



Implementasi Algoritma *Random Forest* Regression untuk Memprediksi Hasil Panen Padi di Desa Minanga

Nahya Nur^{1*}, Farid Wajidi², Sulfayanti³, Wildayani⁴

^{1,2,3,4}Universitas Sulawesi Barat, Informatika, Indonesia

¹nahya.nur@unsulbar.ac.id, ²faridwajidi@unsulbar.ac.id, ³sulfayanti@unsulbar.ac.id, ⁴wildayani0201@gmail.com

*Corresponding Author

Diserahkan: 23 Januari 2023

Diterima: 21 Juni 2023

Diterbitkan: 21 Juni 2023

ABSTRAK

Desa Minanga, Kecamatan Bambang, Kabupaten Mamasa penduduknya melakukan budidaya tanaman padi yang biasanya hasil panen setiap musimnya mengalami fluktuasi yang seringkali terjadi penurunan atau pun peningkatan yang tidak stabil. Penelitian ini diharapkan dapat membantu dalam memprediksi hasil panen padi sesuai dengan kriteria dan data yang ada sebelumnya seperti luas lahan, jumlah bibit, jenis pupuk, curah hujan, hama dan gulma, pengendalian hama dan gulma, dan sistem penanaman padi yang digunakan (jajar legowo), dengan menerapkan algoritma *Random Forest Regression*. Evaluasi kinerja algoritma diukur dengan menggunakan *Root Mean Squared Error (RMSE)*, *Mean Absolute Percentage Error (MAPE)* dan koefisien determinasi (R^2), hasil dari model *Random Forest* yang didapatkan dari 9 pohon, variabel yang memiliki nilai paling tinggi pada variabel *importance* adalah variabel luas lahan. Sehingga dari model tersebut diperoleh nilai akurasi 95,11%, Nilai MAPE pada model ini yaitu 4,884%, nilai RMSE yaitu 0,250 dan nilai R^2 yaitu 0.99.

Kata kunci: Panen Padi, Prediksi, *Random Forest*

ABSTRACT

Minanga Village, Bambang District, Mamasa Regency, the people of which cultivate rice plants, usually the yields fluctuate each season, which often decreases or increases unstable. This research is expected to assist in predicting rice yields in accordance with pre-existing criteria and data such as land area, number of seeds, type of fertilizer, rainfall, pests and weeds, pest and weed control, and the rice planting system used (jajar legowo), by applying the *Random Forest Regression* algorithm. Evaluation of algorithm performance is measured using *Root Mean Squared Error (RMSE)*, *Mean Absolute Percentage Error (MAPE)* and the coefficient of determination (R^2), the results of the *Random Forest* model obtained from 9 trees, the variable that has the highest value on the variable *importance* is the variable land area. So that from this model an accuracy value of 95.11% is obtained, the MAPE value in this model is 4.884%, the RMSE value is 0.250 and the R^2 value is 0.99.

Keywords: Forecasting, *Random Forest*, Rice Harvest

1. PENDAHULUAN

Padi (*Oryza Sativa*) termasuk salah satu kebutuhan primer penduduk dalam memenuhi kebutuhan karbohidrat. Padi merupakan tanaman komoditas utama yang menghasilkan beras dan menjadi bahan makanan pokok di Indonesia [1]. Indonesia menempati posisi ketiga negara penghasil beras terbesar di

dunia, setelah China sekitar 30% dan India sekitar 21%. Dengan produksi padi sebesar 9,67% dari total produksi padi di dunia [2]. Meskipun demikian dengan peningkatan jumlah penduduk tiap tahunnya yang mengakibatkan kebutuhan produksi meningkat, Indonesia masih tetap perlu mengimpor beras hampir tiap tahun walau hanya untuk menjaga tingkat cadangan beras. Oleh karena itu dalam upaya mendukung ketahanan pangan sangat diperlukan adanya manajemen produksi padi. Salah satu daerah penghasil padi yang ada di daerah Sulawesi Barat yaitu di Desa Minanga, Kecamatan Bambang, Kabupaten Mamasa.

