Perancangan Kerangka Decision Support System Terintegrasi Untuk Prediksi Produksi Dan Estimasi Waktu Panen Mangga Menggunakan Random Forest Dan Growing Degree Days

Dian Pramadhana[*1], Darsih[2], Riyan Farismana[3]

[1,2,3] Jurusan Teknik Informatika, Politeknik Negeri Indramayu

E-mail: dianpramadhana@polindra.ac.id[1], darsih@polindra.ac.id[2], riyanfarismana@polindra.ac.id[3]

Abstrak. Prediksi produksi dan estimasi waktu panen merupakan aspek penting dalam mendukung pengelolaan tanaman mangga. Variasi karakteristik tanaman, kondisi kebun, dan lingkungan menyebabkan kebutuhan akan informasi produksi dan waktu panen yang dapat mendukung perencanaan kegiatan budidaya. Penelitian ini bertujuan merancang kerangka *Decision Support System* (DSS) terintegrasi untuk prediksi produksi dan estimasi waktu panen mangga dengan memanfaatkan Random Forest dan *Growing Degree Days* (GDD). Random Forest digunakan untuk merancang proses prediksi produksi berdasarkan karakteristik kebun, tanaman, dan kondisi lingkungan, sedangkan GDD digunakan untuk mengestimasi perkembangan tanaman berdasarkan akumulasi suhu. Kedua pendekatan dirancang sebagai komponen terpisah yang hasilnya diintegrasikan pada tingkat sistem untuk menghasilkan informasi estimasi produksi dan waktu panen. Perancangan mencakup tahapan pengumpulan dan preprocessing data, pembentukan model prediksi, perhitungan GDD, integrasi hasil prediksi, serta perancangan sistem pendukung keputusan berbasis web. Evaluasi model dirancang menggunakan MAE, RMSE, dan R^2 untuk prediksi produksi serta selisih hari antara estimasi dan tanggal panen untuk estimasi waktu panen. Kerangka yang dihasilkan memberikan rancangan integrasi antara data kebun, data lingkungan, model prediksi, dan penyajian informasi sebagai dasar pengembangan sistem informasi yang dapat mendukung perencanaan kegiatan panen mangga.

Kata kunci: Mangga, Random Forest, Growing Degree Days, Decision Support System, Prediksi Produksi, Waktu Panen.

Abstract. Production prediction and harvest time estimation are important aspects of supporting mango crop management. Variations in plant characteristics, orchard conditions, and environmental factors create a need for production and harvest time information that can support agricultural planning. This study aims to design an integrated *Decision Support System* (DSS) framework for mango production prediction and harvest time estimation using Random Forest and *Growing Degree Days* (GDD). Random Forest is used to design the production prediction process based on orchard, plant, and environmental characteristics, while GDD is used to estimate plant development based on accumulated temperature. The two approaches are designed as separate components whose outputs are integrated at the system level to provide estimated production and harvest time information. The design includes data collection and preprocessing, predictive model development, GDD calculation, prediction result integration, and the design of a web-based decision support system. Model evaluation is designed using Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and R^2 for production prediction, while harvest time estimation is evaluated based on the difference in days between the estimated and actual harvest dates. The resulting framework provides an integrated design that connects orchard data, environmental data, predictive models, and information presentation as a foundation for developing an information system to support mango harvest planning.

Keywords: Mango, Random Forest, Growing Degree Days, Decision Support System, Production Prediction, Harvest Time.

1. Pendahuluan

Mangga merupakan salah satu komoditas buah yang telah lama dikembangkan di Indonesia dan memiliki nilai ekonomi dalam sektor hortikultura. Pengembangan mangga berkaitan dengan varietas, kondisi agroklimat, wilayah produksi, serta pengelolaan budidaya [1]–[4]. Pada tingkat kebun, informasi mengenai perkiraan produksi dan waktu panen dibutuhkan untuk membantu perencanaan tenaga kerja, penanganan hasil, distribusi, dan pemasaran. Namun, kedua informasi tersebut dipengaruhi oleh variasi tanaman dan lingkungan sehingga memerlukan pendekatan berbasis data.

Perkembangan machine learning memberikan peluang untuk memanfaatkan data agronomi dan lingkungan dalam pemodelan hasil tanaman [6]. Random Forest merupakan metode ensemble yang dapat digunakan untuk regresi dan menggabungkan prediksi dari sejumlah decision tree [7]. Pada komoditas mangga, Fukuda et al. menunjukkan penggunaan Random Forest untuk mengestimasi hasil mangga di bawah perlakuan irigasi yang berbeda, dengan masukan yang mencakup data hujan dan irigasi [8]. Penelitian yang lebih baru oleh [9] juga menggunakan pendekatan machine learning, termasuk Random Forest, dalam pemodelan prediksi hasil mangga. Studi tersebut menunjukkan bahwa struktur data dan kondisi lapangan merupakan bagian penting dalam pemodelan hasil, sehingga model yang dikembangkan untuk suatu lokasi atau varietas tetap memerlukan validasi pada konteks penerapannya.

