

BAB I

PENDAHULUAN

1.1 Latar Belakang

Provinsi Jawa Tengah merupakan salah satu provinsi dengan kontribusi produksi padi terbesar di Indonesia. Produksi tanaman padi di Indonesia, khususnya di Jawa Tengah, merupakan salah satu aspek penting dalam mendukung ketahanan pangan nasional. Jawa Tengah dikenal sebagai salah satu daerah penghasil padi terbesar di Indonesia nomor dua setelah Jawa Timur pada tahun 2023 dengan menyumbang sebesar 17,5% kebutuhan beras nasional. Namun dalam beberapa tahun terakhir, produksi padi di Jawa Tengah mengalami fluktuasi produksi padi yang menjadi perhatian serius[1].

Pada tahun 2023, produksi padi di Provinsi Jawa Tengah mencapai 9,08 juta ton Gabah Kering Giling (GKG), mengalami penurunan sebesar 0,27 juta ton atau 2,91% dibandingkan tahun 2022 yang mencapai 9,36 juta ton GKG. Penurunan ini dikarenakan berkurangnya luas panen padi yang mencapai 1,64 juta hektare pada tahun 2023, turun 0,05 juta hektar atau 2,86% dibandingkan tahun sebelumnya yang mencapai 1,69 juta hektar. Produksi beras untuk konsumsi pangan penduduk pada tahun 2023 juga mengalami penurunan sebesar 0,16 juta ton atau 2,91%, dengan total produksi mencapai 5,22 juta ton dibandingkan 5,38 juta ton pada tahun 2022[2].

Menurut Badan Pusat Statistik (BPS)[2], terdapat kenaikan dan penurunan produksi padi yang signifikan pada tahun-tahun tertentu, yang disebabkan oleh berbagai faktor seperti perubahan luas tanam, luas panen, dan produktivitas yang kurang optimal. Hal ini menimbulkan permasalahan, di mana petani dan pemangku kepentingan lainnya memerlukan informasi atau gambaran yang akurat untuk merencanakan dan mengelola hasil produksi padi. Variabel-variabel seperti luas tanam, luas panen, dan produktivitas menjadi kunci dalam memahami dinamika produksi padi[3].

Ada banyak faktor yang dapat mempengaruhi hasil produksi padi, namun dalam penelitian ini hanya berfokus pada tiga variabel tersebut, berdasarkan data yang didapatkan[3]. Luas tanam merupakan salah satu faktor utama yang mempengaruhi produksi padi. Luas tanam mengacu pada total lahan yang digunakan untuk menanam padi dalam suatu periode tertentu. Semakin luas lahan yang ditanami, maka potensi hasil produksi padi yang diperoleh juga semakin besar. Namun, luas tanam yang besar tidak selalu menjamin hasil panen yang tinggi, karena masih terdapat faktor lain yang berpengaruh, seperti kondisi cuaca, kualitas benih, dan ketersediaan air. Oleh karena itu, luas tanam menjadi salah satu variabel penting dalam prediksi produksi padi[4].

Selain luas tanam, luas panen juga menjadi variabel penting dalam penelitian ini. Luas panen mengacu pada total lahan yang berhasil dipanen setelah melalui siklus pertumbuhan padi. Perbedaan antara luas tanam dan luas panen dapat terjadi akibat gagal panen yang disebabkan oleh berbagai faktor, seperti hama, penyakit tanaman, atau kondisi cuaca ekstrem. Semakin besar luas panen, semakin tinggi pula jumlah produksi padi yang dihasilkan. Oleh karena itu, luas panen digunakan sebagai salah satu indikator utama dalam memprediksi produksi padi[5].

Produktivitas adalah faktor ketiga yang digunakan dalam penelitian ini. Produktivitas dipengaruhi oleh berbagai faktor, seperti teknik budidaya, penggunaan pupuk, sistem irigasi, serta kualitas benih yang digunakan. Meskipun luas tanam dan luas panen memiliki peran penting dalam menentukan produksi, tingkat produktivitas menjadi faktor penentu seberapa efisien lahan pertanian dalam menghasilkan padi. Produktivitas yang tinggi menunjukkan bahwa lahan pertanian digunakan secara optimal dan menghasilkan panen yang lebih banyak. Oleh karena itu, pemodelan yang tepat juga diperlukan untuk memprediksi produksi padi berdasarkan variabel-variabel tersebut[6].

