



## Laporan Project Data Mining

---

### Abstrak

Kesejahteraan mahasiswa merupakan faktor penting dalam menciptakan lingkungan akademik yang sehat dan produktif. Fenomena ini bersifat multidimensional, mencakup aspek akademik, finansial, fisik, psikologis, dan relasional. Selama ini, analisis kesejahteraan mahasiswa umumnya masih terbatas pada penilaian per aspek tanpa melihat keterkaitan antar domain secara menyeluruh.

Proyek ini mengembangkan aplikasi analisis kesejahteraan mahasiswa berbasis data mining yang mampu memetakan pola kesejahteraan antar aspek serta memberikan rekomendasi aksi nyata bagi pihak kampus. Tahapan analisis mengikuti kerangka CRISP-DM, meliputi preprocessing (cleaning, integration, transformation, reduction, discretization), clustering dengan K-Means, DBSCAN, dan Hierarchical, hingga visualisasi, insight, dan action plan.

Aplikasi dikembangkan dalam dua versi: Flask yang lebih sederhana dan user-friendly, serta Streamlit yang lebih lengkap dan interaktif. Hasil analisis menunjukkan pola kesejahteraan yang beragam antar kelompok mahasiswa serta hubungan erat antara aspek finansial dan psikologis dalam menentukan tingkat kesejahteraan. Temuan ini diharapkan membantu rektor, dosen, dan pihak kampus dalam mengambil kebijakan berbasis data untuk meningkatkan kesejahteraan mahasiswa.

**Kata kunci:** Data Mining, Kesejahteraan Mahasiswa, Clustering, CRISP-DM, K-Means, DBSCAN, Hierarchical, Streamlit, Flask.

### I. Business Understanding

Kesejahteraan mahasiswa merupakan aspek penting yang mencerminkan kualitas kehidupan akademik dan lingkungan kampus secara keseluruhan. Fenomena ini bersifat multidimensional, mencakup lima domain utama: akademik, finansial, fisik, psikologis, dan relasional. Namun, analisis konvensional yang hanya berfokus pada nilai rata-rata tiap aspek belum mampu mengungkap keterkaitan antar domain secara komprehensif.

Proyek ini bertujuan mengembangkan **aplikasi analisis kesejahteraan mahasiswa berbasis data mining** menggunakan pendekatan *unsupervised learning* dengan algoritma **K-Means, Hierarchical Clustering, dan DBSCAN**. Melalui tahapan CRISP-DM, aplikasi ini tidak hanya memetakan pola kesejahteraan mahasiswa, tetapi juga memberikan **insight antar aspek** serta **rekomendasi aksi nyata** bagi rektor, dosen, dan pihak kampus.

Hasil analisis diharapkan menjadi dasar dalam **perumusan kebijakan dan program dukungan mahasiswa** yang lebih tepat sasaran, berbasis data, dan berorientasi pada peningkatan kesejahteraan di lingkungan kampus.

### II. Data Understanding

Data yang digunakan dalam penelitian ini diperoleh melalui survei kesejahteraan mahasiswa yang disebarluaskan secara daring kepada seluruh mahasiswa FMIPA Universitas Padjadjaran melalui grup komunikasi fakultas. Survei ini dirancang untuk mengukur lima domain utama kesejahteraan

mahasiswa, yaitu akademik, finansial, fisik, psikologis, dan relasional. Total terkumpul **119 responden dengan 35 variabel pertanyaan**, yang masing-masing diukur menggunakan **skala Likert 1–5**, di mana skor 1 menunjukkan kondisi terendah dan skor 5 kondisi tertinggi. Pendekatan ini memungkinkan pengukuran persepsi kesejahteraan mahasiswa secara kuantitatif dan terstandar.

Hasil eksplorasi awal menunjukkan bahwa data memiliki **beberapa missing value**, namun **tidak terdapat duplikasi maupun outlier**, sehingga dapat digunakan untuk analisis lebih lanjut. Berdasarkan nilai rata-rata tiap domain, aspek **Physical Well-Being** tercatat sebagai yang **paling rendah**, diikuti oleh **Academic Well-Being**, sementara aspek relasional dan finansial cenderung stabil. Temuan ini menjadi indikator awal adanya ketimpangan kesejahteraan antar domain dan memperkuat perlunya analisis lanjutan melalui proses preprocessing dan clustering untuk memahami pola hubungan antar aspek tersebut secara lebih mendalam.

