

# **ANALISIS KLUSTER PENGGUNA CHATBOT AI**

Muhammad Exfo Albanjari - 140810220057

Bagas Diatama Wardoyo - 140810230061

Arya Muhammad Rafi R. - 140810230072

## **1. Business Understanding**

Tujuan dari proyek ini adalah untuk melakukan segmentasi (clustering) terhadap mahasiswa berdasarkan pola pemanfaatan dan persepsi mereka terhadap ChatBot AI. Dengan meningkatnya penggunaan teknologi AI generatif di lingkungan akademik, institusi pendidikan perlu memahami bagaimana mahasiswa berinteraksi dengan alat ini. Pemahaman ini krusial untuk merancang panduan penggunaan AI yang etis, mengembangkan materi literasi digital yang relevan, serta mengidentifikasi potensi risiko seperti plagiarisme dan penurunan kemampuan berpikir kritis.

Tujuan data mining yang ingin dicapai adalah mengelompokkan data mahasiswa ke dalam beberapa profil yang berbeda dan dapat diidentifikasi. Setiap kluster akan merepresentasikan satu kelompok mahasiswa dengan karakteristik perilaku dan persepsi yang serupa terhadap AI. Hasil analisis ini dapat menjadi dasar bagi pemangku kepentingan di universitas untuk membuat kebijakan yang lebih tepat sasaran.

## **2. Data Understanding**

Data yang digunakan dalam proyek ini adalah data primer yang dikumpulkan melalui survei kuesioner online. Responden survei adalah mahasiswa dari berbagai program studi dan angkatan.

### **2.1. Deskripsi Atribut**

Dataset asli terdiri dari 17 atribut. Untuk keperluan pemodelan, 14 atribut numerik yang relevan dipilih. Atribut ini mengukur tiga dimensi utama:

- Frekuensi Penggunaan: Seberapa sering mahasiswa menggunakan ChatBot AI (1 atribut).
- Tujuan Pemanfaatan: Untuk tugas spesifik apa AI digunakan, seperti brainstorming, coding, atau meringkas (8 atribut).
- Persepsi dan Sikap: Pandangan mahasiswa mengenai dampak AI terhadap pembelajaran, etika, dan akurasi (5 atribut).

Semua atribut ini menggunakan skala Likert (1-5), di mana 1 menunjukkan frekuensi/persepsi terendah dan 5 menunjukkan yang tertinggi. Data awal ditemukan dalam kondisi baik, tanpa

nilai yang hilang (missing values), sehingga proses imputasi data tidak diperlukan.

### 3. Data Preparation

Tahap persiapan data adalah langkah krusial untuk memastikan model clustering dapat bekerja secara optimal. Semua proses pada tahap ini diimplementasikan secara menggunakan NumPy.

#### 3.1. Seleksi Fitur

Dari dataset, hanya 14 atribut numerik yang dijelaskan pada tahap (Data Understanding) yang dipilih sebagai dasar pemodelan. Atribut demografis seperti 'Major' dan 'Year' sengaja tidak disertakan dalam analisis untuk menjaga fokus klustering murni pada aspek perilaku dan persepsi penggunaan AI.

#### 3.2. Standarisasi Data (Scaler)

Algoritma K-Means sangat sensitif terhadap skala data karena berbasis pada perhitungan jarak Euclidean. Fitur dengan rentang nilai yang lebih besar dapat secara tidak proporsional mendominasi hasil klustering. Untuk mengatasi hal ini, dilakukan standarisasi data.

$$Z = \frac{x - \mu}{\sigma}$$

di mana  $x$  adalah nilai data asli,  $\mu$  adalah rata-rata (mean) dari fitur tersebut, dan  $\sigma$  adalah standar deviasi. Proses ini mengubah setiap fitur sehingga memiliki rata-rata 0 dan standar deviasi 1, memastikan semua fitur memiliki bobot yang setara dalam proses pemodelan.

### 4. Modeling

Algoritma **K-Means Clustering** diterapkan. Sesuai dengan tujuan proyek, implementasi algoritma K-Means, serta **Principal Component Analysis (PCA)** untuk visualisasi. Analisis kluster dilakukan secara terpisah untuk Use Features dan Perception Features guna mendapatkan wawasan yang lebih spesifik dari masing-masing dimensi. Untuk menentukan jumlah kluster ( $k$ ) yang optimal, digunakan Metode Elbow, yang secara visual membantu mengidentifikasi titik di mana penambahan jumlah kluster tidak lagi memberikan penurunan Within-Cluster Sum of Squares (WCSS) yang signifikan.

