

Penerapan Algoritma *K-Means* Untuk Mengelompokan Daerah Asal Pada Siswa SMK Kristen Harapan Rantepao Toraja Utara

¹Lewi Yusrel Winarto, ²Sondy C. Kumajas, ³Alfiansyah Hasibuan

^{1,2,3}Prodi Teknik Informatika, Universitas Negeri Manado, Indonesia

lewi Yusrel winarto@gmail.com, sondykumajas@unima.ac.id, alfiansyahhasibuan@unima.ac.id

Submit : 13 Okt 25 | Diterima : 19 Okt 2025 | Terbit : 21 Okt 2025

ABSTRAK

Tujuan penelitian ini adalah memanfaatkan teknik data mining dengan menggunakan algoritma clustering K-means untuk mengkategorikan siswa berdasarkan tempat asal mereka di SMK Kristen Harapan yang terletak di Rantepao, Toraja Utara. Pengelompokan ini dilakukan untuk membantu pihak sekolah dalam memahami distribusi geografis siswa dan mendukung perencanaan strategis, seperti promosi penerimaan siswa baru serta pemerataan fasilitas pendidikan. Data yang digunakan berasal dari data pendaftaran siswa baru angkatan 2022–2025, dengan atribut meliputi daerah asal, jenis kelamin, dan asal sekolah. Proses penelitian meliputi tahapan preprocessing data, dengan menggunakan metode K-Means di Microsoft Excel, serta evaluasi hasil dengan metode *Elbow* dan perhitungan *Sum of Squared Errors* (SSE). Temuan penelitian menunjukkan bahwa jumlah kluster yang ideal adalah tiga ($K=3$), yang terdiri dari: kluster pertama berisi kecamatan dengan jumlah siswa terbanyak seperti Rantepao dan Buntu Pepsan, kluster kedua dengan jumlah siswa sedang seperti Sanggalangi dan Awan Rantekarua, serta kluster ketiga dengan jumlah siswa sedikit seperti Rindingallo dan Balusu. Penerapan metode K-Means telah terbukti efisien dalam menampilkan pola distribusi siswa secara visual dan informatif melalui aplikasi berbasis web, sehingga dapat dijadikan dasar pengambilan keputusan dalam pengelolaan dan pengembangan sekolah.

Kata Kunci: *K-Means Clustering*, data mining, pengelompokan siswa, SMK Kristen Harapan Rantepao, analisis data pendidikan

PENDAHULUAN

Pendidikan adalah salah satu fondasi penting dalam upaya membangun sumber daya manusia yang bermutu. Di Indonesia, Sekolah Menengah Kejuruan (SMK) berperan penting dalam menyediakan keterampilan dan pengetahuan praktis yang dibutuhkan oleh dunia industri (Diah, 2021). Salah satu SMK yang berkomitmen dalam meningkatkan kualitas pendidikan adalah SMK Kristen Harapan Rantepao di Toraja Utara. Sekolah ini tidak hanya diminati oleh siswa dari daerah setempat, tetapi juga oleh pelajar yang berasal dari berbagai wilayah sekitarnya. Peran SMK dalam pembangunan nasional semakin krusial di era globalisasi, di mana lulusan SMK diharapkan menjadi tenaga kerja siap pakai dengan kompetensi sesuai kebutuhan industri (Andayani, 2021). Namun, terdapat kesenjangan antara kompetensi lulusan dan kebutuhan industri, sehingga pengelolaan pendidikan tidak cukup berfokus pada kurikulum, melainkan juga pada pengelolaan data siswa secara komprehensif. Data mengenai asal daerah siswa dapat memberikan perspektif tambahan dalam memahami faktor sosial dan kultural yang memengaruhi motivasi belajar, akses informasi, serta peluang siswa setelah lulus.

Keberagaman daerah asal siswa menjadi salah satu ciri khas SMK Kristen Harapan Rantepao. Siswa di sekolah ini tidak hanya berasal dari Rantepao dan sekitarnya, tetapi juga dari berbagai wilayah di Toraja Utara maupun luar daerah. Keberagaman tersebut menciptakan lingkungan belajar yang heterogen dan dinamis, tetapi juga menghadirkan tantangan bagi pihak sekolah. Perbedaan latar belakang daerah asal dapat memengaruhi kebutuhan, kebiasaan, serta pola interaksi sosial siswa. Apabila tidak dikelola dengan tepat, kondisi ini dapat berpotensi

menimbulkan kesenjangan dalam pelayanan pendidikan, ketidakseimbangan dalam perencanaan kegiatan, hingga kurang optimalnya strategi pembelajaran.

Oleh karena itu, sekolah perlu memahami karakteristik siswa berdasarkan daerah asal mereka. Informasi ini dapat membantu pihak manajemen dalam penyusunan kebijakan penerimaan siswa baru, perencanaan fasilitas asrama atau transportasi, serta pengembangan program ekstrakurikuler yang sesuai dengan keberagaman siswa. Dengan pemetaan yang tepat, sekolah dapat merespons kebutuhan spesifik siswa dari kelompok tertentu sehingga pelayanan pendidikan menjadi lebih adil, efisien, dan efektif.

