

## BAB 5

### SIMPULAN DAN SARAN

#### 5.1 Simpulan

Berdasarkan hasil penelitian ini, klasifikasi antara Stadium I dan Stadium III kanker payudara terbukti memiliki tingkat efektivitas yang lebih tinggi dibandingkan klasifikasi antara Stadium II dan Stadium III. Hal ini ditunjukkan dari perolehan nilai akurasi, presisi, *recall*, dan F1-score yang secara konsisten lebih tinggi pada skenario Stadium I vs. III, disertai waktu komputasi yang lebih efisien. Model terbaik dalam skenario ini, yaitu ANOVA+RFE dengan Logistic Regression, mampu mencapai F1-score sebesar 0,97 hanya dalam durasi komputasi 8 detik. Sebaliknya, klasifikasi Stadium II vs. III menunjukkan performa yang lebih rendah dan tidak stabil, dengan F1-score tertinggi hanya sebesar 0,88. Temuan ini mengindikasikan bahwa perbedaan ekspresi genetik antara Stadium I dan III lebih mudah terdeteksi secara komputasional, dibandingkan perbedaan antara Stadium II dan III yang cenderung lebih samar dan heterogen.

Hasil evaluasi penelitian ini juga menunjukkan bahwa data ekspresi gen RNA-seq memberikan performa klasifikasi yang lebih unggul dibandingkan dengan data ekspresi miRNA. Model berbasis RNA-seq mampu mencapai F1-score hingga 0,99 dalam skenario seleksi fitur tertentu, sedangkan performa tertinggi miRNA hanya mencapai F1-score 0,82, disertai waktu komputasi yang lebih lama, khususnya pada penggunaan algoritma SVM. Perbedaan ini mencerminkan kedalaman informasi yang lebih besar dalam data transkriptomik RNA-seq, yang memungkinkan algoritma klasifikasi mengenali pola stadium kanker dengan lebih akurat. Sementara itu, data miRNA, meskipun memiliki nilai biologis tersendiri, belum cukup memberikan kontribusi prediktif yang kompetitif dalam konteks pemisahan stadium kanker berdasarkan pendekatan machine learning.

Penelitian ini juga berhasil mengidentifikasi 35 gen kandidat biomarker yang memiliki kontribusi signifikan dalam membedakan stadium kanker payudara, khususnya antara Stadium I dan III. Gen-gen ini diperoleh melalui proses seleksi fitur ANOVA+RFE dan diimplementasikan dalam model Logistic Regression yang menunjukkan performa klasifikasi sangat tinggi. Dari keseluruhan kandidat, 10 gen telah memiliki dukungan literatur ilmiah yang relevan, sedangkan 25 lainnya berpotensi menjadi biomarker baru yang layak dieksplorasi lebih lanjut. Namun,

analisis ROC AUC individual menunjukkan bahwa sebagian besar gen tidak cukup kuat untuk berdiri sendiri sebagai prediktor, dengan nilai AUC tertinggi hanya sebesar 0,639. Temuan ini menegaskan pentingnya pendekatan multivariat, di mana kekuatan prediktif tidak bergantung pada satu gen tertentu, melainkan pada kombinasi antar gen dalam ruang berdimensi tinggi. Oleh karena itu, meskipun gen-gen ini belum dapat digunakan secara klinis sebagai biomarker tunggal, mereka sangat menjanjikan jika digunakan secara kolektif dalam sistem klasifikasi kanker berbasis pembelajaran mesin.

## 5.2 Saran

Berdasarkan hasil evaluasi dan temuan dalam penelitian ini, terdapat beberapa hal yang dapat disarankan untuk pengembangan penelitian selanjutnya. Salah satu temuan penting menunjukkan bahwa pemilihan metode seleksi fitur sangat memengaruhi performa model, baik dari segi akurasi maupun efisiensi komputasi. Oleh karena itu, disarankan untuk mengeksplorasi lebih lanjut teknik seleksi fitur yang lebih adaptif dan stabil, seperti pendekatan berbasis ansambel atau metode pembelajaran mendalam (*deep learning*), guna memperoleh fitur-fitur yang secara prediktif lebih kuat dalam membedakan stadium kanker payudara. Mengingat kompleksitas data ekspresi genetik dan tingginya dimensi fitur, pemilihan teknik seleksi yang mampu menangkap interaksi kompleks antar gen menjadi aspek penting yang layak ditelusuri lebih dalam.

Selain itu, peningkatan performa model klasifikasi dapat dicapai melalui evaluasi terhadap penggunaan algoritma pembelajaran mesin yang lebih *robust* dalam menangani data berdimensi tinggi serta mengandung *noise*, seperti arsitektur *deep learning*. Pertimbangan ini menjadi relevan mengingat fitur-fitur yang dihasilkan dalam penelitian ini sangat beragam, dan tidak semua algoritma klasifikasi mampu mengolah serta mengoptimalkannya secara efektif. Oleh karena itu, pemilihan model yang tepat tidak hanya berpotensi meningkatkan akurasi prediksi, tetapi juga memungkinkan pencapaian performa tinggi dengan jumlah fitur yang lebih minimal. Strategi ini diharapkan dapat mengurangi beban komputasi yang ditimbulkan oleh kompleksitas data, sekaligus meningkatkan efisiensi dan interpretabilitas sistem klasifikasi dalam konteks prediksi stadium kanker payudara.