

Klasifikasi Data untuk Prediksi Hasil Panen Tanaman Padi di Wilayah Kabupaten Sukoharjo Menggunakan Algoritma C4.5

Siti Rokhmah*¹, Ari Susilowati², Mutiara Intan P³

^{1,2,3} Informatika, Institut Teknologi Bisnis AAS Indonesia Sukoharjo,
e-mail: *sitirokhmah.itbaas@gmail.com, arisusilowati656@gmail.com,
mutiaraintanp34@gmail.com

Abstrak

Indonesia merupakan salah satu negara penghasil beras terbesar di dunia. Penduduk Indonesia, hampir 95% mengonsumsi beras sebagai bahan pangan pokok wajib, sehingga setiap tahunnya permintaan beras semakin meningkat seiring dengan bertambahnya jumlah penduduk. Sukoharjo dikenal dengan penghasil tanaman padi di wilayah Jawa Tengah. Untuk melihat terpenuhinya kebutuhan beras di wilayah Sukoharjo dan sekitarnya maka diperlukan adanya prediksi hasil panen padi di wilayah Sukoharjo, selain itu hasil prediksi juga dapat dijadikan dasar dalam upaya peningkatan hasil panen padi di wilayah Sukoharjo untuk periode berikutnya. Dalam penelitian ini digunakan algoritma C.45 dalam klasifikasi data untuk prediksi hasil panen di wilayah Kabupaten Sukoharjo. Data yang digunakan adalah data statistik pertanian yang didapat dari website sukoharjo.bps.go.id. Data-data tersebut kemudian dianalisis dengan teknik data mining, sehingga menghasilkan data training yang siap untuk diolah dengan algoritma C.45, pengujian dilakukan dengan split validation untuk menghasilkan validitas klasifikasi data. Hasil dari penelitian ini berupa pola prediksi dengan pohon keputusan yang dapat dijadikan dasar prediksi dalam memperkirakan hasil panen pada periode selanjutnya.

Kata kunci—3-5 panen padi, klasifikasi, algoritma C.45, Kabupaten Sukoharjo

1. PENDAHULUAN

Pertanian di Indonesia memiliki peran yang sangat strategis, sebagian besar penduduk Indonesia menggantungkan hidupnya pada sektor pertanian. Indonesia menjadi salah satu negara yang maju pada sektor pertanian, sehingga menjadikan Indonesia sebagai penghasil padi urutan ketiga di dunia [1]. Salah satu daerah yang memiliki potensi besar bagi usaha pertanian padi adalah Kabupaten Sukoharjo. Hal tersebut didukung dengan sistem irigasi dimana merupakan kawasan yang berada dalam aliran sungai Bengawan Solo sehingga mampu memenuhi kebutuhan air dalam produksi pertanian khususnya padi. Berdasarkan data statistik tahun 2017, hasil panen yang diproduksi Kabupaten Sukoharjo sebesar 3.926.600 kuintal dengan luas panen 52,596 ha dan produktivitas 74,66 kg/Ha. Hal tersebut menunjukkan dengan luas lahan yang sempit, Kabupaten Sukoharjo dapat menghasilkan hasil panen tanaman padi yang lebih besar dibandingkan dengan daerah-daerah yang lain yang memiliki luas yang lebih besar [2].

Pemenuhan kebutuhan beras di wilayah Indonesia bergantung kepada tingkat produktivitas hasil panen padi di wilayah Indonesia, dimana penyumbang hasil panen padi terbesar adalah di wilayah Jawa, dan salah satunya adalah Kabupaten Sukoharjo. Menurunnya produktivitas hasil panen padi berpengaruh terhadap pemenuhan kebutuhan beras dan harga penjualan beras, sehingga diperlukan prediksi dalam menentukan potensi hasil panen padi [3].