Di era digital saat ini, data pendidikan sering kali hanya dimanfaatkan untuk pencatatan administrasi seperti jumlah siswa, nilai akademik, atau kehadiran. Padahal, melalui pendekatan analisis data, informasi tersebut dapat menjadi dasar pengambilan keputusan yang lebih strategis (Iraza, 2023). Misalnya, dengan mengetahui konsentrasi asal daerah siswa, sekolah dapat menyusun strategi promosi yang lebih terarah di daerah dengan potensi besar dan melakukan pendekatan khusus di wilayah yang masih minim representasi. Analisis data dengan demikian tidak hanya mendukung pengelolaan internal, tetapi juga memperkuat keberlanjutan sekolah dalam menghadapi persaingan antar lembaga pendidikan.

Salah satu pendekatan untuk mengkategorikan siswa berdasarkan latar belakang geografis adalah melalui algoritma K-Means. Teknik ini banyak digunakan untuk pengelompokan dalam bidang analisis data (Pratama et al., 2018). Melalui penerapan K-Means, data siswa dapat diorganisasikan ke dalam beberapa kelompok berdasarkan ciri-ciri yang sama, seperti daerah asal. Penelitian ini menggunakan pendekatan berbasis web agar hasil klusterisasi dapat diakses dengan mudah oleh pihak sekolah, guru, dan pemangku kepentingan lainnya. Aplikasi berbasis web ini dirancang dengan antarmuka yang intuitif dan menampilkan hasil analisis secara visual dan informatif.

Melalui penelitian ini diharapkan diperoleh informasi yang bermanfaat untuk pengambilan keputusan dalam pengelolaan sekolah, seperti penyusunan program pengajaran yang sesuai dengan kebutuhan siswa dari berbagai daerah serta perencanaan kegiatan yang memperkuat kerjasama antar siswa. Selain itu, penelitian ini seharusnya menjadi panduan bagi organisasi pendidikan lain dalam mengelola keberagaman demografis siswa. Analisis berbasis data membantu sekolah beradaptasi terhadap perubahan pola rekrutmen dan minat masyarakat terhadap jenis sekolah tertentu. Secara akademis, penelitian ini memperkaya penerapan algoritma K-Means di bidang pendidikan, sementara secara praktis, hasilnya dapat dijadikan model implementasi analisis data yang sederhana namun bermanfaat bagi sekolah lain di Toraja maupun daerah lain di Indonesia.

TINJAUAN PUSTAKA

Data Mining

Data mining melibatkan identifikasi pola, koneksi, atau wawasan berharga dari kumpulan data yang luas dengan menggunakan metode statistika, matematika, kecerdasan buatan, dan pembelajaran mesin. Tujuan utama penambangan data adalah guna menemukan informasi yang bermanfaat dan sebelumnya tersembunyi di dalam data yang tersedia. Teknik ini diterapkan secara luas di berbagai sektor, termasuk bisnis, layanan kesehatan, keuangan, pemasaran, dan pendidikan, untuk meningkatkan pengambilan keputusan berdasarkan data.

Tahapan *Knowledge Discovery In Database (KDD)*

Menurut Fayyad dalam Kusri (2009), istilah penambangan data dan penemuan pengetahuan dalam basis data (KDD) sering digunakan secara sinonim untuk merujuk pada metode mengungkap wawasan tersembunyi dalam basis data yang ekstensif. Umumnya, proses KDD terdiri dari lima tahap utama: pemilihan data, prapemrosesan/pembersihan, transformasi, penambangan data, dan interpretasi/evaluasi.



Gambar 1. Tahapan KDD

Machine Learning

Machine Learning (ML) adalah bagian dari kecerdasan buatan (AI) yang didedikasikan untuk menciptakan sistem yang dapat belajar secara mandiri tanpa perlu pengkodean yang konstan (Alfarizi dkk., 2023). Salah satu pendekatan yang banyak digunakan dalam ML adalah *support vector machine (SVM)*, yang merupakan metode klasifikasi yang terkait dengan pembelajaran terbimbing.

K-means

K-Means adalah teknik pengelompokan data datar yang mengelompokkan data ke dalam berbagai kelompok (Andiani dkk., 2022). Teknik ini mengkategorikan data berdasarkan kesamaan sifatnya, memastikan bahwa data dengan sifat yang berbeda ditempatkan dalam kelompok terpisah. Tujuan utama pendekatan pengelompokan ini adalah untuk mengurangi fungsi objektif dalam setiap kelompok sekaligus meningkatkan disparitas, atau perbedaan, antar kelompok (Purnamasari dkk., 2023).

Teknik *K-Means* adalah algoritma yang mengurutkan data ke dalam beberapa kluster berdasarkan kemiripannya, atau berdasarkan fiturnya. Prosedurnya dimulai dengan menentukan jumlah kluster (*K*) yang akan dibuat dan memilih pusat kluster awal secara acak. Setiap titik data individual kemudian dihubungkan ke kluster terdekat berdasarkan jarak, yang sebagian besar dinilai menggunakan jarak Euclidean. Selanjutnya, pusat kluster diperbarui dengan menghitung rata-rata semua titik data dalam kluster tersebut. Proses pengelompokan ini diulang hingga tidak ada lagi perubahan pada pusat kluster, yang menunjukkan bahwa hasil pengelompokan telah stabil.

