BAB V

KESIMPULAN DAN SARAN

5.1 KESIMPULAN

Kesimpulan yang dapat diambil dari hasil penelitian yang telah dilakukan pada bab-bab sebelumnya yaitu;

1. Penelitian ini menggunakan dataset yang berasal dari Kaggle.com dengan jumlah data 1376 dengan atribut, gender, age, smoking, yellow fingers, anxiety, peer, ressure, chronic disease, fatigue, allergy, wheezing, alcohol consuming, coughing, shortness of breath, swallowing difficulty, chest pain, lung cancer.
2. Presentasi hasil akurasi menggunakan 3 seleksi fitur dengan cara simulasi data 100%,75%.50% atribut .
3. Hasil seleksi fitur adalah bahwa model mencapai akurasi tertinggi sebesar 0.998 ketika semua fitur (100%) digunakan. Akurasi tetap stabil saat jumlah fitur dikurangi menjadi 75%, menunjukkan bahwa model masih mampu memanfaatkan fitur-fitur utama secara efektif. Namun, pengurangan fitur hingga 50% menyebabkan penurunan akurasi yang signifikan menjadi 0.993, menandakan bahwa penghapusan fitur penting mulai berdampak pada performa model. Oleh karena itu, keberadaan semua fitur, terutama fitur-fitur utama, sangat penting untuk memastikan akurasi model yang optimal.

4. fitur-fitur utama yang sangat memengaruhi risiko kanker paru adalah smoking, chronic disease, chest pain, dan shortness of breath. Fitur-fitur ini memiliki kontribusi yang signifikan dalam memprediksi kemungkinan kanker paru, sehingga harus diprioritaskan dalam proses seleksi fitur untuk menjaga akurasi prediksi yang tinggi.

5.2 SARAN

Adapun saran yang dapat penulis berikan dari hasil penelitian ini adalah sebagai berikut:

1. Peneliti menggunakan dataset pada website Kaggle.com, diharapkan para peneliti lainnya dapat menggunakan website tersebut sebagai sumber penelitiannya,
2. Diharapkan kedepannya penelitian ini menggunakan data langsung dari suatu objek serta mencakup atribut lainnya
3. Diharapkan data ini diujikan menggunakan algoritma klasifikasi xgboost.