Selain produksi, waktu panen merupakan informasi yang berkaitan dengan perkembangan fenologi. Growing Degree Days (GDD) menggunakan akumulasi panas untuk merepresentasikan perkembangan tanaman. Penentuan suhu dasar merupakan salah satu komponen penting dalam penerapan GDD karena nilai suhu dasar berkaitan dengan karakteristik perkembangan tanaman [10]. Penerapan GDD juga telah digunakan pada berbagai komoditas pertanian untuk menganalisis kesesuaian kondisi suhu terhadap perkembangan tanaman. Yanti et al. [12] menerapkan pendekatan GDD pada tanaman ubi kayu, sedangkan Yanti et al. [13] menggunakannya dalam pemetaan potensi pengembangan tanaman kedelai. Hal tersebut menunjukkan bahwa GDD dapat digunakan sebagai pendekatan berbasis suhu, tetapi parameter yang digunakan tetap perlu disesuaikan dengan karakteristik tanaman. Penelitian Amaral et al. pada beberapa kultivar mangga menunjukkan bahwa target heat units berkaitan dengan kultivar dan kondisi lokasi, sehingga parameter GDD tidak tepat diperlakukan sebagai angka universal [16]. Penelitian Pramudia et al. pada mangga Arumanis 143 di Situbondo menggunakan suhu dasar 10°C dalam simulasi akumulasi GDD, tetapi hasil tersebut tetap merupakan konteks lokasi dan varietas tertentu [17]. Dengan demikian, nilai suhu dasar dan target GDD pada penelitian ini diposisikan sebagai parameter yang harus dikalibrasi menggunakan data fenologi aktual, bukan sebagai asumsi yang langsung dianggap berlaku pada semua kebun.

Permasalahan metodologis berikutnya adalah hubungan antara unit pengamatan dan waktu. Data pertanian dapat berulang pada pohon, kebun, musim, dan tahun yang sama. Jika observasi yang sangat berkaitan masuk sekaligus ke data training dan testing, skor evaluasi dapat mencerminkan kemiripan unit pengamatan, bukan kemampuan generalisasi ke kondisi yang belum pernah dilihat. Oleh karena itu, rancangan validasi penelitian ini menggunakan grouped holdout atau temporal holdout berdasarkan unit kebun/pohon dan musim/tahun, sesuai struktur data aktual.

Selain itu, prediksi produksi harus memiliki prediction horizon yang jelas. Variabel seperti fase pertumbuhan hanya dapat digunakan jika informasi tersebut memang telah tersedia sebelum waktu prediksi. Penggunaan informasi yang baru diketahui mendekati panen dapat menyebabkan data leakage dan tidak sesuai untuk skenario prediksi dini. Atas dasar itu, penelitian ini mendefinisikan waktu prediksi (t0) dan hanya memperbolehkan fitur yang tersedia sampai t0.

Berdasarkan kondisi tersebut, kontribusi penelitian tidak dinyatakan sebagai pembuktian bahwa RF dan GDD telah mencapai akurasi tertentu di lapangan. Kontribusi penelitian adalah perancangan kerangka DSS yang memisahkan fungsi model prediksi produksi dan estimasi waktu panen, menetapkan protokol validasi yang lebih ketat, serta menghubungkan keluaran model dengan kebutuhan keputusan pengguna. Tujuan penelitian adalah merancang kerangka metodologi dan prototype konseptual yang dapat diuji menggunakan data produksi, fenologi, dan suhu aktual dari kebun mangga.

2. Metode Penelitian

2.1 Posisi dan Tahapan Penelitian

Penelitian ini menggunakan pendekatan perancangan sistem dan metodologi prediktif. Tahapan penelitian meliputi: (1) perumusan kebutuhan keputusan, (2) spesifikasi kebutuhan data, (3) preprocessing dan pemeriksaan potensi leakage, (4) perancangan model RF dan baseline, (5) perancangan serta kalibrasi GDD, (6) perancangan arsitektur DSS, (7) skenario penggunaan end-to-end, dan (8) validasi lapangan yang menjadi tahap pengujian empiris. Dengan posisi ini, angka-angka pada bagian ilustrasi tidak diperlakukan sebagai hasil performa penelitian.

Tabel 1. Tahapan penelitian

Tahap	**Keluaran utama**
1. Kebutuhan keputusan	Kebutuhan informasi produksi dan waktu panen
2. Kebutuhan data	Struktur data kebun, tanaman, lingkungan, fenologi, produksi
3. Preprocessing	Dataset bersih dan pemeriksaan leakage
4. Model produksi	RF + baseline + protokol grouped/temporal holdout
5. Model waktu panen	GDD dengan parameter yang dikalibrasi
6. Integrasi DSS	Arsitektur data–model–backend–antarmuka
7. Skenario penggunaan	Alur informasi dari input sampai keputusan
8. Validasi empiris	Evaluasi model dan sistem menggunakan data aktual

2.2 Kebutuhan Data dan Status Ketersediaan

Penelitian ini menggunakan struktur data yang menjadi spesifikasi kebutuhan sebelum model dilatih dan performa dievaluasi. Data produksi aktual, data fenologi, dan data lingkungan diperlukan untuk membangun serta menguji model secara empiris.

Tabel 2. Spesifikasi data penelitian

Kelompok data	**Variabel yang diperlukan**	**Kegunaan**
Kebun	ID kebun, lokasi, luas, identitas unit pengamatan	Grouping dan konteks kebun
Tanaman	ID pohon, varietas, umur, jumlah pohon	Fitur RF dan grouping
Lingkungan	Tmin, Tmax, kelembapan, curah hujan	Fitur RF dan perhitungan GDD
Fenologi	Tanggal pembungaan, fase, tanggal panen aktual	Kalibrasi dan evaluasi GDD
Produksi	Produksi aktual per unit dan musim	Target RF
Waktu	Tanggal pengamatan, musim, tahun	Temporal split dan horizon

Produksi aktual dan tanggal panen aktual merupakan data kunci dalam evaluasi empiris. Pada tahap perancangan ini, angka dataset dan nilai evaluasi yang belum berasal dari pengukuran lapangan tidak digunakan sebagai hasil performa model. Dengan demikian, hasil yang disajikan berfokus pada rancangan metodologi, validasi, integrasi sistem, dan skenario penggunaan.