Data mining juga berperan penting dalam prediksi ini, di mana teknik ini digunakan untuk menggali informasi dari kumpulan data besar. Melalui data mining, peneliti dapat mengidentifikasi pola dan hubungan yang mungkin tidak terlihat secara langsung. Machine learning terdapat dua jenis yaitu *supervised*

learning dan *unsupervised learning*. *Supervised learning* adalah metode pembelajaran yang melibatkan supervisi, sehingga dalam datasetnya selalu terdapat fitur target atau label. Salah satu tugas utama dari metode ini adalah melakukan prediksi, memungkinkan pembuatan model prediksi berdasarkan data historis dengan variabel-variabel yang sudah diketahui. Sebaliknya, *unsupervised learning* merupakan metode pembelajaran tanpa supervisi, yang berarti dataset yang digunakan tidak memiliki label atau target[7].

Penelitian ini berfokus pada *supervised learning*. *Supervised learning* terdiri dari dua jenis utama, yaitu klasifikasi dan prediksi. Klasifikasi digunakan untuk mengelompokkan data ke dalam kategori tertentu, seperti menentukan apakah email termasuk spam atau tidak, sedangkan prediksi digunakan untuk memperkirakan nilai kontinu berdasarkan data historis[8]. Prediksi adalah proses memperkirakan nilai suatu variabel berdasarkan pola dalam data sebelumnya dengan memodelkan hubungan antara variabel independen dan dependen. Beberapa algoritma yang termasuk dalam *supervised learning* antara lain *Linear Regression*, *Logistic Regression*, *Decision Tree*, dan *Random Forest*[9].

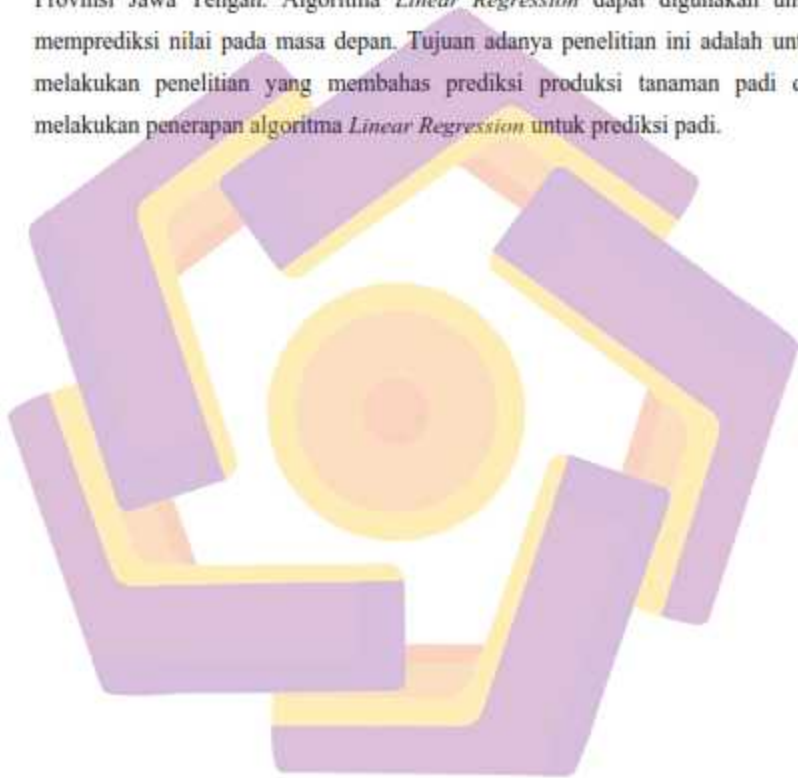
Salah satu pendekatan yang dapat diterapkan untuk memprediksi produksi padi pada penelitian ini adalah dengan menggunakan algoritma *Linear Regression*. Algoritma ini berfungsi untuk memodelkan hubungan antara variabel independen dan dependen, serta merupakan teknik statistik yang digunakan untuk menganalisis keterkaitan antara dua variabel. Kelebihan dari algoritma *Linear Regression* yaitu setiap koefisien dalam model ini menunjukkan pengaruh variabel independen terhadap variabel dependen, bekerja optimal jika data memiliki hubungan linier, sehingga dapat menghasilkan prediksi yang akurat apabila asumsi linearitas terpenuhi. Untuk kekurangannya yaitu sensitif terhadap outlier yang dapat mempengaruhi akurasi model dan kurang efektif jika data memiliki pola non-linear dan jika terdapat terlalu banyak fitur yang tidak relevan, performa model dapat menurun akibat *overfitting* atau *underfitting*[10].

