

Website Profil Sekolah menggunakan Algoritma K-Means di Dinas Pendidikan Kabupaten Minahasa

School Profile Website using the K-Means Algorithm at the Minahasa Regency Education Office

¹Tiara Pongmangatta, ²Kristofel Santa*, ³Sondy Kumajas

^{1,2,3}Program Studi Teknik Informatika, Fakultas Teknik, Universitas Negeri Manado
^{1,2,3}Jl. Kampus Unima, Tonsaru, Kec. Tondano Selatan, Kab. Minahasa, Sulawesi Utara, 95618
*e-mail: kristofelsanta@gmail.com

(received: 5 January 2026, revised: 30 January 2026, accepted: 2 February 2026)

Abstrak

Latar Belakang: Dinas Pendidikan sering mengelola pemprofilan sekolah dengan data yang minimal dan terfragmentasi, seperti nama sekolah, kecamatan, jenjang, total siswa, dan total guru. Kondisi ini membuat proses perencanaan menjadi reaktif dan rentan kesalahan, sehingga dibutuhkan alur kerja yang ringan namun tepercaya untuk mengubah data agregat tersebut menjadi bukti yang dapat ditindaklanjuti. Metode: Penelitian ini mengembangkan alat pemprofilan berbasis web yang mengintegrasikan desain antarmuka menggunakan pendekatan *User-Centered Design* (UCD) dengan algoritma pengelompokan *K-Means* yang transparan. Pengembangan dilakukan melalui prototyping iteratif yang didokumentasikan, di mana fitur distandardisasi menggunakan z-score dan validitas kluster dinilai menggunakan metode silhouette yang dilengkapi indeks validitas internal lainnya. Hasil: Fitur pemasukan data dengan pratinjau header dan pemeriksaan numerik terbukti mengurangi pengerjaan ulang. Nilai silhouette mencapai puncak pada $k=2$, menghasilkan dua segmen yang mudah ditafsirkan (beban ketenagaan moderat vs tinggi), dengan opsi $k=3$ sebagai eksplorasi tambahan. Evaluasi kegunaan menggunakan *System Usability Scale* (SUS) menghasilkan skor ≈ 82 , yang menunjukkan tingkat penerimaan pengguna yang baik dengan waktu respons sistem dalam hitungan detik. Kesimpulan: Sistem ini menawarkan alur kerja yang ramping dan memadai untuk perencanaan rutin, serta meletakkan dasar bagi pengembangan *longitudinal* di masa depan, seperti pelacakan jumlah rombongan belajar dan akreditasi per periode.

Kata kunci: *k-means clustering, school profile, system usability scale, user-centered design, web-based analytics*

Abstract

Background: The Education Office often manages school profiling with minimal and fragmented data, such as school name, district, level, total students, and total teachers. This situation makes planning reactive and error-prone, highlighting the need for a lightweight yet reliable workflow to transform aggregated data into actionable evidence. **Method:** This study developed a web-based profiling tool that integrates a user interface designed using a *User-Centered Design* (UCD) approach with a transparent *K-Means* clustering algorithm. Development was carried out through iterative prototyping, with features standardized using z-scores and cluster validity assessed using the silhouette method along with other internal validity indices. **Results:** Data entry features with header previews and numeric checks effectively reduced rework. The silhouette value peaked at $k=2$, producing two interpretable segments (moderate vs. high staffing load), with an optional $k=3$ for exploratory purposes. Usability evaluation using the *System Usability Scale* (SUS) yielded a score of approximately 82, indicating good user acceptance, with system response times measured in seconds. **Conclusion:** The system provides a streamlined and sufficient workflow for routine planning and establishes a foundation for future longitudinal developments, such as tracking the number of study groups and accreditation per period.

Keywords: *k-means clustering, school profile, system usability scale, user-centered design, web-based analytics*

1 Pendahuluan

Pemanfaatan sistem informasi pada institusi pendidikan berperan penting dalam mendukung pengelolaan data serta penyediaan informasi yang cepat dan akurat. Secara umum, teknologi ini membantu mengurangi kesalahan penginputan data dan meningkatkan efisiensi kerja dalam lingkungan operasional pendidikan [22]. Namun, kondisi di lapangan menunjukkan tantangan spesifik. Dinas pendidikan saat ini semakin membutuhkan analitik yang ringan untuk memprofilkan sekolah menggunakan atribut minimal yang rutin dikumpulkan seperti total siswa dan guru per sekolah, agar perencanaan didorong oleh fakta, bukan impresi ad-hoc.

Masalahnya, Dinas Pendidikan sering kali mengelola data yang terfragmentasi. Di sisi lain, ekspektasi mutu layanan menuntut antarmuka yang dibangun dengan prinsip berpusat pada pengguna (User-Centered Design atau UCD). Pendekatan ini krusial untuk mengurangi kesalahan dan mempercepat tugas harian yang dilakukan operator serta analisis kebijakan, mengingat sistem yang ada sering kali kurang ramah pengguna [1], [4]. Dalam konteks pendidikan Indonesia, jalur data lazimnya diorganisasi di sekitar sistem pelaporan nasional, menghasilkan agregat tingkat sekolah yang memadai untuk segmentasi tahap awal dan pemantauan tingkat tinggi.

Berbagai penelitian terdahulu telah mengeksplorasi penggunaan algoritma K-Means dalam pendidikan. Metode K-Means clustering banyak diterapkan untuk mengelompokkan data siswa berdasarkan kategori tertentu, seperti pemetaan pemahaman materi [24], pengelompokan siswa berprestasi [25], serta hasil pembelajaran mahasiswa [27]. Selain itu, studi lokal juga menunjukkan bahwa pengelompokan rasio sederhana seperti siswa-per-guru dapat menyingkap disparitas antarkawasan yang penting bagi penataan ketenagaan dan alokasi sumber daya [10], [21]. Ada pula karya lain pada data mining pendidikan di Indonesia yang menggunakan K-Means bersama klasifikator sederhana untuk mendukung pengambilan keputusan akademik [8], serta studi yang menunjukkan efektivitas algoritma ini dalam mengelompokkan data siswa berdasarkan atribut non-akademik seperti daerah asal [23].

Meskipun penerapan K-Means dalam pendidikan telah meluas, terdapat perbedaan mendasar antara karya-karya terdahulu dengan pendekatan yang kami usulkan. Mayoritas penelitian sebelumnya, seperti [24], [25], dan [27], berfokus secara eksklusif pada pengelompokan data level siswa untuk tujuan pedagogis, dengan sedikit perhatian pada aspek antarmuka pengguna pengelolanya. Sementara itu, studi yang menerapkan K-Means berbasis web seperti [28] cenderung menyajikan algoritma sebagai proses "kotak hitam" (black-box) di mana pengguna hanya menerima hasil akhir tanpa kendali atas parameter analisis. Dasbor yang ada di kantor-kantor lokal pun kerap hanya menyediakan ringkasan deskriptif tanpa alur analitik terpandu yang terhubung erat dengan validasi data, sehingga meningkatkan beban kognitif dan risiko kesalahan interpretasi.

