

BAB 5

SIMPULAN DAN SARAN

5.1 Simpulan

Berdasarkan dari hasil implementasi, maka disimpulkan sebagai berikut

1. Dalam menjalankan klasifikasi menggunakan metode *K-Nearest Neighbor* dengan *dataset* sebanyak 72 data yang didapatkan melalui situs bernama *Mendeley Dataset*. Untuk menjalankan klasifikasi ini membutuhkan beberapa tahapan yaitu mengumpulkan *dataset* yang masih berbentuk *raw data*, melakukan *pre-processing*, pembagian data, hingga melakukan evaluasi. Dalam melakukan pembagian data dibagi menjadi dua yaitu data uji dan data latih. Dalam melakukan pemodelan di evaluasi pun menjalankan klasifikasi dengan model KNN.
2. Berdasarkan hasil evaluasi menggunakan keseluruhan data, dapat disimpulkan bahwa dengan menggunakan rasio data *training* sebesar 90% dan 80%, model KNN mencapai akurasi 100% berdasarkan class report. Ini menunjukkan bahwa model memiliki kemampuan yang sangat baik untuk mempelajari pola dari sebagian besar data *training* dan menggeneralisasikannya dengan sempurna pada data validasi terpisah.
3. Daftar uji coba parameter K terhadap perbandingan data *training* dan *testing* sebanyak 90:10, 80:20, 75:25, 70:30, 60:40, dan 50:50 memberikan hasil yang berbeda sehingga dapat memilih hasil uji coba yang terbaik dan menghasilkan akurasi yang tinggi. Setiap rasio data di uji dengan nilai $K = [1, 3, 5, 7, 9, 11, 13]$. Dapat disimpulkan bahwa hasil uji dengan $K=3$ mendapatkan akurasi tertinggi yaitu 100% dengan proporsi data 75:25 dan 70:30, untuk $K=11$ mendapatkan akurasi tertinggi yaitu 100% dengan proporsi data 75:25 dan 70:30, dan terakhir dengan nilai $K=13$ mendapatkan akurasi tertinggi yaitu 100% dengan proporsi data 60:40. Hasil evaluasi menggunakan *k-fold cross-validation* menunjukkan bahwa rasio data *training* 75%, 70%, dan 60% memberikan akurasi tertinggi yang mencapai 100% pada nilai rata-rata akurasi. Hal ini menunjukkan bahwa penggunaan parameter K yang tepat dalam model KNN dapat menghasilkan performa yang sangat baik pada kisaran rasio data *training* tersebut. Akan tetapi, semua ini bisa menjadi

overfitting dikarenakan jumlah *dataset* yang terlalu sedikit sehingga masih ada kemungkinan besar terjadinya *overfitting* untuk melakukan klasifikasi.

4. Hasil klasifikasi menggunakan *dataset* yang berupa angka perhitungan seperti pada *dataset* yang digunakan tidak dapat dijadikan sebagai acuan yang akurat sebagai prediksi terkena penyakit PCOS atau tidak. Hal ini dikarenakan khususnya untuk kolom "Status" bisa secara akurat didapatkan berdasarkan hasil teknik pencitraan seperti *ultrasound* yang memungkinkan melakukan visualisasi ovarium dan identifikasi kelainan lainnya.

5.2 Saran

Berdasarkan hasil yang sudah di teliti, peneliti memiliki saran yang ke depannya dapat dilanjutkan untuk pengembangan yaitu

1. Untuk mengembangkan penelitian ini, selanjutnya bisa menggunakan metode *elbow method* untuk mencari dan memilih nilai optimal dalam parameter.
2. Untuk penelitian selanjutnya bisa melakukan pencarian *dataset* yang lebih banyak lagi dan terbaru agar bisa dilakukan penelitian dengan jumlah *dataset* yang besar sehingga bisa menangkap lebih banyak pola.
3. Penelitian terkait dengan penyakit PCOS baiknya menggunakan *dataset* berupa gambar agar mendapatkan prediksi dan klasifikasi yang lebih akurat.

U M M N
U N I V E R S I T A S
M U L T I M E D I A
N U S A N T A R A