

BAB 5 SIMPULAN DAN SARAN

5.1 Simpulan

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan hal-hal berikut:

1. Model klasifikasi yang dilatih menggunakan fitur-fitur genetik telah berhasil dikembangkan. Model tersebut dibangun dengan menggunakan data ekspresi gen yang dikombinasikan dengan pendekatan seleksi fitur Lasso dan RFE serta algoritma klasifikasi *Logistic Regression*. Model tersebut dibangun dengan melakukan *hyperparameter tuning* terhadap nilai *C*, *solver* dan *penalty*.
2. Penelitian ini menggunakan dua pendekatan dalam seleksi fitur yaitu pendekatan statistik dan pendekatan berbasis DEG. Berdasarkan hasil yang didapatkan, pendekatan seleksi fitur berbasis statistik Lasso+RFE dapat menemukan 25 fitur yang dapat membedakan antara kanker prostat stadium II dan III. Dari 25 fitur tersebut, lima fitur (BACH1-AS1, LINC00616, EFNA3, CSNK1A1L, DPT) di antaranya telah dilaporkan memiliki hubungan dengan kanker prostat. Empat fitur lainnya (THNSL2, SOHLH2, MCCC1-AS1, ARHGEF39) telah dilaporkan memiliki hubungan dengan berbagai kanker lainnya. Pendekatan seleksi fitur berbasis statistik Lasso+RFE juga menunjukkan peningkatan performa yang sebesar 10-15% dibandingkan pendekatan berbasis DEG lima.
3. Algoritma *Random Forest* dan *Logistic Regression* menunjukkan ketergantungan dengan metode seleksi fitur yang digunakan. Hal tersebut disebabkan oleh perbedaan antara fitur yang dihasilkan dari berbagai metode seleksi fitur. Kedua algoritma tersebut menunjukkan performa yang serupa pada saat seleksi fitur lima digunakan. Pada pendekatan tersebut, kedua algoritma menghasilkan akurasi di antara 60-75%. Pada pendekatan seleksi fitur statistik seperti Lasso+RFE, kedua algoritma menghasilkan akurasi diantara 70% sampai 99% seiring berkembangnya jumlah fitur yang digunakan. Menggunakan fitur yang dihasilkan pendekatan tersebut, algoritma *Logistic Regression* menunjukkan performa yang lebih

baik dibandingkan *Random Forest*, dengan rata-rata peningkatan sebesar 5-10%. Hasil terbaik dicapai dari penggunaan algoritma *Logistic Regression* dengan metode seleksi fitur Lasso+RFE yang menghasilkan akurasi sebesar 99,17%, precision sebesar 99,05%, recall sebesar 90%, dan F1-score sebesar 90%. Model tersebut menggunakan 25 fitur terpilih dengan konfigurasi *hyperparameter C* bernilai 1, *penalty l2*, dan *solver liblinear*.

5.2 Saran

Sebagai evaluasi dari temuan penelitian ini, sejumlah hal yang dapat menjadi perhatian dalam penelitian-penelitian selanjutnya dijabarkan sebagai berikut:

1. Pengembangan model ke depan sebaiknya mempertimbangkan eksplorasi kombinasi metode seleksi fitur dan algoritma klasifikasi lain yang lebih kompleks, termasuk pendekatan berbasis *deep learning* atau metode *ensemble* yang lebih canggih
2. Validasi eksternal menggunakan dataset dari populasi atau sumber yang berbeda juga penting untuk menguji stabilitas dan kemampuan generalisasi model terhadap variasi biologis yang lebih luas
3. Integrasi data ekspresi gen dan miRNA dengan data klinis seperti usia pasien atau riwayat medis dapat memberikan pendekatan yang lebih holistik dalam pengambilan keputusan klinis.

U M N
UNIVERSITAS
MULTIMEDIA
NUSANTARA