

Prediksi Hasil Panen Kakao Di Desa Minanga Menggunakan Algoritma *Random Forest Regression*

Andrian¹, Arnita Irianti*², Nahya Nur³, Adi Heri⁴

^{1,2,3}Program Studi Teknik Informatika, Universitas Sulawesi Barat

E-mail: ¹riansam12345@gmail.com, *²arnitairianti@unsulbar.ac.id, ³nahya.nur@unsulbar.ac.id

⁴adiheri@unsulbar.ac.id

Abstrak

Hasil produksi tanaman kakao di Desa Minanga sering mengalami fluktuasi, baik dalam bentuk penurunan maupun peningkatan yang tidak menentu disetiap musimnya. Kondisi ini berdampak pada ketidakpastian pendapatan, sehingga para petani kesulitan dalam merencanakan keuangan mereka dengan baik. Penelitian ini menggunakan algoritma *random forest regression* dengan tujuan untuk membantu dalam memprediksi hasil panen kakao pada petani Desa Minanga, dengan menggunakan kriteria luas lahan, jumlah tanaman kakao, jenis bibit, jenis pupuk, hama dan penyakit, penanggulangan hama dan penyakit, tingkat curah hujan, tenaga kerja, dan hasil. Data yang digunakan dalam penelitian ini merupakan data hasil panen kakao dari tahun 2019 hingga 2023, dengan total 2980 data yang akan diolah. Berdasarkan hasil pengujian jumlah pohon yang paling optimal adalah 100 pohon, dengan tingkat akurasi sebesar 98.95% dengan nilai *Mean Absolute Percentage Error* (MAPE) diperoleh adalah 1.04%, *Root Mean Square Error* (RMSE) yaitu 6.26 dan Koefisien Determinasi (R²) sebesar 0.99 pada rasio data 80:20. Variabel yang memiliki nilai *importance* paling tinggi yaitu variabel hama dan penyakit.

Kata kunci— Hasil Panen Kakao, Prediksi, *Random Forest Regression*

Abstract

Cocoa crop production in Minanga Village often fluctuates, either in the form of a decrease or an increase that is uncertain in each season. This condition has an impact on income uncertainty, so that farmers have difficulty in planning their finances properly. This study uses a random forest regression algorithm with the aim of helping to predict cocoa harvests for farmers in Minanga Village, using the criteria of land area, number of cocoa plants, type of seedlings, type of fertilizer, pests and diseases, pest and disease control, rainfall levels, labor, and results. The data used in this study are cocoa harvest data from 2019 to 2023, with a total of 2980 data to be processed. Based on the test results, the most optimal number of trees is 100 trees, with an accuracy level of 98.95% with a Mean Absolute Percentage Error (MAPE) value obtained is 1.04%, Root Mean Square Error (RMSE) is 6.26 and a Determination Coefficient (R²) of 0.99 at a data ratio of 80:20. The variables that have the highest importance value are pests and diseases.

Keywords— Cocoa Harvest Results, Prediction, *Random Forest Regression*

1. PENDAHULUAN

Pertanian merupakan fondasi utama bagi perekonomian sebuah negara. Dengan pertumbuhan sektor pertanian yang kuat, maka akan memberikan dampak positif bagi stabilitas ekonomi secara keseluruhan [1]. Indonesia sebagai negara agraris banyak mengandalkan kebutuhan hidupnya dari hasil pertanian, karena itu sektor pertanian menjadi salah satu sektor yang terus diprioritaskan untuk menunjang laju pertumbuhan ekonomi nasional [2]. Sumber daya alam hayati yang dimiliki oleh Indonesia sangat beragam, salah satunya adalah jenis tumbuhan atau tanaman yang dapat dijadikan sebagai sumber bahan pangan salah satunya adalah kakao yang masih menjadi komoditas andalan sebagian besar masyarakat Indonesia.

Kakao (*Theobroma Cacao*) merupakan satu tanaman yang memberikan nilai ekonomis yang tinggi. Di Indonesia, pertumbuhan devisa juga dipicu oleh peningkatan produksi kakao. Beragam produk dapat dibuat melalui proses pengolahan dari biji hingga lemaknya, seperti cokelat yang bahan bakunya berasal dari kakao. Tidak hanya memiliki nilai jual tinggi kakao juga memiliki manfaat yang baik dalam bidang kesehatan karna buah kakao memiliki atau mengandung antioksidan yaitu *fenol* dan *flavonoid* yang berkhasiat meningkatkan kekebalan tubuh [3]. Kakao, minyak dan gas termasuk dalam komoditas ekspor Indonesia yang berkontribusi terhadap penerimaan devisa negara. Indonesia menduduki peringkat ketiga sebagai produsen dan pengekspor kakao terbesar di dunia, setelah Ghana dan Pantai Gading [4].

Kakao sendiri merupakan tanaman yang banyak dibudidayakan oleh masyarakat Indonesia sebagai komoditas perkebunan yang memiliki prospek menjanjikan. Salah satu daerah yang banyak membudidayakan tanaman kakao adalah Desa Minanga yang berada di Kabupaten Mamasa. Namun berdasarkan hasil wawancara yang dilakukan oleh peneliti dengan beberapa petani yang ada di Desa Minanga, disebutkan bahwa hasil panen mereka mengalami fluktuasi dari musim ke musim, baik berupa peningkatan maupun penurunan yang tidak stabil. Masalah utama yang dihadapi para petani adalah tidak adanya alat bantu atau sistem yang mampu memprediksi hasil panen mereka secara akurat, sehingga menyulitkan dalam hal pengelolaan biaya perawatan tanaman dan perencanaan keuangan jangka panjang. Ketidakpastian ini berdampak pada pendapatan petani dan menurunkan efektivitas pengambilan keputusan dalam usaha tani mereka. Oleh karena itu, dibutuhkan strategi berbasis teknologi yang dapat membantu memprediksi hasil panen kakao mereka, demi mendukung stabilitas ekonomi petani dan kesejahteraan masyarakat desa. Dalam upaya menjawab tantangan ini, penelitian memfokuskan pada pengembangan model prediksi hasil panen kakao dengan memanfaatkan metode *machine learning* yang mempertimbangkan berbagai variabel seperti luas lahan, jumlah tanaman kakao, jenis pupuk, jenis bibit, tenaga kerja, hama dan penyakit, penanggulangan hama dan penyakit, tingkat curah hujan, dan hasil sebelumnya. Diharapkan penelitian ini dapat memberikan kontribusi positif bagi para petani dalam meningkatkan produktivitas dan keberlanjutan usaha mereka.