Unified Modelling Language (UML)

Menurut Sukamto & Shalahuddin pada penelitian Fadila (2019) bahwa “UML merupakan suatu bahasa pemodelan yang digunakan untuk menggambarkan dan menganalisis sistem berbasis objek (Mubarak et al., 2019). UML terdiri dari berbagai macam diagram yang digunakan untuk menggambarkan berbagai aspek dari sistem, seperti struktur kelas, hubungan antar kelas, diagram aktivitas, diagram *use case*, dan lain sebagainya” (Fadilla, 2019).

METODE PENELITIAN

A. Waktu dan Tempat Penelitian

Penelitian ini dilaksanakan selama tiga bulan, yakni dalam rentang waktu dari Juli hingga Oktober 2024, dilakukan di SMK Kristen Harapan Rantepao, yang terletak di wilayah Rantepao, Kabupaten Toraja Utara, Provinsi Sulawesi Selatan

B. Alat dan Bahan Penelitian

- 1) Perangkat Keras (*Hardware*)
 - a. Laptop dengan spesifikasi Processor Intel Celeron N4020, Up To 2.8 GHz X144M.
 - b. Memory RAM 4GB
- 2) Perangkat Lunak (*Software*)
 - a. Sistem Operasi Windows 11

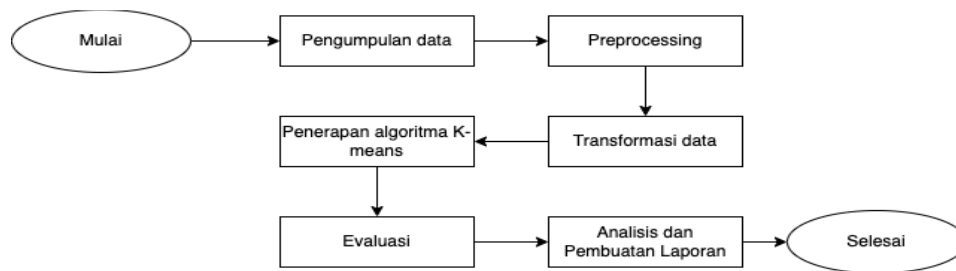
- b. Browser Google Chrome
- c. XAMPP
- d. Sublime Text
- e. MYSQL
- f. PHP

C. Teknik Pengumpulan Data

Pada tahap ini, peneliti menggunakan beberapa metode dalam proses pengumpulan data, yaitu wawancara, studi pustaka, dan observasi (Tiauw et al., 2024).

D. Metode Penelitian

Adapun tahapan-tahapan yang dilakukan dalam penelitian ini adalah sebagai berikut:



Gambar 2. Alur Penelitian

1) Pengumpulan Data

Pengumpulan data dilakukan dengan menduplikasi dokumen Excel yang disediakan oleh divisi akademik institusi pendidikan. Dokumen-dokumen ini berisi data yang belum diolah dan belum dianalisis. Data yang diperoleh terdiri atas satu file dengan dua lembar kerja (sheet), yang memuat informasi mengenai daerah asal siswa kelas XII tahun ajaran 2022–2025 dan kelas XI tahun ajaran 2023–2025.

2) *Preprocessing*

Persiapan data merupakan fase penting dalam alur kerja penambangan data yang dimaksudkan untuk mengubah data yang belum diproses menjadi data yang telah disempurnakan, konsisten dan siap untuk dianalisis. Kualitas data input sangat menentukan keakuratan hasil analisis, sehingga proses ini perlu dilakukan secara teliti untuk menghindari kesalahan seperti duplikasi, kekosongan, atau ketidakkonsistenan data.

Tahapan preprocessing meliputi tiga langkah utama:

- a. *Data Reduction*, yaitu proses pengurangan atribut atau data yang dianggap tidak relevan agar hanya informasi penting yang digunakan dalam analisis, yakni atribut kecamatan asal siswa.
- b. *Data Cleaning*, yaitu proses memperbaiki ketidakkonsistenan data, mengisi nilai yang hilang, serta menghapus data yang bersifat duplikat. Tahapan ini dilakukan menggunakan *Microsoft Excel* dengan memeriksa setiap kolom guna memastikan keseragaman dan ketepatan penulisan data.
- c. *Data Transformation*, yaitu mengubah format data agar sesuai kebutuhan analisis. Karena metode K-Means hanya mampu menangani data yang bersifat numerik, maka data teks pada atribut kecamatan diubah menjadi bentuk angka agar dapat diproses lebih lanjut.

3) Penerapan Algoritma *K-means Clustering*

Setelah tahap praproses selesai dan kumpulan data siap digunakan, langkah berikutnya melibatkan penerapan teknik pengelompokan *K-Means* menggunakan *Microsoft Excel*. Proses penerapan ini dilakukan melalui beberapa tahapan utama.

Pertama, menentukan titik pusat *cluster (centroid)* secara acak dalam rentang nilai setiap atribut. Centroid berfungsi sebagai titik acuan untuk mengukur jarak antara data dengan setiap cluster. Kedua, menghitung jarak data terhadap setiap centroid menggunakan rumus

Euclidean Distance, yang mengukur seberapa dekat atau jauh setiap data dari pusat cluster. Ketiga, menempatkan data ke dalam cluster yang memiliki jarak paling dekat dengan data *centroid*. Setiap data akan dimasukkan ke *cluster* dengan nilai jarak paling kecil.