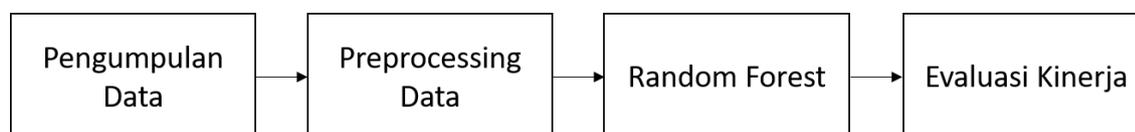
Di Kabupaten Mamasa, Provinsi Sulawesi Barat khususnya di Desa Minanga, Kecamatan Bambang, penduduknya sudah melakukan budidaya tanaman padi. Berdasarkan hasil wawancara yang dilakukan oleh peneliti dengan penyuluh pertanian, aparat desa dan beberapa petani di Desa Minanga, bahwa biasanya hasil panen setiap musimnya mengalami fluktuasi yang seringkali terjadi penurunan atau pun peningkatan yang tidak stabil.

Dari uraian tersebut maka penulis melakukan sebuah penelitian yang dapat membantu dalam memprediksi hasil panen padi sesuai dengan kriteria dan data yang ada sebelumnya. Untuk memperoleh hasil panen padi, diperlukan informasi mengenai luas lahan; jumlah bibit yang ditanam; jenis pupuk yang digunakan seperti Urea dengan NPK Phonska dan Urea dengan ZA; curah hujan; tingkat penyerangan hama dan gulma; pengendalian hama dan gulma menggunakan pestisida dan penyiangan; serta sistem penanaman padi yang digunakan (jajar legowo). Luas lahan yang digunakan dalam penanaman padi sangat berpengaruh terhadap hasil panen yang didapatkan. Begitupun dengan jumlah bibit, apabila luas lahan yang digunakan semakin besar, maka jumlah bibit padi yang ditanam semakin banyak, dan potensi hasil panen padi yang diperoleh juga semakin besar. Begitupun dengan pemeliharaan tanaman padi seperti, pemupukan dan pengendalian dari hama akan mempengaruhi hasil panen suatu tanaman [3]. Dari data dengan kriteria tertentu, maka dapat dilakukan sebuah peramalan atau prediksi sesuatu yang baru berdasarkan data yang ada sebelumnya.

Proses prediksi yaitu memperkirakan sesuatu yang akan terjadi di masa depan secara sistematis, berdasarkan informasi yang dimiliki di masa lalu dan masa sekarang [4]. Untuk memprediksi hasil panen padi, dalam penyelesaian diperlukan sebuah metode, dalam hal ini menggunakan algoritma *Random Forest* Regression. *Random Forest* merupakan salah satu metode machine learning yang bisa digunakan dalam kasus klasifikasi dan regresi, yang menggunakan pohon keputusan dengan pengambilan data sampel secara acak [5]. Pada perhitungan *Random Forest* nantinya akan memberikan pengetahuan baru bagi masyarakat dan petani desa Minanga, dengan adanya model prediksi yang memberikan gambaran dan pengalaman dalam penentuan hasil panen padi.

2. METODE PENELITIAN

Secara umum, penelitian ini dilakukan melalui empat tahapan utama yang dapat ditunjukkan pada gambar 1.



Gambar 1. Tahapan Penelitian

2.1 Pengumpulan Data

Data yang digunakan dalam penelitian ini sebanyak 298 data yang bersumber dari Data Hasil Panen Desa Minanga pada tahun 2021 yang terbagi atas dua musim. Beberapa variabel yang dijadikan sebagai acuan untuk mengetahui prediksi hasil panen dapat dilihat pada tabel 1.