Berdasarkan kesenjangan tersebut, penelitian ini bertujuan merancang dan mengimplementasikan alat analitik profil berbasis web yang mengintegrasikan praktik User-Centered Design (UCD) dengan algoritma pengelompokan K-Means. Berbeda dengan sistem yang kaku, sistem kami memberikan transparansi dan kendali kepada operator dinas untuk melakukan validasi pra-pemrosesan dan memilih jumlah kluster (\$k\$) sesuai konteks lapangan. Kontribusi artikel ini ada tiga: (1) alur kerja end-to-end berbasis UCD untuk klasterisasi profil sekolah; (2) prototipe web yang mengoperasionalkan unggah–analisis–visualisasi–laporan dalam satu tempat; dan (3) cuplikan kegunaan empiris melalui System Usability Scale (SUS) yang dilengkapi validitas kluster internal. Pendekatan ini diharapkan menghasilkan segmentasi yang tidak hanya valid secara komputasional tetapi juga dapat ditindaklanjuti secara efektif oleh Dinas Pendidikan Kabupaten Minahasa.

2 Tinjauan Literatur

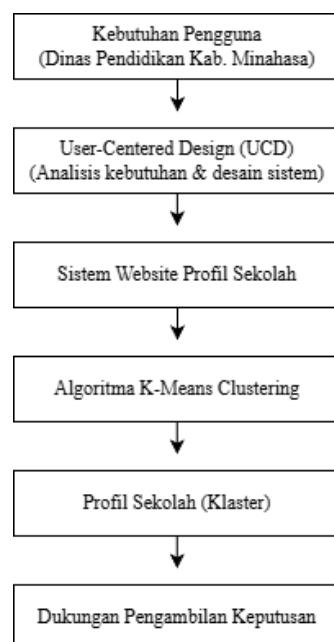
Pengembangan sistem informasi di sektor pendidikan mendorong pengembangan yang tidak hanya berfokus pada kemajuan teknologi, tetapi juga pada keselarasan sistem dengan kebutuhan dan karakteristik pengguna. Desain Berpusat pada Pengguna (UCD) adalah pendekatan desain yang memprioritaskan pengguna sebagai fokus utama di semua tahap pengembangan sistem, mulai dari identifikasi masalah dan pengembangan solusi hingga evaluasi penggunaan secara berulang. Tujuan proyek ini adalah memastikan bahwa sistem yang dikembangkan sesuai dengan lingkungan kerja, tujuan pengguna, dan konteks operasional [1], [4]. Dalam konteks sistem berbasis web, implementasi

UCD dapat meningkatkan efisiensi interaksi, mengurangi kesulitan pengguna, dan meningkatkan penggunaan sistem oleh pengguna non-teknis, termasuk dalam administrasi pendidikan.

Diperlukan alat evaluasi yang sederhana namun andal untuk menentukan apakah sistem melanggar aspek kegunaan. Skala Kegunaan Sistem (SUS) merupakan salah satu metode evaluasi kegunaan yang paling banyak digunakan karena kemudahan penggunaannya dan konsistensi dalam menghasilkan skor penggunaan secara global. SUS terdiri dari serangkaian pernyataan pada skala Likert 1 hingga 5 poin yang digunakan untuk mengukur persepsi pengguna terkait kenyamanan sistem, konsistensi, dan kemudahan penggunaan [2]. Banyak studi menunjukkan bahwa SUS efektif dalam mengevaluasi sistem informasi berbasis web dan dapat memberikan gambaran umum tentang tingkat kegunaan tanpa memerlukan tanggapan [9]. Karena itu, SUS sering digunakan sebagai alat evaluasi dalam penelitian yang mengadopsi UCD.

Menurut analisis data, algoritma k-means adalah metode pengelompokan yang sering digunakan untuk mengidentifikasi pola atau struktur dalam data numerik dengan mengelompokkan objek berdasarkan kedekatan jarak [5], [6]. Algoritma ini bekerja secara iteratif dengan meminimalkan jarak di dalam kluster dan memaksimalkan jarak antara kluster. Dalam penerapannya, k-means sensitif terhadap perbedaan skala fitur, sehingga prosedur normalisasi data, seperti standarisasi z-score, sering digunakan untuk mengevaluasi kontribusi setiap variabel [19]. Untuk mengevaluasi kualitas kluster yang dihasilkan, beberapa indeks validitas internal telah dikembangkan, termasuk koefisien siluet, yang mengukur kohesi dan pemisahan kluster dan sering digunakan dalam penelitian sebelumnya [7].

Banyak penelitian telah menunjukkan bahwa algoritma k-means yang efektif digunakan dalam pendidikan untuk mengelompokkan data siswa atau sekolah berdasarkan karakteristik akademik dan non-akademik, seperti jumlah siswa, jumlah guru, dan rasio siswa-guru [8], [10], dan [21]. Pengambilan keputusan pendidikan, seperti pemetaan beban tenaga kerja, pengelompokan siswa berprestasi, dan segmentasi sekolah untuk alokasi sumber daya, juga didukung oleh penerapan k-means [23]–[28]. Implementasi algoritma ini dalam sistem berbasis web bertujuan untuk menyediakan hasil analisis secara lebih interaktif dan mudah dipahami oleh pengguna. Meskipun demikian, sebagian besar penelitian sebelumnya lebih fokus pada hasil analisis daripada mengintegrasikan proses pengumpulan data dengan analisis data berbasis pengguna dan evaluasi kegunaan sistematis. Karena itu, penelitian ini menggabungkan clustering k-means dengan Desain Berpusat pada Pengguna (User-Centered Design) dan evaluasi kegunaan menggunakan Skala Kegunaan Sistem (System Usability Scale) untuk menghasilkan sistem analisis profil sekolah berbasis web yang tidak hanya secara teknis valid tetapi juga mudah diterapkan dan relevan dalam mendukung pengambilan keputusan di lingkungan pendidikan.



Gambar 1 Kerangka konseptual penelitian

Kerangka konseptual penelitian ditampilkan pada Gambar 1, yang menggambarkan hubungan antara kebutuhan pengguna, desain berpusat pada pengguna, sistem berbasis web, proses pengelompokan menggunakan algoritma K-Means, dan profil sekolah yang mendukung pengambilan keputusan.

