswa sekolah. proses interpretasi manual menjadi tidak efisien dan rentan bias. oleh karena itu, diperlukan pendekatan berbasis data untuk menganalisis dan mengelompokkan siswa berdasarkan pola gejala yang mereka alami.  **Rumusan Masalah:** Berdasarkan latar belakang di atas, dapat ditemukan rumusan masalah dalam penelitian sebagai berikut:  Bagaimana penerapan metode K-Means Clustering dapat digunakan untuk mengelompokkan siswa berdasarkan gejala psikologi yang diukur menggunakan SRQ-20, sehingga dapat membantu proses identifikasi risiko psikologis secara lebih efisien dan akurat?  **Pendekatan Pemecahan Masalah:** Penelitian ini mengusulkan penggunaan **K-Means Clustering**, metode *unsupervised learning,* untuk menganalisis data hasil pengisian SRQ-20. Dataset yang telah tersedia mencakup jawaban dari sejumlah siswa terhadap 20 pertanyaan dalam bentuk biner (1 untuk “ya”, 0 untuk “tidak”), serta jumlah total gejala yang dialami setiap siswa.  Tahapan pendekatan teknis meliputi:   1. **Pembersihan Data**: Menghapus baris-baris meta, merapikan kolom, dan memastikan kelengkapan data. 2. **Normalisasi dan Analisis Awal**: Memeriksa distribusi jawaban dan variabilitas antar responden. 3. **Penerapan K-Means Clustering**: Mengelompokkan siswa berdasarkan pola jawaban mereka terhadap SRQ-20, dengan jumlah klaster ditentukan menggunakan metode Elbow atau Silhouette score. 4. **Validasi Internal**: Menggunakan total skor SRQ-20 (kolom “TOTAL YA”) sebagai tolak ukur awal untuk melihat konsistensi hasil klaster. 5. **Visualisasi dan Interpretasi**: Menggunakan metode reduksi dimensi (misalnya PCA atau t-SNE) untuk memetakan hasil klaster dalam visual dua dimensi.     **State-of-the-Art dan Kebaruan**: Penelitian ini mengambil posisi unik dalam integrasi antara data psikologis berbasis SRQ-20 dan teknik *machine learning*. Meskipun SRQ-20 sudah luas digunakan skrining gejala psikologis, pengolahan datanya secara konvensional masih terbatas pada perhitungan total skor dan interpretasi manual. Begitu pula, metode K-Means sering digunakan di dunia pendidikan untuk mengelompokkan siswa berdasarkan nilai akademik atau preferensi, namun jarang digunakan untuk mengolah data kesehatan mental berbasis SRQ-20.  Beberapa penelitian relevan sebelumnya:   1. Prasetyo et al. (2020) penerapan K-Means untuk pengelompokan mahasiswa berdasarkan tingkat stres, namun tidak menggunakan SRQ-20. 2. Dewi dan Nugroho (2022) menggunakan SRQ-20 untuk mendukung keputusan konseling, tetapi pendekatannya bersifat rule-based, bukan *unsupervised*. 3. WHO (1994) menyatakan SRQ-20 cocok untuk konteks komunitas, namun belum banyak diintegrasikan dengan analistik data modern.   Kebaruan dari penelitian ini terletak pada:   1. penggunaan SRQ-20 sebagai fitur numerik untuk K-Means, yang memungkinkan pengelompokkan berdasarkan pola gejala psikologis. 2. validasi tidak langsung dengan data empiris menggunakan skor “TOTAL YA”. 3. basis pengembangan sistem skrining psikologis otomatis berbasis klaster untuk kebutuhan sekolah.   Peta Jalan (Roadmap) Project:  Tahun 1: Riset Dasar dan Eksperimen Model   1. Bulan 1–2: Pengumpulan dan Pembersihan Data  * Mengumpulkan data hasil kuesioner SRQ-20 dari siswa sekolah mitra. * pembersihan data (missing value, duplikasi) dan encoding jawaban.  1. Bulan 3–5: Implementasi K-Means Clustering  * Eksperimen nilai k dan visualisasi hasil klaster (PCA/t-SNE). * Analisis awal tiap kelompok siswa berdasarkan gejala umum.  1. Bulan 6–8: Validasi Internal dan Pakar  * validasi hasil klaster terhadap skor total SRQ-20. * diskusi dan riview hasil dengan konselor atau guru BK.  1. Bulan 9–12: Dokumentasi dan Publikasi Awal  * penulisan laporan riset tahap awal. * persiapan publikasi seminar nasional dan pengajuan jurnal.  