### 5. Evaluation

Dilakukan melalui dua pendekatan. Secara kuantitatif, metrik Inertia (WCSS) digunakan untuk mengukur kekompakan kluster; nilai yang lebih rendah mengindikasikan kluster yang lebih padat dan lebih

baik. Secara kualitatif, profil rata-rata setiap kluster dianalisis secara mendalam. Visualisasi seperti PCA Scatter Plot dan Radar Chart digunakan untuk memvalidasi pemisahan antar kluster dan mempermudah interpretasi.

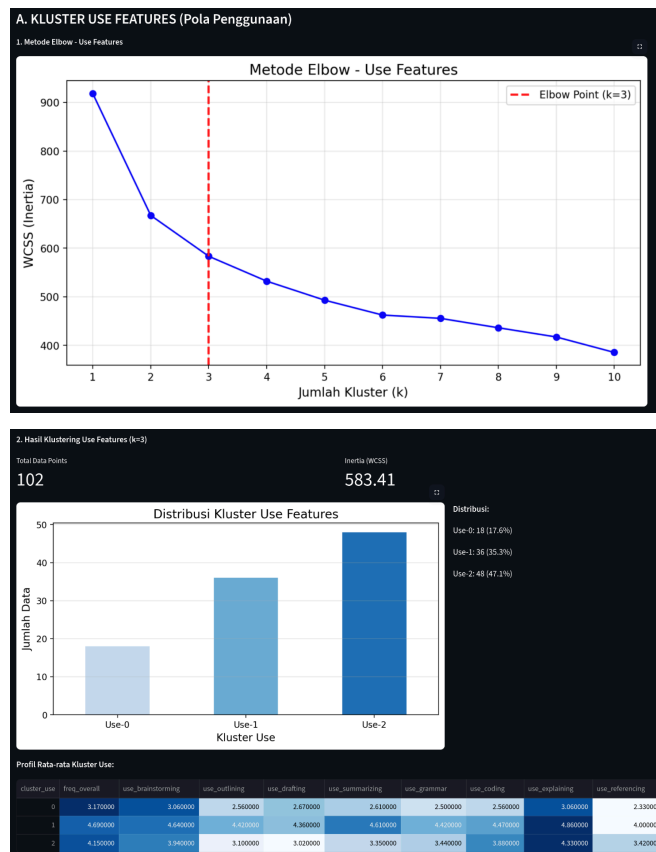
## 6. Deployment

Hasil analisis ini diwujudkan dalam bentuk aplikasi web interaktif yang dibangun menggunakan Streamlit. Aplikasi ini menyajikan visualisasi data, profil kluster, dan interpretasi hasil.

## 7. Analisis Dan Hasil

### 7.1. Kluster Berdasarkan Pola Penggunaan (Use Features)

Analisis terhadap pola penggunaan menghasilkan tiga kluster berbeda yang menunjukkan cara mahasiswa berinteraksi dengan ChatBot AI. Penentuan tiga kluster ini didasarkan pada grafik Metode Elbow, yang menunjukkan titik "siku" yang jelas pada  $k=3$ . Visualisasi PCA mengonfirmasi bahwa ketiga kluster ini terpisah dengan cukup baik dalam ruang dua dimensi.

