

BAB V PENUTUP

5.1 Kesimpulan

Berdasarkan penelitian yang telah dilakukan tentang pengaruh *balancing* data dan normalisasi data pada performa *K-Nearest Neighbor* dalam klasifikasi kualitas *wine* dapat disimpulkan bahwa pengaruh *balancing* data dan normalisasi data pada performa *K-Nearest Neighbor* serta metode yang mendapatkan akurasi paling tinggi, dan menentukan berapa jumlah *K-Neighbor* yang diperlukan dalam klasifikasi *dataset wine* yang memiliki 12 atribut dan 1599 total data dapat diketahui dengan hasil evaluasi *confusion matrix* sebagai berikut:

1. Metode yang mendapatkan akurasi tertinggi adalah K-NN dengan *Min-max* yang mendapatkan akurasi sebesar 80.75% dengan jumlah $K = 17$.
2. *Balancing* data saja pada *dataset wine* membuat performa K-NN turun 0.5% dari metode yang hanya menggunakan K-NN saja yaitu dari 75% menjadi 74.5% metode *balancing* data tidak banyak mempengaruhi dikarenakan *dataset* ini tidak terlu *imbalance* hanya 6.5% saja perbedaannya.
3. Normalisasi data meningkatkan performa K-NN sebanyak 5.75% dari 75% menjadi 80.75% dikarenakan nilai pada atribut telah menjadi setara bobotnya sehingga tidak ada bias.
4. Kombinasi antara *balancing* data dan normalisasi data mendapatkan peningkatan performa sebanyak 4.75% dari metode K-NN saja dari 75% menjadi 79.75% metode ini memiliki hasil lebih banyak dari metode K-NN saja akan tetapi lebih sedikit dibandingkan dengan metode normalisasi dikarenakan pada metode normalisasi *dataset* tersebut masih sedikit *imbalance* jadi data mayoritas memiliki pengaruh besar pada hasil akurasi akhir.

5.2 Saran

Dari kesimpulan yang di dapatkan dari penelitian tentang pengaruh *balancing* data dan normalisasi data pada performa *K-Nearest Neighbor* dalam klasifikasi kualitas wine ini, peneliti mengetahui bahwa masih ada kekurangan dan keterbatasan pada saat melakukan penelitian ini. Oleh karena itu peneliti dapat memberikan saran pada penelitian yang telah di lakukan ini yaitu sebagai berikut:

1. Menggunakan algoritma yang lebih banyak lagi sehingga dapat dilihat pengaruh *balancing* data dan normalisasi data pada tiap-tiap algoritma yang berbeda
2. Menggunakan algoritma *balancing* data lain agar dapat diketahui algoritma *balancing* data mana yang lebih efektif
3. Menggunakan algoritma normalisasi data yang lain agar dapat dibandingkan dengan metode normalisasi data yang lain.