Peramalan merupakan upaya untuk memproyeksikan atau mengestimasi peristiwa yang akan terjadi di masa yang akan datang berdasarkan informasi yang relevan dari masa lampau menggunakan metode ilmiah. Tujuannya adalah untuk memperoleh informasi tentang kemungkinan peristiwa di masa depan dengan tingkat probabilitas tertinggi [5]. Dalam melakukan prediksi hasil panen kakao dibutuhkan sebuah metode berupa algoritma. Algoritma yang digunakan dalam penelitian ini yaitu algoritma *random forest regression*.

Hasil penelitian perbandingan algoritma metode regresi linier, *random forest regression* dan *gradient boosted trees regression method* untuk prediksi harga rumah, menyatakan bahwa penggunaan metode *random forest* dalam memprediksi harga rumah menghasilkan tingkat akurasi tertinggi sebesar 81,5% [6]. Perbandingan metode *Random Forest* dan *Naïve Bayes* dalam prediksi keberhasilan klien *telemarketing* menyatakan bahwa penggunaan metode *Random Forest* memiliki performa paling baik dibandingkan dengan metode *Naïve Bayes* [7]. Perbandingan algoritma Regresi Linier dan Regresi *Random Forest* dalam memprediksi kasus

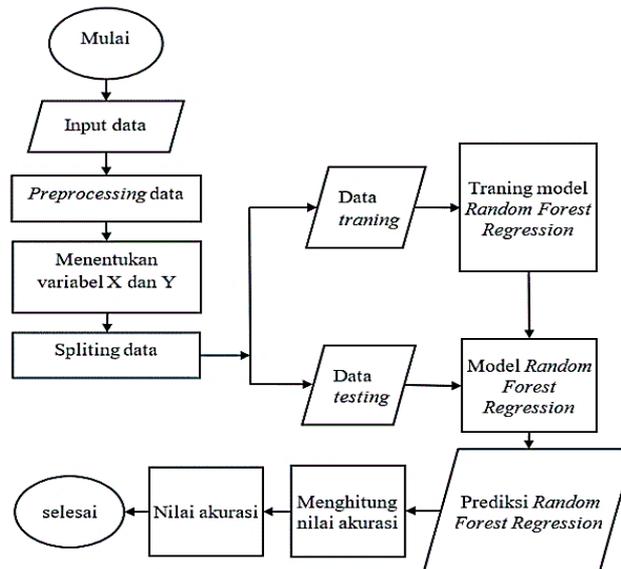
positif *Covid-19*, menyatakan bahwa model *Random Forest* lebih baik dari pada Regresi Linier, karena tingkat akurasi 97.7 % lebih tinggi dari model Regresi Linier [8]. Pada penelitian analisis Perbandingan Algoritma *Decision Tree*, *Random Forest*, dan *Naive Bayes* untuk Prediksi Banjir di Desa Dayeuhkolot menyatakan bahwa algoritma dengan akurasi dan performa yang paling baik didapatkan oleh algoritma *Random Forest* jika dibandingkan dengan algoritma *Decision Tree* dan *Naive Bayes* [9].

Berdasarkan latar belakang masalah yang telah diuraikan diatas, maka penulis mengangkat judul penelitian tentang Prediksi Hasil Panen Kakao di Desa Minanga Menggunakan Algoritma *Random forest regression*.

2. METODE

Pada penelitian ini metode penelitian yang digunakan yaitu metode eksperimen yakni suatu pendekatan penelitian yang digunakan untuk mengeksplorasi pengaruh suatu perlakuan terhadap variabel lain dalam kondisi yang terkendali. Penelitian eksperimen merupakan satu-satunya jenis penelitian yang lebih akurat dan teliti dibandingkan dengan metode penelitian lainnya, terutama dalam menentukan hubungan sebab-akibat. Hal ini disebabkan oleh kemampuan peneliti untuk melakukan pengawasan (kontrol) terhadap variabel bebas baik sebelum maupun selama proses penelitian berlangsung [10].

Metode pengumpulan data yang digunakan dalam penelitian ini yaitu menggunakan jenis data sekunder, yang merupakan data hasil panen kakao petani Desa Minanga pada tahun 2019 sampai dengan 2023 sebanyak 2980 data, yang diperoleh dari kantor Desa Minanga, Kecamatan Bambang, Kabupaten Mamas. Data pendukung diperoleh dari buku, jurnal, dan literatur lainnya yang relevan dengan penelitian ini. Selanjutnya data diolah menggunakan algoritma *random forest regression* untuk menghasilkan prediksi terkait hasil panen. Berikut skema penelitian yang digunakan terlihat pada Gambar 1.



Gambar 1. Tahapan penelitian

2.1. Input Data

Dataset yang digunakan merupakan data sekunder yang diperoleh dari kantor Desa Minanga dari tahun 2019-2023. Sebelum dilakukan *preprocessing*, dataset ini memiliki 11 atribut dengan jumlah data yaitu 2980 baris.