2.3 Prediction Horizon dan Pencegahan Data Leakage

Waktu prediksi didefinisikan sebagai t0, yaitu tanggal ketika sistem menghasilkan estimasi produksi. Dalam rancangan prototype, horizon produksi ditetapkan pada H−30 sebagai skenario awal, yaitu estimasi produksi dibuat sekitar 30 hari sebelum tanggal target panen. Nilai H−30 merupakan keputusan desain yang belum divalidasi dan dapat diubah setelah karakteristik data lapangan diketahui.

Aturan utama adalah seluruh fitur yang digunakan pada t0 harus sudah tersedia sebelum atau pada t0. Variabel fase pertumbuhan tidak otomatis dimasukkan sebagai fitur; variabel tersebut hanya boleh digunakan apabila fase telah dicatat secara operasional sebelum t0. Jika fase baru diketahui setelah pengamatan yang dekat dengan panen, fitur tersebut dikeluarkan dari model prediksi dini. Aturan yang sama berlaku untuk agregasi suhu, curah hujan, dan informasi lain: hanya data sampai t0 yang boleh digunakan.

Tabel 3. Aturan prediction horizon dan leakage

Komponen	**Aturan**
Waktu prediksi	t0; prototype menggunakan skenario H−30
Target produksi	Produksi aktual pada periode panen
Fitur historis	Hanya informasi yang tersedia ≤ t0
Fase pertumbuhan	Digunakan hanya jika telah diketahui ≤ t0
Data setelah t0	Tidak boleh digunakan sebagai fitur
Evaluasi	Test set berasal dari kebun/musim/tahun yang tidak digunakan untuk training

2.4 Preprocessing dan Skema Pembagian Data

Preprocessing meliputi pemeriksaan missing value, pemeriksaan nilai tidak valid, penyesuaian format tanggal dan numerik, konsistensi satuan, serta pemeriksaan duplikasi berdasarkan ID pohon/kebun dan musim. Setelah itu dilakukan audit temporal untuk memastikan bahwa fitur tidak mengambil informasi dari periode setelah t0.

Pembagian acak 80:20 tidak digunakan sebagai skema utama karena belum menjamin pemisahan unit pertanian yang berkorelasi. Skema utama yang dirancang adalah grouped holdout berdasarkan kebun atau pohon. Apabila tersedia data beberapa tahun atau musim, temporal holdout berdasarkan tahun/musim digunakan sebagai validasi tambahan. Dengan demikian, model diuji pada kelompok atau periode yang tidak pernah dilihat ketika training.

Tabel 4. Skema validasi yang direvisi

Skema	**Training**	**Testing**	**Tujuan**
Grouped holdout	Kebun/pohon pada kelompok A	Kebun/pohon pada kelompok B	Menguji generalisasi antar-unit
Temporal holdout	Musim/tahun lebih awal	Musim/tahun lebih baru	Menguji generalisasi antar-waktu
Random split	Tidak menjadi skema utama	—	Hanya untuk analisis tambahan jika diperlukan

2.5 Prediksi Produksi dengan Random Forest

Random Forest Regression digunakan sebagai model utama untuk memetakan fitur kebun, tanaman, dan lingkungan menjadi estimasi produksi. Secara umum, prediksi akhir merupakan rata-rata prediksi dari B decision tree:

$$\hat{y}(x) = (1/B) \sum_{\beta=1}^{B} T_{\beta}(x)$$

Parameter model harus dilaporkan secara eksplisit pada saat eksperimen. Rancangan awal menggunakan random_state = 42 untuk reprodusibilitas. Nilai n_estimators, max_depth, min_samples_split, dan min_samples_leaf ditentukan melalui tuning pada data training/validation dan tidak boleh ditentukan berdasarkan data testing.

Tabel 5. Konfigurasi dan pembanding RF

Komponen	**Rancangan**
Model utama	Random Forest Regressor
Baseline 1	Prediksi rata-rata produksi pada data training
Baseline 2	Decision Tree Regressor [14]
random_state	42 untuk reprodusibilitas
Parameter yang dilaporkan	n_estimators, max_depth, min_samples_split, min_samples_leaf
Tuning	Dilakukan hanya pada training/validation set
Data test	Disimpan sampai tahap evaluasi akhir

Baseline digunakan untuk mengetahui apakah Random Forest memberikan peningkatan dibandingkan pendekatan sederhana. Perbandingan dilakukan menggunakan MAE, RMSE, dan R^2 pada test set yang sama. Tidak ada nilai performa yang dilaporkan pada naskah ini karena data aktual belum disediakan.

