

BAB V

KESIMPULAN

5.1. KESIMPULAN

Setelah melakukan analisis terhadap permasalahan klasifikasi pada dataset bakteri *E.Coli* menggunakan algoritma *naïve bayes* dan KNN, dimana hasil evaluasi yang didapatkan yaitu berupa tabel perbandingan hasil matik evaluasi *naïve bayes* dan KNN, maka dapat ditarik kesimpulan, yaitu:

1. *Dataset* yang digunakan merupakan *dataset public* yaitu *dataset* mengenai bakteri *E.Coli*, dimana dapat diakses melalui *website UCI repository*. *Pre-processing* pada dataset bakteri *E.Coli* yang digunakan yaitu *data selection* untuk pemilihan data menggunakan *information gain*, *data cleaning* untuk melihat serta membersihkan data dari *missing value* dan duplikat data, berikutnya yang terakhir yaitu *data transformation* menggunakan *Discretize* untuk mengelompokkan nilai atribut data yang sangat banyak agar dapat memudahkan membaca data serta memvisualisasikan data tersebut.
2. Penerapan klasifikasi algoritma *naïve bayes* dan KNN menggunakan seluruh atribut pada dataset bakteri *E.Coli*. Pada masing-masing algoritma *naïve bayes* dan KNN sama-sama menggunakan *5-fold cross validation* dan *10-fold cross validation* untuk mendapatkan hasil evaluasi model yang lebih stabil.
3. Hasil evaluasi model klasifikasi algoritma *naïve bayes* dan KNN yaitu berupa *confusion matrix* pada dataset bakteri *E.Coli* berdasarkan nilai

class distribution. Berdasarkan hasil klasifikasi pada tabel evaluasi dapat dilihat bahwa tingkat matrik tertinggi yaitu algoritma KNN menggunakan *tools RapidMiner* dengan nilai akurasi 85.71% (*5-fold cross validation*), 87.50% (*10-fold cross validation*) serta rata-ratanya 86.60%. Sedangkan algoritma *naïve bayes* menggunakan *tools RapidMiner* dengan nilai akurasi 80.95% (*5-fold cross validation*), 80.95% (*10-fold cross validation*) serta rata-ratanya 80.95%. Untuk *excel* sendiri tingkat matrik tertinggi yaitu algoritma KNN dengan nilai akurasi 85.71% sedangkan algoritma *naïve bayes* yaitu 80.95%.

5.2. SARAN

Dalam penyusunan laporan tugas akhir ini, dimana klasifikasi menggunakan algoritma *naïve bayes* dan KNN pada bakteri *E.Coli* terdapat beberapa saran yang bisa diberikan oleh penulis terhadap penelitian selanjutnya yaitu sebagai berikut:

1. Untuk penelitian selanjutnya menggunakan jenis algoritma yang lainnya atau metode selain klasifikasi agar mengetahui algoritma atau metode apa yang terbaik dalam *dataset* bakteri *E.Coli*.
2. Untuk pengembangan selanjutnya dapat melakukan teknik *pre-processing* yang lebih lengkap atau bahkan menggunakan serangkaian teknik EDA (*exploration data analysis*) sebagai pendekatan menganalisis kumpulan data bakteri *E.Coli* untuk merangkum karakteristik utamanya.