Setelah proses pengelompokan selesai, *centroid* diperbarui hingga posisinya stabil dan tidak berubah lagi. Hasil penerapan K-Means menggunakan Microsoft Excel menunjukkan bahwa Cluster 3 memiliki jumlah data terbanyak, meskipun frekuensi kecamatan di dalamnya paling sedikit dibanding cluster lainnya. Hasil ini menunjukkan distribusi yang bervariasi antar cluster dan menjadi dasar analisis pola daerah asal siswa di SMK Kristen Harapan Rantepao.

4) Evaluasi

Pada tahap ini dilakukan evaluasi untuk mencari kluster optimal yang dihasilkan dari algoritma k-means menggunakan grafik elbow dan perhitungan SSE (Sum Square Error).

5) Analisis dan Pembuatan Laporan

Hasil implementasi algoritma K-means akan dibuatkan laporan akhir dari penelitian ini

6) Proses Contoh Penerapan *K-means*

Pada tahap ini, metode pengelompokan K-means digunakan untuk mengkategorikan informasi siswa berdasarkan distrik asal mereka. Prosedur dimulai dengan menetapkan $K = 3$ untuk jumlah total kluster. *Sentroid* awal yang dipilih secara acak kemudian diambil dari data yang ada. Selanjutnya, rumus jarak Euclidean digunakan untuk menghitung jarak dari setiap titik data ke setiap *sentroid*, dan setiap titik data dialokasikan ke kluster yang memiliki jarak terpendek.

Setelah seluruh data terkluster, dilakukan pembaruan posisi *centroid* dengan menghitung rata-rata nilai tiap cluster. Proses iterasi ini terus diulang hingga posisi *centroid* stabil dan tidak berubah lagi. Hasil akhir menunjukkan bahwa terdapat tiga kelompok (cluster), yaitu:

Cluster 1 (Sedang) berisi 9 siswa dengan *centroid* 1,44

Cluster 2 (Kurang) tidak memiliki anggota

Cluster 3 (Tinggi) berisi 6 siswa dengan *centroid* 12,67

Dengan demikian, hasil penerapan K-Means berhasil mengelompokkan siswa SMK Kristen Harapan Rantepao ke dalam kategori Tinggi, Sedang, dan Kurang berdasarkan kesamaan daerah asal, yang dapat membantu pihak sekolah dalam memahami distribusi dan karakteristik siswa secara lebih terstruktur.

HASIL DAN PEMBAHASAN

A. Hasil Penelitian

1) Analisa

Data awal yang diperoleh dari sekolah diproses terlebih dahulu untuk menghasilkan dataset yang dapat digunakan. Selanjutnya, algoritma pengelompokan *K-Means* digunakan pada dataset ini. Dalam penelitian ini, *Microsoft Excel* digunakan untuk menjalankan algoritma pengelompokan *K-Means*, dengan mengikuti langkah-langkah yang diuraikan di bawah ini:

a. *Data Transformation*

Transformasi data melibatkan perubahan data ke dalam struktur yang diperlukan untuk analisis. Setelah data direduksi dan dibersihkan, peneliti mengidentifikasi atribut-atribut terkait, khususnya distrik asal siswa. Karena metode pengelompokan *K-Means* hanya dapat menangani input numerik, maka data teks pada atribut kecamatan diubah menjadi bentuk angka. Proses transformasi ini dilakukan dengan menghitung frekuensi kemunculan masing-masing kecamatan dalam dataset, kemudian mengurutkannya dari yang paling sering hingga paling sedikit untuk digunakan dalam proses klusterisasi.

b. Menentukan Titik Pusat *Cluster*

Titik tengah cluster yang dikenal dengan sebutan *centroid*, adalah titik rata-rata yang digunakan untuk mengukur jarak atau distance antara setiap titik data dalam sebuah *cluster*. Penentuan lokasi *Centroid* ini dapat dilakukan secara acak berdasarkan preferensi peneliti,

asalkan nilai *Centroid* tersebut berada dalam jangkauan nilai data pada setiap atribut. Dalam penentuan titik pusat cluster ini, peneliti menggunakan tiga titik awal pusat cluster sebagai berikut.

Tabel 1. Penentuan Pusat *Cluster*

NO	TITIK AWAL PUSAT CLUSTER	Id Siswa	Kode Asal
1	Cluster 1	1	1
2	Cluster 2	400	7
3	Cluster 3	796	5

c. Menghitung Jarak Data ke Setiap *Cluster*

Setelah titik pusat untuk setiap kelompok teridentifikasi, langkah berikutnya melibatkan pengukuran seberapa jauh setiap titik data dari setiap klaster, yang umumnya disebut sebagai jarak. Perhitungan jarak ini dilakukan di *Microsoft Excel* menggunakan rumus SQRT atau fungsi akar kuadrat. Adapun rumus yang digunakan adalah sebagai berikut:

$$=SQRT((data_kecamatan_asal - centroid_kecamatan_asal)^2)$$

Dengan menggunakan rumus tersebut, peneliti melakukan perhitungan pada satu baris data terlebih dahulu. Setelah hasilnya sesuai, rumus kemudian di-drag ke bawah hingga seluruh dataset terhitung secara otomatis.