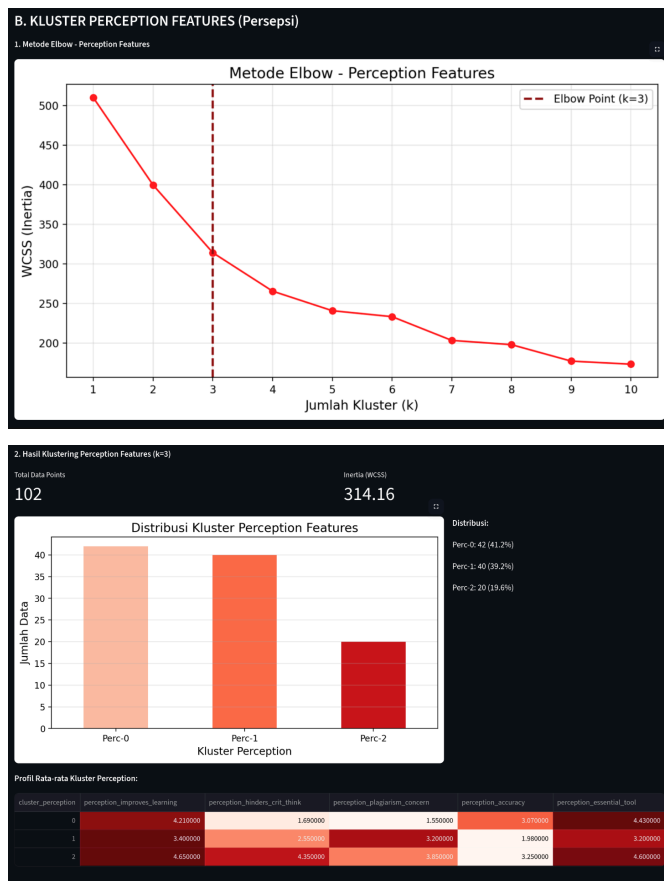


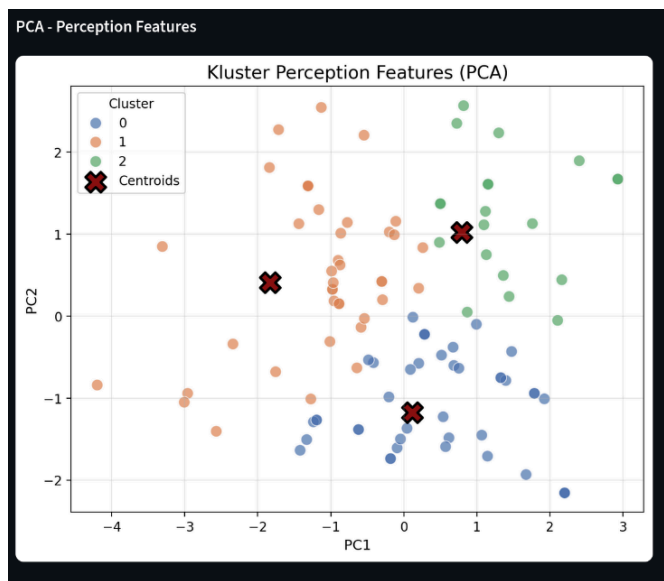
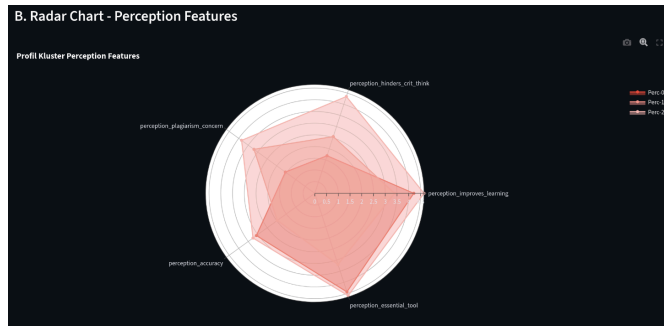


**Cluster 2: The Power Coders & Explainers.** Kluster ini menonjol dengan nilai rata-rata tertinggi pada `use_coding` dan `use_explaining`. Mereka adalah pengguna mahir yang memanfaatkan AI untuk tugas-tugas teknis yang kompleks, seperti menulis dan men-debug kode, serta untuk meminta penjelasan mendalam tentang konsep-konsep sulit. Profil ini sangat identik dengan mahasiswa dari jurusan teknis, seperti Teknik Informatika, yang menjadikan AI sebagai asisten pemrograman dan tutor virtual.

## 7.2. Kluster Berdasarkan Persepsi (Perception Features)

Analisis terhadap data persepsi juga mengidentifikasi tiga kelompok berbeda dengan pandangan yang kontras terhadap ChatBot AI. Visualisasi PCA menunjukkan pemisahan yang jelas antara kelompok-kelompok ini, yang mencerminkan perbedaan mendasar dalam sikap mereka.





**Cluster 0:** The Enthusiasts. Kelompok ini memiliki pandangan yang sangat positif. Mereka menunjukkan skor tertinggi pada perception\_improves\_learning dan perception\_essential\_tool, sementara skornya rendah pada perception\_hinders\_crit\_think. Mereka sangat percaya bahwa AI adalah alat esensial yang secara signifikan meningkatkan kualitas pembelajaran dan tidak terlalu khawatir akan potensi dampak negatifnya terhadap kemampuan berpikir kritis.

**Cluster 1:** The Cautious Optimists. Kluster ini dapat dianggap sebagai kelompok yang seimbang. Mereka mengakui manfaat AI untuk pembelajaran tetapi juga memiliki tingkat kekhawatiran yang cukup pada isu-isu seperti akurasi dan plagiarisme (perception\_accuracy, perception\_plagiarism\_concern). Mereka adalah pengguna yang pragmatis, memanfaatkan AI sambil tetap waspada terhadap keterbatasannya.

**Cluster 2: The Skeptics.** Kelompok ini menunjukkan tingkat kekhawatiran tertinggi, terutama pada `perception_hinders_crit_think` dan `perception_plagiarism_concern`. Mereka adalah individu yang paling kritis terhadap teknologi AI, memandang bahwa ketergantungan pada AI dapat menghambat pengembangan pemikiran orisinal dan menimbulkan risiko etis. Persepsi mereka menyoroti adanya kebutuhan akan panduan penggunaan AI yang bertanggung jawab di lingkungan akademik.