### III. Data Preparation

Tahap ini bertujuan menyiapkan data agar layak digunakan dalam pemodelan. Proses *data cleaning* dilakukan dengan menghapus atau mengimputasi nilai yang hilang serta mengidentifikasi *outlier* yang tidak konsisten. Setiap domain kemudian dirata-ratakan guna memperoleh skor komposit per responden.

Selanjutnya dilakukan **standarisasi menggunakan Z-Score** agar seluruh variabel memiliki skala yang sebanding. Untuk efisiensi dan kemudahan visualisasi, diterapkan **reduksi dimensi menggunakan PCA dan/atau UMAP** hingga

menjadi dua dimensi utama. Hasil akhir berupa dataset bersih, terstandarisasi, dan siap untuk proses *clustering*.

### IV. Modelling

Tahap pemodelan dilakukan dengan menerapkan tiga algoritma *unsupervised learning*: **K-Means**, **Hierarchical Clustering**, dan **DBSCAN**. Masing-masing algoritma digunakan untuk mengekstraksi struktur dan pola tersembunyi dalam data dengan karakteristik berbeda:

- **K-Means** untuk segmentasi berdasarkan jarak centroid, dengan jumlah *cluster* ( $k$ ) ditentukan melalui metode *Elbow Dan Silhouette Score*.
- **Hierarchical Clustering** untuk membentuk struktur hierarki antar data, divisualisasikan dalam bentuk dendogram tanpa perlu menentukan jumlah *cluster* di awal.
- **DBSCAN** untuk mengidentifikasi pola tidak beraturan dan mendeteksi *outlier* berdasarkan kepadatan data.

Keluaran dari tahap ini berupa hasil *clustering* dari masing-masing algoritma yang kemudian akan dievaluasi secara kuantitatif dan interpretatif pada tahap berikutnya.

### V. Evaluation

Tahap evaluasi berfokus pada penilaian performa dan validitas model *clustering*. Evaluasi kuantitatif dilakukan menggunakan tiga metrik utama:

- **Silhouette Score** untuk mengukur kohesi dan separasi antar *cluster*,

- **Davies-Bouldin Index** untuk menilai rasio jarak antar *cluster* terhadap ukuran dalam *cluster*, dan
- **Calinski-Harabasz Index** untuk mengukur rasio dispersi antar dan dalam *cluster*.

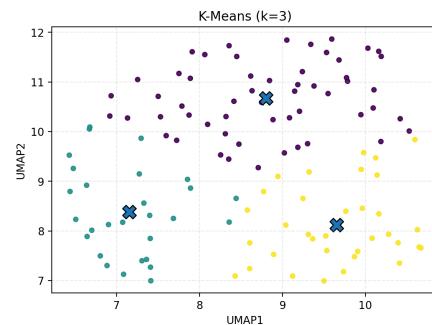
Model terbaik dipilih berdasarkan kombinasi nilai metrik yang optimal. Selain itu, dilakukan evaluasi interpretatif untuk memastikan hasil *clustering* memiliki makna konseptual yang sesuai dengan fenomena kesejahteraan mahasiswa.

## VI. Deployment

Tahap **Deployment** bertujuan untuk menerapkan hasil analisis ke dalam konteks nyata sehingga temuan dari proses data mining dapat memberikan manfaat praktis bagi kampus. Hasil clustering digunakan sebagai dasar untuk **merancang strategi dan program dukungan yang sesuai dengan karakteristik setiap kelompok mahasiswa**. Misalnya, mahasiswa dengan skor finansial rendah dapat diarahkan pada program bantuan ekonomi, sedangkan kelompok dengan kesejahteraan psikologis rendah memerlukan prioritas dalam layanan konseling dan dukungan kesehatan mental.

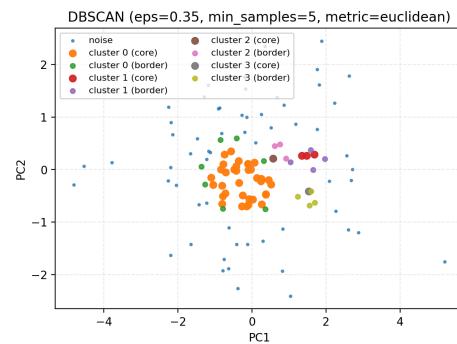
Selain implementasi kebijakan, hasil analisis ini juga diintegrasikan ke dalam **aplikasi berbasis web** yang dikembangkan dalam dua versi: **Streamlit** dan **Flask**. Aplikasi ini berfungsi sebagai **dashboard interaktif** yang menampilkan hasil analisis, visualisasi cluster, serta rekomendasi aksi secara dinamis. Dengan demikian, pihak kampus dapat **memantau perubahan tingkat kesejahteraan mahasiswa secara periodik**, mengevaluasi efektivitas kebijakan yang diterapkan, dan melakukan penyesuaian strategi berbasis data secara berkelanjutan.

