

BAB I

PENDAHULUAN

1.1 Latar Belakang

Produksi pertanian bukan hanya soal jumlah panen, tetapi juga menyangkut stabilitas pasokan, kesejahteraan petani, dan ketahanan pangan daerah. Komoditas tomat di Provinsi Jawa Barat memiliki peran strategis karena permintaannya relatif stabil dan kontribusinya terhadap sektor hortikultura cukup signifikan. Namun kenyataannya, produksi tomat dari tahun ke tahun tidak selalu berada pada kondisi yang stabil. Perubahan luas panen, variasi produktivitas lahan, serta faktor lingkungan sering kali menyebabkan fluktuasi produksi yang sulit diprediksi secara tepat. Jika kondisi ini tidak dikelola dengan pendekatan yang lebih terukur, maka perencanaan produksi akan cenderung bersifat reaktif dan kurang berbasis data. Penelitian sebelumnya menunjukkan bahwa perbandingan beberapa metode regresi dapat membantu meningkatkan kualitas prediksi hasil panen di Jawa Barat, sehingga pendekatan berbasis model statistik dan komputasi menjadi semakin relevan untuk diterapkan [1].

Dalam praktiknya, regresi linear sering dijadikan metode awal dalam membangun model prediksi produksi pertanian karena sederhana dan mudah diinterpretasikan. Model ini mampu menunjukkan hubungan antara variabel seperti luas panen dan hasil produksi. Namun, ketika data memiliki pola yang lebih kompleks dan tidak sepenuhnya linier, kemampuan regresi linear menjadi terbatas. Hal ini terlihat pada studi prediksi produksi jagung di Jawa Barat, di mana meskipun regresi linear memberikan hasil yang cukup baik, masih terdapat potensi peningkatan akurasi apabila digunakan metode yang lebih adaptif terhadap variasi data [2]. Kondisi ini menunjukkan bahwa ketergantungan pada satu metode saja belum tentu menghasilkan prediksi yang optimal.

Di sisi lain, kebijakan keterbukaan data melalui *Open Government Data* (OGD) memberikan peluang besar untuk mengembangkan model prediktif berbasis data resmi pemerintah. Data produksi pertanian yang terdokumentasi secara historis sebenarnya dapat menjadi sumber informasi yang sangat berharga apabila diolah dengan pendekatan yang tepat. Pemanfaatan OGD dalam analisis produksi pertanian telah menunjukkan bahwa data

terbuka mampu mendukung proses pengambilan keputusan yang lebih objektif dan transparan [3]. Artinya, permasalahan bukan terletak pada ketersediaan data, melainkan pada bagaimana data tersebut dianalisis dan dimodelkan secara optimal.

Perkembangan machine learning kemudian menghadirkan alternatif metode yang lebih fleksibel dalam menangani data dengan karakteristik kompleks. Algoritma seperti *Random Forest* dan *Support Vector Regression* (SVR) dirancang untuk menangkap hubungan non-linier yang sulit dijelaskan oleh regresi linear sederhana. Beberapa penelitian menunjukkan bahwa pendekatan ini mampu memberikan performa yang lebih stabil dan akurat dalam berbagai kasus prediksi, terutama ketika variabel yang digunakan memiliki interaksi yang saling memengaruhi [4]. Hal ini memperkuat asumsi bahwa metode berbasis machine learning layak dipertimbangkan dalam konteks prediksi produksi pertanian.

Secara lebih spesifik, *Random Forest Regressor* memiliki keunggulan dalam mengurangi risiko overfitting melalui pendekatan ensemble yang menggabungkan banyak pohon keputusan. Dengan mekanisme tersebut, model menjadi lebih robust terhadap variasi data dan cenderung menghasilkan prediksi yang konsisten. Studi yang membandingkan *Random Forest* dengan metode lain pada prediksi hasil panen menunjukkan bahwa algoritma ini mampu memberikan tingkat akurasi yang lebih tinggi dalam berbagai kondisi pengujian [5]. Temuan ini menjadi dasar bahwa penggunaan metode ensemble dapat meningkatkan reliabilitas model prediksi.

Selain *Random Forest*, *Support Vector Regression* juga terbukti efektif dalam memodelkan pola produksi pertanian yang tidak sepenuhnya linier. Melalui penggunaan fungsi kernel, SVR mampu memetakan data ke ruang berdimensi lebih tinggi sehingga pola yang kompleks dapat diidentifikasi dengan lebih baik. Penelitian mengenai prediksi produksi pertanian lintas wilayah menunjukkan bahwa SVR mampu menghasilkan performa yang kompetitif dan layak dibandingkan dengan metode regresi lainnya [6]. Hal ini menunjukkan bahwa SVR memiliki potensi untuk diterapkan pada konteks produksi tomat berbasis data regional.

