BAB V

PENUTUP

5.1 KESIMPULAN

Berdasarkan hasil penelitian ini, maka dapat disimpulkan :

1. Penelitian ini berhasil menghasilkan model klasifikasi yang efektif dalam mendiagnosa kanker payudara menggunakan algoritma *Naïve Bayes* yang dikombinasikan dengan metode seleksi fitur *Information Gain*. Dataset yang digunakan diperoleh dari situs Kaggle dan mengandung 569 sampel dengan 30 fitur yang diekstrak dari gambar digital hasil aspirasi jarum halus (*Fine Needle Aspiration*, FNA) pada massa payudara. Dataset ini mencakup dua klasifikasi kanker payudara, yaitu *Benign* (B) dengan 357 sampel dan *Malignant* (M) dengan 212 sampel. Melalui proses seleksi fitur, berhasil diidentifikasi 8 fitur paling relevan dan informatif, yang meliputi *perimeter_worst*, *concave_points_mean*, *concave_points_worst*, *area_worst*, *radius_worst*, *perimeter_mean*, *concavity_mean*, dan *area_mean*. Penggunaan fitur terpilih ini berhasil mengurangi jumlah fitur dari 30 menjadi 8, mengurangi kompleksitas model, dan mempercepat proses pelatihan tanpa mengurangi akurasi.
2. Model klasifikasi yang dikembangkan dengan algoritma *Naïve Bayes* dan seleksi fitur *Information Gain* berhasil mencapai tingkat akurasi sebesar 96.49%, yang lebih tinggi dibandingkan dengan akurasi sebelumnya yang hanya 94.15%. Peningkatan ini menunjukkan bahwa model yang dikembangkan mampu mengidentifikasi kanker payudara, baik *benign*

maupun *malignant*, dengan tingkat akurasi yang sangat baik, yang sejalan dengan tujuan penelitian untuk mengetahui sejauh mana akurasi model dalam mengidentifikasi kanker payudara berdasarkan dataset yang tersedia.

3. Evaluasi kinerja model yang dikembangkan menunjukkan bahwa penggunaan seleksi fitur *Information Gain* memberikan peningkatan yang signifikan pada semua metrik evaluasi, termasuk presisi, recall, dan skor F1, khususnya dalam mengidentifikasi klasifikasi kanker payudara yang ganas (*malignant*). Hasil ini mengonfirmasi bahwa algoritma *Naïve Bayes* yang dikombinasikan dengan seleksi fitur *Information Gain* memiliki kinerja yang lebih baik dibandingkan dengan metode klasifikasi lainnya yang diuji dalam penelitian ini. Penerapan seleksi fitur juga berperan penting dalam meningkatkan efisiensi dan kecepatan model, yang sangat relevan untuk praktik medis dalam diagnosis kanker payudara.

5.2 SARAN

Adapun saran yang diberikan mengenai penelitian ini tentang peningkatan performa metode *Naïve Bayes* dengan seleksi fitur *Information Gain* menggunakan *machine learning* untuk klasifikasi diagnosa kanker payudara, terdapat beberapa aspek yang masih perlu diperbaiki dan dikembangkan untuk penelitian mendatang. Penelitian ini masih memiliki ruang untuk penyempurnaan, sehingga beberapa saran berikut ini diharapkan dapat membantu dalam pengembangan pada penelitian mendatang, adapun beberapa sarannya :

1. Mengajukan dalam penelitian selanjutnya menggunakan metode yang berbeda dalam menganalisis data.

2. Penelitian selanjutnya bisa mengintegrasikan lebih banyak atribut yang berkaitan dengan karakteristik *histopatologis* dan genetik tumor, yang dapat memberikan wawasan lebih dalam dan potensi meningkatkan akurasi klasifikasi kanker payudara.
3. Disarankan untuk menggunakan dataset yang lebih besar dan lebih beragam yang mencakup variasi geografis, etnis, dan jenis kanker payudara yang lebih luas, untuk meningkatkan *generalisabilitas* dan keandalan model klasifikasi.