Pada penelitian ini dilakukan prediksi terhadap hasil panen di wilayah Sukoharjo yang merupakan salah satu kabupaten yang menghasilkan padi terbesar di wilayah Jawa Tengah. Pola prediksi dilakukan dengan memanfaatkan teknik klasifikasi data pada data mining dengan menggunakan algoritma C.45. Algoritma C.45 merupakan algoritma yang digunakan untuk menghasilkan prediksi dengan membuat pohon keputusan [4]. Penelitian ini bertujuan untuk memberikan prediksi tingkat produktivitas panen padi di wilayah Kabupaten Sukoharjo, sehingga hasil prediksi tersebut dapat dijadikan sebagai dasar dalam upaya evaluasi untuk meningkatkan hasil panen di wilayah Sukoharjo, sehingga stabilitas harga dan kebutuhan bahan pokok beras dapat terjamin dengan baik.

Penelitian terkait pemanfaatan data mining untuk prediksi panen telah banyak dilakukan baik dengan metode klasifikasi maupun dengan metode klusterisasi. Salah satu penelitian tentang prediksi produksi padi adalah penelitian yang dilakukan oleh Rudi Hariyanto dan Anang Aris Widodo, dalam penelitian tersebut dilakukan penelitian klasifikasi data untuk memprediksi hasil panen padi dengan menggunakan algoritma *naïve bayes*, pada penelitian ini dilakukan klasifikasi berdasarkan fisiologi tanaman padi untuk memprediksi hasil panen [5]. Penelitian lain adalah penelitian tentang pemanfaatan data mining untuk memprediksi produktivitas pertanian padi, metode yang digunakan adalah metode *K-nearest neighbour*, pada penelitian ini di analisis faktor-faktor penyebab penurunan produktivitas padi, kemudian dilakukan klasifikasi data untuk mendapatkan prediksi [6]. Selain itu penelitian pemanfaatan data mining juga digunakan untuk memprediksi hasil panen lain, diantaranya adalah prediksi jumlah hasil panen sawit menggunakan metode *naïve bayes*, pada penelitian ini dilakukan prediksi dengan teknik klasifikasi data dengan menganalisis faktor-faktor yang mempengaruhi hasil panen sawit untuk menghasilkan pola prediksi [7]. Pada penelitian ini merupakan pengembangan dari penelitian sebelumnya dengan mengembangkan penggunaan algoritma dan analisis hasil panen di wilayah Kabupaten Sukoharjo.

2. METODE PENELITIAN

Penelitian ini dilakukan dengan beberapa tahapan penelitian, beberapa tahapan dalam penelitian ini adalah :

2.1 Studi pustaka

Dalam penelitian ini terlebih dahulu dilakukan studi terhadap referensi baik berupa jurnal, buku maupun artikel lain yang relevan. Adapun jurnal yang dijadikan referensi diantaranya adalah jurnal tentang potensi pertanian di Indonesia khususnya di Kabupaten Sukoharjo, data mining, teknik prediksi dan jurnal terkait algoritma C.45.

2.2 Pengumpulan data.

Data yang digunakan dalam penelitian ini adalah data hasil panen padi di wilayah Sukoharjo pada tahun 2021 yang diperoleh dari website badan statistik Kabupaten Sukoharjo dengan alamat <https://www.sukoharjo.bps.go.id>. Setelah data-data tersebut terkumpul kemudian dilakukan pra pemrosesan untuk menghasilkan data training dan data testing [8].

2.3 Pengolahan data

Pengolahan data menggunakan teknik klasifikasi data mining. Analisa data mining meliputi tahapan dalam *Knowledge Discovery in Database* (KDD). Dalam tahapan KDD melibatkan algoritma untuk mengeksplorasi data, mengembangkan dan menemukan pola data untuk melakukan prediksi di masa yang akan datang [9]

2.4 Analisis data mining

Pada analisis data mining dilakukan pengamatan data, data yang diperoleh adalah data luas lahan, curah hujan, kelengkapan pertanian dan hasil panen padi, data curah hujan dan data irigrasi berdasarkan kecamatan di Kabupaten Sukoharjo. Faktor-faktor yang mempengaruhi hasil panen di wilayah sukoharjo diantaranya

1. Kondisi cuaca atau curah hujan
2. Luas lahan pertanian yang ditanami padi
3. Kelengkapan alat pertanian

Dari data dan faktor yang mempengaruhi tersebut kemudian dilakukan analisis untuk membangun data training. Data training merupakan kumpulan data hasil survey tentang panen padi di wilayah Kabupaten Sukoharjo. Hasil analisis tersebut dirangkai menjadi atribut-atribut yang diambilkan dari data hasil survey dan faktor-faktor yang mempengaruhi. Tabel hasil analisis dapat dilihat pada tabel 1.