Beberapa penelitian sebelumnya menunjukkan efektivitas algoritma *Linear Regression* dalam memprediksi hasil produksi pertanian. Salah satu penelitian yang

menggunakan algoritma *Linear Regression* adalah penelitian[11], bertujuan untuk memprediksi hasil panen padi di Lamongan dengan menggunakan algoritma *Linear Regression* guna memperkirakan jumlah produksi berdasarkan variabel seperti luas lahan, jumlah bibit, dan jenis pupuk yang digunakan. Hasil penelitian menunjukkan bahwa model *Multiple Linear Regression* yang diterapkan memiliki tingkat keandalan sebesar 94,51%, dengan nilai akurasi yang diuji menggunakan metode *Root Mean Squared Error (RMSE)*. Penelitian[12], mengimplementasikan algoritma Regresi Linier guna meningkatkan akurasi prediksi hasil panen. Data yang digunakan mencakup variabel suhu, kelembaban, pH tanah, dan curah hujan, yang dikumpulkan dari 10 sampel pertanian. Model evaluasi dilakukan menggunakan metrik *Mean Squared Error (MSE)* dan R^2 untuk mengukur tingkat akurasi prediksi. Hasil penelitian menunjukkan bahwa model Regresi Linier mampu memberikan estimasi hasil panen yang cukup akurat, dengan prediksi pada data baru mencapai 479,5 kg/ha. Penelitian[5], juga mengindikasikan bahwa algoritma Regresi Linear Berganda dengan data luas tanam, luas panen, dan produksi GKP dari 2018–2022. Hasil peramalan menunjukkan produksi padi menurun menjadi 685.403 Ton/GKP dari 701.059 Ton/GKP di tahun sebelumnya. Evaluasi model dengan R^2 -Score 0,99 menunjukkan akurasi tinggi, dengan MSE sebesar 1.691.845,33 dan RMSE 1.300,70.

Dalam penelitian ini, dataset yang digunakan bersumber dari situs website Dinas Pertanian dan Perkebunana Provinsi Jawa Tengah dari tahun 2014 sampai 2023 terdiri dari 35 Kabupaten/Kota. Proses penelitian melibatkan beberapa tahapan seperti pengumpulan data, pre-processing, pemodelan data, *hyperparameter tuning*, evaluasi model, pembuatan model prediksi. *Hyperparameter tuning* dilakukan dengan Teknik *GridSearchCV* guna untuk meningkatkan akurasi algoritma yang dihasilkan. Model yang dihasilkan dievaluasi menggunakan metrik *R-Squared (R2)* dan *Root Mean Squared Error (RMSE)*, *Mean Squared Error (MSE)*, dan *Mean Absolute Error (MAE)* untuk mengukur tingkat akurasi. Selain itu, model prediksi ini di deploy dengan menggunakan platform *Streamlit Cloud*, sehingga pengguna dapat dengan mudah mengakses prediksi hasil produksi padi secara online[13].

Penelitian ini bertujuan untuk menerapkan algoritma *Linear Regression* untuk memprediksi produksi padi di Provinsi Jawa Tengah. Dilakukannya penelitian ini dikarenakan belum adanya penelitian yang membahas prediksi produksi tanaman padi pada wilayah Jawa Tengah. Belum diterapkannya algoritma *Linear Regression* yang membahas tentang prediksi produksi tanaman padi di Provinsi Jawa Tengah. Algoritma *Linear Regression* dapat digunakan untuk memprediksi nilai pada masa depan. Tujuan adanya penelitian ini adalah untuk melakukan penelitian yang membahas prediksi produksi tanaman padi dan melakukan penerapan algoritma *Linear Regression* untuk prediksi padi.



1.2 Rumusan Masalah

Berdasar uraian latar belakang, maka dapat disimpulkan peneliti merumuskan permasalahan yang akan dikaji sebagai berikut :

1. Bagaimana menerapkan Algoritma *Linear Regression* dalam memprediksi produksi tanaman padi di Provinsi Jawa Tengah?
2. Berapa nilai akurasi *cross validation* dari *hyperparameter* Teknik *GridSearchCV*?
3. Seberapa akurat prediksi produksi tanaman padi yang dihasilkan oleh Algoritma *Linear Regression*?

1.3 Batasan Masalah

Batasan masalah yang digunakan dalam penelitian ini adalah sebagai berikut:

1. Penelitian ini berfokus pada sistem prediksi produksi padi di Provinsi Jawa Tengah.
2. Penelitian ini menggunakan dataset padi di Provinsi Jawa Tengah yang diperoleh dari Dinas Pertanian dan Perkebunan.
3. Metode yang digunakan terbatas pada algoritma *Linear Regression* berganda.
4. Penelitian ini hanya berfokus pada untuk mengetahui hubungan antara variabel independen ('Luas Tanam', 'Luas Panen', 'Produktivitas') dengan variabel dependen ('Produksi')
5. Penelitian diterapkan dengan menggunakan bahasa pemrograman *python*.
6. Data yang digunakan mencakup 350 data historis.
7. Aplikasi website hanya menggunakan *streamlit* untuk memprediksi produksi padi.
8. Hasil prediksi berupa angka hasil produksi padi yang belum pasti benar.