Tabel 1 Ringkasan penelitian terdahulu

No. Penulis & Tahun	Metode	Temuan Utama
1. Abras et al. (2004)	User-Centered Design	Pendekatan UCD meningkatkan kesesuaian sistem dengan kebutuhan pengguna
2. Brooke (1996)	System Usability Scale	SUS efektif digunakan untuk mengukur tingkat usability sistem
3. Rousseeuw (1987)	Silhouette Coefficient	Silhouette mampu menilai kualitas pemisahan kluster
4. Wulandari (2021)	K-Means Clustering	K-Means efektif mengelompokan data pendidikan
5. Kumajas (2025)	K-Means Clustering	Clustering membantu identifikasi pola distribusi pendidikan

Berdasarkan sintesis literatur pada Tabel 1, teridentifikasi adanya kesenjangan penelitian di mana belum ditemukan studi yang mengintegrasikan pendekatan *User-Centered Design* (UCD) dengan algoritma *K-Means* secara simultan dalam pengembangan sistem profil sekolah berbasis web. Oleh karena itu, penelitian ini mengisi kekosongan tersebut dengan menawarkan kontribusi kebaruan (*novelty*) melalui integrasi strategis antara aspek desain berpusat pengguna dan analitik data pendidikan.

3 Metode Penelitian

3.1. Desain Penelitian

Studi ini mengadopsi model prototipe iteratif yang berlandaskan *User-Centered Design* (UCD), sebagaimana diilustrasikan pada Gambar 2. Diagram tersebut memperlihatkan alur penelitian yang dimulai dari identifikasi masalah dan analisis kebutuhan, kemudian berlanjut ke perancangan dan implementasi secara siklikal. Terdapat mekanisme keputusan (*decision point*) "Sistem Sesuai?" yang memungkinkan perbaikan desain kembali ke tahap analisis jika hasil belum memuaskan, guna menjaga keterkaitan erat antara tujuan pengguna yang berkembang dan keputusan desain [10].

Aktivitas analitik dalam alur tersebut disejajarkan dengan siklus hidup CRISP-DM, sehingga setiap iterasi secara eksplisit merekam transisi dari pemahaman bisnis hingga penerapan. Hal ini mempertahankan keterlacakan antara kebutuhan pemangku kepentingan, pilihan penyiapan data, dan keluaran pemodelan [3]. Untuk memastikan rancangan dapat dikomunikasikan kepada peserta teknis maupun non-teknis, *Unified Modeling Language* (UML) digunakan untuk menspesifikasikan sistem pada beberapa lapis abstraksi: *use case* untuk cakupan, *activity diagram* untuk logika proses, *class diagram* untuk struktur informasi, dan *sequence diagram* untuk kolaborasi antarmodul [6]. Dengan memosisikan UCD dan CRISP-DM sebagai kerangka yang saling melengkapi (lihat Gambar 2), kami dapat mengatur granularitas iterasi: siklus pendek menekankan umpan balik formatif terkait kegunaan, sementara siklus lebih panjang menambatkan kurasi data dan validitas analitik.

3.2. Akuisi Data

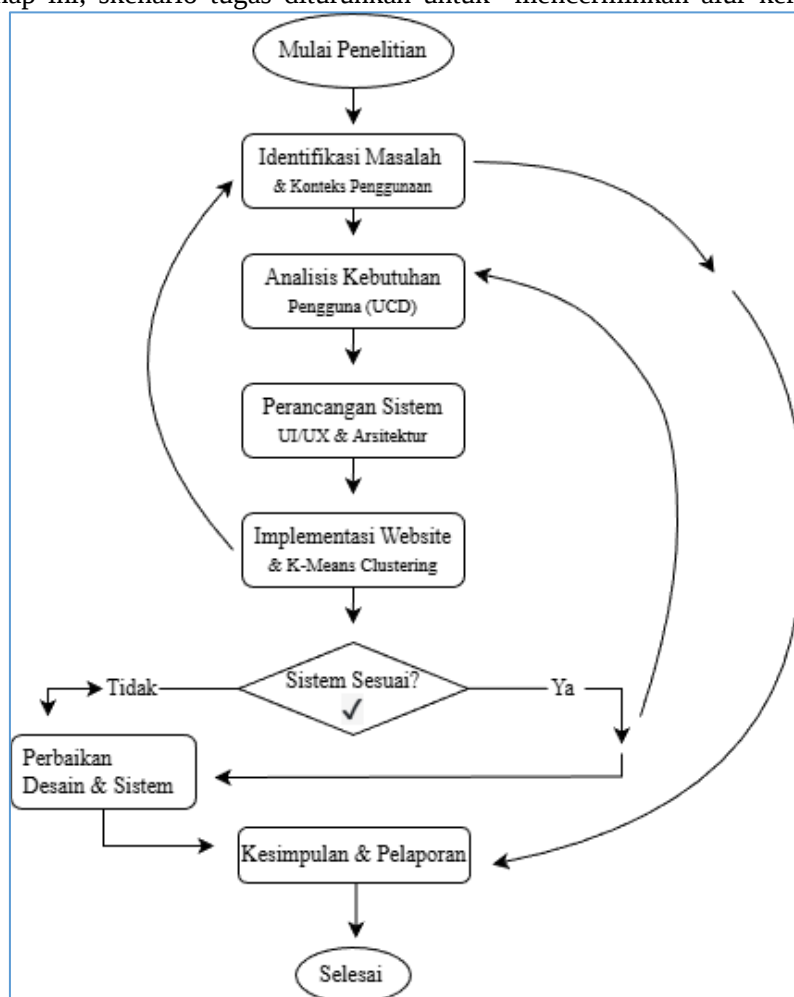
Dataset terdiri atas agregat tingkat sekolah yang disediakan oleh dinas pendidikan dalam templat CSV ber-pembatas titik koma yang menetapkan urutan header dan tipe data agar logika pemasukan dapat diverifikasi secara deterministik di setiap iterasi [3]. Pada fase awal siklus hidup, kami mencatat asal-usul (*provenance*), cakupan temporal, dan definisi medan untuk mendukung reproduktibilitas serta menstabilkan perbandingan analitik hilir ketika periode tambahan ditambahkan ke repositori [6]. Sebelum pemodelan, kami menegaskan pengetikan numerik pada jumlah siswa dan guru, mengevaluasi missingness, serta menghapus artefak pemformatan guna membatasi propagasi

kebisingan input ke tahap klasterisasi di mana metrik jarak sensitif terhadap pencilan dan skala [9]. Variabel turunan rasio siswa terhadap guru dihitung untuk setiap sekolah sebagai indikator ringkas beban ketenagaan yang melengkapi hitungan mentah dan meningkatkan keterbandingan antarsekolah saat merencanakan intervensi pada tingkat kecamatan [8]. Menjelaskan penelitian secara kronologis, termasuk desain penelitian, prosedur penelitian (dalam bentuk algoritme, pseudokode, atau lainnya), cara pengujian, dan akuisisi data [1–3]. Uraian jalannya penelitian harus didukung oleh referensi agar penjelasan dapat diterima secara ilmiah [2], [4].

2.3. Prosedur Penelitian

Prosedur penelitian ini mengikuti model prototipe iteratif yang menyeimbangkan umpan balik cepat dengan dokumentasi yang disiplin, dengan mengintegrasikan praktik *User-Centered Design* (UCD) dan artefak UML sepanjang siklus pengembangan [10]. Tahap awal dimulai dengan elicitasi dan analisis kebutuhan, di mana peran operasional dikonsolidasikan ke dalam satu persona Administrator untuk keperluan pengujian.

