

## BAB V

### KESIMPULAN DAN SARAN

#### 5.1 Kesimpulan

Hasil penelitian menunjukkan bahwa jumlah cluster optimal adalah dua ( $k = 2$ ), baik pada data primer maupun data sekunder. Pada data primer yang memiliki distribusi nilai homogen, kedua algoritma menghasilkan struktur cluster yang hampir identik dengan perbedaan nilai evaluasi yang kecil. Hal ini menunjukkan bahwa pada data dengan variasi terbatas, kedua metode memiliki performa yang setara.

Pada pengujian menggunakan data sekunder dengan jumlah sampel lebih besar dan variasi nilai lebih beragam, tetap terbentuk dua cluster dengan distribusi 613 siswa (51,1%) pada Cluster 0 dan 587 siswa (48,9%) pada Cluster 1. Analisis rata-rata nilai menunjukkan bahwa Cluster 0 memiliki nilai lebih tinggi pada sebagian besar variabel Semester 1 dan satu variabel Semester 2, sedangkan Cluster 1 memiliki nilai lebih tinggi pada beberapa variabel Semester 2. Pengelompokan ini merepresentasikan perbedaan kombinasi performa antar semester, bukan penentuan kelompok lebih baik atau lebih buruk.

Perbandingan kualitas algoritma menunjukkan bahwa pada data sekunder, K-Means menghasilkan Silhouette Score sebesar 0.421 dan Davies-Bouldin Index sebesar 1.035, sedangkan Hierarchical Clustering menghasilkan Silhouette Score sebesar 0.317 dan Davies-Bouldin Index sebesar 1.494. Selisih nilai Silhouette sebesar 0.104 dan selisih DBI sebesar 0.459 menunjukkan bahwa K-Means menghasilkan cluster yang lebih kompak dan memiliki separasi yang lebih baik pada dataset ini.

Analisis perubahan pola antar semester menunjukkan bahwa dari 1200 siswa, sebanyak 1031 siswa (85,92%) tetap berada pada cluster yang sama, sedangkan 169 siswa (14,08%) berpindah cluster dari Semester 1 ke Semester 2. Hasil ini membuktikan bahwa sebagian besar siswa mempertahankan pola kombinasi nilai yang sama antar semester, sementara sebagian siswa mengalami perubahan

kombinasi nilai yang menyebabkan perpindahan segmentasi. Clustering mampu mengidentifikasi struktur pola sekaligus perubahan pola prestasi akademik antar semester secara kuantitatif.

Secara keseluruhan, penelitian ini membuktikan bahwa pendekatan clustering efektif untuk mengidentifikasi pola kombinasi prestasi akademik siswa serta membandingkan performa algoritma berdasarkan karakteristik distribusi data yang berbeda.

## 5.2 Saran

Penelitian selanjutnya dapat menggunakan dataset dengan atribut tambahan seperti kelas, gender, atau karakteristik demografis untuk memperkaya interpretasi hasil cluster. Selain itu, eksplorasi algoritma clustering lain seperti DBSCAN atau Gaussian Mixture Model dapat dilakukan untuk membandingkan sensitivitas metode terhadap bentuk distribusi data. Penelitian lanjutan juga dapat mengintegrasikan hasil clustering dengan metode supervised learning untuk menganalisis implikasi segmentasi terhadap prediksi performa akademik.

