

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan pembahasan yang telah diuraikan mengenai peningkatan nilai akurasi prediksi stroke menggunakan algoritma *Decision Tree C4.5*, maka dapat disimpulkan bahwa

1. Algoritma *Decision Tree C4.5* Akurasi 98,02%
2. *Precision* Algoritma C4.5 94,44%
3. Hasil *Recall* oleh algoritma *decision tree C4.5* mempunyai nilai nilai 77,72%
4. AUC Algoritma *Decision Tree C4.5* 0.96

Nilai akurasi yang tinggi pada suatu model klasifikasi dapat diartikan bahwa model tersebut mampu mengklasifikasikan data dengan benar dalam proporsi yang tinggi. Nilai AUC berkisar antara 0 dan 1, di mana nilai 0.5 menunjukkan kinerja model yang sama dengan model acak, dan nilai 1 menunjukkan kinerja model yang sempurna dalam membedakan kedua kelas. Semakin tinggi nilai AUC, semakin baik kinerja model dalam membedakan antara kedua kelas yang berbeda.

Nilai precision yang tinggi pada suatu model klasifikasi dapat diartikan bahwa model tersebut mampu mengklasifikasikan data positif dengan benar dalam proporsi yang tinggi, atau dengan kata lain, memiliki kemampuan untuk menghindari banyaknya prediksi false positive.

Nilai recall yang tinggi pada suatu model klasifikasi dapat diartikan bahwa model tersebut mampu mengidentifikasi sebagian besar data positif yang seharusnya diklasifikasikan sebagai positif, atau dengan kata lain, memiliki kemampuan untuk menghindari banyaknya prediksi false negative.

Rumus AUC (Area Under the Curve) adalah sebagai berikut:

AUC = $\int \text{TPR} \, d\text{FPR}$ di mana:

TPR (True Positive Rate) adalah proporsi data positif yang benar diklasifikasikan sebagai positif (True Positive) terhadap semua data positif yang seharusnya diklasifikasikan sebagai positif.

FPR (False Positive Rate) adalah proporsi data negatif yang salah diklasifikasikan sebagai positif (False Positive) terhadap semua data negatif yang seharusnya diklasifikasikan sebagai negatif.

dFPR adalah selisih nilai FPR pada threshold yang berbeda-beda.

ROC Curve yang digunakan dalam perhitungan AUC menggambarkan hubungan antara TPR dan FPR pada berbagai nilai threshold. Integral pada rumus AUC menghitung luas area yang dibentuk oleh ROC curve dan sumbu FPR. Semakin besar n...

$$AUC = \int_{0}^{1} (TPR - FPR) dFPR$$

$$AUC = 0,765 - 0,5876$$

$$UAC = 0,96$$

di mana:

TPR (True Positive Rate) adalah proporsi data positif yang benar diklasifikasikan sebagai positif (True Positive) terhadap semua data positif yang seharusnya diklasifikasikan sebagai positif.

FPR (False Positive Rate) adalah proporsi data negatif yang salah diklasifikasikan sebagai positif (False Positive) terhadap semua data negatif yang seharusnya diklasifikasikan sebagai negatif.

dFPR adalah selisih nilai FPR pada threshold yang berbeda-beda.

ROC Curve yang digunakan dalam perhitungan AUC menggambarkan hubungan antara TPR dan FPR pada berbagai nilai threshold. Integral pada rumus AUC menghitung luas area yang dibentuk oleh ROC curve dan sumbu FPR. Semakin besar nilai AUC, semakin baik kinerja model dalam membedakan antara kedua kelas yang berbeda.

5.2 Saran

Dataset public masih terdapat eror sehingga diharapkan nantinya akan diperoleh analisis yang lebih tepat. Selain itu untuk meningkatkan nilai akurasi pada prediksi Prestasi Peserta Didik Berdasarkan Sosial Ekonomi, Motivasi, Kedisiplinan dan Prestasi Masa Lalu dengan algoritma *Decision Tree C4.5* dalam pengimprove akurasi dapat mengubah parameter sampling linier disesuaikan dengan dataset, Efisien dan menghasilkan akurasi yang optimum. Namun apabila data missing value tersebut tidak dapat dihindarkan maka untuk penelitian selanjutnya dapat dilakukan pengembangan metode yang lain untuk data missing value dan penentuan parameter tanpa trial and error yang diharapkan nantinya akan memberikan akurasi yang lebih tinggi.