Tahun 2: Pengembangan Sistem Interaktif  1. Bulan 1–3: Desain dan Prototipe Website  * rancang UI/UX sistem berbasis web. * integrasi hasil clustering ke sistem backend menggunakan python/flask.  1. Bulan 4–6: Uji Coba Terbatas di Sekolah Mitra  * uji coba fungsionalitas sistem dan pengumpulan masukkan pengguna. * pengukuran performa visualisasi dan interpretabilitas hasil.  1. Bulan 7–9: Revisi dan Penambahan Fitur  * tambah fitur filter berdasarkan demografi (usia, kelas). * integrasi grafik dan histori hasil siswa.  1. Bulan 10–12: Finalisasi dan Sosialisasi Awal  * penyusunan panduan pemakaian sistem untuk konselor sekolah. * presentasi hasil ke pihak sekolah dan stakeholders pendidikan lokal.   Tahun 3: Ekspansi Dataset dan Eksperimen Model ALternatif   1. Bulan 1-3: Akuisisi Data dari Sekolah Tambahan  * bekerja sama dengan sekolah baru untuk memperluas dataset. * penyesuaian format dan harmonisasi struktur data.  1. Bulan 4-6: Eksperimen Fuzzy C-Means dan DBSCAN  * bandingkan performa clustering alternatif dengan K-Means. * gunakan metrik Silhouette Score dan Davies-Bouldin Index.  1. Bulan 7-10: Pengembangan Model Prediktif  * buat model prediksi risiko psikologis dengan supervised learning. * validasi akurasi model dan uji coba sistem klasifikasi.  1. Bulan 11-12: Evaluasi Akhir Tahun  * laporan perbandingan metode. * persiapan publikasi jurnal dan pertemuan akademik.   Tahun 4: Integrasi Intervensi dan Kolaborasi Multidisiplin   1. Bulan 1-4: Kolaborasi dengan Psikolog dan Dinas Terkait.  * konsultasi dalam merancang konten intervensi berbasis klaster. * penyesuaian sistem agar mendukung rekomendasi personal.  1. Bulan 5-8: Pembuatan Modul Intervensi Digital  * buat konten intervensi berbentuk tips, video, dan kegiatan mandiri. * uji coba kecil konten ke siswa berdasarkan klaster risiko.  1. Bulan 9-10: Umpan Balik dan Revisi Modul  * wawancara guru dan siswa terkait efektifitas intervensi. * revisi konten agar lebih relevan dan mudah digunakan  1. Bulan 11-12: Finalisasi Manual Intervensi Sekolah  * susun panduan pengguna sistem dan modul untuk guru BK. * rilis dokumentasi sistem dan kurikulum mini berbasis data.   Tahun 5: Konsorsium dan Nasionalisasi Sistem   1. Bulan 1-3: Pembentukan Konsorsium Nasional  * jalin kerja sama dengan universitas, BRIN, dan Kementerian Pendidikan. * bahas format distribusi nasional dan lisensi.  1. Bulan 4-7: Finalisasi Sistem dan Pendaftaran HKI.  * patenkan sistem dan dokumentasi ke DJKI/HKI. * integrasi database dan penyesuaian untuk skala besar.  1. Bulan 8-10: Pelatihan Nasional untuk Guru BK  * pelatihan daring dan tatap muka untuk implementasi sistem. * distribusi modul pelatihan dan simulasi kasus.  1. Bulan 11-12: Peluncuran Nasional dan Publikasi Internasional  * soft-launch sistem nasional. * publikasi hasil akhir jurnal bereputasi dan forum kebijakan. |
| **E. METODE**  *Isian metode atau cara untuk mencapai tujuan yang telah ditetapkan tidak lebih dari 1000 kata. Pada bagian metode wajib dilengkapi dengan diagram alir project yang menggambarkan apa yang sudah dilaksanakan dan yang akan dikerjakan selama waktu yang diusulkan. Format diagram alir dapat berupa file JPG/PNG. Metode penelitian harus memuat sekurang-kurangnya prosedur penelitian project, dataset yang digunakan, hasil yang diharapkan, indikator capaian yang ditargetkan.* |
