

BAB V

PENUTUP

5.1 KESIMPULAN

Berdasarkan penelitian yang telah dilakukan, maka dapat diambil kesimpulan dari hasil penelitian pada bab-bab sebelumnya, yaitu :

- a. Penelitian ini menggunakan *online dataset* yang bersumber dari *website* yang bernama *kaggle* (<https://www.kaggle.com/datasets/rabieelkharoua/students-performance-dataset>), berjumlah 2.392 data, dan memiliki 15 atribut (*feature*) termasuk atribut target (*class*) yaitu *StudentID*, *Age*, *Gender*, *Ethnicity*, *ParentalEducation*, *StudyTimeWeekly*, *Absences*, *Tutoring*, *ParentalSupport*, *Extracurricular*, *Sports*, *Music*, *Volunteering*, *GPA*, *GradeClass*.
- b. Pada *dataset* performa pelajar yang digunakan pada penelitian ini terdapat 15 atribut, kemudian dilakukan proses seleksi fitur agar mendapatkan hasil yang baik dan optimal, sehingga atribut tersebut menjadi 10 dari 15 atribut. Berikut ini merupakan atribut yang berpengaruh terhadap klasifikasi performa pelajar :
StudyTimeWeekly, *Absences*, *Tutoring*, *ParentalSupport*, *Extracurricular*, *Sports*, *Music*, *Volunteering*, *GPA*, *GradeClass* (target/class).
- c. Hasil evaluasi dan validasi untuk membandingkan kinerja terbaik yang dihasilkan dari 2 model *machine learning* dengan 3 metode pengujian model adalah sebagai berikut :

- 1) Hasil dari Algoritma *C4.5 (Testing Set 30%)*
Accuracy = 83.28%
- 2) Hasil dari Algoritma *C4.5 (Testing Set 20%)*
Accuracy = 83.08%
- 3) Hasil dari Algoritma *C4.5 (10 Fold-Cross Validation)*
Accuracy = 85.82%
- 4) Hasil dari Algoritma *Naïve Bayes (Testing Set 30%)*
Accuracy = 77.85%
- 5) Hasil dari Algoritma *Naïve Bayes (Testing Set 20%)*
Accuracy = 76.82%
- 6) Hasil dari Algoritma *Naïve Bayes (10 Fold-Cross Validation)*
Accuracy = 78.34%

Dapat disimpulkan bahwa model terbaik yang memperoleh hasil akurasi terbesar untuk klasifikasi performa pelajar adalah menggunakan Algoritma *C4.5* dengan *10 Fold-Cross Validation* mendapatkan akurasi sebesar 85.82%.

5.2 SARAN

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

- a. Penelitian mendatang diharapkan dapat memanfaatkan *dataset* performa pelajar yang lebih besar dan mencakup atribut tambahan untuk mencapai tingkat akurasi yang lebih optimal dan maksimal.

- b. Data ini diharapkan dapat diuji menggunakan berbagai metode dan algoritma pembelajaran mesin (*machine learning*) lainnya.
- c. Untuk kedepannya diharapkan akan ada penelitian yang membandingkan metode dan algoritma lain untuk memperoleh hasil yang lebih bervariasi.