Tabel 1. Faktor yang mempengaruhi hasil panen padi

Aribut	Intances	keterangan
Luas tanah	Luas	>5000 ha
	Sedang	3000-5000 ha
	Kecil	<20000
Ketersediaan alat	Baik	>200
	Cukup	50-200
	kurang	<50
Curah hujan	Rendah	0-100 mm
	Cukup	100 mm – 300
	Tinggi	>300

2.5 Algoritma C.45

Algoritma C.45 merupakan yang digunakan untuk melakukan prediksi dengan membentuk pohon keputusan. Algoritma C.45 merubah fakta menjadi pohon keputusan yang merepresentasikan aturan-aturan agar mudah dipahami. Algoritma C.45 merupakan pengembangan dari algoritma ID3 [10]. Secara umum algoritma C.45 memiliki pola sebagai berikut :

1. Memilih atribut sebagai akar
2. Membuat cabang untuk tiap-tiap nilai.
3. Membagi kasus menjadi cabang pohon
4. Ulangi proses untuk setiap cabang sampai semua kasus pada cabang memiliki kelas.

Untuk memilih atribut dengan akar, didasarkan pada nilai gain tertinggi dari atribut - atribut yang ada. Untuk menghitung gain digunakan rumus seperti tertera dalam persamaan 1

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

Keterangan:

S: Himpunan kasus

A: Atribut

N: Jumlah partisi atribut A

- | Si : Jumlah kasus pada partisi ke-i
- | S |: Jumlah kasus dalam S

Sehingga akan diperoleh nilai gain dari atribut yang paling tertinggi. Gain adalah salah satu *attribute selection measure* yang digunakan untuk memilih *test attribute* tiap node pada *tree*. Atribut dengan *information gain* tertinggi dipilih sebagai *test attribute* dari suatu node. Sementara itu, penghitungan nilai entropi dapat dilihat pada persamaan 2.

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

- Keterangan:
- S: Himpunan kasus
 - A: Atribut
 - N: Jumlah partisi S
 - Pi: Proporsi dari Si terhadap

3. HASIL DAN PEMBAHASAN

3.1 Analisis data training

Dari hasil analisis pada proses KDD diperoleh data training yang merupakan hasil analissi data dan faktor yang mempengaruhi, data training dapat dilihat pada tabel 2.

Tabel 2. Data training faktor dan hasil panen

kecamatan	luas	Hujan	alat	Hasil panen
010. W e r u	sedang	Tinggi	cukup	baik
020. B u l u	kecil	Sedang	cukup	kurang
030. Tawang Sari	sedang	Sedang	cukup	baik
040. Sukoharjo	luas	Rendah	baik	baik
050. Nguter	luas	Sedang	baik	baik
060. Bendosari	luas	Sedang	baik	baik
070. Polokarto	luas	Sedang	baik	baik
080. Mojolaban	luas	tinggi	cukup	baik
090. Grogol	kecil	sedang	cukup	kurang
100. B a k i	sedang	rendah	cukup	kurang
110. G a t a k	sedang	rendah	cukup	baik
120. Kartasura	kecil	sedang	kurang	kurang

3.2 Penentuan simpul node pohon keputusan

Dengan menggunakan rumus perhitungan entropi, maka diperoleh nilai gain tertinggi yang digunakan untuk menentukan node awal. Nilai gain dari perhitungan entropi yang paling besar di jadikan sebagai awal node [11].