## VII. Hasil Visualisasi & Insight



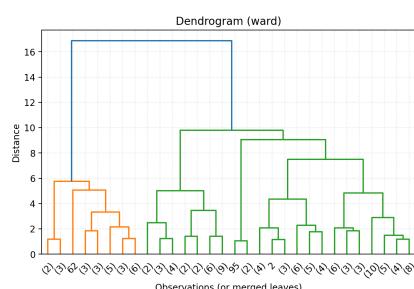
Parameter K-Means		
Silhouette ↗	Calinski-Harabasz ↗	Davies-Bouldin ↗

Gambar 7.1 Visualisasi K-Means dan Parameter



Parameter DBSCAN		
Silhouette ↗	Calinski-Harabasz ↗	Davies-Bouldin ↗

Gambar 7.2 Visualisasi DBSCAN dan Parameter

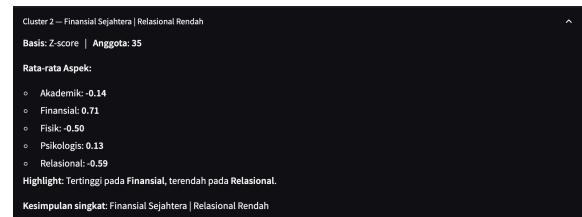
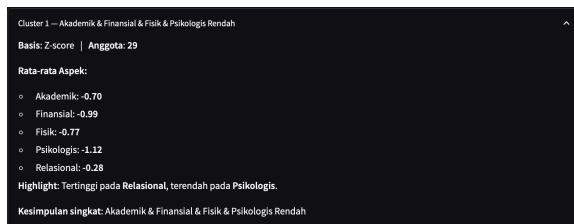
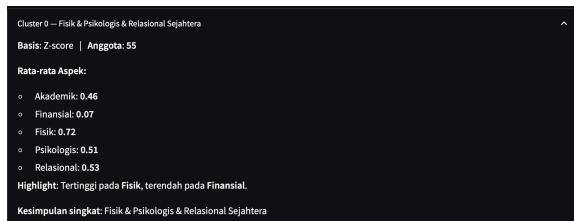


Silhouette ↗	Calinski-Harabasz ↗	Davies-Bouldin ↘
0.4157	114.04	0.6090

Gambar 7.3 Visualisasi Hierarchical dan Parameter

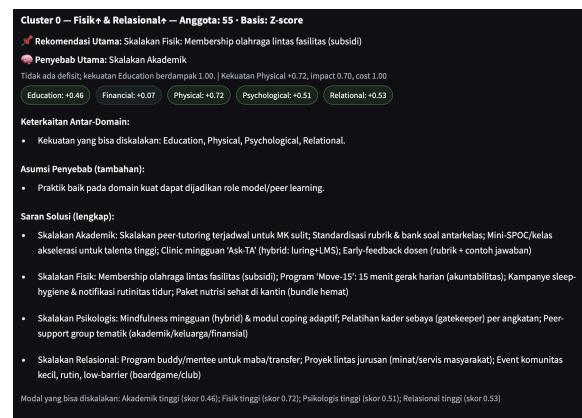
Ketiga algoritma—**K-Means**, **DBSCAN**, dan **Hierarchical**—menunjukkan hasil yang berbeda dalam mengelompokkan data kesejahteraan mahasiswa. Secara umum, **K-Means** dan **Hierarchical** memberikan hasil paling stabil dan terpisah dengan baik, ditunjukkan oleh nilai **Silhouette** sekitar 0.42 dan **Calinski-Harabasz** di atas 110. Sementara itu, **DBSCAN** menghasilkan lebih banyak *noise* dan nilai evaluasi yang lebih rendah (Silhouette 0.37, CH 39.33), menandakan struktur klaster yang kurang jelas.