2.6 Perhitungan dan Kalibrasi Growing Degree Days

GDD digunakan sebagai komponen estimasi waktu perkembangan menuju panen. Persamaan sederhana yang digunakan pada tahap awal adalah:

$$GDD_{harian} = \max(0, ((T_{max} + T_{min})/2) - T_{base})$$

$$GDD_{kumulatif} = \Sigma\, GDD_{harian}$$

Namun, Tbase dan target GDD tidak ditetapkan sebagai angka final berdasarkan contoh simulasi. Literatur menunjukkan bahwa kebutuhan heat units dapat berbeda menurut kultivar dan lokasi. Amaral et al. melaporkan target heat units yang berbeda pada kultivar mangga Australia dan menggunakan batas suhu rendah serta batas suhu tinggi [16]. Pramudia et al. melaporkan penggunaan suhu dasar 10°C pada Arumanis 143 di Situbondo [17]. Kedua temuan tersebut digunakan sebagai dasar metodologis bahwa parameter GDD harus dikalibrasi, bukan dipindahkan secara langsung ke lokasi penelitian.

Kalibrasi GDD dirancang dengan mencari kombinasi parameter yang menghasilkan kesalahan tanggal panen terendah pada data training/kalibrasi. Parameter yang diuji mencakup Tbase dan target GDD; jika data dan literatur lokal mendukung, batas suhu atas juga dapat dipertimbangkan. Evaluasi akhir dilakukan pada data musim/kebun yang tidak digunakan saat kalibrasi.

Tabel 6. Rancangan kalibrasi GDD

Parameter	**Status pada penelitian**
Tbase	Parameter kalibrasi; 10°C hanya kandidat awal jika sesuai varietas/lokasi
Target GDD	Harus diestimasi dari tanggal fenologi dan panen aktual
Tanggal awal akumulasi	Didefinisikan berdasarkan fase fenologi yang dapat diamati
Tanggal panen aktual	Ground truth untuk evaluasi
Metrik	MAE tanggal (hari), median absolute error, dan proporsi prediksi dalam toleransi hari yang ditetapkan

2.7 Integrasi sebagai Kerangka DSS

RF dan GDD tidak membentuk satu persamaan hybrid. RF menghasilkan estimasi kuantitas produksi, sedangkan GDD menghasilkan estimasi waktu perkembangan menuju panen. Keduanya diintegrasikan pada tingkat DSS melalui data layer, model layer, service/backend layer, dan presentation layer. Istilah yang digunakan dalam naskah ini adalah kerangka DSS terintegrasi agar kontribusi tidak melebihi implementasi yang tersedia.

Tabel 7. Arsitektur konseptual DSS

Lapisan	**Fungsi**
Data layer	Menyimpan data kebun, tanaman, lingkungan, fenologi, produksi, dan pengguna
Model layer	Menjalankan RF untuk produksi dan GDD untuk waktu panen
Service/backend	Mengatur input, validasi, pemanggilan model, dan penyimpanan keluaran
Presentation layer	Dashboard, estimasi produksi, kalender/perkiraan panen, notifikasi
Decision layer	Mengubah keluaran menjadi informasi untuk perencanaan panen, tenaga kerja, distribusi, dan pemasaran

Alur data yang dirancang adalah: data kebun dan lingkungan → preprocessing → RF/GDD → hasil prediksi → penyimpanan hasil → dashboard → keputusan pengguna. Pada tahap implementasi, arsitektur ini perlu dibuktikan dengan diagram komponen, struktur basis data, API/backend, antarmuka, dan hasil pengujian aplikasi.

2.8 Evaluasi Model dan Sistem

Evaluasi RF menggunakan MAE, RMSE, dan R^2. MAE menyatakan rata-rata selisih absolut antara prediksi dan nilai aktual, RMSE memberikan penalti lebih besar terhadap kesalahan yang besar, sedangkan R^2 menggambarkan proporsi variasi target yang dapat dijelaskan model.

$$MAE = (1/n) \sum |y_i - \hat{y}_i|$$

$$RMSE = \sqrt{[(1/n) \sum (y_i - \hat{y}_i)^2]}$$

$$R^2 = 1 - [\sum (y_i - \hat{y}_i)^2 / \sum (y_i - \bar{y})^2]$$

Evaluasi GDD menggunakan selisih absolut tanggal prediksi terhadap tanggal panen aktual dalam satuan hari. Selain itu, sistem perlu dievaluasi secara fungsional untuk memastikan proses input, pemanggilan model, penyimpanan, dan tampilan berjalan sesuai kebutuhan. Karena target jurnal berada pada bidang IT/IS, evaluasi usability dapat dilakukan menggunakan SUS atau instrumen usability yang sesuai setelah prototype dapat digunakan oleh petani atau penyuluh.

Tabel 8. Rencana evaluasi

Jenis evaluasi	**Indikator**	**Data yang dibutuhkan**
RF	MAE, RMSE, R^2	Produksi aktual
Baseline	MAE, RMSE, R^2	Produksi aktual
GDD	Kesalahan tanggal dalam hari	Tanggal pembungaan dan panen aktual
Fungsional	Skenario berhasil/gagal	Prototype aplikasi
Usability	SUS atau instrumen sejenis	Pengguna sasaran
Decision support	Kesesuaian informasi dengan keputusan	Skenario penggunaan dan umpan balik pengguna

3. Hasil Perancangan dan Pembahasan

3.1 Status Hasil dan Batasan

Hasil penelitian pada tahap ini berupa spesifikasi metodologi dan rancangan DSS, sedangkan performa model lapangan belum dilaporkan karena evaluasi empiris memerlukan data produksi dan tanggal panen aktual. Contoh angka yang bersifat simulatif tidak digunakan untuk menyatakan akurasi model. Dengan demikian, hasil yang disajikan tidak menimbulkan interpretasi bahwa model telah terbukti pada kondisi lapangan.