Pada tahap ini, skenario tugas diturunkan untuk mencerminkan alur kerja harian (unggah



Gambar 2 Desain penelitian

\$\rightarrow\$ analisis \$\rightarrow\$ visualisasi \$\rightarrow\$ laporan), sehingga kebutuhan fungsional dan non-fungsional yang teridentifikasi benar-benar muncul dari operasi nyata di lapangan, bukan sekadar daftar keinginan yang abstrak [7].

Berdasarkan kebutuhan tersebut, perancangan sistem dilakukan menggunakan spesifikasi UML untuk menjembatani komunikasi teknis. Diagram *Use Case* digunakan untuk mematok batas cakupan, *Activity Diagram* untuk mengungkap jalur pengecualian (seperti ketidaksesuaian *header* atau kolom kosong), *Class Diagram* untuk struktur entitas basis data, serta *Sequence Diagram* untuk menetapkan tanggung jawab antara *controller* dan *service* [6]. Rancangan ini kemudian dieksekusi melalui pengembangan prototipe inkremental yang mengimplementasikan modul inti secara berurutan, mulai dari fitur unggah dan pratinjau CSV, analitik *K-Means* (dengan normalisasi *z-score* dan seleksi k), hingga visualisasi dan ekspor. Pendekatan bertahap ini memastikan antarmuka mampu mengomunikasikan struktur klaster dan asumsi kualitas data secara efektif di titik penggunaan [3].

Setelah prototipe fungsional tersedia, siklus berlanjut ke tahap pengujian dan evaluasi kegunaan. Uji fungsional berbasis skenario dan *heuristic walkthrough* dilakukan pada iterasi formatif untuk mendeteksi masalah lebih dini. Setelah alur sistem stabil, evaluasi sumatif dilakukan menggunakan kuesioner *System Usability Scale* (SUS) untuk mengukur efisiensi dan tingkat keterpelajaran yang dirasakan pengguna [7]. Penelitian diakhiri dengan tahap finalisasi dan dokumentasi rilis, yang mencakup penghalusan elemen antarmuka (kontras, spasi, terminologi) serta konsolidasi dokumentasi teknis seperti konfigurasi, skema, templat CSV, dan parameter analitik agar sistem dapat dioperasikan dan diaudit secara konsisten lintas instalasi [3].

2.4. Algoritme dan Pseudokode

Klasterisasi dilakukan menggunakan *k-means Lloyd* dengan opsi *seeding k-means++* yang mengurangi ketidakstabilan akibat inisialisasi centroid yang buruk dan menurunkan jumlah iterasi menuju konvergensi dalam praktik [2]. Karena jarak Euclidean bergantung pada skala, fitur (jumlah siswa, jumlah guru, dan rasio turunannya) distandardisasi dengan *z-score* agar tidak ada satu besaran pun yang mendominasi langkah penugasan, dan agar pembaruan centroid merefleksikan struktur alih-alih level absolut [9]. Validitas internal dinilai dengan koefisien *silhouette* dan dilengkapi indeks *Calinski–Harabasz* serta *Davies–Bouldin* [21] untuk menyeimbangkan kohesi dan separasi dalam pemilihan *k*, sekaligus secara sadar menghindari ketergantungan eksklusif pada kriteria *elbow* sesuai rekomendasi terbaru [8]. Pseudokode di bawah ini (Algoritma 1) merangkum prosedur yang digunakan pada prototipe dan menandai titik kendali pengguna yakni pemilihan *k* dan kemampuan menjalankan ulang pipeline setelah pratinjau data sehingga keluaran analitik dapat ditriangulasi dengan ekspektasi domain sebelum temuan diformalkan dalam laporan [4].

Algoritma 1. Pseudokode Klasterisasi K-Means

Input : Baris CSV dengan {jumlah_siswa, jumlah_guru} dan rasio turunan

Output : label klaster, centroid, dan indeks validitas

```
1 X ← select({students, teachers, ratio})
2 Xs ← z_score(X) // normalization [9]
3 For k in {2..5} do // candidate k
4 C ← kmeanspp_init(Xs, k) // seeding [2]
5 repeat
6 assign xi to nearest centroid in C // Lloyd [2]
7 update each centroid as mean of members
8 until assignments stable or max_iter reached
9 compute silhouette(k), CH(k), DB(k) // validity [8]
10 select k via validity trade-offs and interpretation
11 return labels, centroids, validity for k
```

2.5. Strategi Pengujian

Pengujian mencakup aspek fungsional, integrasi, kegunaan, dan performa dalam satu narasi untuk mencerminkan urutan tugas nyata yang dilakukan pengguna saat mengoperasikan sistem secara end-to-end [5]. Uji fungsional memvalidasi integritas pemasukan data, kebenaran keluaran klasterisasi pada pengaturan parameter yang representatif, serta kestabilan visualisasi saat filter diterapkan berdasarkan kecamatan dan jenjang yang krusial untuk mendukung keputusan harian terkait ketenagaan dan pelaporan [5]. Uji integrasi menggunakan CSV yang realistis untuk menelusuri alur

<http://sistemasi.ftik.unisi.ac.id>

unggah, pratinjau, impor, analitik, visualisasi, dan ekspor sehingga transisi status dapat diamati pada batas modul dan regresi dapat direproduksi secara andal sepanjang iterasi prototipe [3]. Kegunaan ditinjau melalui heuristic walkthrough selama pengembangan dan diringkas menggunakan SUS setelah alur stabil, sehingga masukan dapat ditindaklanjuti tanpa mengorbankan kecepatan iterasi dan dengan meminimalkan beban responden [7]. Patokan performa (baseline) mencatat waktu pemasukan untuk lima hingga sepuluh ribu baris dan latensi klasterisasi untuk kandidat k , menghasilkan titik acuan yang menginformasikan affordance antarmuka seperti indikator pemuatan dan prompt pilihan parameter pada rilis berikutnya [3].

4 Hasil dan Pembahasan

Bagian ini menguraikan hasil yang diperoleh dari setiap tahapan prosedur penelitian yang telah ditetapkan, mulai dari temuan kebutuhan hingga dokumentasi akhir sistem.

3.1. Elicitasi & Temuan Kebutuhan (UCD)

Sesuai dengan tahap pertama prosedur penelitian, proses elicitasi kebutuhan dilakukan untuk memetakan konteks penggunaan yang nyata. Berdasarkan hasil lokakarya dengan pemangku kepentingan, peran operasional berhasil disarikan menjadi satu persona Administrator terkonsolidasi. Hal ini dilakukan untuk merampingkan proses validasi sistem sekaligus tetap mengakomodasi kebutuhan operator di tingkat dinas maupun sekolah. Selanjutnya, skenario tugas (task scenarios) dikembangkan untuk mencerminkan urutan kerja harian operator, yaitu: unggah $\rightarrow$ analisis $\rightarrow$ visualisasi $\rightarrow$ laporan. Pendekatan berbasis skenario ini memastikan bahwa diskusi pengembangan berlabuh pada alur kerja konkret, bukan fitur yang abstrak. Prioritas kebutuhan yang dihasilkan berfokus pada mekanisme umpan balik kualitas data, transparansi pemilihan parameter k , serta kemudahan ekspor bukti untuk rapat koordinasi. Rincian kebutuhan fungsional yang diperoleh dirangkum dalam Tabel 2.

