

BAB V

PENUTUP

5.1 KESIMPULAN

Berdasarkan tujuan penelitian yang telah ditetapkan, yaitu untuk mengimplementasikan metode *Random Forest* dan *Extra Trees* dalam klasifikasi dataset kanker paru-paru serta untuk membandingkan kinerja kedua metode dalam melakukan klasifikasi, dapat disimpulkan sebagai berikut:

1. Pada penelitian ini, kedua metode *Random Forest* dan *Extra Trees* berhasil diimplementasikan untuk melakukan klasifikasi dataset kanker paru-paru. Kedua metode tersebut menggunakan teknik *cross-validation* dengan $k=5$ untuk memastikan evaluasi yang lebih andal dan mengurangi potensi *overfitting*. Selain itu, kedua model juga diterapkan dengan teknik penyeimbangan data menggunakan *SMOTE* (*Synthetic Minority Over-sampling Technique*) dan penyesuaian bobot kelas untuk menangani masalah ketidakseimbangan data, yang ditemukan pada dataset kanker paru-paru.
2. Berdasarkan hasil evaluasi, metode *Random Forest* menunjukkan kinerja yang sangat baik dengan akurasi 95% setelah penerapan SMOTE. *Precision*, *recall*, *F1-score*, dan ROC-AUC menunjukkan bahwa model *Random Forest* sangat sensitif dalam mendeteksi kasus kanker paru dengan *precision* 0.78 dan *recall*

0.88 untuk kelas minoritas (tidak kanker), serta precision 0.98 dan recall 0.96 untuk kelas mayoritas (kanker paru). Sementara itu, metode *Extra Trees* juga menunjukkan performa yang baik setelah penerapan SMOTE dengan akurasi 93%. *Precision* dan *recall* untuk kelas 1 (kanker paru) mencapai 0.98 dan 0.94, namun untuk kelas 0 (tidak kanker) menunjukkan precision 0.70 dan recall 0.88.

3. Dari hasil evaluasi, meskipun kedua metode (*Random Forest* dan *Extra Trees*) menunjukkan hasil yang sangat baik setelah penyeimbangan data, *Random Forest* dengan kombinasi SMOTE memiliki performa yang lebih unggul dengan akurasi 95% dibandingkan *Extra Trees* 93%. *Random Forest* menunjukkan konsistensi yang lebih baik dengan F1-score 0.82 untuk kelas minoritas dan 0.97 untuk kelas mayoritas, dibandingkan *Extra Trees* yang mencapai 0.78 dan 0.96. Nilai ROC-AUC *Random Forest* (0.9661) juga lebih tinggi dibandingkan *Extra Trees* (0.9622), menunjukkan kemampuan diskriminasi yang superior. Dengan demikian, berdasarkan hasil penelitian ini, metode *Random Forest* dapat dianggap lebih efektif dalam melakukan klasifikasi dataset kanker paru-paru dibandingkan dengan metode *Extra Trees*.

Secara keseluruhan, penerapan teknik SMOTE pada kedua metode klasifikasi berhasil mengatasi ketidakseimbangan data dan meningkatkan kemampuan model dalam mendeteksi kelas minoritas tanpa mengorbankan kinerja kelas mayoritas. *Random Forest* menunjukkan keunggulan khususnya dalam *precision* untuk kelas

minoritas yang sangat penting dalam konteks deteksi medis untuk meminimalkan *false negative*. Penelitian ini memberikan kontribusi pada pengembangan model prediksi untuk deteksi kanker paru yang dapat menjadi acuan dalam pengembangan sistem diagnosis berbasis *machine learning* di masa depan.

5.2 SARAN

Berdasarkan hasil penelitian yang telah dilakukan, terdapat beberapa hal yang dapat menjadi saran untuk penelitian selanjutnya serta aplikasi praktis dalam pengembangan model deteksi kanker paru-paru:

1. Meskipun metode *Random Forest* dan *Extra Trees* menunjukkan kinerja yang baik, masih terdapat banyak metode *machine learning* lainnya yang dapat dieksplorasi, seperti *Support Vector Machines (SVM)*, *XGBoost*, atau *Neural Networks*. Penelitian lebih lanjut dengan membandingkan berbagai metode ini dapat memberikan wawasan lebih lanjut mengenai metode mana yang lebih unggul dalam menangani dataset kanker paru.
2. Teknik *SMOTE* yang diterapkan dalam penelitian ini berhasil menyeimbangkan jumlah sampel antara kelas mayoritas dan minoritas. Namun, terdapat beberapa teknik penyeimbangan data lainnya yang dapat dipertimbangkan, seperti *ADASYN (Adaptive Synthetic Sampling)*, *Random Undersampling*, atau *Borderline-SMOTE*.

3. Dari segi implementasi praktis, perlu dikembangkan antarmuka pengguna yang *user-friendly* untuk memudahkan tenaga medis dalam menggunakan model untuk diagnosis, serta mengintegrasikan model dengan sistem informasi rumah sakit (SIMRS) untuk implementasi yang seamless dalam praktik klinis. Validasi klinis dengan melibatkan tenaga medis ahli perlu dilakukan untuk memastikan relevansi dan akurasi model dalam kondisi nyata, disertai dengan implementasi sistem monitoring untuk memantau performa model dalam jangka panjang dan melakukan update model secara berkala sesuai dengan data terbaru.
4. Penelitian ini menggunakan dataset yang terbatas untuk mengklasifikasikan kanker paru-paru. Untuk meningkatkan generalisasi model, sangat disarankan untuk menggunakan dataset yang lebih besar dan lebih bervariasi. Hal ini dapat membantu model untuk lebih robust dalam menghadapi variasi data yang lebih luas dan meningkatkan kemampuannya dalam mendeteksi kanker paru pada populasi yang lebih besar.
5. Meskipun model yang dibangun berhasil menunjukkan performa yang baik dalam kondisi terkendali, perlu adanya uji coba pada data dunia nyata untuk memastikan keandalan dan efektivitas model dalam situasi klinis. Data yang lebih kompleks, seperti gambar medis atau data pasien dengan berbagai faktor

risiko, dapat memberikan tantangan lebih besar bagi model dan menguji ketahanan serta akurasi prediksi dalam konteks medis nyata.

6. Model yang dikembangkan dalam penelitian ini memiliki potensi untuk diterapkan dalam sistem diagnosis kanker paru berbasis *machine learning*. Oleh karena itu, sangat disarankan untuk melakukan pengembangan lebih lanjut dalam bentuk aplikasi klinis yang dapat membantu dokter dalam mendeteksi kanker paru lebih cepat dan akurat. Penelitian lebih lanjut mengenai interpretabilitas model juga sangat penting agar hasil prediksi model dapat dipahami oleh tenaga medis dengan lebih mudah.

Dengan mengikuti saran-saran tersebut, penelitian mengenai deteksi kanker paru-paru melalui algoritma *machine learning* dapat lebih ditingkatkan untuk menghasilkan model yang lebih efektif dan bermanfaat dalam mendukung sistem diagnosis medis di masa depan.