Dengan posisi tersebut, penelitian dapat dibaca sebagai tahap desain dan persiapan validasi. Hasil empiris baru dapat dilaporkan setelah data produksi dan tanggal panen aktual dikumpulkan, model dilatih, parameter dituning tanpa menggunakan test set, dan pengujian dilakukan pada kebun/musim/tahun yang belum dilihat model.

3.2 Rancangan Validasi Random Forest

Rancangan validasi membandingkan RF dengan baseline rata-rata dan Decision Tree Regressor. Pemisahan berdasarkan kebun, pohon, musim, atau tahun dipilih sesuai struktur dataset aktual. Contohnya, apabila terdapat data beberapa kebun selama tiga musim, satu atau beberapa kelompok kebun/musim dapat disimpan sebagai test set dan tidak digunakan dalam training. Jika fokus utama adalah generalisasi antar-tahun, musim/tahun terbaru dapat digunakan sebagai temporal holdout.

Konfigurasi model dan hasil tuning harus dicatat agar eksperimen dapat direplikasi. Selain skor MAE, RMSE, dan R^2, pelaporan sebaiknya mencakup jumlah observasi per kelompok, rentang target produksi, distribusi train-test, fitur yang tersedia pada t0, serta parameter akhir model. Dengan cara ini, pembaca dapat membedakan performa model dari efek kemiripan data.

Pada tahap ini tidak diberikan nilai akurasi RF karena tidak ada data aktual yang dapat digunakan untuk menghitungnya secara sah. Nilai performa pada versi sebelumnya dihapus dari kesimpulan dan pembahasan.

3.3 Rancangan Kalibrasi GDD

GDD tidak diperlakukan sebagai model yang langsung menggunakan Tbase=10°C dan target 1.200 GDD. Nilai tersebut pada naskah sebelumnya hanya merupakan ilustrasi. Literatur menunjukkan variasi parameter antar-kultivar dan lokasi [16], sedangkan penelitian Arumanis 143 di Situbondo memberikan contoh penggunaan Tbase 10°C pada konteks tertentu [17]. Oleh karena itu, penelitian ini menggunakan pendekatan kalibrasi lokal: tanggal awal perkembangan ditetapkan dari catatan fenologi, GDD diakumulasikan berdasarkan data suhu, dan target heat unit dicari dari hubungan antara akumulasi suhu dan tanggal panen aktual.

Dengan prosedur tersebut, GDD menjadi komponen yang dapat diuji dan disesuaikan. Hasil evaluasi nantinya harus dilaporkan sebagai kesalahan tanggal dalam hari pada data uji, bukan sebagai contoh tanggal yang dibuat secara simulatif.

3.4 Integrasi RF dan GDD pada DSS

Integrasi pada penelitian ini bersifat integrasi tingkat sistem. RF menghasilkan satu keluaran kuantitatif, yaitu estimasi produksi, sedangkan GDD menghasilkan keluaran temporal, yaitu perkiraan waktu panen. Kedua keluaran tidak saling menjadi input matematis. Informasi tersebut disatukan oleh DSS untuk memberikan satu tampilan keputusan.

Tabel 9. Skenario integrasi end-to-end

Input pengguna/sistem	**Proses**	**Output**	**Pemanfaatan keputusan**

Data kebun, tanaman, lingkungan	RF	Estimasi produksi	Perencanaan kapasitas panen, pengemasan, distribusi
Tmin, Tmax, fenologi	GDD terkalibrasi	Perkiraan waktu panen	Penjadwalan tenaga kerja dan panen
RF + GDD + konteks kebun	DSS	Ringkasan kondisi dan kalender	Perencanaan operasional dan pemasaran

3.5 Skenario Penggunaan End-to-End

Skenario penggunaan dimulai ketika petani atau operator memasukkan atau memperbarui data kebun dan tanaman. Sistem mengambil data lingkungan yang tersedia sampai t0. RF kemudian menghitung estimasi produksi berdasarkan fitur yang sah pada horizon prediksi. Secara paralel, GDD menghitung akumulasi panas dari titik awal fenologi yang telah ditetapkan dan menggunakan parameter hasil kalibrasi untuk memperkirakan waktu perkembangan menuju panen.

Keluaran tersebut ditampilkan dalam dashboard sebagai dua informasi yang berbeda: jumlah produksi yang diperkirakan dan rentang/tanggal perkiraan panen. Dari sisi keputusan, estimasi produksi dapat digunakan untuk memperkirakan kebutuhan tenaga panen, wadah, kapasitas pengemasan, dan distribusi. Estimasi waktu panen dapat digunakan untuk menyusun jadwal kerja dan komunikasi pemasaran. Penggunaan tersebut merupakan skenario manfaat yang perlu diuji kepada pengguna; naskah ini tidak mengklaim bahwa sistem telah terbukti meningkatkan kualitas keputusan.

Untuk memenuhi kebutuhan publikasi bidang IT/IS, implementasi berikutnya perlu menampilkan bukti prototype berupa arsitektur aplikasi, rancangan database, alur API/backend, halaman antarmuka, dan hasil pengujian fungsional. Evaluasi usability dengan petani atau penyuluh juga diperlukan agar keluaran model tidak hanya benar secara komputasional, tetapi dapat dipahami dan digunakan.