Tabel 1. Sampel data

No	Luas Lahan (Ha)	Jumlah Bibit (L)	Jenis Pupuk	Curah Hujan	Hama & Gulma	Pengendalian Hama & Gulma	Jajar Legowo	Hasil (Ton)
1	1.4	74	Urea, NPK	Normal	14,7%	Pestisida	Ya	4,691
2	0.9	54	Urea, NPK	Normal	3,21%	Pestisida Dan Penyiangan	Tidak	3,353
3	0.86	50.5	Urea, ZA	Normal	23%	Pestisida	Tidak	2,477
4	1.4	74	Urea, NPK	Rendah	15%	Pestisida Dan Penyiangan	ya	4,675
5	1	60	Urea, NPK	Rendah	30,5%	Pestisida	Tidak	2,675

Tabel 1 menunjukkan sampel data yang digunakan pada penelitian ini. Keseluruhan data yang diperoleh akan dibagi menjadi dua bagian, yaitu data training dan data testing untuk diproses pada algoritma *Random Forest*. Adapun variabel yang digunakan terdiri atas 6 variabel prediktor yaitu

- a. luas lahan
Luas lahan pada penanaman padi sangat berpengaruh terhadap hasil panen yang diperoleh. Dimana luas lahan diasumsikan berbanding lurus dengan hasil panen yang akan dihasilkan [3].
- b. jumlah bibit
Sejalan dengan luas lahan, jumlah bibit juga berdampak pada hasil panen disebabkan jumlah bibit yang ditanam akan disesuaikan dengan luas lahan yang tersedia.
- c. jenis pupuk
penggunaan pupuk sangat berpengaruh dalam hasil produksi tanaman padi dibutuhkan, karena unsur hara yang ada di dalam tanah tidak selalu mencukupi untuk memacu pertumbuhan tanaman secara optimal [6]. Jenis pupuk yang biasanya petani di Desa Minanga gunakan adalah pupuk urea, NPK Phonska dan pupuk Urea, ZA.
- d. curah hujan
Pola curah hujan sangat berpengaruh dalam kegiatan budidaya pertanian. Curah hujan seringkali menjadi faktor pembatas terhadap produksi pertanian [7]. Berdasarkan data yang diambil dari kantor Desa Minanga, curah hujan yang ada di Desa Minanga pada tahun 2021 dari bulan januari sampai dengan bulan juli masih masuk dalam kategori normal, yang rata-rata curah hujannya yaitu 100 mm – 250 mm. Pada bulan juli sampai dengan oktober, rata-rata curah hujan di Desa Minanga rendah, yaitu 80 mm – 95 mm.
- e. hama dan gulma
Faktor lainnya yang mempengaruhi hasil panen adalah hama dan gulma. Pada penelitian ini akan menggunakan persentase tingkat hama dan gulma yang terjadi pada lahan yang dijadikan sebagai objek penelitian.
- f. pengendalian hama dan gulma
Usaha petani di Desa Minanga dalam melakukan pengendalian hama dan gulma pada tanaman padi yaitu dengan menggunakan pestisida yang tidak mengandung racun, dan meningkatkan pengontrolan rumput liar berupa penyiangan.
- g. jajar legowo
Sistem tanam jajar legowo yaitu sistem tanam yang berselang seling antara dua atau lebih baris tanaman padi dan satu baris kosong. Jarak tanam pada budidaya ini menjadikan lebih banyak tanaman yang menjadi tanaman pinggir. Tanaman pinggir akan memperoleh air yang cukup, unsur hara yang lebih merata, mempermudah pemeliharaan tanaman, serta sinar matahari yang lebih banyak dalam proses fotosintesis [8]. Sistem tanam padi di Desa Minanga yang digunakan oleh petani, tidak semuanya menggunakan sistem tanam jajar legowo.

Selain variabel prediktor, terdapat satu variabel respon yaitu hasil panen. Variabel tersebut yang akan menjadi data aktual untuk memastikan bagaimana kinerja dari algoritma yang digunakan dengan membandingkan hasil panen yang diperoleh dari proses *Random Forest* dengan hasil panen pada data aktual tersebut.

2.2 Preprocessing

Pada penelitian ini, sebelum data yang diperoleh diolah menggunakan algoritma *Random Forest*, terlebih dahulu dilakukan preprocessing. Tahapan ini bertujuan untuk memastikan bahwa seluruh data yang diperlukan dapat diproses oleh sistem. Seluruh variabel yang diperoleh dianalisis terkait tipe data dari masing-masing variabel tersebut. Hal ini diperlukan mengingat algoritma yang digunakan hanya dapat memproses data berupa data numerik [9].

