

BAB 5

SIMPULAN DAN SARAN

5.1. Simpulan

Pada penelitian kali ini, himpunan data fragmen metagenomik diolah dan dikelompokkan menggunakan algoritma *agglomerative hierarchical clustering*, dengan algoritma *linear discriminant analysis* sebagai metode reduksi dimensi dan metode *k-mer* sebagai metode ekstraksi fitur dalam tahap praproses data sampel. Nilai *silhouette index* sebagai tolak ukur validitas pengelompokan data oleh model pembelajaran dalam konteks *3-mer* berkisar antara 0,6022 ~ 0,8029 untuk himpunan latih dan 0,4397 ~ 0,6356 untuk himpunan uji, sementara dalam konteks *4-mer*, nilai *silhouette index* berkisar antara 0,0162 ~ 0,1634 untuk himpunan latih dan 0,0572 ~ 0,3724 untuk himpunan uji. Ini berarti bahwa dalam konteks *3-mer*, hasil pengelompokan data sampel oleh model pembelajaran sudah tergugus dengan cukup baik dan cukup terpisah antar gugusannya. Namun, dalam konteks *4-mer*, terdapat banyak gugusan data yang saling bertimpa satu sama lain, menyebabkan nilai *silhouette index*-nya mendekati nol.

5.2. Saran

Beberapa saran untuk penelitian selanjutnya adalah sebagai berikut.

1. Himpunan data mikroorganisme yang dipilih lebih divariasikan antar genusnya, atau jumlah sampel serta panjang total fragmen yang diteliti diperbesar.

2. Algoritma dan metode lainnya yang belum pernah atau jarang digunakan untuk penelitian serupa dicoba untuk ditelusuri, seperti *divisive hierarchical clustering*.