- a. Tahap 1 menentukan node awal

Untuk menentukan dilakukan dengan menghitung nilai entropi pada masing-masing atribut, setelah diperoleh nilai entropi kemudian dilakukan perhitungan untuk memperoleh nilai gain. Gain dengan nilai tertinggi dijadikan sebagai node awal dalam pohon keputusan.

Menghitung entropi

$$\text{Entropy (S)} = \sum_{i=1}^n -p_i * \log_2 p_i$$

- Entropy (total)
 - = $- 8/12 * \log_2 (8/12) + -4/12 * \log_2 (4/12)$
 - = 0,918295834

- Entropi luas lahan
 - Entropy(luas)
 - = $- 5/5 * \log_2 (5/5) + -0/5 * \log_2 (0/5)$
 - = 0
 - Entropy(sedang)
 - = $- 3/4 * \log_2 (3/4) + -1/4 * \log_2 (1/4)$
 - = 0,811278124
 - Entropy(kecil)
 - = $- 0/3 * \log_2 (0/3) + -3/3 * \log_2 (3/3)$
 - = 0

- Entropi curah hujan
 - Entropy(tinggi)
 - = $- 2/2 * \log_2 (2/2) + -0/2 * \log_2 (0/2)$
 - = 0
 - Entropy(sedang)
 - = $- 4/7 * \log_2 (4/7) + -4/7 * \log_2 (4/7)$
 - = 0,985228136
 - Entropy(rendah)
 - = $- 2/3 * \log_2 (2/3) + -1/3 * \log_2 (1/3)$
 - = 0,918295834

- Entropi kelengkapan pertanian
 - Entropy(baik)
 - = $- 4/4 * \log_2 (4/4) + -0/4 * \log_2 (0/4)$
 - = 0
 - Entropy(cukup)
 - = $- 4/7 * \log_2 (4/7) + -3/7 * \log_2 (3/7)$
 - = 0,922691339
 - Entropy(kurang)
 - = $- 2/2 * \log_2 (2/2) + -0/2 * \log_2 (0/2)$
 - = 0

Menghitung gain dari atribut

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

- gain untuk atribut luas lahan

Gain (luas lahan) = (0,918295834 – (2/12) * 0) + (7/12) *0,811278124) + (3/12) * 0) = **0,647869793**

- gain untuk atribut curah hujan

Gain (luas lahan) = (0,918295834 – (5/12) * 0) + (4/12) *0,985228136) + (5/12) *0,918295834) = **0,114005463**

- gain untuk atribut kelengkapan pertanian

Gain (luas lahan) = (0,918295834 – (2/12) * 0) + (7/12) *0,922691339) + (3/12) * 0) = 0,380059219

Untuk hasil keseluruhan perhitungan pada tahap 1, dapat melihat pada tabel 3

Tabel 3 Hasil perhitungan tahap 1

atribut		Jumlah Kasus	baik	kurang	entropi	Gain
	Total data	12	8	4	0,918295834	
Luas lahan						0,647869793
	luas	5	5	0	0	
	sedang	4	3	1	0,811278124	
	kecil	3	0	3	0	
Curah hujan						0,114005463
	tinggi	2	2	0	0	
	sedang	7	4	3	0,985228136	
	rendah	3	2	1	0,918295834	
Alat						0,380059219
	baik	4	4	0	0	
	cukup	7	4	3	0,922691339	
	kurang	1	0	1	0	

Dari perhitungan nilai tersebut didapatkan nilai gain tertinggi adalah luas dengan nilai gain **0,647869793**. Sehingga luas lahan dijadikan sebagai node awal dalam pohon keputusan

