

# **BAB I**

## **PENDAHULUAN**

### **1.1 Latar Belakang**

Padi merupakan komoditas pangan yang utama di Indonesia dan berperan penting dalam kestabilan pangan nasional (Ghaffar et al., 2025). Produktivitas padi dipengaruhi oleh berbagai faktor internal dan eksternal seperti, teknik budidaya, varietas benih, penggunaan pupuk dan pestisida, curah hujan, suhu, kelembaban, dan serangan hama. Menurut Susanti et al., (2019), penggunaan pupuk dan pestisida berkontribusi sebesar 56% terhadap hasil panen, namun penggunaannya belum cukup efisien di kalangan petani. Penggunaan pestisida secara berlebihan dapat berpengaruh terhadap lingkungan sekitar dan membahayakan kesehatan (Fathur et al., 2023; Simambela, 2024), pemupukan yang tidak tepat mengakibatkan pemborosan dan tidak efektif. Kondisi ini menunjukkan bahwa keputusan budidaya yang tidak didukung oleh informasi yang memadai berpotensi menurunkan efisiensi dan keberlanjutan produksi padi.

Menurut Badan Pusat Statistik (BPS, 2025), produksi padi nasional diperkirakan mencapai 2,16 juta ton dengan luas panen sekitar 0,42 juta hektar. Namun, capaian ini belum merata di seluruh wilayah, termasuk Kabupaten Nagekeo, Nusa Tenggara Timur. Data BPS NTT (Pura, 2025) mencatat penurunan produksi padi di Nagekeo sebesar 7,81% pada 2024. Penurunan ini tidak semata-mata disebabkan oleh berkurangnya luas panen, tetapi juga dipengaruhi oleh gangguan seperti serangan hama, jamur, serta ketidakpastian pola cuaca. Laporan dari RRI Ende (Meli, 2025) menyebutkan bahwa serangan hama dan jamur menjadi ancaman utama bagi hasil panen. Penyuluh pertanian merekomendasikan penggunaan pupuk dan pestisida secara efisien untuk menjaga kesehatan tanaman dan mengurangi dampak lingkungan.

Kondisi ini menunjukkan pentingnya pengelolaan faktor-faktor pertanian, seperti penggunaan pupuk dan pestisida, ketersediaan air, jenis

benih, serta teknik budidaya, yang didasarkan pada data empiris dari kondisi lapangan. Beberapa penelitian menunjukkan bahwa tidak tepatnya penggunaan input pertanian dapat menurunkan efisiensi produksi dan meningkatkan risiko kerugian hasil panen, terutama pada tingkat petani (Salam et al., 2021; Susanti et al., 2019). Jika kondisi tersebut terus berlanjut, petani berpotensi mengalami kerugian produksi secara berulang, yang pada akhirnya dapat berdampak pada pendapatan petani dan stabilitas produksi pertanian (BPS, 2025). Oleh karena itu, dibutuhkan pendekatan berbasis data lapangan yang dianalisis secara terukur agar faktor-faktor penentu hasil panen dapat dipahami secara lebih sistematis, sehingga pengambilan keputusan budidaya tidak hanya bergantung pada pengalaman atau percobaan berulang.

Seiring dengan perkembangan teknologi, berbagai penelitian telah memodelkan prediksi hasil panen padi menggunakan metode statistik dan algoritma *machine learning*, mulai dari regresi linier hingga algoritma *ensemble* seperti *Random Forest*, *XGBoost* dan *Gradient Boosting Tree Regression*. Pendekatan ini terbukti mampu meningkatkan akurasi prediksi dibandingkan metode tradisional, sebagaimana ditunjukkan oleh penelitian (Hutahaean et al., 2024; Nur et al., 2023). Namun, sebagian besar penelitian prediksi hasil panen padi masih memiliki beberapa keterbatasan. Dari aspek metode, penelitian umumnya berfokus pada peningkatan akurasi model tanpa menjadikan analisis faktor-faktor yang mempengaruhi hasil panen sebagai tujuan utama penelitian (Mugemangango et al., 2024; Nur et al., 2023). Dari aspek data, penelitian masih banyak menggunakan data sekunder atau data agregat berskala regional hingga nasional sehingga belum sepenuhnya merepresentasikan variasi praktik budidaya di tingkat petani (Ansari et al., 2023; Mugemangango et al., 2024). Dari aspek analisis, sebagian penelitian belum mengintegrasikan faktor internal dan eksternal secara bersamaan dalam satu kerangka analisis yang komprehensif (Ansari et al., 2023). Selain itu, dari aspek konteks lokal, penelitian berbasis data primer petani yang menggambarkan kondisi nyata di lapangan, khususnya

pada wilayah seperti Kabupaten Nagekeo, masih relatif terbatas. Oleh karena itu, penelitian ini dilakukan untuk mengisi kesenjangan tersebut dengan membandingkan beberapa algoritma prediksi serta mengidentifikasi faktor-faktor yang paling berpengaruh terhadap hasil panen padi berdasarkan data primer petani.

