

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Kemajuan dan pengenalan teknologi komputasi telah terbukti sangat efektif dalam menghasilkan produksi data dalam jumlah besar untuk dianalisis (Wang Haoxiang, 2021). Data yang tepat dan analisis yang cermat dari data tersebut sangat penting dalam pengambilan keputusan yang efisien dan efektif. Salah satu teknik penting dalam pengolahan data adalah metode clustering, yang memungkinkan untuk mengelompokkan data ke dalam kelompok-kelompok berdasarkan karakteristik atau fitur tertentu. Clustering adalah alat penting dalam penelitian dan aplikasi data mining. Clustering merupakan subjek penelitian aktif dalam banyak bidang studi, seperti ilmu komputer, ilmu data, statistik, pengenalan pola, kecerdasan buatan, dan pembelajaran mesin (Ezugwu, 2022).

Meskipun metode *clustering* telah lama digunakan dan terbukti bermanfaat dalam berbagai aplikasi, ada tantangan penting yang perlu diatasi. Salah satu tantangan tersebut adalah variasi dalam skala fitur-fitur yang digunakan dalam proses clustering. Variasi ini dapat mengakibatkan kontribusi yang tidak seimbang dari setiap fitur terhadap proses pengelompokan, sehingga menghasilkan *cluster* yang tidak efisien dan efektif.

Banyak penelitian yang sudah dilakukan dalam bidang clustering (Yudhistira & Andika, 2023). Namun masih ada kekurangan dalam pemahaman yang mendalam tentang dampak normalisasi skala fitur terhadap efisiensi dan efektivitas metode clustering. Terdapat sedikit literatur yang secara khusus membahas topik ini dengan mendalam. Hal ini menciptakan peluang untuk penelitian yang lebih lanjut.

Penelitian ini memiliki relevansi yang signifikan dalam konteks pengolahan data dan analisis data. Penelitian ini berusaha memahami bagaimana normalisasi skala fitur dapat memengaruhi efisiensi dan efektivitas metode clustering akan membantu praktisi data science dalam mengambil keputusan yang lebih baik dalam

memilih teknik clustering yang paling sesuai untuk data mereka. Selain itu, penelitian ini juga dapat memiliki dampak pada pengembangan algoritma clustering yang lebih canggih dan efektif di masa depan.

Dalam penelitian ini, akan digunakan berbagai metode clustering yang umum seperti K-Means, Hierarchical Clustering, dan DBSCAN. Selain itu juga akan diterapkan teknik normalisasi skala fitur yang berbeda seperti *Min-Max Scaling* dan *Z-Score Standardization*. Untuk mengukur efisiensi dan efektivitas metode clustering, akan digunakan metrik evaluasi *clustering* seperti *Silhouette Score* dan *Davies-Bouldin Index*.

Melalui penelitian ini, diharapkan dapat memberikan pemahaman yang lebih baik tentang peran normalisasi skala fitur dalam konteks pengelompokan data dan memberikan wawasan yang berharga bagi komunitas data science dan pengolahan data secara lebih luas. Dengan memahami dampak normalisasi skala fitur, hasil penelitian ini dapat meningkatkan kualitas analisis data dan pengambilan keputusan yang didasarkan pada data.

## **1.2 Rumusan Masalah**

Masalah yang menjadi fokus penelitian dalam proposal ini adalah:

1. Bagaimana pengaruh normalisasi skala fitur terhadap efisiensi dan efektivitas metode clustering dalam pengolahan data?
2. Bagaimana hasil perbandingan berbagai metode clustering untuk mengukur efisiensi dan efektivitas dalam pengelompokan data ?

## **1.3 Tujuan**

Penelitian ini bertujuan untuk:

1. Meninvestigasi dampak normalisasi skala fitur terhadap efisiensi metode *clustering*, yaitu sejauh mana normalisasi skala fitur memengaruhi waktu komputasi dan penggunaan sumber daya dalam proses *clustering*.

2. Mengukur efektivitas metode *clustering* dalam menghasilkan kelompok yang lebih homogen dan terpisah secara jelas setelah normalisasi skala fitur dilakukan.

#### **1.4 Manfaat Penelitian**

Penelitian ini memiliki beberapa manfaat yang signifikan, yaitu:

1. **Pemahaman yang Lebih Baik:** Memberikan pemahaman yang lebih baik tentang dampak normalisasi skala fitur dalam pengelompokan data, membantu praktisi data science dalam memahami peran kritis normalisasi dalam proses *clustering*.
2. **Pemilihan Metode yang Lebih Tepat:** Membantu praktisi dalam memilih metode *clustering* yang paling sesuai untuk dataset mereka dengan mempertimbangkan normalisasi skala fitur, sehingga meningkatkan kualitas analisis data.
3. **Pengembangan Metode *Clustering*:** Dapat memberikan kontribusi terhadap pengembangan teknik *clustering* yang lebih efisien dan efektif dengan mempertimbangkan normalisasi skala fitur.

#### **1.5 Batasan Penelitian**

Penelitian ini akan memiliki beberapa batasan, yaitu:

1. **Batasan Dataset:** Penelitian akan menggunakan dataset simulasi dan dataset publik yang tersedia. Data yang digunakan mungkin tidak mencakup semua jenis data yang ada di dunia nyata.
2. **Metode Clustering Terbatas:** Fokus penelitian akan pada beberapa metode *clustering* umum seperti K-Means, Hierarchical Clustering, dan DBSCAN. Metode *clustering* lainnya tidak akan dieksplorasi dalam ruang lingkup penelitian ini.
3. **Normalisasi Skala Fitur Tertentu:** Penelitian ini akan memfokuskan pada teknik normalisasi skala fitur tertentu seperti *Min-Max Scaling* dan *Z-*

*Score Standardization*. Teknik normalisasi lainnya mungkin tidak dipergunakan.

4. Evaluasi Terbatas: Evaluasi akan menggunakan metrik clustering tertentu seperti *Silhouette Score* dan *Davies-Bouldin Index*. Ada metrik lain yang mungkin dapat digunakan untuk evaluasi yang lebih khusus. Tetapi di focus penelitian kali ini peneliti menggunakan *Silhouette Score*.

Dengan memahami batasan-batasan ini, penelitian ini diharapkan dapat memberikan wawasan yang berharga dalam menggali pengaruh normalisasi skala fitur terhadap metode clustering dalam pengolahan data.