2.2. Preprocessing Data

Pada tahapan ini, data yang suda terkumpul akan dilakukan *preprocessing* untuk meningkatkan kualitas data. *Preprocessing* data dilakukan dengan beberapa langkah meliputi *Exploratory data Analysis*, menghilangkan data yang tidak dibutuhkan, dan transformasi data berupa konversi angka persentase ke desimal serta label *encoding* dengan tujuan agar data dapat di proses oleh *machine learning*.

2.2.1. Proses *Exploratory Data Analysis* (EDA)

Proses *Exploratory data Analysis* (EDA) merupakan langkah awal sebelum dataset digunakan, dimana dataset tersebut akan dilakukan analisis dengan cara memeriksa data yang kosong atau *missing value*. Jika terdapat data yang kosong dalam dataset, maka dapat dilakukan *cleaning* data dengan menghapus data tersebut. Data awal yang digunakan dalam proses EDA ini merupakan data sekunder yang diperoleh dari kantor Desa Minanga dari tahun 2019-2023. Sebelum dilakukan proses EDA, dataset ini memiliki 11 atribut dengan jumlah data yaitu 2980 baris kemudian setelah dilakukan *cleaning* data, jumlah data berkurang menjadi 2910 baris.

2.2.2. *Cleaning* Data

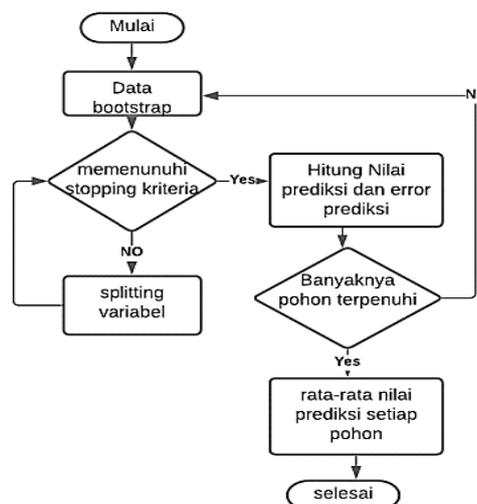
Pembersihan data (*cleaning* data) yaitu menghapus data yang tidak relevan atau tidak diperlukan seperti data pada kolom nama petani, kolom nomor, serta keterangan tambahan seperti satuan (Ha), (Pohon), (Kg) dan tanda persen (%).

2.2.3. Transformasi Data

Transformasi data meliputi konversi angka persentase ke desimal serta label *encoding* (perubahan data bertipe *string* ke data *numerik*). Adapun atribut yang akan dilakukan transformasi data yaitu jenis pupuk, jenis bibit, penanggulangan hama dan penyakit, dan tingkat curah hujan yang akan diubah menggunakan label *encoding* dimana setiap nilai yang ada di dalam atribut tersebut akan diberi label berupa angka 0 dan 1. Sementara itu, atribut hama dan penyakit yang sebelumnya memiliki nilai dalam bentuk persentase akan dikonversi ke angka desimal.

2.3. Implementasi Algoritma *Random Forest Regression*

Random forest regression merupakan kumpulan beberapa pohon keputusan (*decision tree*) yang dibangun secara acak dengan teknik *bootstrap*. Hasil prediksi dari pohon-pohon ini kemudian dikombinasikan sesuai dengan jenis masalah (klasifikasi atau regresi).



Gambar 3. Diagram alir *random forest regression*

2.3.1. Data Bootstrap

Pada tahap *bootstrap*, sampel data (*training*) diambil secara acak untuk membangun satu pohon acak (*random forest regression*). Data *bootstrap* ini kemudian digunakan untuk melatih model pohon dan diuji dengan data testing guna memperoleh prediksi.

2.3.2. Stopping Kriteria

Pada tahap ini jika *sub-node* hanya memiliki satu sampel, maka proses *splitting* berhenti karena memenuhi kriteria berhenti, dan *sub-node* menjadi *leaf node* yang menyajikan hasil prediksi. Jika belum memenuhi, *splitting* variabel akan dilanjutkan.

2.3.3. Splitting Variabel

Pada langkah ini akan dilakukan penentuan atribut *node (root)* sebagai pemisah. Terdapat beberapa kriteria pemisahan yang umum digunakan dalam membangun model *random forest regression*, seperti nilai *entropy* atau *information gain* dan MSE (*Mean Square Error*). Adapun kriteria pemisahan yang digunakan yaitu *entropy* atau *information gain*, dan MSE untuk melihat perbandingan akurasi model yang dihasilkan dari keduanya. Pada kriteria pemisahan MSE, atribut yang akan dipilih sebagai variabel pemisah adalah variabel yang memiliki nilai MSE terkecil [11]. Sementara pada kriteria pemisahan *entropy* atau *information gain* akan dipilih *information gain* tertinggi sebagai variabel pemisah [12]. Untuk mendapatkan nilai dari MSE dapat digunakan rumus :

$$MSE_n = \frac{1}{N} \sum_{i=1}^n (Y_i - Y_n)^2 \quad (1)$$

Jika ingin mendapatkan nilai dari *entropy* dapat digunakan rumus :

$$Entropy = - \sum_{i=1}^n p_i \log_2(p_i) \quad (2)$$

$$Gain(S,A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} Entropy(S_i) \quad (3)$$

Information gain pada atribut dengan nilai *kontinu* atau *numerik* ditentukan oleh nilai pembelah (*split point*) terbaik untuk mengelompokkan nilai. Proses mencari *split point* terbaik dilakukan dengan mengurutkan data terlebih dahulu. *Split point* terbaik dapat ditemukan dengan menganggap median atau nilai tengah dari setiap pasangan nilai yang saling berdekatan. Jika atribut A adalah atribut dengan nilai kontinu, langkah selanjutnya adalah mengurutkan seluruh nilai A dan menentukan nilai tengahnya. Dengan demikian, jumlah partisi yang mungkin pada persamaan 2.2 adalah dua, atau secara matematis dapat dinyatakan sebagai $v = 2$ (dengan $j=1$ dan 2).