Berdasarkan kondisi tersebut, penelitian ini disusun untuk mengisi celah penelitian yang masih ada dengan membandingkan beberapa algoritma prediksi guna memperoleh model terbaik, sekaligus mengidentifikasi faktor-faktor internal dan eksternal yang paling berpengaruh terhadap hasil panen padi. Penelitian ini menggunakan data primer yang dikumpulkan langsung dari petani, sehingga variabel yang dianalisis mencerminkan praktik budidaya padi secara nyata di Kabupaten Nagekeo. Untuk mencapai tujuan tersebut, penelitian ini membandingkan empat algoritma regresi, yaitu Regresi Linier Berganda, *Random Forest*, *XGBoost*, dan *Gradient Boosted Tree Regression*. Pemilihan keempat algoritma ini didasarkan pada perbedaan karakteristik masing-masing metode dalam memodelkan data. Regresi Linier Berganda digunakan untuk mengidentifikasi hubungan linier (Murugan et al., 2020), sementara algoritma berbasis *ensemble* seperti *Random Forest* mampu menangkap hubungan non-linear dan interaksi antar variabel pada data pertanian. Selain itu, algoritma *boosting* seperti *XGBoost* dan *Gradient Boosted Tree Regression* dipilih karena memiliki kemampuan meningkatkan akurasi model melalui proses pembelajaran bertahap yang telah terbukti efektif pada berbagai penelitian sebelumnya (Mishra et al., 2020; Mugemangango et al., 2024). Model dengan performa terbaik selanjutnya dianalisis untuk menentukan faktor-faktor yang paling berpengaruh terhadap hasil panen padi di Kabupaten Nagekeo.

Dengan pendekatan analisis yang mengombinasikan perbandingan algoritma prediksi dan identifikasi faktor-faktor yang paling berpengaruh, penelitian ini diharapkan tidak hanya menghasilkan model prediksi hasil panen padi yang relevan dengan kondisi lokal, tetapi juga memberikan

pemahaman yang lebih jelas mengenai faktor-faktor budidaya yang paling menentukan hasil panen. Temuan penelitian ini diharapkan dapat menjadi dasar bagi petani dan penyuluh pertanian dalam menentukan praktik budidaya yang lebih efisien, seperti penggunaan pupuk dan pestisida yang sesuai kebutuhan, pemilihan benih, serta pengelolaan lahan dan air yang tepat, berdasarkan pola yang diperoleh dari data lapangan petani di Kabupaten Nagekeo.

## **1.2 Rumusan Permasalahan**

Berdasarkan pada latar belakang diatas, dapat dirumuskan beberapa permasalahan utama sebagai berikut:

1. Belum diketahui algoritma prediksi yang paling optimal dalam memprediksi hasil panen padi melalui perbandingan beberapa algoritma regresi, khususnya metode statistik dan *machine learning* berbasis ensemble, berdasarkan faktor internal dan eksternal.
2. Belum diketahui faktor-faktor internal dan eksternal yang paling berpengaruh terhadap hasil panen padi berdasarkan analisis kepentingan variabel (*feature importance*) dari model prediksi terbaik.

## **1.3 Pertanyaan Penelitian**

Berdasarkan pada latar belakang dan juga rumusan permasalahan tersebut, dapat dirumuskan beberapa pertanyaan penelitian sebagai berikut:

1. Di antara algoritma Regresi Linear Berganda serta algoritma machine learning berbasis ensemble seperti *Random Forest*, *XGBoost*, dan *Gradient Boosted Tree Regression*, algoritma manakah yang menunjukkan performa paling optimal dalam memprediksi hasil panen padi berdasarkan faktor internal dan eksternal menggunakan metrik evaluasi model?
2. Faktor-faktor internal dan eksternal apa saja yang paling berpengaruh terhadap hasil panen padi berdasarkan analisis kepentingan variabel (*feature importance*) dari model prediksi terbaik?

