

BAB V PENUTUP

5.1 Kesimpulan

Berdasarkan hasil penelitian dan pembahasan yang telah dilakukan, dapat ditarik beberapa kesimpulan sebagai berikut:

1. Kinerja pada model regresi hybrid BERT-LightGBM memberikan hasil hubungan demografi berupa faktor fitur demografi yang dominan ada pada 'country' dan 'gender'.
2. Model regresi *hybrid* BERT-LightGBM dengan fitur *embedding* teks dan demografi memberikan performa terbaik dibanding model tunggal. Pada data uji diperoleh nilai MAE 3.12, RMSE 4.32, dan R^2 0.938, lebih baik daripada model demografi saja (MAE 7.91, RMSE 10.36, R^2 0.602) dan model teks saja (MAE 11.40, RMSE 14.62, R^2 0.208). Penggabungan fitur teks dan demografi memberikan performa prediksi yang lebih baik dan lebih stabil dibandingkan model tunggal berbasis salah satu jenis fitur saja.
3. Proses pemodelan dilakukan melalui tahapan *cleaning*, *encoding*, *embedding* BERT, PCA, dan *train-test split* sebelum standarisasi untuk mencegah data *leakage*, serta menunjukkan hasil evaluasi yang konsisten antara data latih dan data uji.
4. Model berhasil diimplementasikan pada *dashboard* Streamlit dan digunakan untuk prediksi nilai agregat tingkat kebenaran, bukan untuk penilaian individu.

5.2 Saran

Berdasarkan proses dan hasil penelitian yang telah dilakukan, terdapat beberapa saran untuk pengembangan penelitian selanjutnya. Penelitian berikutnya dapat menggunakan dataset dengan cakupan yang lebih luas atau data yang lebih mutakhir untuk meningkatkan generalisasi model. Penambahan sumber data lain seperti data survei regional atau data longitudinal juga dapat membantu menangkap dinamika perubahan pola kebenaran dari waktu ke waktu.

Dari sisi pemodelan, penelitian lanjutan dapat membandingkan pendekatan *hybrid* BERT–LightGBM dengan metode *hybrid* regresi lainnya, seperti kombinasi *embedding* transformer dengan model *neural network* regresi atau *boosting* lain. Eksperimen terhadap variasi dimensi reduksi *embedding*, teknik *encoding* kategorikal yang berbeda, serta strategi *feature selection* juga dapat dilakukan untuk menguji pengaruhnya terhadap performa model.

Pada aspek evaluasi, pengujian *fairness* yang lebih mendalam antar kelompok demografi dan wilayah dapat dikembangkan dengan metrik bias yang lebih spesifik. Selain itu, pengembangan dashboard dapat diperluas dengan fitur analisis skenario dan monitoring performa model secara berkala. Penelitian ini masih terbatas pada data agregat survei dan representasi teks pertanyaan, sehingga pengembangan pada data tingkat individu (dengan kontrol etika dan privasi yang ketat) dapat menjadi arah penelitian lanjutan yang potensial.