3.6 Pembahasan

Berdasarkan tahapan perancangan yang telah dilakukan, Random Forest dan Growing Degree Days (GDD) memiliki fungsi yang berbeda dalam kerangka sistem prediksi tanaman mangga. Random Forest dirancang untuk menghasilkan estimasi jumlah produksi berdasarkan hubungan antara karakteristik kebun, karakteristik tanaman, dan kondisi lingkungan dengan data produksi aktual. Sementara itu, GDD digunakan untuk memperkirakan perkembangan tanaman berdasarkan akumulasi suhu hingga mencapai nilai heat unit tertentu yang nantinya dikalibrasi menggunakan data fenologi aktual.

Pada tahap penelitian ini, nilai-nilai hasil prediksi yang ditampilkan masih digunakan sebagai ilustrasi metodologis untuk menunjukkan bagaimana proses perhitungan dan integrasi dilakukan. Oleh karena itu, nilai MAE, RMSE, R^2, target GDD, serta estimasi tanggal panen yang digunakan dalam contoh belum dapat dianggap sebagai performa empiris model. Performa Random Forest baru dapat dinyatakan setelah model dilatih dan diuji menggunakan data produksi mangga aktual, sedangkan akurasi GDD baru dapat ditentukan setelah parameter suhu dasar dan target akumulasi panas dibandingkan dengan data fenologi serta tanggal panen aktual.

Random Forest pada penelitian ini ditempatkan sebagai komponen prediksi produksi. Input yang dirancang meliputi informasi kebun, tanaman, dan lingkungan, sedangkan produksi aktual menjadi target prediksi. Dalam implementasi selanjutnya, proses evaluasi tidak hanya menggunakan pembagian data secara acak, tetapi perlu mempertimbangkan struktur data pertanian yang memiliki keterkaitan berdasarkan kebun, pohon, musim, dan tahun. Apabila data dari kebun atau pohon yang sama digunakan pada data training dan testing, model berpotensi memperoleh pola yang sangat mirip antara kedua kelompok data sehingga hasil evaluasi dapat terlihat lebih baik daripada kemampuan generalisasi sebenarnya.

Oleh karena itu, pengujian Random Forest pada tahap implementasi dirancang menggunakan grouped split atau temporal holdout. Grouped split dapat digunakan apabila data dikelompokkan berdasarkan kebun atau pohon, sehingga data dari kelompok yang sama tidak muncul pada data training

dan testing secara bersamaan. Sementara itu, temporal holdout dapat digunakan apabila tersedia data dari beberapa musim atau tahun, dengan periode yang lebih awal digunakan untuk pembentukan model dan periode berikutnya digunakan sebagai pengujian. Strategi tersebut diharapkan dapat memberikan gambaran yang lebih realistis mengenai kemampuan model ketika digunakan pada kondisi kebun atau periode yang belum pernah digunakan selama proses pelatihan.

Selain Random Forest, penggunaan baseline juga diperlukan agar kemampuan model dapat dibandingkan dengan pendekatan yang lebih sederhana. Model Random Forest nantinya dapat dibandingkan dengan baseline produksi rata-rata dan algoritma regresi pohon keputusan. Perbandingan tersebut diperlukan untuk mengetahui apakah penggunaan ensemble Random Forest memberikan peningkatan performa dibandingkan pendekatan sederhana. Konfigurasi model seperti jumlah pohon (n_estimators), kedalaman pohon (max_depth), batas minimum pembentukan percabangan, serta random_state juga perlu dicatat dan dilaporkan sehingga proses eksperimen dapat direproduksi.

Aspek lain yang perlu diperhatikan adalah waktu ketika prediksi dilakukan. Sistem prediksi produksi tidak hanya membutuhkan model dengan performa baik, tetapi juga harus memastikan bahwa seluruh variabel yang digunakan memang telah tersedia pada saat prediksi dilakukan. Dalam rancangan penelitian ini, prediction horizon ditentukan sebagai bagian dari desain sistem. Sebagai contoh, prediksi dapat dilakukan pada periode tertentu sebelum panen, misalnya sekitar 30 hari sebelum waktu panen yang ditargetkan. Dengan demikian, sistem hanya menggunakan informasi yang sudah tersedia sampai waktu tersebut dan tidak menggunakan data yang baru diketahui setelah periode prediksi. Pendekatan ini penting untuk menghindari data leakage dan memastikan bahwa prediksi yang dihasilkan benar-benar dapat digunakan sebagai prediksi dini.

Berbeda dengan Random Forest, GDD tidak dirancang sebagai model machine learning yang menghasilkan prediksi produksi. GDD digunakan sebagai pendekatan berbasis fenologi untuk memperkirakan perkembangan tanaman berdasarkan akumulasi suhu. Perhitungan GDD memerlukan suhu minimum, suhu maksimum, suhu dasar, serta target akumulasi panas. Nilai suhu dasar dan target GDD tidak dapat langsung dianggap sebagai nilai universal karena perkembangan tanaman dapat dipengaruhi oleh varietas, lokasi, kondisi lingkungan, dan fase pertumbuhan.

Dengan demikian, nilai Tbase = 10°C dan target 1.200 GDD yang digunakan pada contoh dalam penelitian ini hanya berfungsi sebagai ilustrasi proses perhitungan. Pada tahap pengembangan berikutnya, parameter tersebut perlu dikalibrasi menggunakan data aktual. Data yang diperlukan meliputi tanggal pembungaan atau awal fase fenologi, data suhu harian, fase perkembangan buah, dan tanggal panen aktual. Dari data tersebut dapat dilakukan pengujian beberapa nilai suhu dasar dan target GDD untuk mengetahui parameter yang paling sesuai dengan varietas dan kondisi kebun penelitian.

