

# BAB I

## PENDAHULUAN

### 1.1. Latar Belakang

*Clustering* termasuk ke dalam *unsupervised learning* yang bertujuan untuk mengelompokkan objek atau entitas ke dalam suatu kelompok. *Clustering* digunakan di dalam berbagai bidang seperti *data mining*, *machine learning*, *pattern recognition*, dan *image processing*(Flores-Vidal et al., 2019). Berdasarkan pada tingkat kemiripan atau ketidakmiripan fitur, *clustering* akan mempartisi *instance* di dalam *dataset* ke dalam suatu kelompok(Laohakiat & Sa-ing, 2021).

Terdapat banyak algoritma *clustering*, yang secara umum dapat dikelompokkan ke dalam 4 (empat) kategori, yaitu: *Centroid Based Clustering*(Pietrzykowski, 2017), *Herarchical Clustering*(Bouguettaya et al., 2015), *Distribution-Based Clustering*(Zhong et al., 2020), dan *Density-Based Clustering*(Castro Heredia & Rodrigo Mor, 2019). Diantara banyak algoritma *clustering* tersebut, maka *K-Means* yang termasuk di dalam *Centroid Based Clustering* yang paling banyak diterapkan khususnya karena kesederhanaan dan efektifitasnya(Nidheesh et al., 2017).

Cara kerja *K-Means* adalah bahwa ditentukan sejumlah cluster, kemudian tiap *instance* atau data akan ditempatkan ke dalam suatu *cluster* berdasarkan pada kedekatan jarak antara tiap *instance* atau data dengan *centroid* dari masing-masing *cluster*(Yang et al., 2020). Di dalam *K-Means* yang menjadi permasalahan adalah penentuan *centroid* untuk tiap *cluster* yang dilakukan dengan menggunakan bilangan acak yang mengakibatkan akurasi hasil klasifikasi yang diperoleh kurang baik(Xu et al., 2017).

Penentuan centroid yang baik dapat meningkatkan akurasi klasifikasi dimana nilai error dapat berkurang sebesar 60%(Fränti, 2018). Namun upaya untuk menentukan centroid yang baik, cenderung dicapai dengan meningkatkan jumlah iterasi, padahal seharusnya jumlah iterasi dan waktu eksekusi dapat dikurangi(Zubair et al., 2022).

Pendekatan yang dapat digunakan untuk meningkatkan *performance* dari *K-Means* sehubungan dengan penentuan *centroid* yaitu: menggabungkan *K-Means* dengan metode lain yang digunakan untuk menentukan *centroid* dan pendekatan yang lain adalah meningkatkan *performance K-Means* dengan tetap mempertahankan konsep *K-Means* secara umum dengan hanya memberikan perhatian khusus pada aspek penentuan *centroid* tetapi konsep dasar penentuan *centroid* dengan menggunakan bilangan acak tetap dipertahankan. Hasil yang diperoleh menunjukkan pendekatan yang tetap mempertahankan konsep *K-Means* secara umum memperoleh hasil yang lebih baik (Pourahmad et al., 2020).

Sehubungan penggunaan pendekatan dengan tetap mempertahankan konsep *K-Means* secara umum maka sejumlah peneliti telah memberi perhatian pada pemecahan suatu *cluster* ke dalam beberapa *grid point* (Zhu et al., 2019). Dasar pemikirannya adalah bahwa dengan memecah suatu *cluster* ke dalam beberapa *grid point* yang mana masing-masing *grid point* memiliki *centroid* maka peluang untuk memperoleh hasil klasifikasi yang lebih baik akan semakin besar. (Ni et al., 2021) mengemukakan metode *cluster merging* dan menggunakan konsep pemecahan suatu *cluster* menjadi beberapa *cluster* dan menggabungkan hasil yang diperoleh.

Akan menarik untuk diteliti apakah konsep yang dikemukakan oleh (Ni et al., 2021) yang memecah suatu *cluster* menjadi beberapa *cluster* dan kemudian menggabungkan hasilnya digabungkan dengan konsep yang dikemukakan oleh (Zhu et al., 2019) yang memecah suatu *cluster* ke dalam beberapa *grid point*. Sehingga, penelitian ini akan mengkaji bagaimana suatu *cluster* di bagi ke dalam beberapa *sub cluster* dan masing-masing *sub cluster* tersebut dibagi ke dalam beberapa *grid point*. Sehingga penelitian ini dilakukan dengan mengambil judul “Peningkatan Performa K-Means Melalui Cluster and Grid Examining Algorithm (CGEA)”

## 1.2. Rumusan Masalah

Konsep *Cluster Merging* membagi suatu *cluster* ke dalam beberapa *cluster* dan kemudian menggabungkan hasil yang diperoleh. Adapun konsep *Grid Point*

membagi suatu *cluster* ke dalam beberapa *Grid Point*. Algoritma *Cluster and Grid Examining Algorithm* (CGEA) akan menggabungkan kedua konsep tersebut. Adapun masalah yang akan dibahas di dalam penelitian ini adalah Bagaimana Akurasi *K-Means* dapat ditingkatkan melalui *Cluster and Grid Examining Algorithm* (CGEA)?.

### 1.3. Tujuan Penelitian

Adapun tujuan penelitian dari ini adalah menganalisis perbandingan akurasi *K-Means* yang diperoleh melalui *Cluster and Grid Examining Algorithm* (CGEA).

### 1.4. Manfaat Penelitian

Adapun manfaat dari penelitian adalah sebagai berikut.

1. Melalui penelitian ini peneliti akan memperoleh perbandingan akurasi dari *K-Means* dalam bentuk peningkatan akurasi dengan menggunakan CGEA.
2. Mengetahui jumlah *Cluster* dan jumlah *Grid Point*.

### 1.5. Ruang Lingkup Dan Batasan Penelitian

Sehubungan dengan luasnya permasalahan dan adanya keterbatasan waktu dan pengetahuan peneliti, maka peneliti membatasi masalah yang akan dibahas di dalam penelitian ini sebagai berikut:

1. Akurasi *K-Means* yang digunakan di dalam penelitian ini didasarkan pada akurasi hasil klasifikasi dengan menggunakan *Mean Square Error* (MSE).
2. Perhitungan *distance* dilakukan dengan menggunakan *Euclidean Distance*.
3. Fitur pada Dataset Iris didapat dari Sepal Length : Diukur dari dasar sepal (bagian bawah kelopak bunga) hingga ujung sepal. Sepal adalah bagian dari bunga yang biasanya berwarna hijau dan melindungi kelopak bunga saat masih kuncup. Sepal Width : Diukur pada lebar sepal di bagian yang terlebar. Lebar ini diukur secara tegak lurus terhadap panjang sepal. Petal Length : Diukur dari dasar petal (bagian dasar bunga yang berwarna) hingga ujung petal. Petal adalah bagian dari bunga yang biasanya berwarna cerah dan menarik perhatian serangga penyerbuk.

Petal Width : Diukur pada lebar petal di bagian yang terlebar, seperti halnya lebar sepal, diukur secara tegak lurus terhadap panjang petal.