

BAB V

PENUTUP

5.1 KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan terkait komparasi algoritma *Naïve Bayes classifier* dan *Decision Tree* untuk mengklasifikasi penyakit *Stroke* pada dataset *Stroke Prediction*, dapat diambil beberapa kesimpulan utama sebagai berikut.

1. Algoritma *Naïve Bayes classifier* dan *Decision Tree* berhasil diterapkan dengan baik untuk mengklasifikasi penyakit *Stroke* pada dataset *Stroke Prediction*. Kedua algoritma ini memberikan kemudahan dalam memodelkan hubungan antar fitur, sehingga memungkinkan klasifikasi data dengan efisien. Melalui analisis fitur-fitur seperti *Hypertension*, *Avg Glucose level*, *heart disease*, *BMI*, dan *smoking status* menggunakan dataset *Stroke Prediction* dari *Repository Kaggle*, ditemukan hubungan signifikan antara variabel-variabel tersebut dengan kelas *Stroke* dan *non-Stroke*.
2. Berdasarkan hasil penelitian, algoritma *Decision Tree* mencapai tingkat akurasi sebesar 91%, yang merupakan peningkatan sebesar 4% dibandingkan dengan algoritma *Naïve Bayes classifier* yang memiliki tingkat akurasi sebesar 89.04%. Hal ini menunjukkan bahwa dalam konteks klasifikasi penyakit *Stroke* pada dataset *Stroke Prediction*,

algoritma *Decision Tree* cenderung memberikan hasil yang lebih baik dalam memprediksi kasus penyakit *Stroke*.

3. Hasil evaluasi kinerja model algoritma *Naïve Bayes Classifier* dengan menggunakan *Confusion Matrix* memberikan gambaran lebih komprehensif terkait performa model. Presisi sebesar 14,43% menunjukkan sejauh mana hasil positif yang diberikan model benar-benar relevan, sementara *recall* sebesar 35,89% menunjukkan sejauh mana model dapat mengidentifikasi keseluruhan kasus yang sebenarnya. *F1 score* sebesar 20,59% menyatukan keseimbangan antara presisi dan *recall*.

5.2 SARAN

Dalam konteks penelitian ini, beberapa saran dapat diajukan untuk pengembangan lebih lanjut dan peningkatan kualitas penelitian di masa mendatang.

1. Eksplorasi lebih lanjut terhadap penggunaan fitur tambahan yang mungkin meningkatkan akurasi model, seperti faktor risiko tambahan atau informasi pasien yang lebih detail.
2. Evaluasi performa model dengan dataset yang lebih beragam atau dari sumber yang berbeda untuk memahami generalitas model di luar dataset penelitian.
3. Studi komparatif dengan lebih banyak metode klasifikasi lainnya seperti *KNN*, *Random Forest* atau *Support Vector Machines* untuk

mendapatkan pemahaman yang lebih komprehensif mengenai performa algoritma *Naïve Bayes classifier* dan *Decision Tree*.