d. Mengalokasikan Data ke dalam Cluster

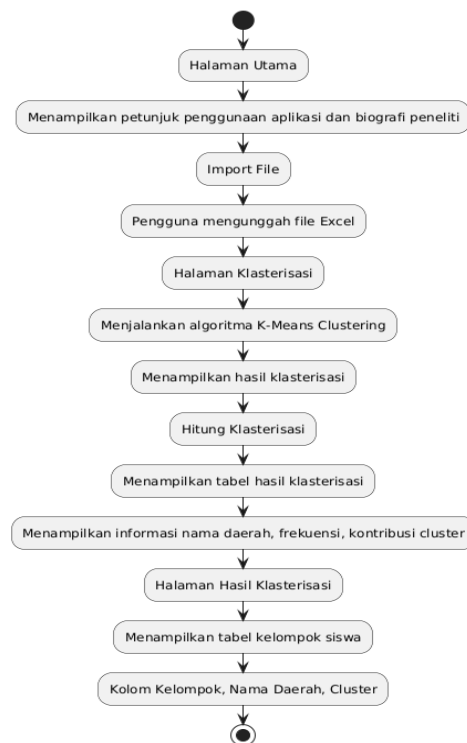
Setelah menghitung jarak untuk setiap tanggal dalam kaitannya dengan semua klaster, tugas berikutnya adalah menghubungkan setiap tanggal ke klaster dengan jarak terpendek ke data tersebut. Proses ini dilakukan menggunakan fungsi MATCH di *Microsoft Excel* untuk menentukan cluster dengan nilai jarak terkecil. Hasil pengalokasian menunjukkan data yang masuk ke cluster 1, 2, atau 3, yang kemudian digunakan untuk menentukan titik pusat (*centroid*) baru pada tahap berikutnya.

e. Hasil Penerapan K-means Menggunakan Microsoft Exel

Hasil analisis klasterisasi menggunakan algoritma K-means pada data siswa SMK Kristen Harapan Rantepao menunjukkan adanya sebaran geografis yang terbagi menjadi tiga klaster utama. Cluster 1 berisi kecamatan dengan jumlah siswa terbanyak, seperti Rantepao, Buntu Pepasan, dan Sesean Suloara, yang masing-masing mengirimkan lebih dari 60 siswa. Cluster 2 terdiri dari kecamatan dengan jumlah siswa sedang, seperti Sanggalangi, Awan Rantekarua, dan Rantebua, dengan jumlah siswa berkisar antara 20 hingga 30 siswa per kecamatan. Cluster 3 mencakup kecamatan dengan jumlah siswa sedikit, seperti Rindingallo, Balusu, dan Bonggakaradeng, yang hanya mengirimkan sekitar 2 hingga 4 siswa.

Evaluasi menggunakan metode Elbow dan perhitungan *Sum of Squared Errors* (SSE) menunjukkan bahwa tiga klaster (K=3) adalah jumlah yang optimal, dengan SSE yang stabil setelah beberapa iterasi. Hal ini menunjukkan bahwa pengelompokan data berdasarkan daerah asal siswa sudah maksimal dan cukup representatif. Sebaran siswa di SMK Kristen Harapan Rantepao sebagian besar berasal dari kecamatan yang lebih dekat dengan sekolah, seperti Rantepao, sementara kecamatan yang lebih jauh atau berpenduduk lebih sedikit memiliki jumlah siswa yang lebih rendah, seperti terlihat pada Cluster 3.

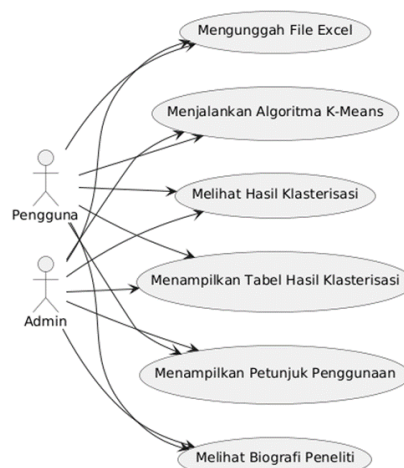
d. *Flowchart* Sistem



Gambar 3. *Flowchart* Sistem Berjalan

Flowchart ini menggambarkan alur kerja aplikasi berbasis web untuk mengelompokkan siswa menggunakan algoritma *K-Means Clustering*. Proses dimulai dari Halaman Utama yang berisi petunjuk penggunaan dan profil peneliti, dilanjutkan dengan Import File untuk mengunggah data siswa dalam format *Excel*. Setelah data diunggah, sistem menjalankan proses klasterisasi pada halaman khusus dan menampilkan hasil perhitungan dalam bentuk tabel berisi daerah asal, jumlah siswa per klaster, serta frekuensinya. Tahap akhir, Halaman Hasil Klasterisasi, menampilkan pembagian siswa berdasarkan klaster yang terbentuk. *Flowchart* ini memberikan gambaran ringkas tentang alur kerja aplikasi dari awal hingga hasil akhir.