## **1.4 Tujuan Penelitian**

Tujuan dari penelitian ini adalah sebagai berikut:

1. Menganalisis dan membandingkan performa algoritma Regresi Linear Berganda serta algoritma *machine learning* berbasis *ensemble* seperti *Random Forest*, *XGBoost*, dan *Gradient Boosted Tree Regression* dalam memprediksi hasil panen padi menggunakan metrik evaluasi RMSE, MAE, dan koefisien determinasi ( $R^2$ ), guna menentukan model prediksi dengan kinerja terbaik.
2. Mengidentifikasi dan menganalisis faktor-faktor internal dan eksternal yang paling berpengaruh terhadap hasil panen padi berdasarkan hasil analisis *feature importance* dari model prediksi terbaik.

### **1.5 Batasan Permasalahan**

Penelitian ini hanya berfokus pada beberapa permasalahan yang telah ditetapkan sebagai berikut:

1. Penelitian ini menggunakan studi kasus pada wilayah Kabupaten Nagekeo, Provinsi Nusa Tenggara Timur (NTT).
2. Penelitian ini membatasi analisis pada faktor internal dan eksternal yang mempengaruhi hasil panen padi. Faktor internal meliputi penggunaan pupuk, pestisida, jenis bibit padi, pengolahan lahan, luas lahan, teknik budidaya, dan musim tanam, sedangkan faktor eksternal mencakup curah hujan, serangan hama, serta ketersediaan air melalui irigasi atau tadah hujan.
3. Data yang digunakan dalam penelitian ini terdiri dari data primer dan data sekunder. Data primer dikumpulkan melalui survei menggunakan kuesioner berbasis *Google Form* yang disebarkan kepada petani di Kabupaten Nagekeo. Untuk menjangkau petani yang memiliki keterbatasan akses internet, kuesioner juga disediakan dalam bentuk cetak dan dibagikan secara langsung. Data sekunder diperoleh dari Dinas Pertanian Kabupaten Nagekeo digunakan sebagai data pendukung
4. Data yang digunakan dalam penelitian ini berdasarkan musim tanam terakhir yang dilakukan oleh petani, sehingga tidak secara khusus

membedakan musim tanam dan tidak mempertimbangkan tren jangka panjang atau perubahan iklim.

5. Analisis model hanya menggunakan Regresi Linier Berganda, *Random Forest*, *XGBoost* dan *Gradient Boosted Tree Regression*.

## 1.6 Keaslian Penelitian

Keaslian penelitian ini terletak pada:

1. Penggabungan faktor internal dan eksternal dalam satu analisis  
Penelitian ini menggabungkan dan menganalisis bersama faktor internal dan faktor eksternal yang mempengaruhi hasil panen padi dalam satu model prediksi. Faktor internal meliputi penggunaan pupuk dan pestisida, jenis bibit, teknik budidaya, luas dan pengolahan lahan, serta musim tanam, sedangkan faktor eksternal mencakup curah hujan, ketersediaan air (irigasi atau tadah hujan), serta serangan hama. Pendekatan ini berbeda dari sebagian penelitian sebelumnya yang umumnya hanya menelaah salah satu kelompok faktor secara terpisah.
2. Penggunaan data primer langsung dari petani  
Data dalam penelitian ini diperoleh langsung dari petani melalui survei, sehingga mencerminkan kondisi nyata praktik budidaya padi di Kabupaten Nagekeo. Pendekatan ini membedakan penelitian ini dari banyak studi terdahulu yang lebih mengandalkan data sekunder atau data pertanian dasar dengan cakupan umum.
3. Perbandingan algoritma dengan fokus pemilihan model terbaik  
Penelitian ini membandingkan empat algoritma regresi, yaitu Regresi Linier Berganda, *Random Forest*, *XGBoost*, dan *Gradient Boosted Tree Regression*, dengan tujuan utama menentukan algoritma yang menunjukkan performa paling optimal dalam memprediksi hasil panen padi pada konteks lokal Kabupaten Nagekeo.
4. Analisis faktor paling berpengaruh sebagai *output* penelitian  
Berbeda dari penelitian terdahulu yang umumnya berhenti pada evaluasi akurasi prediksi, penelitian ini menjadikan analisis

kepentingan variabel (*feature importance*) dari model terbaik sebagai bagian dari penelitian. Pendekatan ini memungkinkan identifikasi faktor-faktor internal dan eksternal yang paling berpengaruh terhadap hasil panen padi.

5. Fokus pada konteks lokal Kabupaten Nagekeo

Penelitian ini secara khusus difokuskan pada Kabupaten Nagekeo, Provinsi Nusa Tenggara Timur, yang memiliki karakteristik budidaya dan permasalahan pertanian yang berbeda dari wilayah lain. Fokus lokal ini memberikan nilai tambah dalam memahami hasil panen padi berdasarkan kondisi spesifik daerah.

