

BAB V

PENUTUP

5.1 Kesimpulan

Berdasarkan hasil penelitian dan pembahasan tentang penerapan klasifikasi *Random Forest* dengan menggunakan metode *Principal Component Analysis* (PCA) pada dataset Gangguan Spektrum Autisme pada Anak – anak, dapat ditarik kesimpulan sebagai berikut:

1. Penelitian ini telah berhasil menerapkan metode seleksi fitur *Principal Component Analysis* (PCA) untuk mereduksi dimensi data dan transformasi data yaitu melalui beberapa tahapan seperti standarisasi data, menghitung matriks kovarian, mencari nilai *eigenvectors* dan *eigenvalues*, menentukan nilai K atau jumlah PC dengan metode *Kaiser's Stopping Rule* kemudian akan menghasilkan matriks proyeksi dan terakhir mentransformasi data, kemudian hasil data dari proses PCA tersebut akan dilakukan klasifikasi *random forest* dengan membagi data menjadi data latih dan data uji, data latih akan digunakan untuk membuat model, kemudian data uji akan dilakukan uji coba pada pemodelan yang telah dibuat.
2. Pada penelitian ini, hasil evaluasi terhadap data yang melalui proses PCA menghasilkan nilai akurasi sebesar 98%, dengan presisi sebesar 96%, *recall* sebesar 100% dan *specificity* sebesar 96% dari data penelitian Gangguan Spektrum Autisme (ASD) pada Anak-anak.
3. Terdapat perbedaan terhadap data yang tidak menggunakan atau tanpa melalui proses *Principal Component Analysis* (PCA) dengan yang menggunakan metode tersebut. Setelah diuji coba pada model, data yang tanpa melalui proses PCA didapat hasil evaluasi yaitu nilai akurasi sebesar 91%, *precision* sebesar 92%, *recall* sebesar 84% dan *specificity* sebesar 100%. Berdasarkan hasil tersebut, diketahui bahwa penerapan metode *Principal Component Analysis* (PCA) terhadap data yang digunakan pada penelitian ini dapat meningkatkan performa klasifikasi *random forest*.

5.2 Saran

Penelitian ini memiliki banyak keterbatasan, jauh dari sempurna dan masih banyak yang belum dieksplorasi serta dapat dikembangkan lagi. Untuk memperbaiki dan mengembangkannya, penelitian selanjutnya dapat dilakukan dengan mempertimbangkan beberapa hal berikut:

- a. Pada praproses data, untuk menangani *missing value* pada data dapat diterapkan metode imputasi lainnya seperti *mean* atau *median*.
- b. Pada tahapan proses *Principal Component Analysis* (PCA), dalam menentukan nilai “K” atau jumlah PC dapat dilakukan dengan menggunakan metode lainnya seperti *scree plot* dan *cumulative variance*. Walaupun penulis telah melakukan percobaan sebelumnya dengan *cumulative variance* yang diterapkan pada dataset ini dan menghasilkan hasil atau nilai akurasi yang sama, itu mungkin bisa jadi suatu kebetulan karena penulis memilih nilai K atau jumlah PC dengan total nilai berkisar 80%, dan masih banyak nilai K lainnya yang belum diuji coba, serta penerapan metode pada dataset berbeda mungkin akan menghasilkan hasil yang berbeda – beda juga.
- c. Pada pembuatan model klasifikasi *random forest*, penulis menetapkan angka 100 pada *n_estimators* untuk pembangunan jumlah pohon (*trees*) dalam pembuatan model. Pada penelitian selanjutnya mungkin dapat menerapkan jumlah pohon yang berbeda untuk melihat perbedaan dan perbandingan diantara masing – masing.
- d. Pembagian data sebelum masuk pada pemodelan dilakukan secara manual dengan membaginya menjadi 80% data latih dan 20% data uji, dan dapat dipertimbangkan untuk menggunakan angka persentase pembagian data lainnya dan atau menggunakan metode pembagian data yang berbeda.
- e. Dataset yang digunakan tidak cukup banyak, jumlahnya masih berkisar ratusan data, hal ini bisa menjadi pertimbangan untuk penelitian selanjutnya dengan menerapkan metode yang sama pada dataset dengan jumlah berkisar ribuan atau bahkan puluhan ribu dengan lebih banyak lagi fitur yang digunakan.

- f. Selain metode *Principal Component Analysis* (PCA) dan klasifikasi *random forest*, masih terdapat banyak metode seleksi fitur untuk mereduksi dimensi data dan metode klasifikasi lainnya yang dapat dipertimbangkan dan dieksplorasi pada penelitian selanjutnya.