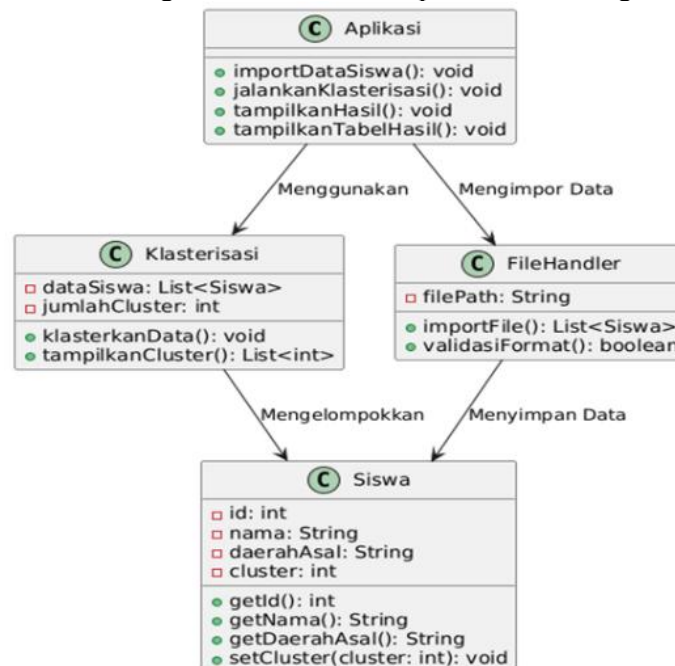
e. UML



Gambar 4. *UseCase* Diagram

Use Case Diagram ini menunjukkan interaksi antara dua aktor, yaitu Pengguna dan Admin, dengan sistem aplikasi berbasis web yang menerapkan algoritma *K-Means Clustering*. Baik Pengguna maupun Admin dapat mengunggah file *Excel*, menjalankan proses klasterisasi,

serta melihat hasil dan tabel klasterisasi. Selain itu, pengguna juga dapat mengakses petunjuk penggunaan dan biografi peneliti. Admin memiliki tanggung jawab tambahan untuk memastikan format file sesuai dan memantau jalannya proses. Diagram ini menggambarkan fungsi utama sistem serta hubungan antara aktor dan aplikasi secara ringkas dan terstruktur.



Gambar 5. *Class Diagram*

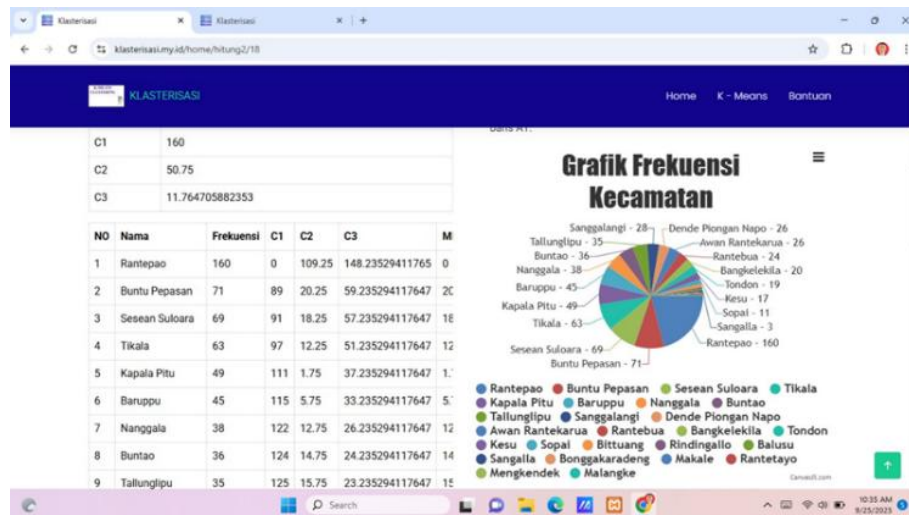
Class Diagram ini menggambarkan struktur aplikasi berbasis web yang menggunakan teknik pengelompokan K-Means untuk mengkategorikan informasi siswa berdasarkan daerah asal mereka. Diagram ini terdiri dari empat kelas utama, yaitu Kelas Aplikasi yang mengatur alur sistem seperti impor data, proses klasterisasi, dan penampilan hasil; Kelas Siswa yang menyimpan data siswa meliputi ID, nama, daerah asal, serta hasil klaster; Kelas Klasterisasi yang menjalankan proses pengelompokan dan menampilkan hasilnya; serta Kelas FileHandler yang berfungsi untuk mengunggah dan memvalidasi file *Excel* berisi data siswa. Keempat kelas ini saling berhubungan untuk memastikan proses pengelompokan berjalan efisien dan hasilnya mudah dipahami pengguna.