Tabel 2. Variabel

Jenis Variabel	Nama Variabel	Tipe Data
Variabel prediktor	Luas lahan	Numerik
	Jumlah bibit	Numerik
	Jenis pupuk	Kategorik
	Curah hujan	Kategorik
	Hama dan gulma	Numerik
	Pengendalian hama dan gulma	Kategorik
	Jajar legowo	Kategorik
Variabel respon	Hasil	Numerik

Tabel 1 menunjukkan variable yang digunakan pada penelitian ini. Variabel dengan tipe data kategorik akan diubah kedalam bentuk numerik (0 dan 1) agar dapat terbaca diproses lebih lanjut. Adapun variabel independen yang masuk kedalam proses pelabelan yaitu, Jajar Legowo, Curah Hujan, Pengendalian Hama dan Penyakit, serta Jenis Pupuk. Berikut adalah sample data yang digunakan dalam penelitian ini.

2.3 *Random Forest*

Random Forest adalah salah satu metode *supervised learning* sebagai pengembangan dari konsep *Decision Tree*. *Decision Tree* termasuk pohon klasifikasi tunggal maka pada *Random Forest* dibuatkan banyak pohon untuk menentukan hasil prediksinya [4]. *Random Forest* merupakan pohon keputusan dengan menggunakan strategi ensemble (kumpulan) bagging (sub-dataset) dengan pengacakan variabel input yang digunakan untuk membagi node. Menurut Braiman, strategi ensemble bagging dapat mengatasi masalah overfitting yang terjadi jika data train yang dimiliki berukuran kecil. Pada setiap node (akar) dilakukan pengambilan secara acak diantara semua variabel independen sebagai pemilah terbaik dalam pembentukan pohon sehingga diharapkan menghasilkan prediksi yang lebih akurat.

Pada tahun 2012, Thais Mayumi Oshiro, Pedro Santoro Perez, dan Jos´e Augusto Baranauskas melakukan peneliitian untuk mengetahui hubungan antara jumlah pohon yang digunakan dengan hasil yang diperoleh. Hasil penelitian tersebut menyatakan bahwa hampir tidak ada literatur yang menunjukkan berapa banyak pohon yang harus digunakan dalam membangun sebuah model *Random Forest* untuk memperoleh hasil yang maksimal. Oleh karena itu, jumlah pohon yang banyak tidak dapat menjadi tolak ukur bahwa kinerja model *Random Forest* akan lebih baik dibandingkan dengan jumlah pohon yang lebih sedikit [10].

2.4 Evaluasi Kinerja

Evaluasi kinerja dilakukan untuk menguji atau mengevaluasi kebenaran dan mengukur kinerja yang diinginkan pada sistem yang dibuat. Teknik pengujian yang digunakan pada penelitian ini adalah *Root Mean Squared Error* (RMSE), *Mean Absolute Percentage Error* (MAPE) dan *koefisien determinasi* (R²), untuk mengukur ketepatan ramalan untuk mengetahui hasil Algoritma *Random Forest* Regression dalam memprediksi hasil panen padi di Desa Minanga. RMSE yaitu besarnya angka kesalahan antara nilai aktual dan nilai prediksi yang telah ditentukan. Rumus untuk mencari nilai RMSE yaitu:

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (Y_i - Y_n)^2}, \tag{1}$$

Semakin rendah nilai MAPE, maka semakin baik model yang dihasilkan karena tingkat akurasinya akan tinggi. Nilai Akurasi diperoleh berdasarkan persamaan

$$Akurasi = 100\% - MAPE, \tag{2}$$

sedangkan nilai MAPE ditunjukkan pada persamaan

$$MAPE = \frac{1}{N} \sum_{i=1}^N \frac{Y_i - Y_n}{Y_i} \times 100\% \tag{3}$$

dimana,

MAPE : Mean Absolute Percentage Error

Y_n : nilai prediksi

Y_i : nilai aktual

N : jumlah dataset sampel

3. HASIL DAN PEMBAHASAN

Data testing yang digunakan pada penelitian ini berjumlah 90 data. Adapun sampel data hasil pengujian dengan membandingkan nilai aktual dan prediksi dapat dilihat pada tabel 3. Nilai aktual adalah nilai sebenarnya yang diperoleh dari data yang digunakan, sedangkan nilai prediksi adalah nilai yang diperoleh dari hasil pengimplementasian algoritma *Random Forest*. Kedua nilai tersebut dibandingkan untuk melihat seberapa baik kinerja dari algoritma yang digunakan.