Tabel 2 Temuan kebutuhan (requirements findings)

ID	Functional Requirement (FR)	Alasan / Sumber
FR-01	Unggah CSV (pemisah ';') dengan pratinjau header & validasi	Mencegah kesalahan pemasukan yang tak terdeteksi; selaras praktik CSV yang ada
FR-02	Analitik: K-Means dengan k yang dipilih pengguna dan skala z-score	Mendukung segmentasi eksploratif dengan parameter transparan
FR-03	Visualisasi: sebar (scatter) dan ringkasan per klaster	Mengomunikasikan struktur ke pemangku kepentingan non-teknis
FR-04	Ekspor hasil ke CSV	Membuat <i>audit trail</i> dan memungkinkan berbagi luring
FR-05	Filter berdasarkan kecamatan, jenjang, dan status	Mendukung triase dan perbandingan lokal
ID	Functional Requirement (FR)	Alasan / Sumber
FR-01	Unggah CSV (pemisah ';') dengan pratinjau header & validasi	Mencegah kesalahan pemasukan yang tak terdeteksi; selaras praktik CSV yang ada
FR-02	Analitik: K-Means dengan k yang dipilih pengguna dan skala z-score	Mendukung segmentasi eksploratif dengan parameter transparan
FR-03	Visualisasi: sebar (scatter) dan ringkasan per klaster	Mengomunikasikan struktur ke pemangku kepentingan non-teknis
FR-04	Ekspor hasil ke CSV	Membuat <i>audit trail</i> dan memungkinkan berbagi luring

3.2. Desain Sistem (UML)

Rancangan alur sistem secara keseluruhan diperlihatkan pada Gambar 3, yang mengilustrasikan urutan proses operasional mulai dari autentikasi hingga pelaporan. Berdasarkan alur tersebut, model

use-case mendefinisikan cakupan sistem pada lima kapabilitas utama: kelola pengguna, impor data, jalankan analitik, visualisasikan hasil, dan ekspor laporan. Detail logika dalam *activity diagram* kemudian mengungkap jalur pengecualian penting, seperti penanganan kolom wajib yang kosong atau kesalahan label *header* yang secara otomatis ditangkap pada tahap pratinjau.

Untuk mendukung alur kerja tersebut, pandangan *class diagram* menstabilkan skema basis data di sekitar entitas *users* dan *schools*, termasuk akomodasi kolom opsional untuk rasio siswa-guru. Sementara itu, interaksi antar komponen diperjelas melalui *sequence diagram* yang memetakan tanggung jawab spesifik antara *controller* dan *service*, memastikan pipa proses impor–analisis–ekspor berjalan secara terstruktur dan valid.



Gambar 3 Alur sistem

Selain alur sistem, spesifikasi teknis diperjelas menggunakan empat jenis diagram UML utama yang rincian tujuan dan bukti kuncinya dirangkum dalam Tabel 3. Sebagaimana tercantum dalam tabel tersebut, model *use-case* digunakan untuk mematok cakupan pada lima kapabilitas inti (kelola pengguna, impor, analitik, visualisasi, ekspor). Detail logika operasional, termasuk penanganan pengecualian seperti kesalahan *header*, dipetakan dalam *activity diagram*. Struktur informasi distabilkan melalui *class diagram* yang mencakup entitas pengguna dan sekolah, sedangkan interaksi teknis antar-modul (*controller-service*) diperjelas melalui *sequence diagram*.

Tabel 3 UML

Diagram	Tujuan	Bukti Kunci dari Sistem
Use Case	Mendefinisikan cakupan & aktor (Administrator)	Login, Import CSV, Run K-Means, Visualize, Export, Manage Users
Activity	Alur operasional & pengecualian	Upload → Preview (validate) → Import → Analyze → Visualize → Export
Class	Entitas & atribut	users(id, email, name, password_hash), schools(id, kecamatan, nama_sekolah, siswa, guru, jenjang, status, created_at)
Sequence	Interaksi modul	Controller→Service: validasi CSV, hitung k-means, simpan/sajikan hasil

3.3. Iterasi Prototipe

Pengembangan sistem dilaksanakan melalui serangkaian siklus pendek yang secara bertahap mematangkan fitur, mulai dari validasi data hingga penyempurnaan antarmuka. Rincian fokus pengembangan dan perubahan teknis yang diperkenalkan pada setiap iterasi dirangkum dalam Tabel 4. Sebagaimana terlihat pada tabel tersebut, iterasi awal berpusat pada penguatan alur unggah dan kontrol analitik, sedangkan iterasi lanjutan memprioritaskan visualisasi, fitur ekspor, dan harmonisasi elemen (*polishing*). Umpan balik pengguna sepanjang siklus ini sangat krusial, di mana permintaan akan legenda yang lebih jelas dan fitur penyimpanan parameter terakhir menjadi kunci untuk mempercepat analisis berulang dalam satu sesi.

Tabel 4 Fokus iterasi

Iterasi	Fokus	Perubahan yang Diperkenalkan
---------	-------	------------------------------

I-1	Upload & Preview	Penegakan urutan header; pesan validasi <i>inline</i>
I-2	Analytics Panel	Pemilih <i>k</i> , <i>toggle</i> z-score (aktif bawaan), tampilan nilai silhouette
I-3	Visualization	Scatter dengan legenda klaster; filter per kecamatan/jenjang
I-4	Export	Ekspor CSV untuk label klaster & centroid
I-5	Polish	Harmonisasi terminologi; spasi; persistensi parameter

3.4. Pemasukan Data & Kualitas

Fitur pemasukan data dirancang untuk mengenali format CSV dengan pemisah titik koma dan menegakkan validasi urutan *header* yang ketat pada setiap proses eksekusi. Tahap pratinjau berfungsi sebagai gerbang validasi yang secara efektif mencegah isu kualitas data umum, seperti ketidaksesuaian label *header*, karakter non-numerik pada kolom angka, serta nilai wajib yang hilang, sehingga koreksi dapat dilakukan pengguna sebelum data diimpor. Efektivitas mekanisme validasi ini diringkas dalam Tabel 5, yang menyajikan statistik representatif dari uji impor terhadap ribuan baris data beserta jenis kesalahan utama yang berhasil ditangkap oleh sistem.