- b. menentukan simpul node tahap 2

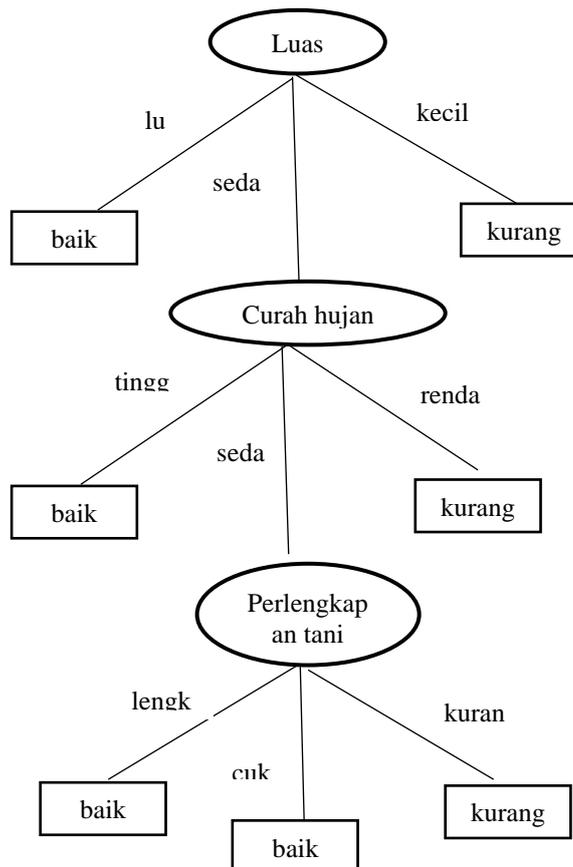
Setelah menentukan node awal selanjutnya dilakukan perhitungan kembali seperti pada tahap pertama untuk menentukan akar pohon berikutnya. Perhitungan tahap 2 dapat dilihat pada tabel

Tabel 4. Hasil perhitungan tahap 2

atribut		Jumlah Kasus	baik	kurang	entropi	Gain
	Total	4	3	1	0,811278124	
curah hujan						0,311278124
	tinggi	1	1	0	0	
	sedang	2	1	1	1	
	rendah	1	0	1	0	
alat						0
	baik	0	0	0	0	
	cukup	4	3	1	0,811278124	
	kurang	0	0	0	0	

Dari hasil perhitungan tahap kedua diperoleh nilai gain yang lebih tinggi adalah curah hujan dengan nilai gain **0,311278124**.

Berdasarkan hasil perhitungan nilai entropi maka didapatkan pohon keputusan. Pohon keputusan model prediksi panen padi di Kabupaten Sukoharjo dapat dilihat pada gambar 1.



Gambar 1. Pohon keputusan klasifikasi data hasil panen

3.3 Penentuan simpul node pohon keputusan

Pengujian dilakukan dengan menggunakan data training yang berjumlah sama dengan data yang berbeda tahun. Pengujian dilakukan dengan menggunakan *software* rapidmanner, dengan melakukan pengujian terhadap model, pengujian dilakukan dengan model *split validation* dan diperoleh hasil yang terdapat pada gambar 2.

accuracy: 100.00%

	true baik	true kurang	class precision
pred. baik	2	0	100.00%
pred. kurang	0	1	100.00%
class recall	100.00%	100.00%	

Gambar 2. Hasil pengujian dengan Rapid manner

Performance model yang dihasilkan adalah 100 %, hal tersebut dikarenakan tidak banyak perubahan kondisi dari tahun ke tahun terhadap variabel yang mempengaruhi hasil.

4. KESIMPULAN

Kesimpulan dari penelitian ini adalah

- Faktor yang mempengaruhi panen padi di wilayah Kabupaten Sukoharjo diantaranya luas lahan, cuah hujan dan kelengkapan pertanian.
- Diperoleh nilai prediksi dengan aturan sebagai berikut
 R1: IF luas lahan=luas Then hasil = baik
 IF luas lahan = kecil THEN hasil=kurang
 R2: IF luas lahan =sedang^curah hujan=tinggi THEN hasil=baik,
 IF luas lahan = sedang ^curah hujan=rendah THEN hasil= kurang
 R3: IF luas lahan =sedang^curah hujan=tinggi ^ alat = baik THEN hasil = baik
 IF luas lahan =sedang^curah hujan=sedang ^ alat = cukup THEN hasil = baik
 R4: IF luas lahan =sedang^curah hujan=rendah ^ alat = kurang THEN hasil = kurang
- Hasil model prediksi dengan algoritma C.45 diuji dengan split validation dengan tingkat performance 100 % terhadap data yang sama dengan tahun yang berbeda.

