

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Penelitian ini bertujuan untuk mengevaluasi kinerja algoritma *Support Vector Machine* (SVM) dalam memprediksi kanker paru-paru berdasarkan data pasien yang bersumber dari platform Kaggle. Permasalahan utama yang diangkat adalah bagaimana efektivitas algoritma SVM dalam klasifikasi penyakit kanker paru-paru, serta sejauh mana hasil prediksi model ini dapat dibandingkan dengan penelitian terdahulu. Untuk menjawab rumusan tersebut, pendekatan CRISP-DM digunakan sebagai kerangka penelitian, dimulai dari tahap pemahaman bisnis hingga evaluasi model. Penelitian ini juga menekankan pada pemanfaatan dua jenis kernel SVM, yaitu kernel linear dan kernel RBF, serta penerapan teknik *tuning hyperparameter* dan validasi silang guna memperoleh model prediksi yang optimal.

Tahapan data preparation melibatkan berbagai proses seperti pengecekan *missing value* dan duplikasi data, deteksi outlier, transformasi label target, normalisasi menggunakan metode *Min-Max*, serta seleksi fitur menggunakan metode *Recursive Feature Elimination* (RFE). Fitur-fitur seperti merokok, paparan polusi, tingkat energi, masalah pernapasan, dan ketidaknyamanan tenggorokan teridentifikasi sebagai faktor signifikan dalam mempengaruhi status penyakit. Dataset yang digunakan terdiri dari 5000 data, yang kemudian dibagi menjadi 80% data latih dan 20% data uji untuk proses pelatihan dan pengujian model secara seimbang.

Hasil modeling menunjukkan bahwa SVM dengan kernel linear menghasilkan akurasi sebesar 89%, sedangkan kernel RBF memberikan sedikit peningkatan dengan akurasi sebesar 89,6%. Namun, peningkatan performa yang paling signifikan terjadi ketika dilakukan *tuning hyperparameter* menggunakan *Grid Search* dengan kombinasi terbaik $C = 100$ dan $\gamma = 1$, serta validasi silang (cross-validation). Model SVM kernel RBF yang dituning ini mencapai akurasi tertinggi sebesar 91,7%, dengan nilai *precision*, *recall*, dan *F1-score* di atas 90% pada kedua kelas target (positif dan negatif). Hasil ini mengindikasikan bahwa model tidak hanya akurat, tetapi juga seimbang dalam melakukan klasifikasi.

Selain menunjukkan kinerja yang lebih unggul dibandingkan penelitian sebelumnya (misalnya, akurasi 90% oleh Iqbal et al., dan 62,3% oleh Septya et al.), model ini juga menunjukkan kemampuan generalisasi yang baik, sebagaimana ditunjukkan oleh visualisasi *learning curve* dan *validation curve*. Dengan demikian, SVM khususnya dengan kernel RBF dan tuning parameter yang tepat memiliki potensi besar untuk diterapkan dalam sistem pendukung keputusan medis, khususnya dalam upaya deteksi dini kanker paru-paru yang lebih cepat, akurat, dan efisien.

5.2 Saran

Beberapa saran untuk pengembangan dan penelitian lebih lanjut adalah sebagai berikut:

1. Penelitian selanjutnya disarankan untuk membandingkan SVM dengan algoritma lain seperti Random Forest, XGBoost, atau metode Deep Learning untuk mendapatkan hasil klasifikasi yang lebih optimal.
2. Penambahan jumlah data dan variasi fitur medis dapat meningkatkan akurasi serta membuat model lebih robust terhadap data nyata.
3. Hasil penelitian ini dapat dijadikan dasar untuk membangun sistem pendukung keputusan yang membantu diagnosis dini kanker paru-paru secara praktis dan efisien.
4. Diperlukan validasi model menggunakan data medis dari rumah sakit atau klinik agar model dapat diuji efektivitasnya dalam konteks dunia nyata.