2.3.4. Hitung Nilai Prediksi dan Hasil Prediksi

Pada bagian ini, jika hasil nilai prediksi dari pohon acak telah diperoleh, selanjutnya dilakukan perhitungan nilai *error*nya dari hasil prediksi berdasarkan nilai aslinya menggunakan validasi model MAPE (*Mean Absolute Percentage Error*) Nilai RMSE menunjukkan besarnya angka *error* pada hasil prediksi. Rumus untuk mencari nilai RMSE sebagai berikut [13].

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{n}} \quad (4)$$

Nilai MAPE dapat dihitung menggunakan rumus berikut :

$$\text{MAPE} = \frac{1}{N} \sum_{i=1}^1 \frac{Y_i - Y_n}{Y_i} * 100 \% \quad (5)$$

2.3.5. Banyaknya Pohon Yang Terpenuhi

Pada tahap ini, langkah 1-3 akan diulang hingga mencapai jumlah *Ntree* pohon yang diinginkan.

2.3.6. Rata-Rata Nilai Prediksi

Menentukan hasil prediksi akhir dengan menggabungkan hasil prediksi pada setiap pohon, kemudian nilai rata-ratanya dijadikan sebagai hasil prediksi sehingga menghasilkan prediksi yang lebih stabil dan mengurangi potensi *overfitting*. Untuk mencari nilai rata-rata seluruh prediksi pohon, dapat menggunakan persamaan berikut [11].

$$Y_i = \frac{1}{N_{\text{tree}}} \sum_{i=1}^{N_{\text{tree}}} Y_n \quad (6)$$

2.3.7. Mengevaluasi Model

Evaluasi adalah parameter yang menggambarkan seberapa dekat nilai prediksi dengan nilai aktual. Suatu peramalan atau prediksi dapat dievaluasi menggunakan berbagai metode untuk mengukur tingkat akurasi, seperti *Root Mean Squared Error* (RMSE), *Mean Absolute Percentage Error* (MAPE), dan *koefisien determinasi* (R^2).

Evaluasi model dalam penelitian ini menggunakan *Root Mean Squared Error* (RMSE) sebagai metode alternatif untuk menilai akurasi peramalan. RMSE mengukur rata-rata kuadrat dari kesalahan prediksi model. RMSE merupakan teknik yang mudah diimplementasikan dan sering digunakan dalam berbagai studi yang terkait dengan peramalan [14].

Mean Absolute Percentage Error (MAPE) mengukur akurasi dengan menghitung persentase perbedaan antara prediksi dan data aktual, sehingga memberikan metrik ketepatan relatif. MAPE dipilih untuk menguji akurasi karena kemampuannya dalam memberikan hasil yang relatif tepat [15].

Tingkat akurasi model prediksi dihitung dengan mengurangi kesalahan relatif terhadap data aktual. Semakin kecil kesalahan, semakin tinggi akurasinya. Nilai MAPE yang rendah menunjukkan akurasi mendekati 100%, sementara nilai MAPE yang tinggi menandakan model kurang efektif dalam prediksi [16]. Oleh karena itu, akurasi dapat dihitung menggunakan rumus berikut:

$$\text{Akurasi} = 100\% - \text{MAPE} \quad (7)$$

Koefisien determinasi (R^2) adalah sebuah ukuran untuk menilai sejauh mana variabel independen berkontribusi bersama-sama terhadap variabel dependen atau respons. Nilai R^2 berada dalam rentang antara 0 hingga 1. Semakin mendekati nilai 1, maka model regresi tersebut dapat memprediksi variabel Y (dependen) dengan akurat [17]. Untuk mendapatkan nilai R^2 digunakan rumus :

$$R^2 = 1 - \sum \frac{(y_i - y_n)^2}{(y_i - y)^2} \quad (8)$$

2.4. Feature impotrance

Dalam random forest regression, feature importance digunakan untuk menilai pengaruh variabel independen terhadap variabel dependen. Namun, metode ini hanya mengukur tingkat

kepentingan tanpa menjelaskan hubungan yang terbentuk di antara keduanya. Salah satu tujuan *feature importance* adalah untuk mencegah *overfitting* dalam model dan memahami faktor-faktor yang mendasari pembentukan model. Pemilihan *feature importance* dilakukan selama pembentukan model.

2.5. Flask

Flask adalah microframework web Python yang berperan sebagai pondasi aplikasi dan antarmuka pengguna. Sasaran *flask* adalah untuk menyediakan dasar dari aplikasi yang sederhana namun dapat dengan mudah diperluas dengan tambahan fitur.

3. HASIL DAN PEMBAHASAN

Hasil penelitian ini diperoleh melalui proses pengujian dataset hasil panen petani kakao di Desa Minanga pada periode 2019 hingga 2023, dengan menggunakan metode yang telah dibahas sebelumnya. Pengujian diawali dengan beberapa tahapan, dimulai dari proses *preprocessing* data hingga pembentukan model *random forest regression*. Selanjutnya, dilakukan evaluasi terhadap model yang telah dibentuk untuk mengukur tingkat akurasi dalam melakukan prediksi. Berdasarkan tahapan-tahapan dalam metode penelitian tersebut, hasil penelitian secara rinci akan dibahas pada bagian berikut.

