

BAB V PENUTUP

5.1 Kesimpulan

Berdasarkan hasil pengembangan dan evaluasi klasifikasi bakteri *Escherichia coli* menggunakan algoritma *Support Vector Machine* atau SVM dan Naïve Bayes, dapat disimpulkan beberapa hal sebagai berikut:

1. Penelitian ini berhasil membangun dan mengimplementasikan klasifikasi berbasis machine learning yang menggunakan algoritma SVM dan Naïve Bayes untuk mengklasifikasikan strain bakteri *E. coli* berdasarkan fitur-fitur biologis numerik. Klasifikasi dikembangkan secara menyeluruh melalui tahapan pengumpulan data, pra-pemrosesan (normalisasi, analisis korelasi, pengecekan data hilang), hingga evaluasi performa model.
2. Hasil evaluasi menunjukkan bahwa algoritma SVM memberikan performa klasifikasi yang lebih baik dibandingkan Naïve Bayes. Hal ini ditunjukkan dari nilai akurasi yang lebih tinggi, yaitu 90.10%, serta F1-score sebesar 89.59%. Sementara itu, model Naïve Bayes hanya mencapai akurasi sebesar 73.27% dan F1-score sebesar 67.19%. Performa unggul SVM dipengaruhi oleh kemampuannya dalam menangani data non-linear dan relasi antar fitur yang kompleks melalui penggunaan kernel. Selain akurasi dan F1-score, evaluasi klasifikasi juga mempertimbangkan waktu pelatihan model. Naïve Bayes menunjukkan keunggulan signifikan dalam efisiensi waktu, dengan rata-rata pelatihan hanya 0.0034 detik, jauh lebih cepat dibandingkan SVM yang memerlukan 0.0308 detik. Perbedaan ini memperkuat karakteristik masing-masing algoritma: Naïve Bayes unggul dari sisi kecepatan, sedangkan SVM lebih unggul dalam akurasi dan kestabilan prediksi.
3. Confusion matrix dari model SVM menunjukkan prediksi yang lebih akurat pada sebagian besar kelas, terutama pada kelas mayoritas, sedangkan Naïve Bayes mengalami lebih banyak kesalahan klasifikasi pada kelas minoritas. Hal ini menunjukkan bahwa SVM lebih stabil dan andal dalam mengklasifikasikan data yang tidak seimbang. Confusion matrix dari model

SVM menunjukkan prediksi yang lebih akurat pada sebagian besar kelas, terutama pada kelas mayoritas, sedangkan Naïve Bayes mengalami lebih banyak kesalahan klasifikasi pada kelas minoritas. Hal ini menunjukkan bahwa SVM lebih stabil dan andal dalam mengklasifikasikan data yang tidak seimbang. Sebagian besar nilai prediksi benar (*true positives*) pada model SVM terkonsentrasi di diagonal matrix, sementara model Naïve Bayes cenderung melakukan misklasifikasi yang signifikan pada kelas-kelas dengan representasi data yang kecil. Ini mengindikasikan bahwa dalam konteks distribusi kelas yang tidak seimbang, algoritma SVM memiliki ketahanan klasifikasi yang lebih baik.

4. Seluruh proses pengembangan klasifikasi tidak menggunakan teknik *boosting* atau tuning lanjutan, tetapi tetap menghasilkan performa yang kuat. Visualisasi data seperti matriks korelasi dan distribusi kelas sangat membantu dalam proses eksplorasi awal dan mendukung pemahaman struktur data yang digunakan dalam klasifikasi.
5. Berdasarkan hasil yang dicapai, dapat disimpulkan bahwa klasifikasi berbasis SVM lebih efektif dan direkomendasikan untuk diterapkan dalam pengembangan sistem deteksi dan klasifikasi bakteri *E. coli*. Klasifikasi ini memiliki potensi aplikasi di bidang mikrobiologi klinis untuk mendukung proses identifikasi bakteri secara cepat, akurat, dan efisien. Meskipun dalam studi sebelumnya terdapat penggunaan metode *boosting* seperti AdaBoost, penelitian ini secara sadar tidak mengadopsi pendekatan *ensemble* tersebut. Hal ini dikarenakan fokus utama penelitian adalah membandingkan performa dari dua metode klasifikasi *single-classifier* secara langsung, yaitu SVM dan Naïve Bayes, tanpa pengaruh teknik *ensemble* yang dapat meningkatkan kompleksitas model. Selain itu, secara teknis dan praktis, metode *boosting* seperti AdaBoost cenderung membutuhkan sumber daya komputasi yang lebih besar, serta belum tentu kompatibel atau efisien jika diterapkan pada lingkungan dengan keterbatasan infrastruktur. Oleh karena itu, pendekatan yang lebih sederhana namun tetap efektif dipilih sebagai dasar pengembangan dalam konteks penelitian ini.

5.2 Saran

Berdasarkan hasil dan keterbatasan dari pengembangan klasifikasi yang telah dilakukan, berikut beberapa saran yang dapat dipertimbangkan untuk penelitian dan pengembangan lanjutan:

1. Penelitian ini menggunakan parameter default tanpa tuning, sehingga pada penelitian lanjutan dapat dipertimbangkan penggunaan teknik optimasi hyperparameter seperti *Grid Search* atau *Random Search* untuk mendapatkan hasil yang lebih optimal. Selain itu, disarankan untuk mengeksplorasi teknik optimasi lainnya seperti seleksi fitur otomatis dan penyederhanaan struktur data guna mengurangi kompleksitas model, mempercepat pelatihan, dan meminimalkan risiko *overfitting*.
2. Untuk meningkatkan kemampuan generalisasi klasifikasi, sebaiknya digunakan dataset yang lebih besar dan lebih beragam, baik dari jumlah instance maupun variasi strain *E. coli*, termasuk kondisi lingkungan atau jenis media yang berbeda.
3. Evaluasi model terhadap kelas-kelas minoritas masih menunjukkan kelemahan, terutama pada algoritma Naïve Bayes. Oleh karena itu, teknik penyeimbangan data seperti SMOTE (*Synthetic Minority Over-sampling Technique*) dapat dieksplorasi pada penelitian berikutnya.
4. Walaupun dalam penelitian ini tidak digunakan teknik ensemble, pendekatan seperti *Voting Classifier*, *Bagging*, atau *Boosting* dapat diteliti lebih lanjut sebagai alternatif untuk meningkatkan performa model jika kompleksitas klasifikasi dapat ditingkatkan.
5. Klasifikasi yang telah dikembangkan berpotensi untuk diintegrasikan dalam bentuk aplikasi berbasis web atau mobile, yang dapat digunakan sebagai sistem pendukung keputusan di laboratorium mikrobiologi untuk membantu proses klasifikasi bakteri secara otomatis, efisien, dan terstandar.
6. Penelitian selanjutnya disarankan untuk mengkaji *trade-off* antara waktu pelatihan dan akurasi klasifikasi, terutama dalam konteks implementasi sistem nyata yang memerlukan respons cepat seperti deteksi dini penyakit berbasis mikroorganisme.