

BAB V PENUTUP

5.1. Kesimpulan

1. Berdasarkan hasil analisis clustering menggunakan algoritma K-Means, dapat disimpulkan bahwa setiap jumlah cluster ($K = 3$, $K = 5$, dan $K = 7$) memberikan hasil pengelompokan yang berbeda terhadap data yang dianalisis. Pada saat jumlah cluster ditentukan sebanyak $K = 3$, hasil evaluasi menunjukkan bahwa nilai *Average Within Centroid Distance* adalah sebesar -0.429. Nilai ini menggambarkan bahwa data dalam satu cluster masih memiliki jarak yang cukup besar terhadap pusat cluster-nya, yang berarti tingkat keseragaman (kohesi) dalam cluster masih tergolong rendah. Hal ini mengindikasikan bahwa tiga cluster belum mampu mengelompokkan data secara optimal.
2. Selanjutnya, pada saat jumlah cluster ditingkatkan menjadi $K = 5$, nilai *Average Within Centroid Distance* menurun menjadi -0.171. Penurunan ini menunjukkan bahwa kualitas pengelompokan mengalami peningkatan, karena data dalam satu cluster menjadi lebih homogen dan lebih dekat dengan centroid-nya. Hal ini menandakan bahwa pembagian data ke dalam lima cluster mampu menghasilkan kelompok yang lebih representatif dibandingkan dengan $K = 3$. Dengan demikian, $K = 5$ memberikan hasil yang lebih optimal dalam hal kohesi antar data dalam cluster.
3. Ketika jumlah cluster ditambahkan menjadi $K = 7$, nilai *Average Within Centroid Distance* semakin menurun menjadi -0.062, yang merupakan nilai terkecil dari ketiga skenario. Hal ini menunjukkan bahwa jarak antar data

dengan pusat cluster menjadi semakin kecil dan kelompok yang terbentuk menjadi lebih padat dan terpusat. Namun, meskipun nilai ini menunjukkan kualitas pengelompokan terbaik secara statistik, penggunaan jumlah cluster yang terlalu banyak dapat mengurangi kejelasan interpretasi hasil dan berisiko terhadap overfitting. Oleh karena itu, dalam memilih jumlah cluster terbaik, perlu mempertimbangkan tidak hanya hasil evaluasi metrik, tetapi juga konteks dan tujuan analisis secara menyeluruh.

5.1.Saran

1. Berdasarkan hasil evaluasi dari proses clustering menggunakan algoritma K-Means dengan variasi jumlah cluster ($K = 3$, $K = 5$, dan $K = 7$), disarankan agar peneliti atau praktisi ke depannya lebih cermat dalam menentukan jumlah cluster yang digunakan. Meskipun nilai *Average Within Centroid Distance* paling rendah diperoleh pada $K = 7$, hal ini tidak serta-merta menjadikannya pilihan terbaik tanpa mempertimbangkan aspek interpretabilitas. Oleh karena itu, sebaiknya pemilihan jumlah cluster juga memperhatikan konteks data, kebutuhan analisis, dan kemudahan dalam menafsirkan hasil pengelompokan.
2. Selanjutnya, untuk meningkatkan kualitas hasil clustering, disarankan agar data yang digunakan menjalani proses pra-pemrosesan yang lebih optimal, seperti normalisasi, penghapusan outlier, dan penyesuaian skala pada setiap atribut. Selain itu, konversi data kategorikal ke bentuk numerik perlu dilakukan dengan bijak agar tidak menghilangkan makna semantik dari data aslinya. Pemilihan atribut yang relevan juga sangat memengaruhi akurasi

pengelompokan, sehingga proses seleksi atribut yang signifikan perlu menjadi perhatian khusus dalam penelitian selanjutnya.

3. Saran lainnya adalah agar penelitian selanjutnya mencoba membandingkan metode clustering lain seperti DBSCAN atau Hierarchical Clustering untuk melihat perbandingan performa dan kesesuaian dengan karakteristik data yang dimiliki. Selain itu, untuk memperkaya analisis, peneliti juga dapat mengintegrasikan hasil clustering dengan data spasial atau visualisasi peta tematik apabila data berkaitan dengan wilayah tertentu. Hal ini akan membantu dalam memahami pola distribusi dan kecenderungan dalam kelompok yang terbentuk secara lebih komprehensif.