### **1.7 Manfaat Penelitian**

Penelitian ini diharapkan dapat memberikan manfaat dalam beberapa aspek sebagai berikut:

1. Manfaat Akademik

Penelitian ini diharapkan dapat menambah referensi ilmiah dalam bidang prediksi hasil panen padi dengan pendekatan data *science*, khususnya yang menggabungkan analisis faktor internal dan eksternal secara bersamaan. Selain itu, penelitian ini dapat menjadi bahan rujukan bagi penelitian selanjutnya yang membahas perbandingan algoritma prediksi serta analisis faktor-faktor yang berpengaruh terhadap produktivitas padi pada konteks pertanian lokal.

2. Manfaat Praktis

Hasil penelitian ini diharapkan dapat memberikan informasi yang lebih terukur mengenai algoritma prediksi yang paling sesuai untuk memodelkan hasil panen padi di Kabupaten Nagekeo. Selain itu, identifikasi faktor-faktor yang paling berpengaruh terhadap hasil panen padi diharapkan dapat membantu petani dan penyuluh pertanian dalam mengambil keputusan budidaya yang lebih efisien, khususnya terkait penggunaan pupuk, pestisida, dan pengelolaan sumber daya lainnya.

### 3. Manfaat bagi pengembangan model prediksi

Penelitian ini diharapkan dapat memberikan kontribusi dalam pengembangan model prediksi hasil panen padi melalui perbandingan beberapa algoritma regresi serta analisis faktor-faktor yang paling berpengaruh. Hasil penelitian ini dapat menjadi dasar dalam penyusunan model prediksi yang lebih relevan dan sesuai dengan kondisi pertanian lokal.

## 1.8 Sistematika Penulisan

Sistematika penulisan tesis ini disusun untuk memberikan gambaran yang terstruktur mengenai alur penelitian, sehingga memudahkan pembaca dalam memahami hubungan antar bab dan tahapan penelitian yang dilakukan. Tesis ini disusun dalam enam bab dengan sistematika sebagai berikut.

### **Bab I Pendahuluan**

Bab ini menguraikan latar belakang penelitian yang menjelaskan pentingnya kajian prediksi hasil panen padi berbasis analisis data. Selain itu, bab ini memuat rumusan masalah, tujuan penelitian, manfaat penelitian, batasan penelitian, serta sistematika penulisan tesis.

### **Bab II Tinjauan Pustaka**

Bab ini membahas kajian pustaka yang relevan dengan penelitian, meliputi penelitian-penelitian terdahulu terkait faktor-faktor yang mempengaruhi hasil panen padi, penggunaan metode statistik dan *machine learning* dalam bidang pertanian, serta posisi penelitian ini dibandingkan dengan penelitian sebelumnya.

### **Bab III Landasan Teori**

Bab ini berisi landasan teori yang digunakan sebagai dasar dalam penelitian, mencakup konsep budidaya padi, faktor internal dan eksternal yang mempengaruhi hasil panen, konsep analisis data, *machine learning*, regresi linear, *Random Forest*, *XGBoost*, *Gradient Boosted Tree Regression*, serta konsep evaluasi model dan *feature importance*.

### **Bab IV Metodologi Penelitian**

Bab ini menjelaskan metode penelitian yang digunakan, meliputi desain penelitian, lokasi dan objek penelitian, sumber dan jenis data, teknik pengumpulan data, penentuan variabel dan jumlah responden, serta tahapan analisis data. Pada bab ini juga dijelaskan alur penelitian mulai dari pra-pemrosesan data, eksplorasi data, pemodelan prediksi, evaluasi model, hingga analisis *feature importance* dan penyusunan rekomendasi berbasis data.

### **Bab V Hasil dan Pembahasan**

Bab ini menyajikan hasil penelitian yang diperoleh dari setiap tahapan analisis, mulai dari hasil pengumpulan data, hasil pra-pemrosesan data, eksplorasi data (EDA), hasil pemodelan dan evaluasi model, pemilihan model terbaik, hingga analisis *feature importance*. Hasil tersebut kemudian dibahas dan diinterpretasikan dalam konteks budidaya padi untuk menjelaskan makna temuan penelitian dan implikasinya terhadap praktik pertanian.

### **Bab VI Penutup**

Bab ini berisi kesimpulan yang merangkum temuan utama penelitian sesuai dengan tujuan yang telah ditetapkan, termasuk model prediksi terbaik dan faktor-faktor dominan yang mempengaruhi hasil panen padi. Selain itu, bab ini juga memuat saran yang ditujukan kepada petani, penyuluh pertanian, dan peneliti selanjutnya sebagai tindak lanjut dari hasil peneliti.