1.4 Tujuan Penelitian

Penelitian ini bertujuan untuk:

1. Menerapkan Algoritma *Linear Regression* untuk memprediksi produksi padi di Provinsi Jawa Tengah, dengan menggunakan data historis yang relevan dari tahun 2014 hingga 2023.
2. Mengetahui nilai akurasi *cross-validation* yang dihasilkan setelah menerapkan *hyperparameter tuning* menggunakan teknik *GridSearchCV* pada algoritma *Linear Regression* dalam prediksi produksi padi di Provinsi Jawa Tengah.
3. Menganalisis tingkat akurasi prediksi produksi tanaman padi yang dihasilkan oleh Algoritma *Linear Regression* dengan menggunakan metrik evaluasi seperti *Mean Squared Error (MSE)*, *Mean Absolute Error (MAE)*, *R-Squared (R2)* dan *Root Mean Squared Error (RMSE)*.

1.5 Manfaat Penelitian

Dalam penelitian diharapkan dapat memberikan manfaat yang dapat digunakan dengan baik, baik secara teoritis dan praktisi antara lain sebagai berikut:

1.5.1 Manfaat Teoritis

- a. Menambah literatur mengenai penerapan machine learning, khususnya algoritma *Linear Regression*, dalam bidang agrikultur.
- b. Hasil penelitian ini dapat menjadi referensi bagi pengembangan metode prediksi produksi berdasarkan data.
- c. Memberikan kontribusi akademik dalam memahami hubungan antara faktor-faktor lingkungan dengan hasil produksi pertanian.

1.5.2 Manfaat Praktisi

- a. Bagi Peneliti

Bagi peneliti mendapatkan wawasan dan pengalaman berharga dalam membuat pengembangan model prediksi, mulai dari tahap pra-pemrosesan data hingga evaluasi model.

- b. Bagi Pembaca

Bagi pembaca yang ingin membuat serta mengembangkan teknologi khususnya dalam machine learning, penelitian ini diharapkan dapat menjadi referensi yang baik untuk memahami proses prediksi hasil produksi pertanian, khususnya padi.

c. Bagi Masyarakat

Tentunya penelitian ini bisa diterapkan oleh masyarakat, terutama petani dan pembuatan kebijakan yang ada di Provinsi Jawa Tengah. Dengan model prediksi yang dihasilkan, para petani dapat memperoleh gambaran mengenai hasil produksi tanaman padi yang akan datang sehingga mempermudah perencanaan.

1.6 Sistematika Penulisan

Berisi sistematika penulisan skripsi yang memuat uraian secara garis besar isi skripsi untuk tiap-tiap bab. Berikut penulisannya :

BAB I PENDAHULUAN, bab ini menjelaskan latar belakang penelitian, perumusan masalah, batasan penelitian, tujuan penelitian, serta manfaat yang diharapkan, baik secara teoritis maupun praktis. Selain itu, bagian ini juga menyajikan sistematika penulisan untuk memberikan gambaran menyeluruh mengenai konteks dan arah penelitian.

BAB II TINJAUAN PUSTAKA, bab ini mengulas teori-teori yang menjadi dasar penelitian, kajian literatur yang relevan, serta penelitian terdahulu yang berkaitan. Selain itu, disusun pula kerangka konseptual yang berfungsi sebagai landasan teori dalam menganalisis permasalahan yang diteliti.

BAB III METODE PENELITIAN, bab ini menguraikan secara rinci metode yang digunakan dalam penelitian, mencakup jenis penelitian, objek penelitian, populasi dan sampel, teknik pengumpulan data, instrumen penelitian, serta metode analisis data. Penjelasan dalam bab ini disusun secara sistematis untuk memastikan transparansi dalam pelaksanaan penelitian.

BAB IV HASIL DAN PEMBAHASAN, bab ini menyajikan hasil penelitian berdasarkan analisis data yang telah dilakukan. Selain itu, disertakan pembahasan

yang menghubungkan temuan penelitian dengan teori serta kajian sebelumnya guna memberikan pemahaman yang lebih mendalam.

BAB V PENUTUP, bab ini berisi kesimpulan dari penelitian yang telah dilakukan serta memberikan saran untuk pengembangan lebih lanjut. Saran yang diberikan mencakup rekomendasi bagi penelitian selanjutnya maupun penerapannya dalam praktik di dunia nyata.