Meskipun berbagai metode regresi dan machine learning telah diterapkan pada sejumlah komoditas pertanian, penelitian yang secara khusus membandingkan Regresi Linear, Support Vector Regression (SVR), dan Random Forest Regressor dalam memprediksi produksi tomat berbasis Open Government Data di tingkat Provinsi Jawa Barat masih relatif terbatas. Padahal, ketersediaan data resmi serta kemajuan algoritma prediktif seharusnya mampu menghasilkan model yang lebih akurat dan teruji secara empiris. Oleh karena itu, penelitian ini difokuskan pada analisis dan implementasi ketiga algoritma tersebut menggunakan data OGD periode 2013–2024, dengan tujuan membandingkan tingkat akurasi berdasarkan metrik R^2 , MSE, dan MAE, serta mengidentifikasi variabel yang paling berpengaruh terhadap produksi tomat. Selanjutnya kombinasi Cross Validation dilihat dengan menggunakan angka std terkecil untuk menentukan algoritma terbaik berdasarkan metrik R^2 tertinggi, MSE, dan MAE terendah. Dengan demikian, penelitian ini diharapkan dapat memberikan kontribusi nyata dalam mendukung perencanaan produksi hortikultura berbasis data di Provinsi Jawa Barat. Selanjutnya kombinasi Cross Validation dilihat dengan menggunakan angka std kecil untuk menentukan algoritma terbaik berdasarkan R

1.2 Rumusan Masalah

Produksi tanaman tomat di Provinsi Jawa Barat dari tahun ke tahun cenderung mengalami naik turun yang dipengaruhi oleh berbagai faktor seperti luas areal tanam, luas panen, curah hujan dan kelembapan. Ketidakstabilan ini membuat perencanaan produksi dan distribusi menjadi tidak selalu berjalan optimal. Data historis produksi telah tersedia melalui Open Government Data, namun pada pemanfaatannya untuk menghasilkan informasi prediksi yang benar-benar akurat masih belum maksimal. Selain itu, belum ada gambaran yang jelas mengenai model prediksi mana yang paling mampu menggambarkan pola hubungan antarvariabel produksi tomat di tingkat kabupaten/kota, serta variabel apa yang sebenarnya paling besar pengaruhnya terhadap perubahan jumlah produksi. Kondisi tersebut menunjukkan bahwa masih terdapat permasalahan dalam memahami pola hubungan antarvariabel dan dalam menilai tingkat akurasi model prediksi produksi tomat berdasarkan data historis yang tersedia di Provinsi Jawa Barat.

1.3 Batasan Masalah

Batasan Masalah yang digunakan dalam penelitian ini adalah sebagai berikut:

1. Sistem prediksi pada penelitian ini hanya berfokus pada produksi tomat di Provinsi Jawa Barat
2. Metode yang digunakan terbatas pada algoritma *Regresi Linear*, *Support Vector Regression (SVR)*, dan *Random Forest Regressor*
3. Penelitian ini menggunakan dataset produksi tomat berdasarkan kabupaten/kota di Provinsi Jawa Barat, luas area tanaman, luas panen, curah hujan, dan kelembapan yang didapat dari open data Jabar
4. Penelitian diterapkan menggunakan bahasa pemrograman *python* dan website yang digunakan hanya menggunakan *streamlit* untuk memprediksi produksi tomat
5. Penelitian ini tidak membahas aspek ekonomi mikro seperti harga jual tomat, biaya produksi, dan pengaruh kebijakan perdagangan, tetapi hanya difokuskan pada prediksi volume produksi berdasarkan variabel lingkungan dan produksi

1.4 Tujuan Penelitian

Tujuan spesifik dari penelitian ini adalah:

1. Menganalisis penerapan algoritma *Regresi Linear*, *Support Vector Regression (SVR)*, dan *Random Forest Regressor* untuk memprediksi produksi tanaman tomat di Provinsi Jawa Barat dengan menggunakan data historis produksi pada tahun 2013–2024
2. Mengevaluasi kinerja model menggunakan kombinasi *train-test split* dan *K-fold Cross Validation* pada algoritma *Regresi Linear*, *Support Vector Regression (SVR)*, dan *Random Forest Regressor* dalam memprediksi produksi tanaman tomat di Provinsi Jawa Barat
3. Membandingkan performa ketiga algoritma untuk menentukan model dengan tingkat akurasi terbaik berdasarkan metrik evaluasi R^2 , Mean Squared Error (MSE) (MSE), dan Mean Absolute Error (MAE) dalam memprediksi produksi tanaman tomat di Provinsi Jawa Barat

4. Mengidentifikasi variabel yang paling berpengaruh terhadap hasil prediksi produksi tanaman tomat di Provinsi Jawa Barat berdasarkan hasil analisis dari model regresi yang digunakan

1.5 Manfaat Penelitian

Dalam penelitian ini diharapkan dapat memberikan manfaat yang berguna baik secara teoritis maupun praktis, antara lain sebagai berikut:

1.5.1 Manfaat Teoritis

Penelitian ini diharapkan dapat memberikan kontribusi baik dari sisi pengembangan ilmu pengetahuan maupun dari sisi penerapan praktis di lapangan. Adapun manfaat penelitian ini dapat dijabarkan sebagai berikut:

- a. Menambah wawasan dan literatur mengenai penerapan algoritma regresi dalam bidang pertanian, khususnya penerapan Regresi Berganda, *Support Vector Regression* (SVR), dan *Random Forest Regressor* untuk memprediksi produksi tanaman tomat
- b. Menjadi referensi akademik bagi peneliti lain dalam pengembangan metode machine learning untuk prediksi hasil produksi pertanian berbasis data historis
- c. Memberikan kontribusi akademik dalam memahami hubungan antara variabel-variabel penentu produksi tomat seperti luas area, luas panen, curah hujan, dan kelembapan terhadap hasil produksi aktual

1.5.2 Manfaat Praktis

Selain memberikan kontribusi secara teoritis, penelitian ini juga diharapkan memberikan manfaat secara praktis bagi berbagai pihak yang berkepentingan, sebagai berikut:

- a. Bagi Peneliti

Bagi Peneliti selanjutnya penelitian ini dapat menambah pengetahuan mengenai penerapan teknik machine learning pada sektor pertanian, khususnya bagaimana pemilihan algoritma dan parameter model memengaruhi tingkat akurasi prediksi

- b. Bagi Masyarakat

Bagi Masyarakat penelitian ini diharapkan dapat memberikan informasi dan wawasan prediktif mengenai potensi hasil produksi tomat di Provinsi Jawa Barat, sehingga dapat dimanfaatkan oleh petani, pelaku usaha, maupun pihak terkait dalam perencanaan dan pengambilan keputusan.

1.6 Sistematika Penulisan

Sistematika penulisan laporan skripsi disusun dalam beberapa bab dan setiap bab terdiri dari sub-sub bab yang tersusun dengan garis besar sebagai berikut:

BAB 1 PENDAHULUAN. Berisi pengantar dari penelitian yang dilakukan dan memberikan gambaran umum mengenai topik yang diteliti secara keseluruhan. Pada bagian ini, penulis menjelaskan latar belakang dari topik yang diteliti dengan mengidentifikasi permasalahan yang ingin diselesaikan, merumuskan masalah yang akan di jawab dalam penelitian, menetapkan batasan masalah, mengungkapkan tujuan dari penelitian tersebut, manfaat penelitian yang diperoleh serta sistematika penulisan yang memuat uraian secara garis besar isi skripsi untuk tiap-tiap bab.

BAB II TINJAUAN PUSTAKA. Berisi perbandingan dari penelitian-penelitian sebelumnya dan konsep dasar beserta teori-teori yang relevan dengan topik penelitian, yang diambil dari berbagai sumber pustaka dan referensi yang menjadi landasan dasar dalam penelitian ini.

BAB III METODE PENELITIAN. Bab ini menjelaskan metode yang digunakan dalam penelitian. Pembahasan meliputi jenis penelitian, objek penelitian, populasi dan sampel, teknik pengumpulan data, serta instrumen penelitian yang digunakan. Selain itu, bab ini juga menguraikan tahapan dalam pengolahan data dan penerapan algoritma yang digunakan, mulai dari Regresi Berganda, *Support Vector Regression* (SVR), hingga *Random Forest Regressor*, serta metode evaluasi yang diterapkan. Uraian dalam bab ini disusun secara sistematis untuk memastikan kejelasan dan keterulangan proses penelitian.

BAB IV HASIL DAN PEMBAHASAN. Bab ini berisi hasil pengolahan data serta implementasi algoritma yang telah dilakukan pada dataset produksi tanaman tomat di Provinsi Jawa Barat. Hasil penelitian ditampilkan dalam bentuk tabel, grafik, dan penjelasan untuk menunjukkan kinerja setiap model regresi. Selanjutnya, dilakukan

pembahasan mendalam yang mengaitkan hasil penelitian dengan teori, penelitian terdahulu, dan faktor-faktor yang memengaruhi tingkat akurasi model, sehingga diperoleh pemahaman yang lebih menyeluruh mengenai masing-masing algoritma.

BAB V PENUTUP, Bab ini berisi rangkuman dari keseluruhan penelitian yang telah dilakukan. Di dalamnya mencakup kesimpulan yang menjawab rumusan masalah berdasarkan hasil analisis, serta saran yang ditujukan untuk pengembangan penelitian selanjutnya. Saran tersebut diharapkan dapat memberikan arah bagi penelitian berikutnya maupun penerapan praktis di bidang pertanian, khususnya dalam upaya meningkatkan akurasi prediksi produksi tanaman tomat di wilayah Provinsi Jawa Barat.