Tabel 3 Perbandingan hasil prediksi dengan data aktual

No	Nilai Aktual	Nilai Prediksi
1	3,595	3,554
2	2,208	2,203
3	2,655	2,741
4	2,125	2,134
5	3,688	3,725
6	2,411	2,414
7	2,172	2,179
8	2,31	2,304
9	1,533	1,623
10	5,544	5,290

Tabel 3 merupakan perbandingan nilai prediksi data testing, dengan data asli dari hasil panen padi Desa Minanga. Berdasarkan tabel tersebut dapat diketahui bahwa nilai hasil prediksi yang dihasilkan oleh model *Random Forest* tidak berbeda jauh dari data aktualnya. Ketika data hasil prediksi mendekati data aktualnya, maka model tersebut masuk dalam kriteria regresi yang baik.

Model *Random Forest* regression digunakan untuk mendapatkan nilai prediksi dengan nilai akurasi terbaik. Penentuan jumlah pohon sangat berpengaruh pada nilai akurasi, dimana semakin kecil nilai

semakin baik model prediksi yang dihasilkan. Hasil akurasi diperoleh berdasarkan persamaan (2). Adapun hasil perbandingan nilai MAPE berdasarkan jumlah pohon dapat dilihat pada tabel 4.

Tabel 4 Nilai MAPE berdasarkan jumlah pohon

Jumlah Pohon	MAPE (%)
5	5.126
9	4.884
19	5.155
29	5.139
39	5.072
50	5.093
100	5.207

Dengan membandingkan jumlah pohon yang terbentuk dengan nilai error terkecil pada MAPE, akan menjadi jumlah pohon terbaik untuk digunakan dalam pembentukan model prediksi. Berdasarkan tabel 4, maka jumlah pohon dengan nilai error terkecil yaitu 9 pohon (*n estimators*) dengan nilai MAPE yaitu 4,884% dibandingkan dengan jumlah pohon yang lainnya. Hal tersebut sejalan dengan penelitian sebelumnya [10] yang melakukan analisis terhadap jumlah pohon pada algoritma *random forest* yang menyatakan bahwa jumlah pohon yang banyak tidak selalu menghasilkan akurasi prediksi yang tinggi.

Tabel 5 Perbandingan hasil akurasi

Ratio training testing (%)	Jumlah data		Hasil akurasi pada n pohon (%)						
	Train	Test	5	9	19	29	39	50	100
60 : 40	178	120	94.48	94.53	94.53	94.55	94.55	94.63	94.77
70 : 30	208	90	94.87	95.11	94.84	94.86	94.42	94.90	94.79

Dari tabel perbandingan akurasi berdasarkan dua ratio data yang berbeda, diketahui bahwa ratio data 70:30 dengan jumlah pohon 9 memberikan hasil yang lebih baik dibandingkan ratio data lainnya dengan akurasi sebesar 95.11%.

Tabel 6 Evaluasi kinerja

RMSE	R ²	MAPE	Akurasi
0,250	0,99	4,884 %	95,11%

Model *Random Forest* yang dibentuk dengan menggunakan percabangan 9 pohon didapatkan RMSE yaitu 0,250 dan nilai R² yaitu 0,99, yang merupakan nilai untuk menunjukkan besarnya pengaruh dari variabel prediktor (independen) terhadap variabel respon (dependen). Nilai R² pada model ini masuk dalam kategori atau interpretasi koefisien sangat kuat karena mendekati angka 1, yaitu 0,99. Sedangkan untuk nilai MAPE yang dihasilkan pada model *Random Forest* yaitu 4,88%, ini menunjukkan bahwa kemampuan model peramalan sangat baik (nilai MAPE < 10%). Nilai hasil prediksi MAPE diambil dari 9 pohon percabangan yang menunjukkan rata-rata dari persentase keseluruhan selisih antara data aktual dan data hasil prediksi. Semakin kecil nilai MAPE, maka semakin kecil tingkat kesalahan yang dihasilkan oleh model dan tingkat akurasi semakin tinggi, yaitu 95,11%