| 1. Dataset. Dataset yang digunakan dalam penelitian ini berasal dari hasil pengisian kuesioner SRQ-20 (Self Reporting Questionnaire) oleh siswa. SRQ-20 terdiri dari 20 item pertanyaan yang dirancang untuk mengidentifikasi gejala psikologis seperti kecemasan, depresi, dan gangguan suasana hati lainnya. Setiap pertanyaan dijawab dengan pilihan “ya” atau “tidak”. Dalam tahap awal, data ini dikumpulkan dalam bentuk file Excel dengan satu baris per responden.2. Pra-pemprosesan Data. Pra-pemrosesan merupakan tahap penting untuk memastikan kualitas data sebelum dilakukan pemodelan. Tahapan yang dilakukan meliputi:  * Penghapusan Kolom Non-Relevan: Seperti kolom “Nama” atau identitas yang tidak diperlukan untuk proses analisis dan dapat menimbulkan bias. * Cek dan Penanganan Missing Values: Mengecek apakah terdapat nilai kosong atau tidak terisi, dan menghapus/mengimputasi nilai tersebut jika perlu. * Transformasi Jawaban: Jawaban “ya” dikonversi menjadi nilai 1 dan “tidak” menjadi 0 untuk setiap pertanyaan SRQ-20. * Normalisasi Data: Apabila diperlukan, normalisasi dilakukan agar setiap fitur memiliki skala yang seragam, terutama jika digunakan algoritma lain untuk validasi.  3. Penentuan Jumlah Cluster (K). Sebelum menerapkan algoritma K-Means, perlu ditentukan jumlah cluster yang optimal. Dalam penelitian ini digunakan metode Elbow, yaitu dengan menghitung inertia (jumlah total kuadrat jarak antara titik data dan pusat klasternya) untuk berbagai nilai K (biasanya dari 1 hingga 10).  Grafik inertia terhadap K akan menunjukkan titik di mana penurunan inertia mulai melambat tajam, yang dikenal sebagai "elbow point" — titik inilah yang diambil sebagai jumlah cluster optimal.4. Clustering dengan K-Means. Setelah jumlah cluster optimal diketahui, algoritma K-Means diterapkan:  * Menentukan titik pusat awal (centroid) secara acak. * Mengelompokkan data berdasarkan kedekatan Euclidean distance ke centroid. * Memperbarui posisi centroid berdasarkan rata-rata titik dalam setiap klaster. * Proses iteratif dilakukan hingga centroid stabil (konvergen) atau jumlah iterasi maksimum tercapai. Hasil dari proses ini adalah label klaster untuk setiap siswa, yang menunjukkan kelompok psikologis berdasarkan kemiripan jawaban SRQ-20.  5. Reduksi Dimensi dan Visualisasi. Data SRQ-20 memiliki 20 fitur, yang membuat visualisasi langsung sulit. Oleh karena itu, digunakan teknik Principal Component Analysis (PCA) untuk mereduksi dimensi menjadi 2 dimensi utama. Hasil reduksi divisualisasikan dalam bentuk scatter plot, dengan warna berbeda untuk setiap klaster. Visualisasi ini membantu memahami distribusi dan hubungan antar kelompok siswa.6. Interpretasi dan Analisis Klaster. Setelah klaster terbentuk, analisis dilakukan untuk memahami karakteristik umum dari setiap kelompok. Misalnya:  * Klaster 0: Siswa dengan skor jawaban “ya” sangat rendah → risiko psikologis rendah. * Klaster 1: Siswa dengan skor sedang → potensi stres ringan. * Klaster 2: Siswa dengan skor tinggi → indikasi awal gangguan psikologis. Hasil ini dapat digunakan oleh pihak sekolah untuk memberikan intervensi awal dan tindak lanjut sesuai kebutuhan.     Gambar . Diagram alir metode penelitian 7. Luaran yang Diharapkan.  * Pembentukan 2–4 klaster siswa berdasarkan gejala psikologis SRQ-20. * Visualisasi distribusi siswa dalam bentuk grafik 2D. * Dataset yang sudah dilengkapi label klaster, yang siap digunakan oleh pihak konselor/guru BK. * Laporan yang menjelaskan temuan utama dari masing-masing klaster.  8. Indikator Capaian  * Penerapan algoritma K-Means berhasil dan hasilnya dapat divisualisasikan dengan jelas. * Visualisasi berbentuk scatter plot menggambarkan distribusi tiap klaster secara terpisah. * Analisis interpretatif menunjukkan perbedaan nyata antara kelompok siswa dalam hal gejala psikologis. * Tersedianya dokumentasi dan data yang dapat digunakan sebagai dasar pengambilan kebijakan oleh sekolah. |
| **F. JADWAL PROJECT**  *Jadwal rencana project disusun berdasarkan pelaksanaan project dan disesuaikan berdasarkan lama tahun pelaksanaan project selama 14 minggu perkuliahan.* |