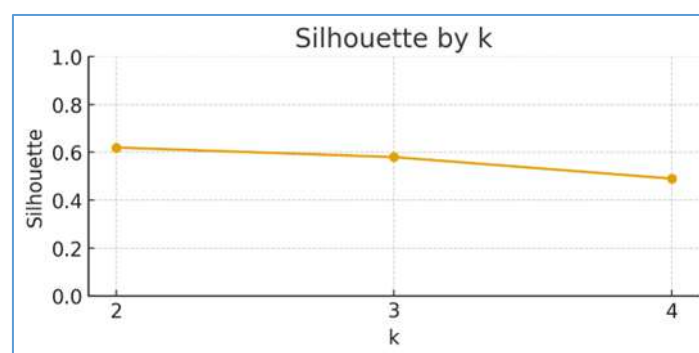
Tabel 5 Data quality

Run	Baris (Total)	Diterima	Ditolak	Isu Utama (Ditolak)
R1	5.000	4.928	72	Ketidaksesuaian label header
R2	7.500	7.442	58	Karakter non-numerik pada hitungan
R3	10.000	9.934	66	Kolom wajib kosong

3.5. Hasil Klasterasi

Klasterisasi dijalankan menggunakan fitur yang telah distandardisasi (total siswa, total guru, dan rasio siswa-guru) agar perhitungan jarak merefleksikan keserupaan struktural yang valid. Evaluasi validitas klaster untuk menentukan jumlah kelompok optimal (k) ditunjukkan pada Gambar 4. Grafik tersebut mengilustrasikan tren nilai *Silhouette* untuk kandidat $k \in \{2,3,4\}$. Terlihat bahwa nilai koefisien memuncak pada $k=2$ (0,62), yang menandakan pemisahan terbersih dan ideal untuk pelaporan manajerial. Meskipun demikian, opsi $k=3$ (0,58) tetap dipertahankan sebagai pilihan yang layak untuk keperluan triase operasional yang lebih halus.

Implementasi hasil analisis ini ke dalam sistem ditampilkan pada Gambar 6. Antarmuka analitik ini dirancang transparan, di mana pengguna dapat melihat langsung indikator skor *Silhouette* (kotak hijau) dan memilih jumlah klaster melalui panel kontrol di sisi kiri. Tabel hasil di bagian bawah secara otomatis menampilkan label klaster untuk setiap sekolah, memudahkan operator Dinas Pendidikan untuk memvalidasi segmen sekolah (misalnya: "Beban Tinggi" vs "Moderat") sebelum diekspor menjadi laporan.



Gambar 4 Silhoutte pada berbagai k

Berdasarkan hasil yang diperoleh, penjelasan data ditampilkan pada Tabel 6.

Tabel 6 Penjelasan silhouette

<i>k</i>	Silhouette	Interpretasi
2	0,62	Dua segmen luas: beban ketenagaan lebih rendah vs lebih tinggi
3	0,58	Kenaikan marginal; kohesi sedikit menurun
4	0,49	Over-segmentation dengan partisi kurang stabil
Cluster	Centroid: Students	Centroid: Teachers
C0	~420	~26
		Centroid: Ratio (Stud/Teach)
		~16.2
		Qualitative Label
		Moderate load

k	Silhouette	Interpretasi
C1	~1,050 ~45	~23.3 High load

3.6. Visualisasi & Interpretasi

Filter per kecamatan dan jenjang dengan cepat menyingkap kantong rasio yang tinggi, mendukung diskusi penataan ulang sementara dan rekrutmen. Penyimpanan parameter per run memungkinkan analis membuka kembali konfigurasi dan mengeksplor tabel konsisten untuk rapat koordinasi.

3.7. Penguji & Evaluasi

Pengujian fungsional dilaksanakan dengan mengikuti skenario tugas yang mencakup validasi aturan pemasukan, akurasi keluaran klasterisasi, serta kestabilan visualisasi saat fitur penyaringan diterapkan. Selain itu, *heuristic walkthrough* dilakukan untuk membimbing perbaikan redaksi dan tata letak antarmuka. Setelah alur sistem stabil, evaluasi kegunaan (*usability*) diukur menggunakan instrumen *System Usability Scale* (SUS). Hasil penilaian rinci dari setiap responden dirangkum dalam Tabel 7. Data tersebut menunjukkan skor rata-rata sebesar **82,0**, angka yang menegaskan bahwa sistem memiliki tingkat penerimaan yang baik (*acceptable*) dan mudah dipelajari tanpa membebani pengguna dalam operasional rutin.

Tabel 7 Skor SUS responden

Respondent	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	SUS Score
R01	4	2	4	2	4	2	4	2	4	2	80.0
R02	5	2	4	2	5	2	4	2	4	2	85.0
R03	4	2	4	2	4	2	4	2	4	2	80.0
R04	4	3	4	2	4	2	4	2	4	2	77.5
R05	5	2	5	2	4	2	4	2	4	2	87.5

Rerata SUS sekitar 82,0 yang menandakan kegunaan dapat diterima dan sejalan dengan penurunan pengerjaan ulang saat penyiapan data.

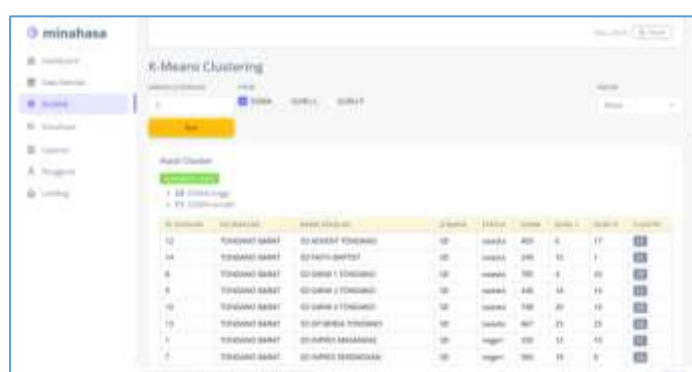
3.8. Finalis & Dokumentasi

Iterasi terakhir berfokus pada harmonisasi terminologi, peningkatan kontras di legenda, dan standarisasi spasi pada halaman Dashboard dan Analytics. Dokumentasi mengonsolidasikan detail konfigurasi, skema dan kamus data, templat CSV, serta parameter analitik. Cuplikan antarmuka final disertakan agar dokumentasi sesuai dengan sistem operasional.



Gambar 5 Unggah CSV pratinjau header dan validasi

Gambar 5 menampilkan halaman Unggah Data Sekolah (CSV). Komponen kunci: pemilih berkas, tombol Preview dan Import, catatan format header, serta panel Sample CSV di sisi kanan.



Gambar 6 Antarmuka *Analytic (K-Means)*

Gambar 6 menampilkan halaman K-Means Analytics dengan kontrol jumlah kluster (k), pemilihan fitur, tombol Run, indikator nilai silhouette, dan tabel hasil kluster.