3.1. Input dataset

Data yang digunakan dalam penelitian ini untuk memprediksi hasil panen kakao menggunakan algoritma *random forest regression* diperoleh dari hasil panen petani kakao pada periode 2019 hingga 2023. Dataset ini memiliki 11 atribut dengan jumlah baris data yaitu 2980 *record*.

3.2. Preprocessing

Pada tahapan ini, data yang suda terkumpul akan dilakukan *Preprocessing* untuk menghasilkan data yang berkualitas sehingga siap digunakan untuk proses analisis lebih lanjut dalam proses menghasilkan prediksi. *Preprocessing* data dilakukan dengan beberapa langkah, dimulai dari melakukan proses *Exploratory data Analysis*, menghilangkan data yang tidak dibutuhkan, dan transformasi data berupa konversi angka persentase ke desimal serta label *encoding*. Pada Tabel 3 merupakan sampel data sesudah dilakukan *preprocessing* dimana data tersebut terdiri dari 9 atribut yang dibagi menjadi variabel prediktor (*independen*) dan variabel respon (*dependen*). Dimana variabel *independen* terdiri dari luas lahan, jumlah tanaman kakao, jenis bibit, jenis pupuk, hama dan penyakit, penanggulangan hama dan penyakit tingkat curah hujan, tenaga kerja dan hasil sebagai variabel *dependen*. Adapun atribut yang dilakukan transformasi data, yaitu jenis pupuk, jenis bibit, penanggulangan hama dan penyakit, dan tingkat curah hujan yang diubah menggunakan label *encoding*, sedangkan atribut hama dan penyakit akan dikonversi dari angka persentase ke angka desimal.

Tabel 3. Sampel data sesudah *preprocessing*

Luas Lahan	Jumlah Tanaman Kakao	Jenis Bibit	Jenis Pupuk	Hama & Penyakit	Penanggulangan Hama & Penyakit	Tingkat Curah Hujan	Tenaga Kerja	Hasil
0.9	985	0	0	0.2170	1	1	7	694
1.1	1169	0	1	0.1460	0	0	10	927
1.02	1085	1	0	0.1850	1	1	9	813
0.62	733	0	0	0.1390	0	1	5	536
0.7	807	0	0	0.1770	1	1	6	551
0.97	979	1	0	0.1650	1	1	8	722
0.37	461	0	1	0.1740	0	2	2	288
0.31	620	0	1	0.0670	0	1	3	322

Luas Lahan	Jumlah Tanaman Kakao	Jenis Bibit	Jenis Pupuk	Hama & Penyakit	Penanggulangan Hama & Penyakit	Tingkat Curah Hujan	Tenaga Kerja	Hasil
0.52	643	0	0	0.1490	0	1	4	431
0.3	441	0	0	0.1040	0	0	3	292

3.3. Pembagian Data *Training* Dan Data *Testing* (*Spilitting* Data)

Data yang telah diproses sebelumnya kemudian dibagi menjadi data *training* dan data *testing*. Penentuan pembagian proporsi data *training* dan *testing* akan berpengaruh terhadap tingkat akurasi. Jumlah keseluruhan data yang akan diolah setelah dilakukan tahap *preprocessing* yakni 2910 data dan akan dibagi dalam 4 rasio.

Tabel 4. Pembagian data *training* dan data *testing*

Rasio Data	Jenis Data	Jumlah Data
60:40	<i>Training</i>	1746
	<i>Testing</i>	1164
70:30	<i>Training</i>	2036
	<i>Testing</i>	874
80:20	<i>Training</i>	2328
	<i>Testing</i>	582
90:10	<i>Training</i>	2619
	<i>Testing</i>	291

3.4. Prediksi dan Validasi Model

Pembentukan model *random forest regression* berdasarkan *training* dataset yang telah ditentukan berdasarkan variabel dependen dan variabel independent. Variabel hasil sebagai variabel dependen dengan tipe data *numerik* yang akan diprediksi nilainya berdasarkan variabel luas lahan, jumlah tanaman kakao, jenis bibit, jenis pupuk, hama dan penyakit, tingkat curah hujan, tenaga kerja sebagai variabel independent.

Model *random forest regression* dimulai dengan penentuan jumlah pohon yang akan digunakan untuk mendapatkan nilai prediksi dengan nilai akurasi terbaik. Penentuan jumlah pohon sangat berpengaruh pada nilai akurasi, dimana semakin kecil nilai *error* yang terbentuk maka semakin baik model prediksi yang dihasilkan. Pada Tabel 5 merupakan penentuan jumlah pohon dari masing-masing kriteria pemisahan.

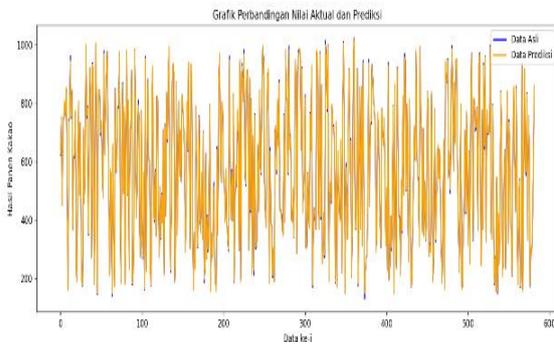
Tabel 5. Penentuan jumlah pohon

Entropy / Information Gain		MSE	
Jumlah Pohon	MAPE%	Jumlah Pohon	MAPE%
5	1.27 %	5	1.18 %
9	1.20 %	9	1.12 %
19	1.18 %	19	1.07 %
29	1.14 %	29	1.08 %
30	1.14 %	30	1.07 %
50	1.13 %	50	1.06 %
80	1.13 %	80	1.05 %
100	1.13 %	100	1.04 %
200	1.12 %	200	1.04 %
300	1.12 %	300	1.18 %