| | No | Nama Kegiatan | Minggu | | | | | | | | | | | | | | | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | | 1 | Pemilihan Topik |  |  |  |  |  |  |  |  |  |  |  |  |  |  | | 2 | Studi Literatur dan SRQ-2 |  |  |  |  |  |  |  |  |  |  |  |  |  |  | | 3 | Pengumpulan dan Pembersihan Data |  |  |  |  |  |  |  |  |  |  |  |  |  |  | | 4 | Penentuan K Optimal |  |  |  |  |  |  |  |  |  |  |  |  |  |  | | 5 | Implementasi K-Means |  |  |  |  |  |  |  |  |  |  |  |  |  |  | | 6 | Reduksi Dimensi dan Visualisasi |  |  |  |  |  |  |  |  |  |  |  |  |  |  | | 7 | Interpretasi dan Analisis |  |  |  |  |  |  |  |  |  |  |  |  |  |  | | 8 | Penyusunan Proposal |  |  |  |  |  |  |  |  |  |  |  |  |  |  | | 9 | Penulisan Laporan Akhir |  |  |  |  |  |  |  |  |  |  |  |  |  |  | |
| G. JOB DESK  *Peran dan Tugas dalam pengerjaan proposal project* |
| | Nama | Peran dan Tugas | | --- | --- | | Muhammad Dyaka Faiz Yuhendra | Menyusun Ringkasan Project, yang mencakup urgensi permasalahan, tujuan, metode yang digunakan, serta luaran yang ditargetkan dari penelitian. | | Muhamad Aziz Zahrul Zamzami | Merangkum bagian Pendahuluan, termasuk penyusunan latar belakang, perumusan masalah yang diangkat, serta merancang pendekatan pemecahan masalah yang akan digunakan. | | Nathanael Christian Arifana | Bertanggung jawab pada bagian Metode Penelitian, merancang diagram alir metode, dan menyusun Jadwal Project selama 14 minggu pelaksanaan. | | Noor Adekah Apriyana | Membuat rincian Langkah-Langkah Teknis dalam metode penelitian, termasuk tahapan preprocessing, pemodelan K-Means, reduksi dimensi, hingga interpretasi hasil. | | Anggita Alya Salsabila | Berperan dalam menyusun dan merangkum Pendahuluan, termasuk latar belakang, rumusan masalah, pendekatan solusi, serta merancang dasar untuk Roadmap Project jangka panjang. | |
| **H. DAFTAR PUSTAKA**  *Sitasi disusun dan ditulis berdasarkan sistem nomor sesuai dengan urutan pengutipan. Hanya pustaka yang disitasi pada usulan project yang dicantumkan dalam Daftar Pustaka menggunakan format sitasi Vancouver.* |
| 1. [1] N. Chayati, F. N. Azizah, S. Wardaningsih, S. I. Syahputri, dan N. D. Pangestika, “EDUKASI DAN KREATIVITAS KARYA TANGAN SEBAGAI UPAYA PEMELIHARAAN KESEHATAN MENTAL REMAJA PUTRI,” vol. 9, no. 2, 2025, doi: 10.31764/jmm.v9i2.28887. 2. [2] N. Fajriyah dkk., “Psychometric Properties Analysis of Self-Reported Questionnaire (SRQ)-20 Instrument with Rasch Model”, doi: 10.20473/jpkm.v8i22023.148. 3. [3] M. Rachmat dkk., “Detection of Mental Emotional Disorder Symptoms using SRQ-20 in Pregnant Women: A Case Example from South Sulawesi, Indonesia,” Jurnal Ilmiah Kesehatan (JIKA), vol. 4, no. 1, hlm. 106–115, Apr 2022, doi: 10.36590/jika.v4i1.231. 4. [4] M. Z. Saefurrohim, I. Zainafree, M. A. Fitrah, dan D. N. D. Utomo, “PANAKEIA JOURNAL OF PUBLIC HEALTH WHO SRQ-20 Outcomes in Mothers of Children 6-23 Months By Fecal Disposal and Drinking Water Source Factors Article Info Abstrak.” [Daring]. Tersedia pada: https://e-journals2.unmul.ac.id/ 5. [5] D. Dara Safitri, A. Widodo, P. Studi Keperawatan, F. Ilmu Kesehatan, dan U. Muhammadiyah Surakarta, “ANALISIS VALIDITAS SELF REPORTING QUESTIONNAIRE (SRQ) TERHADAP KESEHATAN MENTAL PADA REMAJA,” vol. 5, no. 1, 2024. 6. [6] WHO. (1994). A User's Guide to the Self Reporting Questionnaire (SRQ). World Health Organization. 7. [7] Prasetyo, Y., & Rachmawati, D. (2020). Analisis Tingkat Stres Mahasiswa Menggunakan Metode K-Means. Jurnal Teknologi dan Sistem Komputer, 8(2), 112-118. 8. [8] Dewi, N. K., & Nugroho, A. S. (2022). Sistem Pendukung Keputusan Konseling Menggunakan Metode SRQ-20 dan Fuzzy Logic. Jurnal Informatika, 19(1), 33-40. 9. [9] Xu, R., & Wunsch, D. (2005). Survey of clustering algorithms. IEEE Transactions on Neural Networks, 16(3), 645–678. 10. [10] Hastie, T., Tibshirani, R., & Friedman, J. (2009). The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer. |