BAB V

KESIMPULAN

5.1 KESIMPULAN

Berdasarkan hasil dari penelitian yang telah dilakukan, beberapa temuan utama dapat disimpulkan sebagai berikut:

1. Tiga algoritma machine learning (K-*Nearest Neighbors*, *Support Vector Machine*, dan *Random Forest*) telah diuji untuk memprediksi resiko depresi berdasarkan dataset yang tersedia. Hasil evaluasi menunjukkan bahwa SVM (*Support Vector Machine*) memberikan performa terbaik dengan *Accuracy Score* sebesar 98.63%, serta presisi, dan recall yang hampir sempurna, yang membuat model ini sangat efektif dan paling direkomendasikan dalam mendeteksi kedua kelas target. Random Forest juga menunjukkan hasil yang baik yakni pada angka akurasi mencapai 94,92%, meskipun performanya sedikit lebih rendah dibandingkan SVM, terutama dalam mendeteksi kelas depresi. K-Nearest Neighbors (KNN) memiliki performa yang memadai pada angka akurasi 90.43%, tetapi kurang optimal, terutama dalam mendeteksi kelas depresi.
2. Kombinasi metrik "*manhattan*", *n_neighbors* = 5, dan *weights* = "*distance*" ditemukan sebagai parameter terbaik untuk model K-NN oleh *Hyperparameter Tuning*. Sementara itu, kombinasi Kernel = "linear", C = 10, dan gamma = "*scale*" adalah kombinasi terbaik untuk SVM. Adapun kombinasi *max_depth = 20, min_samples_leaf = 1, min_samples_split =*

5, dan *n_estimators* = 100 ditemukan sebagai kombinasi yang paling optimal bagi *Random Forest.*

3. Penelitian ini telah berhasil mengembangkan dan melatih SVM, K-NN, dan *Random Forest,* untuk menentukan tingkat depresi dan kecemasan dari dataset *final depression dataset* berdasarkan nilai aspek kehidupan yang ada pada dataset, yang terdiri dari 2556 data, 18 fitur, dan 1 target, yang telah melalui proses seperti *ordinal encoding*, hingga *SimpleImputer* dan *StandardScaler* untuk memastikan kualitas data.
4. Analisis distribusi target menunjukkan ketidakseimbangan yang signifikan dalam dataset, di mana mayoritas data berasal dari individu yang tidak mengalami depresi (82,1%). Ketidakseimbangan ini dapat memengaruhi performa model prediksi dan perlu diperhatikan dalam implementasi teknik penanganan data untuk hasil yang lebih adil dan akurat.
5. Analisis fitur penting mengungkapkan bahwa variabel "usia" dan "pernah memiliki pikiran bunuh diri" konsisten menjadi indikator utama dalam semua model. Faktor-faktor seperti tekanan akademik, kepuasan belajar, tekanan finansial, dan tekanan kerja juga memiliki kontribusi yang signifikan terhadap hasil prediksi. Sebaliknya, jenis profesi dan beberapa atribut demografis menunjukkan pengaruh yang minimal.

5.2 SARAN

Berdasarkan hasil penelitian dan kesimpulan yang diperoleh, berikut adalah beberapa saran yang dapat dipertimbangkan untuk pengembangan penelitian lebih lanjut dan aplikasi praktis:

1. Ketidakseimbangan dataset antara individu yang mengalami depresi dan yang tidak perlu diperhatikan dalam penelitian berikutnya. Pengumpulan data tambahan pada populasi dengan kategori minoritas (individu yang mengalami depresi) dapat membantu menciptakan dataset yang lebih representatif dan meningkatkan akurasi model pada kategori tersebut.
2. Hasil penelitian ini dapat menjadi dasar untuk mengembangkan alat bantu digital, seperti aplikasi berbasis web atau perangkat lunak untuk skrining awal depresi. Alat ini dapat digunakan oleh individu atau profesional kesehatan mental untuk mengidentifikasi risiko depresi secara dini.
3. Mengingat dataset yang digunakan mencakup rentang usia dan profesi tertentu, penelitian berikutnya dapat memperluas cakupan populasi untuk melihat apakah model tetap efektif pada kelompok usia, budaya, atau kondisi sosioekonomi yang berbeda.