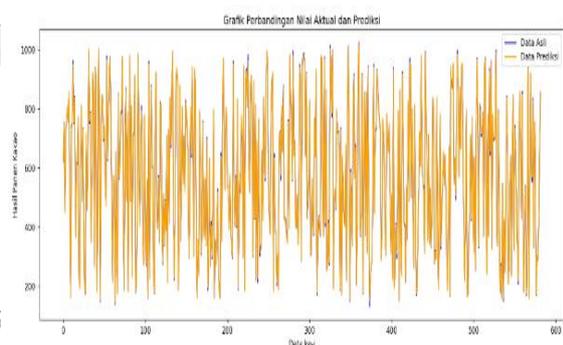
Dengan membandingkan jumlah pohon yang terbentuk dengan nilai *error* terkecil pada *Mean Absolute Percentage Error* (MAPE), akan menjadi jumlah pohon terbaik untuk digunakan dalam pembentukan model prediksi. Berdasarkan Tabel 5, jumlah pohon dengan nilai error terkecil menggunakan kriteria pemisahan *entropy dan information gain* yaitu 200 pohon (n estimators) dengan nilai MAPE yaitu 1.12 % dibandingkan dengan jumlah pohon yang lainnya. Sementara, jumlah pohon dengan nilai error terkecil menggunakan kriteria pemisahan MSE yaitu 100 pohon dengan nilai MAPE sebesar 1.04%. Dengan demikian jumlah pohon yang digunakan untuk mengetahui akurasi adalah 100 dan 200 pohon.

3.5. Pengujian sistem

Ketika data hasil prediksi mendekati data aktualnya, maka model tersebut masuk dalam kriteria regresi yang baik. Dari model regresi yang telah dibangun dengan menggunakan data *training* 2328, *testing* 582 data dan jumlah pohon yang terpilih sebanyak 100 pohon dan 200 pohon, maka didapatkan grafik model regresi dengan data aktual dan data hasil prediksi menggunakan *random forest regression*.



Gambar 4. Grafik perbandingan data asli dan data prediksi (*entropy dan information gain*)



Gambar 5. Grafik perbandingan data asli dan data prediksi (MSE)

Grafik perbandingan data aktual dan prediksi pada Gambar 4 dan Gambar 5, menunjukkan bahwa nilai prediksi yang dihasilkan model memiliki nilai yang hampir mendekati nilai aktual.

Tabel 6. Perbandingan nilai aktual dan nilai prediksi pada sampel data *tasting*

No.	<i>Entropy & Information Gain</i>		MSE	
	Actual Values (Kg)	Predicted Values (Kg)	Actual Values (Kg)	Predicted Values (Kg)
1.	623.000	625.495	623.000	625.760
2.	753.000	751.375	753.000	752.040
3.	421.000	447.810	451.000	448.830
4.	721.000	723.875	721.000	726.480
5.	755.000	748.171	755.000	748.880
6.	796.000	807.422	796.000	808.885
7.	755.000	755.775	755.000	754.580
8.	849.000	856.740	849.000	853.010
9.	478.000	476.000	478.000	478.950
10.	163.000	159.000	163.000	158.610

Hasil persentase dari pembagian data *training* dan data *testing* dengan menggunakan jumlah pohon sebanyak 100 dan 200, dapat kita lihat pada Tabel 7 berikut :

Tabel 7. Perbandingan hasil akurasi menggunakan *entropy/information gain*

Persentase Pembagian Dataset (%)	Jumlah Pembagian Data		Hasil Akurasi Pada n Pohon (%)									
	Train	testing	5	9	19	29	39	50	80	100	200	300
60:40	1571	1048	98.67	98.71	98.73	98.75	98.75	98.76	98.78	98.78	98.78	98.78
70:30	1833	786	98.62	98.72	98.75	98.76	98.77	98.77	98.77	98.77	98.79	98.80
80:20	2095	524	98.73	98.80	98.82	98.86	98.86	98.87	98.87	98.87	98.88	98.88
90:10	2357	262	98.65	98.71	98.80	98.80	98.83	98.83	98.84	98.84	98.84	98.83

Tabel 8. Perbandingan hasil akurasi menggunakan MSE

Persentase Pembagian Dataset (%)	Jumlah Pembagian Data		Hasil Akurasi Pada n Pohon (%)									
	Train	testing	5	9	19	29	39	50	80	100	200	300
60:40	1571	1048	98.75	98.78	98.78	98.81	98.81	98.81	98.82	98.83	98.75	98.83
70:30	1833	786	98.80	98.80	98.81	98.83	98.84	98.85	98.86	98.86	98.80	98.87
80:20	2095	524	98.81	98.87	98.92	98.91	98.92	98.93	98.94	98.95	98.81	98.95
90:10	2357	262	98.80	98.83	98.87	98.87	98.87	98.88	98.89	98.90	98.80	98.91

Dari Tabel 7 dan Tabel 8 perbandingan akurasi pada pembagian data *testing* dan data *training*, maka dalam penulisan ini, penulis menggunakan perbandingan 80 : 20, yang menghasilkan nilai RMSE, R2, MAPE, dan Akurasi sebagai berikut:

Tabel 9. Hasil akurasi 200 pohon

RMSE	R2	MAPE	Akurasi
6.82	0.99	1.12 %	98.88 %

Tabel 10. Hasil akurasi 100 pohon

RMSE	R2	MAPE	Akurasi
6.26	0.99	1.04%	98.95%

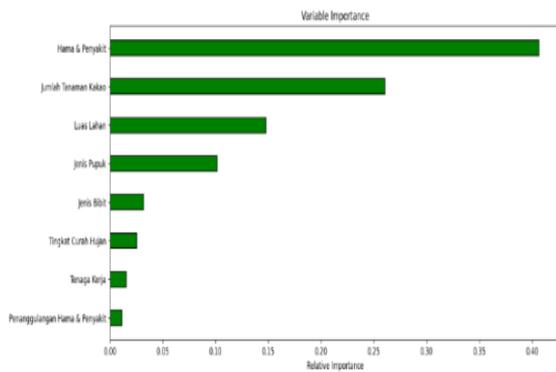
Model *random forest regression* yang dibentuk dengan jumlah 200 pohon menggunakan kriteria pemisahan *entropy/information gain*, diperoleh RMSE yaitu 6.82 dan nilai R^2 yaitu 0,99, serta nilai MAPE sebesar 1.12% dengan akurasi 98.88%. Sementara model yang dibentuk menggunakan MSE dengan jumlah 100 pohon, diperoleh RMSE yaitu 6.26, dan nilai R^2 yaitu 0.99, nilai MAPE sebesar 1.04% dengan akurasi 98.95%. Hasil ini menunjukkan bahwa penggunaan *random forest regression* dengan kriteria pemisahan MSE memberikan performa yang baik dengan akurasi yang tinggi.

3.6. Variabel *importance*

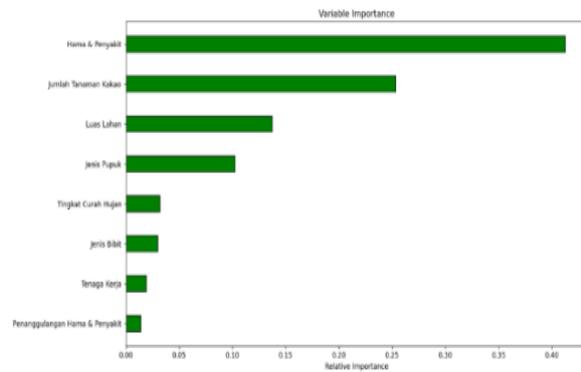
Pada tabel 11, terlihat bahwa kedua kriteria pemisahan *entropy/information gain* dan MSE, variabel hama dan penyakit memiliki pengaruh paling signifikan dalam memprediksi hasil panen dengan nilai sebesar 0.4064 menggunakan 200 pohon sementara untuk jumlah 100 pohon yaitu sebesar 0.4125 dibandingkan dengan tujuh variabel lainnya yang juga digunakan dalam analisis ini.

Tabel 11 . Variabel *importance*

Variabel <i>Importance</i>			
200 pohon		100 pohon	
Luas lahan	0.1479	Luas lahan	0.1374
Jumlah tanaman kakao	0.2606	Jumlah tanaman kakao	0.2530
Jenis bibit	0.0317	Jenis bibit	0.0299
Jenis pupuk	0.1016	Jenis pupuk	0.1024
Hama dan penyakit	0.4064	Hama dan penyakit	0.4125
Penanggulangan hama dan penyakit	0.0113	Penanggulangan hama dan penyakit	0.0140
Tingkat curah hujan	0.0254	Tingkat curah hujan	0.0318
Tenaga kerja	0.0152	Tenaga kerja	0.0190



Gambar 6. Variabel *Importance* (200 pohon)



Gambar 7. Variabel *Importance* (100 pohon)

3.7. Pengembangan Model Prediksi

Model yang sudah didapatkan kemudian disimpan untuk pengembangan *website*, yang dapat digunakan untuk melakukan prediksi dengan menggunakan fungsi *pickle*. Model tersebut dapat juga memprediksi data baru hanya dengan memanggil model tanpa harus melakukan analisis dari awal.

3.7.1. Membuat tampilan halaman *website*

Setelah proses pembentukan model disimpan, model tersebut dikembangkan dalam sebuah rancangan *website* yang digunakan untuk melakukan prediksi hasil panen kakao menggunakan algoritma *Random Forest Regression*. Pada halaman utama *website* ditampilkan

sebuah form yang memuat seluruh variabel independen untuk diisi oleh pengguna. Ketika pengguna mengisi form dan mengirimkannya, sistem akan mengeksekusi model untuk menampilkan hasil prediksi panen kakao secara langsung di halaman web. Rancangan website ini dapat dijalankan secara lokal melalui *localhost* menggunakan peramban seperti Microsoft Edge.

PREDIKSI HASIL PANEN KAKAO DESA MINANGA	
Luas Lahan	<input type="text"/>
Jumlah Tanaman Kakao	<input type="text"/>
Jenis Bibit	Pilih Jenis Bibit
Jenis Pupuk	Pilih Jenis Pupuk
Hama & Penyakit (%)	<input type="text"/>
Penanggulangan Hama & Penyakit	Pilih Jenis Pestisida
Tingkat Curah Hujan	Curah Hujan
Tenaga Kerja	<input type="text"/>
<input type="button" value="PREDIKSI"/>	

Berikut adalah hasil prediksi berdasarkan perhitungan dari kriteria yang telah diinput :
{{\$ result_text }}

Gambar 8. Tampilan web

Gambar 8 menunjukkan tampilan interface dari website yang dibuat untuk melakukan prediksi hasil panen kakao berdasarkan variabel-variabel yang telah ditentukan. Tampilan website tersebut akan menghasilkan output nilai prediksi hasil panen kakao dalam satuan kilogram (Kg).