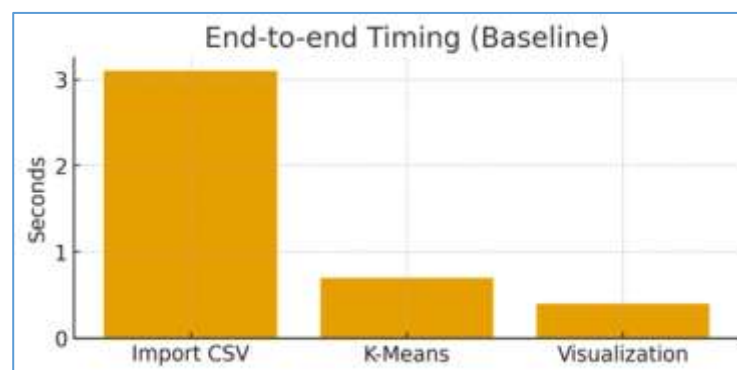


Gambar 7 Antarmuka dashboard

Gambar 7 menampilkan Dashboard final: kartu KPI (total sekolah per jenjang, total siswa, total guru, rasio siswa/guru) dan grafik distribusi guru.

3.9. Diskusi & Implikasi Kebijakan

Hasil yang mengikuti prosedur menunjukkan bahwa pemasukan data yang sederhana namun tervalidasi dengan baik, dipasangkan dengan klusterisasi yang transparan, dapat meningkatkan perencanaan rutin tanpa menambah kompleksitas alat. Pelaporan dua segmen dari $k = 2$ memberi titik awal yang jelas untuk diskusi ketenagaan, sementara antarmuka mendorong pemeriksaan eksploratif pada $k = 3$ ketika kecamatan tertentu membutuhkan triase lebih halus. Tabel yang dapat diekspor dan parameter yang tersimpan meningkatkan reproduisibilitas serta menciptakan bukti yang terlacak untuk rapat koordinasi.



Gambar 8 Antarmuka dashboard

Patokan waktu pada lingkungan uji: ~3,1 detik untuk impor CSV (5–10 ribu baris), ~0,7 detik untuk K-Means ($k \in \{2..5\}$), dan ~0,4 detik untuk visualisasi mendukung interaksi yang responsif dalam analisis rutin.

3.10. Keterbatasan & Pekerjaan Mendatang

Bekerja dengan agregat tingkat sekolah membatasi interpretasi rinci atas heterogenitas per jenjang/kelas dan efek program. Pekerjaan mendatang akan memperluas skema dengan jumlah kelas/rombongan belajar, status akreditasi, dan deret historis agar perbandingan longitudinal dan himpunan fitur yang lebih kaya dapat diintegrasikan. Kami juga berencana menyimpan label kluster per periode dan menambahkan opsi ekspor figur untuk merampingkan pelaporan.

5 Kesimpulan

Penelitian ini berhasil mengembangkan sistem profil sekolah berbasis web yang mengintegrasikan *User-Centered Design* (UCD) dan algoritma *K-Means* untuk mendukung perencanaan berbasis data di Dinas Pendidikan Kabupaten Minahasa. Hasil pengujian menunjukkan

<http://sistemasi.ftik.unisi.ac.id>

bahwa algoritma bekerja optimal pada solusi dua kluster ($k=2$) dengan nilai *Silhouette* 0,62, didukung oleh skor penerimaan pengguna (*System Usability Scale*) rata-rata 82,0 dan waktu respons sistem di bawah 3,1 detik, yang secara efektif menjawab kesenjangan penelitian terdahulu dengan menawarkan transparansi parameter analitik bagi pengguna non-teknis dibandingkan pendekatan "kotak hitam" konvensional. Meskipun sistem telah beroperasi secara fungsional, keterbatasan pada penggunaan data agregat dan ketiadaan fitur pelacakan historis mengisyaratkan perlunya pengembangan lanjutan untuk memperluas atribut data mencakup rombongan belajar dan status akreditasi, serta menambahkan modul analisis longitudinal guna memantau tren kinerja sekolah antar-periode secara lebih komprehensif.

Ucapan Terima Kasih

Kami menyampaikan terima kasih kepada Dinas Pendidikan Kabupaten Minahasa atas kepercayaan, akses data, dan umpan balik konstruktif selama evaluasi iteratif. Kami juga berterima kasih kepada tim operator di tingkat dinas dan sekolah yang partisipasinya dalam walkthrough dan telaah kegunaan sangat meningkatkan kejelasan pesan validasi, legenda, serta pengurutan alur kerja. Apresiasi kami sampaikan kepada rekan-rekan yang memberikan masukan kebahasaan, saran penulisan, dan proofreading yang memperkuat penyajian artikel ini. Segala kekeliruan yang tersisa sepenuhnya menjadi tanggung jawab penulis.

Referensi

- [1] C. Abras, D. Maloney-Krichmar, and J. Preece, "User-Centered Design," in *Encyclopedia of Human-Computer Interaction*, W. S. Bainbridge, Ed., 2004. [Online]. Available: https://aim.johnkeston.com/wp-content/uploads/2012/01/User-centered_design_encyclopedia_chapter.pdf
- [2] J. Brooke, "SUS: A 'Quick and Dirty' Usability Scale," in *Usability Evaluation in Industry*, P. W. Jordan, B. Thomas, I. L. McClelland, and B. Weerdmeester, Eds., 1996. [Online]. Available: https://digital.ahrq.gov/sites/default/files/docs/survey/systemusabilityscale%28sus%29_comp%5B1%5D.pdf
- [3] IBM Corporation, "IBM SPSS Modeler CRISP-DM Guide," 2021. [Online]. Available: https://www.ibm.com/docs/it/SS3RA7_18.3.0/pdf/ModelerCRISPDm.pdf
- [4] International Organization for Standardization, *ISO 9241-210:2019 Ergonomics of human-system interaction Human-centred design for interactive systems*, 2019. [Online]. Available: <https://cdn.standards.iteh.ai/samples/77520/8cac787a9e1549e1a7ffa0171dfa33e0/ISO-9241-210-2019.pdf>
- [5] J. MacQueen, "Some Methods for Classification and Analysis of Multivariate Observations," in *Proc. 5th Berkeley Symp. on Mathematical Statistics and Probability*, Vol. 1, pp. 281–297, 1967. [Online]. Available: <https://www.cs.cmu.edu/~bhiksha/courses/mlsp.fall2010/class14/macqueen.pdf>
- [6] L. Morissette and S. Chartier, "The K-Means Clustering Technique: General Considerations and Implementation," *The Quantitative Methods for Psychology*, Vol. 9, No. 1, pp. 15–24, 2013. [Online]. Available: <https://www.tqmp.org/RegularArticles/vol09-1/p015/p015.pdf>
- [7] P. J. Rousseeuw, "Silhouettes: A Graphical Aid to the Interpretation and Validation of Cluster Analysis," *Journal of Computational and Applied Mathematics*, Vol. 20, pp. 53–65, 1987. [Online]. Available: <https://wis.kuleuven.be/stat/robust/papers/publications-1987/rousseeuw-silhouettes-jcam-sciencedirectopenarchiv.pdf>
- [8] D. A. N. Wulandari, "An Educational Data Mining for Student Academic Prediction (K-Means Bayes)," *Jurnal Pilar*, Vol. 17, No. 2, pp. 67–74, 2021. [Online]. Available: <https://repository.bsi.ac.id/repo/files/295599/download/Jurnal-Pilar-September-2021.pdf>