4. KESIMPULAN

Sistem yang dibangun dengan menggunakan algoritma *Random Forest* mampu melakukan prediksi hasil panen padi berdasarkan 7 variabel. Jumlah pohon yang paling optimal digunakan sebanyak 9 pohon dengan nilai akurasi tertinggi yang diperoleh adalah 95,11% , Nilai MAPE pada model ini yaitu 4.884%, nilai RMSE yaitu 0,250 dan nilai R² yaitu 0.99 dalam memprediksi hasil panen padi di Desa Minanga.

DAFTAR PUSTAKA

- [1] A. Basit, "Implementasi Algoritma Naive Bayes Untuk Memprediksi Hasil Panen Padi," *J. Tek. Inform. Kaputama*, vol. 4, no. 2, pp. 208–213, 2020.

- [2] S. S. Defiyanti Muhammad; Dermawan, Budi Arif, “Algoritma C4.5 Untuk Memprediksi Produktivitas Tanaman Padi Di Karawang,” *Inform. Pertan.*, no. Vol 28, No 2 (2019): DESEMBER 2019, pp. 103–110, 2019, [Online]. Available: <http://ejurnal.litbang.pertanian.go.id/index.php/IP/article/downloadSuppFile/9923/827>
- [3] H. W. Herwanto, T. Widiyaningtyas, and P. Indriana, “Penerapan Algoritme Linear Regression untuk Prediksi Hasil Panen Tanaman Padi,” *J. Nas. Tek. Elektro dan Teknol. Inf.*, vol. 8, no. 4, p. 364, 2019, doi: 10.22146/jnteti.v8i4.537.
- [4] I. L. Mulyahati, “Implementasi Machine Learning Prediksi Harga Sewa Apartemen Menggunakan Algoritma Random Forest Melalui Framework Website Flask Python,” Universitas Islam Indonesia, 2020. [Online]. Available: <papers3://publication/uuid/143B9B2C-186A-4872-8DC1-638AAC4D36C3>
- [5] M. Mirqoatul Mafa’atiah, “Implementasi Artificial Intelligence Untuk Memprediksi Harga AIRBNB Menggunakan Metode Random Forest Dan Penerapan Web Application Menggunakan Flask,” Universitas Islam Indonesia, 2020. doi: 10.53730/ijhs.v6ns6.9986.
- [6] i. Y. Supartha, g. Wijaya, and g. M. Adnyana, “Aplikasi Jenis Pupuk Organik pada Tanaman Padi Sistem Pertanian Organik,” *E-Jurnal Agroekoteknologi Trop.*, vol. 1, no. 2, pp. 98–106, 2012.
- [7] F. Azizah, S. Suwarsito, and E. Sarjanti, “Pengaruh Pola Curah Hujan Terhadap Produktivitas Padi di Kecamatan Bukateja Kabupaten Purbalingga,” *Sainteks*, vol. 18, no. 1, p. 1, 2021, doi: 10.30595/sainteks.v18i1.10567.
- [8] C. V Donggulo, I. M. Lapanjang, and U. Made, “Growth and Yield of Rice (*Oryza sativa* L.) under Different Jajar Legowo System and Planting Space,” *J. Agrol.*, vol. 24, no. 1, pp. 27–35, 2017, [Online]. Available: <http://jurnal.untad.ac.id/jurnal/index.php/AGROLAND/article/view/8569>
- [9] S. Saadah and H. Salsabila, “Prediksi Harga Bitcoin Menggunakan Metode Random Forest: (Studi Kasus: Data Acak Pada Masa Pandemic Covid-19),” *J. Komput. Ter.*, vol. 7, no. 1 SE-, pp. 24–32, Jun. 2021, doi: 10.35143/jkt.v7i1.4618.
- [10] T. M. Oshiro, P. S. Perez, and J. A. Baranauskas, “How many trees in a random forest?,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 7376 LNAI, no. July 2012, pp. 154–168, 2012, doi: 10.1007/978-3-642-31537-4_13.