2) Perancangan

Hasil desain penelitian ini disajikan di bawah ini, khususnya aplikasi web yang memanfaatkan metode pengelompokan K-Means untuk mengkategorikan siswa berdasarkan wilayah geografis mereka asal mereka di SMK Kristen Harapan Rantepao:



Gambar 6. Halaman Utama



Gambar 7. Halaman Hitung Klasterisasi

B. Pembahasan

Hasil klasterisasi menunjukkan bahwa siswa terbagi menjadi tiga cluster, masing-masing mencerminkan karakteristik demografis yang berbeda. *Cluster 1* terdiri dari kecamatan dengan jumlah siswa terbanyak, seperti Rantepao, Buntu Pepasan, dan Sesean Suloara, Sementara itu, Cluster 2 mencakup kecamatan seperti Sanggalangi, Awan Rantekarua, dan Rantebua, yang menunjukkan bahwa meskipun jumlah siswa dari daerah ini jumlahnya lebih rendah dibandingkan dengan yang terdapat pada Cluster 1. Cluster 3 berisi kecamatan dengan jumlah siswa yang lebih sedikit, seperti Rindingallo, Balusu, dan Bonggakaradeng. Hal ini menunjukkan bahwa ada daerah yang kurang terjangkau atau mungkin kurang mendapatkan informasi mengenai SMK Kristen Harapan Rantepao. Dengan mengetahui hal ini, pihak sekolah dapat mempertimbangkan strategi pemasaran yang lebih efisien guna meningkatkan jumlah pendaftaran siswa dari daerah-daerah ini, misalnya melalui sosialisasi yang lebih intensif atau program kemitraan dengan sekolah-sekolah di kecamatan tersebut

KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa dalam studi ini, aplikasi secara efektif memanfaatkan teknik pengelompokan *K-Means* dalam kerangka penambangan data berbasis web untuk mengelompokkan siswa berdasarkan daerah asal mereka di SMK Kristen Harapan Rantepao. Aplikasi ini mampu menampilkan hasil klasterisasi secara visual sehingga memudahkan pihak sekolah dalam memahami distribusi geografis siswa. Dengan demikian, penerapan algoritma K-Means terbukti efektif untuk mendukung analisis data siswa berdasarkan daerah asal.

Hasil klasterisasi menunjukkan bahwa siswa terbagi ke dalam tiga cluster utama yang mencerminkan perbedaan karakteristik demografis. Cluster 1 terdiri dari kecamatan dengan jumlah siswa terbanyak, seperti Rantepao, Buntu Pepasan, dan Sesean Suloara, yang menandakan tingginya akses dan minat pendaftaran dari daerah tersebut. Cluster 2 mencakup kecamatan dengan jumlah siswa menengah, seperti Sanggalangi, Awan Rantekarua, dan Rantebua. Cluster 3 berisi kecamatan dengan jumlah siswa relatif sedikit, seperti Rindingallo, Balusu, dan Bonggakaradeng, yang menunjukkan masih rendahnya keterjangkauan atau kurangnya informasi mengenai keberadaan SMK Kristen Harapan Rantepao

UCAPAN TERIMA KASIH

Penulis mengucapkan rasa syukur kepada Tuhan Yang Maha Esa atas karunia dan bimbingan-Nya sehingga ia dapat menyelesaikan proyek ini. Ia juga menyampaikan rasa terima kasih kepada mentor, kerabat, dan seluruh pihak di SMK Kristen Harapan Rantepao yang telah

memberikan nasihat, dukungan, dan dukungan selama proses penelitian dan penulisan karya ini.

