

BAB V

PENUTUP

5.1 Kesimpulan

Penerapan, pengujian, dan analisis untuk membandingkan algoritma *Regresi Linier* dan *Regresi Random Forest* digunakan untuk memperkirakan Tingkat Pengangguran Terbuka (TPT) di Provinsi Jawa Timur dapat ditarik kesimpulan sebagai berikut:

1. Performa algoritma *Regresi Linear* mampu digunakan untuk memprediksi Tingkat Pengangguran Terbuka (TPT), dengan memperoleh nilai MAE sebesar 0,8518, MSE sebesar 1,2439, RMSE sebesar 1,1153, dan R² sebesar 0,5902. Hasil tersebut menunjukkan bahwa model mampu menjelaskan sekitar 59,02% variasi data TPT, namun tingkat kesalahan prediksinya masih relatif lebih tinggi.
2. Performa algoritma *Random Forest Regression* memberikan performa yang lebih baik dibandingkan *Regresi Linear*. Model menghasilkan nilai MAE sebesar 0,6768, MSE sebesar 0,8211, RMSE sebesar 0,9062, dan R² sebesar 0,7295. Nilai tersebut menunjukkan bahwa model tersebut mampu menjelaskan sekitar 72,95 persen variasi data serta menghasilkan tingkat kesalahan prediksi yang lebih rendah.
3. Berdasarkan perbandingan nilai menggunakan metrik MAE, MSE, RMSE, dan R², *Random Forest Regression* ditetapkan sebagai algoritma terbaik dalam penelitian ini. Hal tersebut dibuktikan dengan nilai kesalahan prediksi yang lebih kecil dan kinerja model untuk menjelaskan variasi data yang lebih baik dibandingkan *Regresi Linear*.
4. Berdasarkan hasil analisis *feature importance*, variabel rata-rata berapa lama sekolah merupakan faktor yang paling berpengaruh terhadap prediksi Tingkat Pengangguran Terbuka dengan nilai kepentingan sebesar 0,3554, diikuti oleh PDRB per kapita (0,2301), TPAK (0,1929), Inflasi (0,1247), dan UMR (0,0969). Hasil tersebut menunjukkan bahwa tingkat pendidikan

memberikan kontribusi paling besar dalam proses prediksi TPT pada penelitian ini.

5.2 Saran

Hasil penelitian ini memberikan beberapa rekomendasi untuk penelitian lanjutan, seperti berikut:

1. Penelitian selanjutnya dapat menambahkan jumlah data dan periode pengamatan yang lebih panjang agar model dapat mempelajari pola perubahan Tingkat Pengangguran Terbuka secara lebih komprehensif dan dapat membandingkan *Random Forest Regression* dengan algoritma *machine learning* lainnya seperti, *Support Vector Regression* (SVR), dan *Artificial Neural Network* (ANN) untuk memperoleh model prediksi yang lebih optimal.
2. Kategori yang digunakan dapat diperluas dengan menambahkan indikator ekonomi dan sosial lainnya, seperti tingkat kemiskinan, jumlah penduduk, investasi daerah, pertumbuhan ekonomi, indeks pembangunan manusia (IPM), serta jumlah industri untuk meningkatkan kemampuan prediksi model.
3. Penggunaan teknik optimasi parameter (*hyperparameter tuning*) dapat dilakukan pada model *Random Forest Regression* agar performa model yang dihasilkan menjadi lebih baik dibandingkan penggunaan parameter standar, contohnya seperti penggunaan XGBoost optimasi *Random Forest*, *Gradient Boosting Regression* untuk optimasi *Regresi Linear*.
4. Temuan penelitian ini dapat digunakan sebagai acuan dalam analisis kondisi ketenagakerjaan daerah serta sebagai pendukung pengambilan keputusan berbasis data bagi pemerintah dan pihak terkait dalam upaya menekan angka pengangguran di Provinsi Jawa Timur.