4. KESIMPULAN

Berdasarkan hasil penelitian dan pembahasan, dapat disimpulkan bahwa dari kedua kriteria pemisahan yang digunakan untuk membangun model *random forest regression*, yakni kriteria pemisahan menggunakan nilai *entropy/information gain* dan kriteria pemisahan menggunakan MSE, jumlah pohon yang menghasilkan model *random forest regression* yang paling optimal yaitu terdiri atas 100 pohon keputusan dengan menggunakan kriteria pemisahan MSE. Variabel yang memiliki nilai paling tinggi pada variabel *Importance* adalah hama dan penyakit. Model ini menunjukkan akurasi sebesar 98,95%. Selain itu, nilai MAPE yang diperoleh adalah 1,04 %, nilai RMSE mencapai 6.26, dan nilai R^2 adalah 0,99, yang semuanya menunjukkan kemampuan model dalam memprediksi hasil panen kakao di Desa Minanga. Model yang telah dibangun kemudian disimpan dan digunakan dalam pengembangan sebuah website, sehingga pengguna dapat melakukan prediksi hasil panen kakao secara mudah tanpa perlu menyusun kembali program secara manual.

REFERENSI

- [1] Puspitasari, R. D. (2020). Pertanian berkelanjutan berbasis revolusi industri 4.0. *Jurnal Layanan Masyarakat (Journal of Public Services)*, 3(1), 26.
- [2] Bramastyo, R. M., Novianti, A. F., Utami, M. E. K. A. S. R. I., Ruszia, Y. E. G. A., Maliatin, R., & Ghozali, D. E. O. N. (2019). Pelatihan Pengolahan Dan Pemasaran Produk

- Keripik Pisang Melalui E-Marketing Dusun Jati Desa Jatidukuh Kec Gondang Mojokerto. *Jurnal Abdi Bhayangkara*, 1(01), 17–26.
- [3] Meniati, L., Gaol, N. Y. L., & Santoso, I. (2022). Sistem Pakar Mendiagnosa Penyakit Tanaman Kakao Menggunakan Metode Certainty Factor. *Jurnal Teknologi Sistem Informasi Dan Sistem Komputer TGD*, 5(1), 83–94.
- [4] Izzah, N., & Damayanti, D. (2023). Pengaruh Jumlah Produksi dan Harga terhadap Nilai Ekspor Kakao Indonesia Tahun 2017-2020. *Transparansi: Jurnal Ilmiah Ilmu Administrasi*, 6(1), 78–85.
- [5] Wanto, A., & Windarto, A. P. (2017). Analisis prediksi indeks harga konsumen berdasarkan kelompok kesehatan dengan menggunakan metode backpropagation. *Sinkron: Jurnal Dan Penelitian Teknik Informatika*, 2(2), 37–43.
- [6] Fitri, E. (2023). Analisis Perbandingan Metode Regresi Linier, *Random forest regression* Dan Gradient Boosted Trees Regression Method Untuk Prediksi Harga Rumah. *Journal of Applied Computer Science and Technology*, 4(1), 58–64.
- [7] Leonardo, R., Pratama, J., & Chrisnatalis, C. (2020). Perbandingan Metode Random Forest Dan Naïve Bayes Dalam Prediksi Keberhasilan Klien Telemarketing. *Jurnal Teknologi Dan Ilmu Komputer Prima (Jutikomp)*, 3(2), 455–459.
- [8] Fachid, S., & Triayudi, A. (2022). Perbandingan Algoritma Regresi Linier dan Regresi Random Forest Dalam Memprediksi Kasus Positif Covid-19. *Jurnal Media Informatika Budidarma*, 6(1), 68–73.
- [9] Darmawan, M. B., Dewanta, F., & Astuti, S. (2023). Analisis Perbandingan Algoritma Decision Tree, Random Forest, dan Naïve Bayes untuk Prediksi Banjir di Desa Dayeuhkolot. *TELKA: Jurnal Telekomunikasi, Elektronika, Komputasi, Dan Kontrol*, 9, 56–61.
- [10] Putri, A. D., Ahman, A., Hilmia, R. S., Almaliyah, S., & Permana, S. (2023). Pengaplikasian Uji T dalam Penelitian Eksperimen. *Jurnal Lebesgue: Jurnal Ilmiah Pendidikan Matematika, Matematika Dan Statistika*, 4(3), 1978–1987.
- [11] Mulyahati, I. L. (2020). *Implementasi Machine Learning Prediksi Harga Sewa Apartemen Menggunakan Algoritma Random Forest Melalui Framework Website Flask Python (Studi Kasus: Apartemen di DKI Jakarta Pada Website mamikos.com)*.
- [12] Dengen, C. N., Kusri, K., & Luthfi, E. T. (2020). Implementasi Decision Tree Untuk Prediksi Kelulusan Mahasiswa Tepat Waktu. *Sisfotenika*, 10(1), 1–11.
- [13] Suliztia, M. L. (2020). *Penerapan Analisis Random Forest pada Prototype Sistem Prediksi Harga Kamera Bekas Menggunakan Flask*.
- [14] Wiranda, L., Sadikin, M., Informatika, J. T., & Komputer, F. I. (2019). Penerapan Long Short Term Memory Pada Data Time Series Untuk Memprediksi Penjualan Produk PT. *Metiska Farma. Jurusan Teknik Informatika, Fakultas Ilmu Komputer*.
- [15] Laksmana, R. D., Santoso, E., & Rahayudi, B. (2019). Prediksi penjualan roti menggunakan metode exponential smoothing (Studi Kasus: Harum Bakery). *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 3(5), 4933–4941.
- [16] Saadah, S., & Salsabila, H. (2021). Prediksi Harga Bitcoin Menggunakan Metode Random Forest:(Studi Kasus: Data Acak Pada Masa Pandemic Covid-19). *Jurnal Komputer Terapan*, 7(1), 24–32.
- [17] Nur, N., Wajidi, F., Sulfayanti, S., & Wildayani, W. (2023). Implementasi Algoritma *Random forest regression* untuk Memprediksi Hasil Panen Padi di Desa Minanga. *Jurnal Komputer Terapan*, 9(1), 58–64.