5. SARAN

Kami menyadari penelitian ini masih banyak kekurangan. Harapan untuk penelitian selanjutnya lebih memfokuskan penelitian dengan teknik dan algoritma pengolahan data yang berbeda dan dengan teknik pengumpulan data yang lebih baik dengan melakukan analisa dan observasi secara langsung.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada semua pihak yang telah membantu dalam kegiatan penelitian ini.

DAFTAR PUSTAKA

- [1] M. N. Fawaiq, A. Jazuli, and M. M. Hakim, "Prediksi Hasil Pertanian Padi Di Kabupaten Kudus Dengan Metode Brown'S Double Exponential Smoothing," *JUPI (Jurnal Ilm. Penelit. dan Pembelajaran Inform.*, vol. 4, no. 2, p. 78, 2019, doi: 10.29100/jipi.v4i2.1421.
- [2] D. W. Prabowo *et al.*, "1* 1* , 2 , 2," vol. 5, pp. 145–155, 2021.
- [3] D. P. Lamondjong and M. Hardjianto, "Data Mining untuk Memprediksi Jumlah Penjualan Hasil Pertanian Menggunakan Algoritma Forecasting (Studi Kasus: Dinas Pertanian Kabupaten Banggai)," *Respati*, vol. XVI, pp. 94–102, 2021, [Online]. Available: <http://jti.respati.ac.id/index.php/jurnaljti/article/view/405%0Ahttp://jti.respati.ac.id/index.php/jurnaljti/article/download/405/351>
- [4] N. V. B. Siahaan, P. Poningsih, D. Suhendro, D. Hartama, and S. Suhada, "Penerapan Data Mining Algoritma C4.5 Terhadap Prediksi Faktor Menurunnya Hasil Panen Padi," *Jurasik (Jurnal Ris. Sist. Inf. dan Tek. Inform.*, vol. 7, no. 1, p. 27, 2022, doi: 10.30645/jurasik.v7i1.412.
- [5] H. Panen, P. Keputusan, and N. Bayes, "Klasifikasi Hasil Prediksi Panen Padi Berdasarkan Fisiologis Menggunakan Metode," no. Ciastech, pp. 237–244, 2019.
- [6] S. Maesaroh and Kusriani, "Sistem Prediksi Produktifitas Pertanian Padi Menggunakan Data Mining," *J. Energi*, vol. 7, no. 2, pp. 25–30, 2017, [Online]. Available: eprints.dinus.ac.id/16925/1/jurnal_16115.pdf
- [7] W. Ananda, M. Safii, and M. Fauzan, "Prediksi Jumlah Hasil Panen Sawit Menggunakan Algoritma Naive Bayes," *TIN Terap. Inform. Nusant. Vol*, vol. 1, no. 10, pp. 513–519, 2021.
- [8] sukoharjo.bps.go.id, - *Data Statistik Pertanian dan Kehutanan*. 2020, p. sukoharjo.bps.go.id. [Online]. Available: <https://pertanian.kulonprogokab.go.id/detil/491/data-statistik-pertanian-dan-kehutanan>
- [9] I. Budiman and R. Ramadina, "Penerapan Fungsi Data Mining Klasifikasi untuk Prediksi Masa Studi Mahasiswa Tepat Waktu pada Sistem Informasi Akademik Perguruan Tinggi," *Ijccs*, vol. x, No.x, no. 1, pp. 1–5, 2015.
- [10] B. Novianti, T. Rismawan, and S. Bahri, "Implementasi Data Mining Dengan Algoritma C4.5 Untuk Penjurusan Siswa (Studi Kasus: Sma Negeri 1 Pontianak)," *J. Coding, Sist. Komput. Untan*, vol. 04, no. 3, pp. 75–84, 2016.
- [11] Y. Mardi, "Data Mining : Klasifikasi Menggunakan Algoritma C4.5," *Edik Inform.*, vol. 2, no. 2, pp. 213–219, 2017, doi: 10.22202/ei.2016.v2i2.1465.