- [9] S. R. Arifin et al., “System Usability Scale (SUS) Implementation in Ruang Perpustakaan,” *Matrix: Jurnal Manajemen Teknologi dan Informatika*, 2024. [Online]. Available: <https://ojs2.pnb.ac.id/index.php/MATRIX/article/download/1263/844/16038>
- [10] M. H. P. Swari, “Pengelompokan Wilayah berdasarkan Rasio Guru-Murid,” *Jurnal Ilmiah Digitalita*, 2022. [Online]. Available: <https://ejournal.itbwigalumajang.ac.id/index.php/jid/article/download/898/519/3363>
- [11] D. Arthur and S. Vassilvitskii, “K-Means++: The Advantages of Careful Seeding,” in *Proc. 18th Annu. ACM-SIAM Symp. Discrete Algorithms (SODA)*, 2007. [Online]. Available: <https://theory.stanford.edu/~sergei/papers/kMeansPP-soda.pdf>
- [12] S. P. Lloyd, “Least Squares Quantization in PCM,” *IEEE Trans. Inf. Theory*, Vol. 28, No. 2, pp. 129–137, 1982. [Online]. Available: <https://www.stat.cmu.edu/~rnugent/PCMI2016/papers/LloydKMeans.pdf>
- [13] P. Chapman et al., “CRISP-DM 1.0: Step-by-Step Data Mining Guide,” SPSS Inc., 2000. [Online]. Available: <https://public.dhe.ibm.com/software/analytics/spss/documentation/modeler/14.2/es/CRISP-DM.pdf>
- [14] E. Schubert and P. J. Rousseeuw, “Stop using the Elbow Criterion for K-Means,” *SIGKDD Explorations*, Vol. 22, No. 1, pp. 36–45, 2020. [Online]. Available: https://kdd.org/exploration_files/p36-elbow.pdf
- [15] M. Fritz, T. Kriegel, P. Schubert, and A. Zimek, “LOG-Means: Efficiently Estimating the Number of Clusters,” *Proc. VLDB Endowment (PVLDB)*, Vol. 13, No. 12, pp. 2118–2132, 2020. [Online]. Available: <https://www.vldb.org/pvldb/vol13/p2118-fritz.pdf>
- [16] C. Schröer, F. Kruse, and J. M. G. Gómez, “A Systematic Literature Review on Applying CRISP-DM,” *Procedia Computer Science*, Vol. 181, pp. 526–534, 2021. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1877050921002416>
- [17] J. Nielsen and R. Molich, “Heuristic Evaluation of User Interfaces,” in *Proc. CHI ’90*, 1990, pp. 249–256. [Online]. Available: <https://concreta.com.uy/wp-content/uploads/nielsenheuristicsCHI.pdf>
- [18] B. Desgraupes, “Clustering Indices,” *ClusterCrit R Package Vignette*, 2017. [Online]. Available: <https://cran.r-project.org/web/packages/clusterCrit/vignettes/clusterCrit.pdf>
- [19] M. M. A. Rahman, M. N. Islam, and M. A. Islam, “Standardization and Its Effects on K-Means Clustering Algorithm,” *Research Journal of Applied Sciences, Engineering and Technology*, Vol. 6, pp. 3299–3303, 2013. [Online]. Available: <https://maxwellsci.com/print/rjaset/v6-3299-3303.pdf>
- [20] M. Maguire, “Methods to Support Human-Centred Design,” *International Journal of Human-Computer Studies*, 2001. [Online]. Available: <https://www.criiasupr.org/multimedia/documents/Human%20Centred%20Design-Martin%20Maguire.pdf>
- [21] A. Duaulu, G. D. P. Maramis, and K. Santa, “Penerapan Algoritma K-Means Clustering untuk Pengelompokan Zona Nilai Tanah berbasis WebGIS,” *SemanTIK: Teknik Informasi*, Vol. 11, No. 2, 2025. [Online]. Available: <https://semantik.uho.ac.id/index.php/journal/article/view/220>
- [22] E. R. S. Moningkey, “Analisis dan Perancangan Sistem Informasi Perpustakaan Ruang Baca Jurusan PTIK Fatek Unima,” *Engineering Education Journal-E2J*, Vol. 4, No. 3, Aug. 2016.
- [23] S. Kumajas, “Penerapan Algoritma K-Means untuk mengelompokan Daerah Asal pada Siswa SMK Kristen Harapan Rantepao Toraja Utara,” 2025. [Online]. Available: <https://jurnal.polgan.ac.id/index.php/jmp/article/view/15431>
- [24] W. Herulambang, M. M. Hidayat, and N. A. Putri, “Educational Data Mining for Mapping the Comprehension of personality subject using K-Means algorithm (case study of SD Insan Mulya),” *Jurnal Mandiri IT*, vol. 11, No. 4, Apr. 2023. [Online]. Available: <https://ejournal.isha.or.id/index.php/Mandiri/article/view/194>

- [25] D. Apriandi, R. M. Sari, and M. I. Sarif, “Analisis Clustering untuk menentukan Siswa Berprestasi di SMK Swasta TI Panca Dharma Stungkit menggunakan Metode *K-Means*,” *Jurnal Minfo Polgan*, vol. 13, no. 1, 2024. [Online]. Available: <https://doi.org/10.33395/jmp.v13i1.13959>
- [26] A. Nisa, M. K. Anam, H. Yenni, P. Kudadiri, and G. Gunadi, “*Clustering Junior Schools in Implementing Smart School using the K-Means in Pekanbaru*,” *Jurnal Sains, Nalar, dan Aplikasi Teknologi Informasi*, Vol. 3, No. 3, 2024. [Online]. Available: <https://doi.org/10.20885/snati.v3.i3.40>
- [27] S. Rahmah, M. Jamil, A. Aslindah, A. B. Fawait, and Y. F. Saputra, “Pengelompokan hasil pembelajaran mahasiswa dengan algoritma *K-Means Clustering*,” *Jurnal Pengabdian Masyarakat dan Riset Pendidikan*, Vol. 3, No. 4, pp. 4215–4221, 2025. [Online]. Available: <https://doi.org/10.31004/jerkin.v3i4.1217>
- [28] V. Maulida, N. Mulyani, and M. F. L. Sibuea, “Sistem Klasifikasi Strata Kelas Peserta Kursus berbasis *Web* menggunakan algoritma *K-Means*,” *Edumatic: Jurnal Pendidikan Informatika*, Vol. 8, No. 2, pp. 477–486, 2024. [Online]. Available: <https://doi.org/10.29408/edumatic.v8i2.27311>