Penggunaan Random Forest dan GDD dalam penelitian ini juga tidak dimaksudkan sebagai pembentukan satu algoritma hybrid secara matematis. Random Forest dan GDD tetap merupakan dua komponen dengan proses perhitungan yang berbeda. Random Forest menghasilkan estimasi jumlah produksi, sedangkan GDD menghasilkan estimasi waktu perkembangan menuju panen. Kedua hasil tersebut kemudian diintegrasikan pada tingkat sistem untuk membentuk suatu kerangka Decision Support System (DSS) terintegrasi.

Integrasi tersebut memberikan alur informasi yang lebih lengkap bagi pengelolaan kebun. Sebagai contoh, hasil Random Forest dapat memberikan informasi mengenai perkiraan jumlah produksi yang dapat digunakan untuk membantu perencanaan tenaga kerja, kapasitas pengemasan, dan kebutuhan distribusi. Pada sisi lain, hasil GDD dapat memberikan informasi mengenai perkiraan waktu panen yang dapat digunakan untuk membantu penjadwalan kegiatan panen. Ketika kedua informasi tersebut ditampilkan secara bersamaan, pengguna dapat memperoleh gambaran mengenai berapa banyak produksi yang diperkirakan dan kapan produksi tersebut diperkirakan dapat dipanen.

Dari perspektif Sistem Informasi, hasil penelitian tidak berhenti pada pengembangan model prediksi. Kedua komponen tersebut dirancang untuk menjadi bagian dari sistem informasi yang menghubungkan data kebun, data tanaman, data lingkungan, proses prediksi, dan informasi pendukung keputusan. Data yang dimasukkan oleh petani atau pengelola kebun akan diproses oleh sistem, kemudian diteruskan ke

komponen prediksi. Hasil prediksi selanjutnya dikembalikan ke aplikasi dalam bentuk informasi yang lebih mudah dipahami oleh pengguna.

Kerangka sistem yang dirancang dapat digambarkan melalui alur data kebun dan tanaman → data lingkungan → preprocessing → model Random Forest/GDD → hasil prediksi → dashboard DSS → informasi pendukung keputusan. Dengan alur tersebut, sistem tidak hanya berfungsi sebagai kalkulator prediksi, tetapi menjadi media yang menghubungkan data lapangan dengan kebutuhan pengambilan keputusan pengguna.

Meskipun demikian, rancangan tersebut masih memerlukan implementasi dan pengujian lebih lanjut. Prototype sistem perlu diuji dari sisi fungsional untuk memastikan setiap proses input, pengolahan data, pemanggilan model, penyimpanan hasil, dan penyajian informasi berjalan sesuai rancangan. Selain pengujian fungsional, usability sistem juga perlu dievaluasi menggunakan instrumen seperti System Usability Scale (SUS) atau metode evaluasi usability lainnya. Evaluasi tersebut diperlukan karena keberhasilan sistem tidak hanya ditentukan oleh kemampuan model menghasilkan prediksi, tetapi juga oleh kemampuan pengguna memahami dan memanfaatkan informasi yang diberikan.

Dengan demikian, kontribusi penelitian ini berada pada perancangan kerangka integrasi antara prediksi produksi berbasis Random Forest, estimasi waktu perkembangan berbasis GDD, dan sistem pendukung keputusan berbasis web. Validasi empiris model dan pengujian prototype menjadi tahap lanjutan yang diperlukan untuk memastikan bahwa kerangka yang dirancang dapat diterapkan pada kondisi kebun mangga sebenarnya

4. Kesimpulan dan Saran

4.1 Kesimpulan

Penelitian ini merancang kerangka Decision Support System terintegrasi yang menggunakan Random Forest untuk estimasi produksi mangga dan Growing Degree Days untuk estimasi waktu perkembangan menuju panen. Kedua pendekatan diposisikan sebagai komponen terpisah yang diintegrasikan pada tingkat sistem, bukan sebagai satu algoritma hybrid. Revisi metodologi menetapkan bahwa klaim performa hanya dapat dibuat setelah tersedia data produksi dan tanggal panen aktual.

Kerangka validasi Random Forest mencakup baseline sederhana dan Decision Tree Regressor, pelaporan konfigurasi model dan tuning, serta grouped atau temporal holdout untuk mengurangi risiko kebocoran informasi akibat keterkaitan pohon, kebun, musim, dan tahun. Prediction horizon didefinisikan melalui waktu prediksi t0 dan skenario awal H−30, dengan aturan bahwa hanya informasi yang tersedia sampai t0 yang dapat digunakan.

Pada komponen GDD, suhu dasar dan target akumulasi panas tidak diperlakukan sebagai nilai final. Parameter tersebut harus dikalibrasi terhadap varietas, lokasi, musim, dan fenologi aktual. Dengan demikian, hasil penelitian pada tahap ini adalah kerangka metodologi, spesifikasi validasi, arsitektur DSS, dan skenario penggunaan yang siap diuji dengan data lapangan.

4.2 Saran

Penelitian selanjutnya disarankan menggunakan data produksi, suhu, dan tanggal panen aktual dari beberapa kebun serta varietas mangga. Model Random Forest selanjutnya perlu divalidasi menggunakan pembagian data berdasarkan kebun, musim, atau waktu serta dibandingkan dengan metode baseline. Parameter GDD, termasuk suhu dasar dan target akumulasi panas, perlu dikalibrasi berdasarkan data fenologi aktual sesuai varietas dan lokasi penelitian.

