



**BAB V**

**TEKNOLOGI INFORMASI**

**UNIVERSITAS DARMA PERSADA**

## BAB V

### PENUTUP

#### 5.1 Kesimpulan

Berdasarkan hasil penelitian yang dilakukan oleh penulis didapati beberapa kesimpulan sebagai berikut:

1. Pembentukan *decision tree* dari algoritma *c4.5* dapat memberikan informasi bahwa atribut *feature* yang mempunyai hubungan yang kuat terhadap atribut *class* adalah program studi, yang menempati *root node* pada *decision tree* yang terbentuk.
2. Hasil analisa tabel *confusion matrix* untuk *naïve bayes* mempunyai rata-rata prediksi benar sebesar 40 serta rata-rata prediksi salah 30, Hasil analisa tabel *confusion matrix* untuk *decision tree* mempunyai rata-rata prediksi benar sebesar 36 serta rata-rata prediksi salah 34.
3. Hasil evaluasi algoritma *naïve bayes* mempunyai nilai rata-rata *accuracy* 56.57%, *error rate* 43.43%, *recall* 38.77%, *precision* 39.66%, dan *f1-score* 39.19%, sementara algoritma *decision tree* mempunyai nilai rata-rata *accuracy* 51.14%, *error rate* 48.86%, *recall* 35.63%, *precision* 34.14%, dan *f1-score* 34.87%.
4. Berdasarkan hasil evaluasi dapat dilihat nilai *f1-score* yang rendah pada kedua metode, hal tersebut disebabkan oleh dataset yang mempunyai karakteristik class yang tidak seimbang (*imbalance dataset*).
5. Berdasarkan hasil evaluasi dengan karakteristik *class* yang tidak seimbang (*imbalance dataset*) dapat diketahui bahwa *naïve bayes* merupakan algoritma terbaik dalam melakukan prediksi predikat

kelulusan mahasiswa, dengan perolehan nilai *accuracy* dan *f1-score* yang lebih baik dibandingkan algoritma *decision tree*.

6. Dari keseluruhan analisis dapat disimpulkan bahwa kedua algoritma belum siap memasuki tahap *deployment* dikarenakan nilai akurasi dan *f1-score* yang buruk.

## 5.2 Saran

Penulis menyadari terdapat beberapa kekurangan dalam penelitian ini.

Berikut beberapa saran untuk penelitian dan pengembangan sistem prediksi:

1. Menambahkan atribut yang bersifat eksternal, seperti alasan mendaftar, gaji orang tua, asal sekolah, dan yang lainnya.
2. Mencari alternatif algoritma selain *naïve bayes* dan *decision tree*, seperti *support vector machine*, *k-nearest neighbor*, dan yang lainnya.
3. Melengkapi data pada atribut *feature* dan atribut *target* yang serupa, seperti nilai tes masuk, pilihan program studi dan predikat kelulusan mahasiswa pada periode kelulusan 2018 dan seterusnya, agar sistem prediksi dapat di *deploy* ke seluruh program studi.
4. Untuk mengatasi F1-Score rendah yang disebabkan oleh *imbalance dataset*, maka lakukan teknik *oversampling* ataupun *undersampling*.
5. Membuat aplikasi *prediction/classification* dengan bahasa pemrograman yang saat ini sedang populer digunakan untuk *data mining/data science* yaitu *python*.