

BAB V PENUTUP

5.1 Kesimpulan

Berdasarkan hasil penelitian dan pembahasan yang telah dilakukan, dapat disimpulkan bahwa pendekatan regresi berbasis machine learning mampu digunakan secara efektif untuk memprediksi produksi tomat di Provinsi Jawa Barat dengan memanfaatkan data *Open Government Data*. Kesimpulan penelitian ini dirumuskan sebagai berikut:

1. Proses pemodelan prediksi produksi tomat dilakukan melalui tahapan yang sistematis, meliputi pengumpulan dan penggabungan dataset luas areal tanam, luas panen, produksi tomat, serta variabel iklim berupa curah hujan dan kelembapan. Dataset kemudian melalui tahap pembersihan data, analisis statistik deskriptif, serta penghapusan outlier menggunakan metode *Interquartile range* (IQR) sehingga diperoleh dataset akhir yang lebih representatif dengan 244 observasi. Selanjutnya, fitur numerik dan kategorikal diproses menggunakan *StandardScaler* dan *OneHotEncoder* yang diintegrasikan dalam *pipeline* pemodelan untuk memastikan konsistensi proses preprocessing pada seluruh algoritma yang diuji.
2. Evaluasi kinerja model dilakukan menggunakan kombinasi train-test split dengan rasio 80:20 serta *K-fold Cross Validation* ($K=5$) untuk menilai kestabilan dan kemampuan generalisasi model. Hasil evaluasi menunjukkan bahwa model *Regresi Linear* memberikan performa terbaik dibandingkan *Support Vector Regression* dan *Random Forest Regressor*, dengan nilai *Koefisien regresi* determinasi (R^2) pada data uji sebesar 0,912. Hasil ini menunjukkan bahwa sekitar 91,2% variasi produksi tomat dapat dijelaskan oleh variabel-variabel yang digunakan dalam model, serta mengindikasikan bahwa hubungan antar variabel dalam dataset cenderung bersifat linier setelah proses pembersihan data dilakukan.
3. Model *Regresi Linear* yang terpilih kemudian disimpan dalam bentuk berkas *pickle* dan diimplementasikan ke dalam aplikasi web berbasis *Streamlit*. Aplikasi ini memungkinkan pengguna melakukan prediksi produksi tomat secara interaktif, baik melalui input manual maupun unggahan data dalam format *CSV* atau *Excel*. Implementasi ini menunjukkan bahwa model yang dibangun tidak hanya bersifat

teoritis, tetapi juga dapat dimanfaatkan secara praktis sebagai alat bantu prediksi dan pengambilan keputusan di bidang pertanian, khususnya dalam perencanaan produksi tomat di Provinsi Jawa Barat.

4. Berdasarkan hasil analisis korelasi *Pearson*, *Koefisien* regresi, serta evaluasi model satu fitur, dapat disimpulkan bahwa luas panen dan luas areal tanam merupakan faktor yang paling berpengaruh terhadap produksi tomat di Provinsi Jawa Barat. Variabel iklim seperti curah hujan dan kelembapan menunjukkan pengaruh yang relatif lemah terhadap produksi tomat dalam model regresi linier yang digunakan. Selain itu, faktor karakteristik wilayah (kabupaten/kota) juga berkontribusi dalam memengaruhi tingkat produksi, yang tercermin dari perbedaan nilai *Koefisien regresi* pada fitur kategorikal wilayah. Temuan ini menunjukkan bahwa faktor lahan dan wilayah menjadi penentu utama produksi tomat, sementara faktor iklim berperan sebagai faktor pendukung.

5.2 Saran

Berdasarkan penelitian yang telah dilakukan, peneliti menyadari bahwa penelitian ini masih memiliki beberapa keterbatasan. Oleh karena itu, saran-saran berikut dapat dipertimbangkan untuk pengembangan penelitian selanjutnya:

1. Penelitian selanjutnya disarankan untuk memperluas rentang waktu data historis yang digunakan agar analisis tren produksi tomat dapat dilakukan secara lebih longitudinal. Penambahan data dalam jangka waktu yang lebih panjang memungkinkan model menangkap dinamika perubahan produksi akibat kebijakan pertanian, perubahan pola tanam, serta fluktuasi iklim antar tahun secara lebih komprehensif.
2. Untuk meningkatkan performa model non-linier, penelitian selanjutnya dapat menerapkan teknik optimasi *hyperparameter* seperti *Grid Search* atau *Random Search*, khususnya pada algoritma *Support Vector Regression* dan *Random Forest Regressor*. Pendekatan ini berpotensi menghasilkan konfigurasi model yang lebih optimal dibandingkan penggunaan *parameter* default, terutama pada dataset dengan karakteristik wilayah yang beragam.
3. Mengingat hasil penelitian menunjukkan bahwa faktor lahan dan wilayah memiliki pengaruh dominan terhadap produksi tomat, penelitian selanjutnya dapat menambahkan variabel lain yang lebih spesifik, seperti jenis lahan, pola tanam,

penggunaan pupuk, atau teknologi budidaya, agar model mampu menangkap karakteristik produksi tomat secara lebih detail.

4. Pengembangan aplikasi prediksi produksi tomat dapat diarahkan pada integrasi dengan sumber data real-time, seperti API cuaca atau sistem sensor pertanian. Integrasi ini diharapkan mampu meningkatkan adaptivitas model terhadap kondisi lingkungan aktual, sehingga hasil prediksi dapat dimanfaatkan secara lebih optimal oleh petani maupun instansi terkait dalam pengambilan keputusan berbasis data.