Tahap berikutnya adalah mengimplementasikan rancangan ke dalam sistem informasi berbasis web yang mengintegrasikan data kebun, prediksi produksi, dan estimasi waktu panen. Sistem kemudian dapat dilengkapi dengan dashboard monitoring dan informasi pendukung keputusan untuk membantu perencanaan kegiatan panen.

5. Referensi

[1] W. Broto, N. Primawati, L. H. Sule, Y. Krisnawati, O. Kusmawardani, and Purwati, Budi Daya dan Pascapanen Mangga. Jakarta: Pusat Perpustakaan Pertanian dan Komunikasi Penelitian, 1994.

[2] S. Edhy, Profil Mangga. Jakarta: Direktorat Buah dan Florikultura, Kementerian Pertanian Republik Indonesia, 2017.

[3] Kementerian Pertanian Republik Indonesia, Statistik Pertanian 2021. Jakarta: Pusat Data dan Sistem Informasi Pertanian, Kementerian Pertanian Republik Indonesia, 2021.

[4] Direktorat Buah dan Florikultura, Buku Lapang Budidaya Mangga. Jakarta: Direktorat Jenderal Hortikultura, 2021.

[5] R. Fitranto, N. D. Wahyono, and Y. Wibisono, "Strategi pengembangan pemasaran buah mangga Arumanis 143 PT. Trigatra Rajasa Situbondo Jawa Timur," Jurnal Agribisnis Indonesia, vol. 8, no. 1, pp. 58–68, 2020, doi: 10.29244/jai.2020.8.1.58-68.

[6] D. Elavarasan, D. R. Vincent, V. Sharma, A. Y. Zomaya, and K. Srinivasan, "Forecasting yield by integrating agrarian factors and machine learning models: A survey," Computers and Electronics in Agriculture, vol. 155, pp. 257–282, 2018, doi: 10.1016/j.compag.2018.10.024.

[7] L. Breiman, "Random forests," Machine Learning, vol. 45, pp. 5–32, 2001, doi: 10.1023/A:1010933404324.

[8] S. Fukuda, W. Spreer, E. Yasunaga, K. Yuge, V. Sardsud, and J. Müller, "Random forests modelling for the estimation of mango (Mangifera indica L. cv. Chok Anan) fruit yields under different irrigation regimes," Agricultural Water Management, vol. 116, pp. 142–150, 2013, doi: 10.1016/j.agwat.2012.07.003.

[9] L. Kouadio, B. Kulanthaivel, T. Mai, T. Nguyen-Huy, Q. J. Wang, V. M. Byrareddy, S. Mushtaq, A. Senthil, N. K. Newlands, and V. Geethalakshmi, "Quantifying error in fine-scale crop yield forecasts to guide data and algorithm improvements: Case study of mango in Tamil Nadu, India," Computers and Electronics in Agriculture, vol. 236, p. 110450, 2025, doi: 10.1016/j.compag.2025.110450.

[10] C. Y. Arnold, "The determination and significance of the base temperature in a linear heat unit system," Proceedings of the American Society for Horticultural Science, vol. 74, pp. 430–445, 1959.

[11] N. Nur, F. Wajidi, Sulfayanti, and Wildayani, "Implementasi Algoritma Random Forest Regression untuk Memprediksi Hasil Panen Padi di Desa Minanga," Jurnal Komputer Terapan, vol. 9, no. 1, pp. 58–64, 2023, doi: 10.35143/jkt.v9i1.5917.

[12] D. Yanti, N. Shalihah, I. Safitri, and R. Rusnam, "Analisis Kesesuaian Lahan Tanaman Ubi Kayu (Manihot esculenta C.) dengan Metode Growing Degree Days," Jurnal Teknologi Pertanian Andalas, vol. 27, no. 2, pp. 185–209, 2023, doi: 10.25077/jtpa.27.2.185-209.

[13] D. Yanti, O. C. Chatib, I. Safitri, and N. N. Hasanah, "Mapping the Potential Development of Soybenas Crops Based on Growing Degree Days," Jurnal Keteknikan Pertanian, vol. 12, no. 1, pp. 57–76, 2024, doi: 10.19028/jtep.012.1.57-76.

[14] R. Faizal, A. Abdullah, and M. W. Pangestika, "Comparison of Random Forest Regressor and Decision Tree Regressor for Crop Yield Prediction," Jurnal CoSciTech (Computer Science and Information Technology), vol. 6, no. 2, 2025, doi: 10.37859/coscitech.v6i2.9966.

[15] F. D. Choirunisa, M. W. Pangestika, and Sucipto, "Prediksi Panen Kelapa Sawit Menggunakan Metode Random Forest Regression," NUSANTARA: Jurnal Ilmu Pengetahuan Sosial, vol. 13, no. 7, 2026.

[16] M. H. Amaral, C. McConchie, G. Dickinson, and K. B. Walsh, "Growing Degree Day Targets for Fruit Development of Australian Mango Cultivars," Horticulturae, vol. 9, no. 4, p. 489, 2023, doi: 10.3390/horticulturae9040489.

[17] A. Pramudia, Affandi, and N. Marta, "Application of the concept of growing degree days to determine the optimal mango harvest schedule in Situbondo," Buitenzorg: Journal of Tropical Science, vol. 1, no. 1, pp. 10–17, 2024, doi: 10.70158/buitenzorg.v1i1.2.