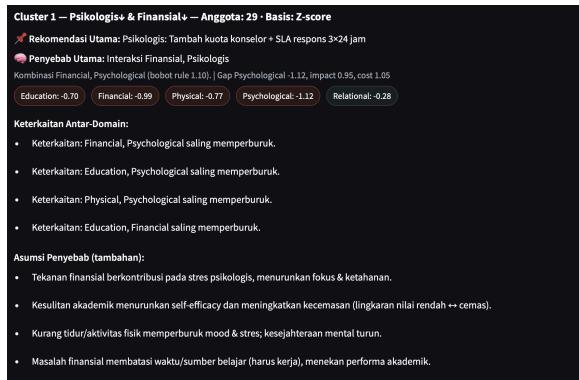
Dengan demikian, **K-Means (k=3)** dapat dianggap sebagai metode paling optimal untuk analisis ini karena memiliki pemisahan antar klaster yang jelas, mudah diinterpretasikan, dan sesuai untuk penerapan pada aplikasi dashboard kesejahteraan mahasiswa.



Gambar 7.4, 7.5, 7.6 Insight Cluster K-Means

Analisis K-Means menghasilkan tiga kelompok utama kesejahteraan mahasiswa. **Cluster 0** (55 anggota) menunjukkan kesejahteraan tinggi pada aspek fisik, psikologis, dan relasional namun rendah secara finansial. **Cluster 1** (29 anggota) mencerminkan kondisi rendah di hampir semua aspek, terutama psikologis dan finansial. Sementara **Cluster 2** (35 anggota) memiliki kesejahteraan finansial tinggi tetapi relasional rendah. Temuan ini menunjukkan bahwa kesejahteraan finansial, psikologis, dan relasional merupakan faktor paling berpengaruh terhadap kesejahteraan mahasiswa secara keseluruhan.





Gambar 7.7, 7.8, 7.9 Action Plan Cluster K-Means

Analisis K-Means mengungkap tiga profil kesejahteraan mahasiswa dengan karakteristik berbeda. **Cluster 0** menunjukkan kesejahteraan tinggi pada aspek fisik, psikologis, dan relasional namun rendah pada finansial, sehingga kelompok ini dapat dijadikan model praktik baik dengan dukungan program olahraga dan mentoring akademik. **Cluster 1** merupakan kelompok paling rentan dengan kesejahteraan psikologis dan finansial rendah, sehingga membutuhkan dukungan terpadu berupa konseling, bantuan ekonomi fleksibel, dan pendampingan akademik intensif. **Cluster 2** memiliki finansial stabil tetapi kesejahteraan fisik dan relasional rendah, yang dapat ditingkatkan melalui kegiatan komunitas, pelatihan komunikasi, dan program kebugaran kampus.

Secara keseluruhan, hasil ini menegaskan bahwa kesejahteraan mahasiswa bersifat multidimensional dan saling berkaitan. Upaya peningkatannya perlu dilakukan secara **holistik, kolaboratif, dan berbasis data**, agar kebijakan dan intervensi kampus dapat lebih tepat sasaran dan berkelanjutan.

## VIII. Kesimpulan

Penerapan CRISP-DM pada survei 119×35 (Likert 1–5) berhasil mengubah data mentah menjadi peta profil kesejahteraan mahasiswa yang operasional. Setelah pembersihan, standarisasi Z-score, dan reduksi dimensi (PCA/UMAP), evaluasi menunjukkan **K-Means (k=3)** sebagai model paling *interpretable*, menghasilkan tiga profil: (1) fisik-psikologis-relasional relatif tinggi namun finansial rendah, (2) kelompok paling rentan dengan psikologis dan finansial rendah yang berdampak ke akademik, dan (3) finansial baik tetapi fisik serta relasional rendah. Temuan ini menegaskan sifat multidimensional kesejahteraan serta perlunya intervensi **klaster-spesifik** (dukungan finansial adaptif, konseling terstruktur, penguatan komunitas dan kebugaran) alih-alih pendekatan seragam. Implementasi melalui laporan berkala, ringkasan klaster, dan survei pulse periodik memungkinkan pemantauan berkelanjutan serta penyesuaian kebijakan berbasis bukti. Secara keseluruhan, pendekatan ini tetap efektif, manusiawi, dan dapat ditindaklanjuti untuk meningkatkan kualitas hidup serta capaian akademik mahasiswa.