

## BAB V SIMPULAN DAN SARAN

### 5.1. Simpulan

Berdasarkan penelitian yang telah dilakukan dalam menerapkan algoritma KNN untuk memprediksi waktu kelulusan *siswa* Binar Academy, dapat disimpulkan beberapa hal sebagai berikut:

- 1) Setelah dilakukannya pengujian terhadap dua metode perhitungan tetangga yakni menggunakan *Eulidean* dan *Manhattan* dalam memprediksi waktu kelulusan siswa Binar Academy, *Euclidean* memiliki performa yang lebih baik daripada *Manhattan*. Metode perhitungan terbaik ditentukan menggunakan penilaian performa algoritma yaitu *Confusion Matrix*. Dalam hal ini yang dinilai adalah *accuracy*, *precision*, dan *recall*. *Accuracy* bagus digunakan sebagai acuan performa algoritma, jika dataset memiliki jumlah data FN dan FP yang sangat mendekati. *Precision* akan bagus saat lebih banyak terjadinya TP dan tidak tidak menginginkan terjadinya FP. *Recall* bagus digunakan ketika memilih FP lebih baik terjadi daripada FN.
- 2) Pengujian *Euclidean* dan *Manhattan* dilakukan dengan parameter  $K= 1, 3, 5, 7, \text{ dan } 9$  dalam memprediksi waktu kelulusan siswa Binar Academy. Penentuan tetangga ini dilakukan sebanyak 10 kali hingga menemukan model terbaik dari kedua metode terhadap nilai  $K$  yang digunakan. Hasil menunjukkan performa *Euclidean* dengan nilai  $K= 1$  adalah yang terbaik.
- 3) Dataset pertama memiliki perbandingan 56 untuk keterangan tepat waktu dan 12 untuk keterangan tidak tepat waktu. Untuk mengatasi hal tersebut perlu dilakukannya metode penyimbangan data. Pada penelitian ini SMOTE dipilih karena diantara dua metode lain yakni *lowersampling* dan *oversampling* keduanya masih belum memiliki performa yang baik. *Lowersampling* akan menghilangkan data mayoritas hingga perbandingannya sama dengan minoritas, menyebabkan beberapa informasi hilang terhadap data yang dihapus. Sementara *oversampling* akan melakukan duplikasi data minoritas hingga perbandingannya sama dengan mayoritas, menyebabkan beberapa data akan mengalami redundansi. Setelah SMOTE di terapkan dataset menjadi 56:56, SMOTE akan melakukan sistetis dataset dengan melihat data minoritas dan memunculkan data baru

Andika Putra Kamula, 2022

PEMBANGUNAN MODEL KLASIFIKASI SMOTE - KNN  
(SYNTHETIC MINORITY OVER-SAMPLING TECHNIQUE -  
K-NEAREST NEIGHBOR) DALAM MEMPREDIKSI WAKTU KELULUSAN SISWA BOOTCAMP BINAR  
ACADEMY

Universitas Pendidikan Indonesia | repository.upi.edu | perpustakaan.upi.edu

berdasarkan kedekatan nilai dengan data mayoritas. Hal ini tidak menyebabkan adanya reduksi data karena data yang muncul merupakan rentang nilai antara data data yang ada. Oleh karena itu, SMOTE dapat mengatasi permasalahan data yang tidak seimbang.

- 4) Akurasi dihitung menggunakan rumus  $(TP + TN) / (TP + FP + TN + FN)$ , berdasarkan tabel *confusion matrix* *True Positive* (TP) adalah nilai banyak data positif yang terprediksi benar oleh model prediksi. *True Negative* (TN) adalah nilai banyak data negatif yang terprediksi benar oleh model prediksi. *False Positive* (FP) adalah nilai banyak data negatif yang terprediksi salah oleh model prediksi. *False Negative* (FN) adalah nilai banyak data positif yang terprediksi salah oleh model prediksi. Tingkat akurasi tertinggi pada pemodelan algoritma KNN dalam memprediksi waktu kelulusan siswa Binar Academy adalah 86.36% untuk *Euclidean Distance* pada K=1, dan 81.82% untuk *Manhattan Distance* pada K=1.

## 5.2. Saran

Berdasarkan hasil penelitian yang telah dilakukan, berikut beberapa saran terhadap penelitian ini:

- 1) Hasil *accuracy*, *precision*, *recall* pada pemodelan ini masih belum dikatakan sempurna, performa algoritma KNN sendiri masih bisa ditingkatkan dengan memaksimalkan dataset. Kualitas dataset yang bagus akan meningkatkan hasil prediksi. Salah satunya dengan melakukan reduksi dimensi fitur pada data.
- 2) Pada penelitian ini hanya menggunakan dua metode perhitungan jarak yaitu *Euclidean* dan *Manhattan*. Pemodelan berikutnya bisa menambahkan metode perhitungan jarak lain seperti *Minkowski* atau *Cosine*. Untuk melihat perbandingan nilai *accuracy*, *precision*, *recall*, dari keempat metode yang ada.
- 3) Untuk menilai kualitas lulusan Binar Academy terdapat beberapa faktor lain yang belum disebutkan dalam penelitian ini, dikarenakan data yang ada bersifat kategori. Untuk data tersebut tidak bisa diimplementasi terhadap algoritma KNN. Oleh karena itu bisa digunakan algoritma klasifikasi lain seperti *Decision Tree*.

- 4) Penelitian menggunakan algoritma klasifikasi tidak hanya pada data bersifat numerik, namun bisa juga terhadap data dengan sifat kategorikal. Penggunaan algoritma KNN bisa diterapkan dalam penelitian lain seperti menemukan rasio penderita diabetes, memprediksi kanker payudara, memprediksi harga beras di Indonesia, memprediksi cuaca di Indonesia, memprediksi perfoma mahasiswa, dan lain sebagainya.