REFERENSI

- Alfarizi, M. R. S., Al-farish, M. Z., Taufiqurrahman, M., Ardiansah, G., & Elgar, M. (2023). Penggunaan Python sebagai bahasa pemrograman untuk Machine Learning dan deep learning. *Karimah Tauhid*, 2(1), 1–6.
- Alpina, D., & Witriyono, H. (2022). Pemanfaatan Framework Laravel Dan Framework[1] D. Alpina and H. Witriyono, “Pemanfaatan Framework Laravel Dan Framework Bootstrap Pada Pembangunan Aplikasi Penjualan Hijab Berbasis Web,” *J. Media Infotama*, vol. 18, no. 1, pp. 36–42, 2022. *Bootstrap Pada P. Jurnal Media Infotama*, 18(1), 36–42.
- Andiani, D., Rahayu, S. D., & Riana, A. (2022). Analisis Teknik non-Hierarki untuk Pengelompokan Kabupaten/Kota di Provinsi Jawa Barat Berdasarkan Indikator Kesejahteraan Rakyat 2020. *JRMST| Jurnal Riset Matematika Dan Sains Terapan*, 2(1), 21–28.
- Aroni, A., Fitri, H., & Prasetyo, E. (2018). Penerapan Metode Clustering dengan Algoritma K-means pada Pengelompokan Data Calon Mahasiswa Baru di Universitas Muhammadiyah Yogyakarta (Studi Kasus: Fakultas Kedokteran dan Ilmu Kesehatan, dan Fakultas Ilmu Sosial dan Ilmu Politik). *Semesta Teknika*, 21(1), 60–64.
- Ayuningtyas, P., Khomsah, S., & Sudioanto, S. (n.d.). Pelabelan Sentimen Berbasis Semi-Supervised Learning menggunakan Algoritma LSTM dan GRU. *JISKA (Jurnal Informatika Sunan Kalijaga)*, 9(3), 217–229.
- Dalimunthe, A. L. (2022). Sistem Informasi E-Learning Di SMA Negeri 1 Rantau Selatan Berbasis Web. *Journal of Student Development Informatics Management (JoSDIM)*, 2(1), 1–11.
- Diah, N. (2021). Sistem informasi administrasi sekolah berbasis web (studi kasus: SMK Negeri 16 Jakarta). *KONSTELASI: Konvergensi Teknologi Dan Sistem Informasi*, 1(2), 216–225.
- Dirgantara, D. A., & Andrian, R. (2023). Pengembangan Responsif Website Untuk Semarang Heritage Run 2022 dengan Framework Bootstrap. *Jurnal Media Infotama*, 19(2), 433–438.
- Fadilla, D. N. (2019). Sistem Informasi Pendaftaran Mahasiswa Baru Di Politeknik Negeri Sriwijaya Studi Kasus Scan Kode QR Lokasi dan Presensi. *Journal of Chemical Information and Modeling*, 53(9), (287).
- Kaunang, T., Wijaya, J., & Kumajas, S. (2020). Sistem Informasi Lembaga Pembinaan Khusus Anak Kelas II Tomohon Berbasis Web. *Jointer - Journal of Informatics Engineering*, 1(02), 30–36. <https://doi.org/10.53682/jointer.v1i02.13>
- Kho, W., & Simanjuntak, P. (2024). Perancangan Sistem Wihara Terintegrasi Berbasis Web Menggunakan Metode Prototype. *Computer and Science Industrial Engineering (COMASIE)*, 10(3), 1–10.
- Mubarak, A., Metro, J. J., & Selatan, K. T. (2019). Rancang Bangun Aplikasi Web Sekolah Menggunakan Uml (Unified Modeling Language) Dan Bahasa Pemrograman Php (Php Hypertext Preprocessor) Berorientasi Objek. *JIKO (Jurnal Inform. Dan Komputer)*, 2(1), 19–25.
- Mustahir, B., Kembuan, O., & Kumajas, S. (2021). Aplikasi Penjualan di Toko Emas Surya Kencana Kota Bitung Menggunakan Framework Codeigniter. *Jointer - Journal of Informatics Engineering*, 2(02), 29–38. <https://doi.org/10.53682/jointer.v2i02.42>
- Pratama, R., Shaufiah, S., & Asror, I. (2018). Deteksi Sms Spam Menggunakan Vector Space Model Dengan K-means. *EProceedings of Engineering*, 5(2).
- Purnamasari, N. M., Syauqi, A., & Pramana, D. A. (2023). Pengelompokan Data Calon Siswa Baru Di Sekolah Menengah Kejuruan menggunakan Algoritma K-means. *Jurnal Sistem Informasi Dan Teknologi Peradaban*, 4(1), 24–30.
- Putra, P. H., Hasibuan, A., & Marpaung, E. A. (2022). Analisis Klasifikasi Metode X-Means Pada Minat dan Bakat Anak Dimasa Pandemi. *SITEKIN: Jurnal Sains, Teknologi Dan Industri*, 19(2), 424–429. <https://ejournal.uin-suska.ac.id/index.php/sitekin/article/download/17889/7834>

-
- Rozi, F., Haryanti, T., & Fahrhani, N. (2022). Rancang Bangun Website Profil Sekolah TAUD-SaQu Ashabul Qurâ€™TM an Surabaya Berbasis HTML. *Computing Insight: Journal of Computer Science*, 4(1), 1–9.
- Sari, D. N. P., & Sukestiyarno, Y. L. (2021). Analisis cluster dengan metode K-means pada persebaran kasus COVID-19 berdasarkan Provinsi di Indonesia. *PRISMA, Prosiding Seminar Nasional Matematika*, 4, 602–610.
- Sari, I. P., Azzahrah, A., Qathrunada, I. F., Lubis, N., & Anggraini, T. (2022). Perancangan sistem absensi pegawai kantor secara online pada Website berbasis HTML dan CSS. *Blend Sains Jurnal Teknik*, 1(1), 8–15.
- Sinlae, F., Rosyad, F. S., Nurhidayat, F., & Jannah, W. (2024). Evolusi Teknologi Web dan Dampaknya Terhadap Masyarakat Digital. *Jurnal Ilmu Multidisplin*, 3(2), 146–154.
- Sudrajat, A., Wardhani, S. K., & Apnena, R. D. (2023). Aplikasi Pembayaran SPP Sekolah Terintegrasi Whatsapp Berbasis Web. *Journal of Informatics and Electronics Engineering*, 3(1), 6–12.
- Susilawati, R., & Sunendiari, S. (2022). Peramalan Jumlah Penumpang Kereta Api Menggunakan Metode Arima dan Grey System Theory. *Jurnal Riset Statistika*, 1–13. <https://doi.org/10.29313/jrs.vi.603>
- Tiauw, M., Rantung, V. P., Kumajas, S. C., Informatika, T., Teknik, F., Manado, U. N., Gaji, K., Pengadilan, B., & Tondano, N. (2024). Aplikasi Kenaikan Gaji Berkala Di Pengadilan Negeri Tondano Menggunakan Metode SDLC. 94–104.
- Tinambunan, M. H., Hasibuan, A., Wahyuni, S., & Wibowo, A. S. (2023). Klasifikasi Tingkat Kepuasan Mahasiswa Terhadap Fasilitas Pada Ftik Universitas Dharmawangsa Medan Dengan Algoritma Naive Bayes. *Bisnis-Net Jurnal Ekonomi Dan Bisnis*, 6(1), 208–215. <https://doi.org/10.46576/bn.v6i1.3356>
- Wahyuni, R. (2024). Penerapan sistem e-arsip surat masuk dan keluar di pengadilan agama tondano berbasis Website. 1(1)