

Penerapan Algoritme C4.5 Pada Klasifikasi Produksi Ubi Jalar di Pulau Jawa

Seno Setiyawan¹, Yuliana Susanti², Tri Atmojo Kusmayadi³

^{1,3} Program Studi Matematika FMIPA, Universitas Sebelas Maret

² Program Studi Statistika FMIPA, Universitas Sebelas Maret

*Email: senosetiyawan23@gmail.com

Abstrak

Keywords:
Produksi Ubi Jalar,
Algoritme C4.5,
Klasifikasi, Pohon
Keputusan

Pada tahun 2015, hasil produksi ubi jalar nasional mencapai 2.297.634 ton, atau turun 3,57% dari tahun 2014. Hal ini disebabkan berkurangnya lahan pertanian ubi jalar. Impor ubi jalar menjadi pilihan jika ketersediaan produksi ubi jalar rendah. Maka diperlukan usaha untuk meningkatkan produksi ubi jalar dengan membuat klasifikasi produksi ubi jalar di pulau Jawa. Produksi ubi jalar di pulau Jawa dapat diklasifikasikan berdasarkan nilai median, hal ini dapat dilihat pada Tabel 1, Sedangkan kategori variabel lainnya dapat dilihat pada Tabel 2. Penelitian ini menggunakan metode Algoritme C4.5. Metode klasifikasi dari data mining yang digunakan untuk mengkonstruksikan pohon keputusan serta mampu mengatasi nilai yang hilang (missing value), mengatasi data bertipe kontinu, dan melakukan pemangkasan pohon (prunning trees). Selain itu, dengan menggunakan algoritme C4.5 dapat diketahui nilai ketepatan klasifikasi. Metode Algoritme C4.5 menunjukkan bahwa terdapat 10 aturan klasifikasi dan nilai tingkat kesalahan adalah 7,21% dan ketepatan klasifikasi sebesar 92,85% dalam memprediksi klasifikasi produksi ubi jalar di kabupaten atau kota di pulau Jawa.

1. PENDAHULUAN

Ubi jalar merupakan komoditas sumber karbohidrat utama, setelah padi, jagung dan ubi kayu. Ubi jalar dikonsumsi sebagai makanan tambahan atau sampingan bahkan dianggap sebagai makanan kampung, kecuali di Irian Jaya dan Maluku, ubi jalar digunakan sebagai makanan pokok [1]. Setiap tahun Indonesia bergelut dengan persoalan pangan karena sangat tergantung dengan beras. Padahal sumber pangan di negeri ini sangat banyak, salah satunya ubi jalar.

Menurut Lingga [2], ubi jalar dapat dimanfaatkan sebagai pengganti makanan pokok karena merupakan sumber kalori yang

efisien. Masyarakat lokal mengonsumsi umbi-umbian ini sebagai makanan pengganti nasi karena memiliki kandungan karbohidrat yang tinggi yaitu 20,1 gram per 100 gram atau setara dengan energi 86 kkal [3].

Berdasarkan data BPS [4] pada tahun 2015, hasil produksi ubi jalar nasional mencapai 2.297.634 ton, atau turun 3,57 % dari tahun 2014. Hal ini disebabkan berkurangnya lahan pertanian ubi jalar. Kabupaten atau kota yang menghasilkan produksi ubi jalar di pulau Jawa tersebar di Provinsi Jawa Barat, Jawa Tengah, Jawa Timur, Yogyakarta dan Banten. Produksi ubi jalar ini dipengaruhi oleh beberapa faktor

antara lain luas lahan panen, suhu udara, curah hujan, dan tinggi wilayah. Produksi ubi jalar di pulau Jawa dapat diklasifikasikan berdasarkan jumlah produksi ubi jalar di setiap kabupaten atau kota di pulau Jawa.

Algoritme C4.5 adalah salah satu metode klasifikasi dari data mining yang digunakan untuk mengkonstruksikan pohon keputusan (decision tree). Menurut Witten et al [5], algoritme C4.5 merupakan pengembangan dari ID3 (Iterative Dichotomiser Three) yang mampu mengatasi nilai yang hilang (missing value), mengatasi data bertipe kontinu, dan melakukan pemangkasan pohon (pruning trees). Menurut Khonsari et al. [6], algoritme C4.5 memiliki beberapa kelebihan yaitu algoritme C4.5 dapat menangani atribut kontinu dan diskrit, dapat menangani training data dengan missing value, hasil pohon keputusan C4.5 dapat dipangkas setelah dibentuk dan pemilihan atribut dilakukan dengan menggunakan gain ratio.

Berbeda dengan penelitian penelitian sebelumnya, pada penelitian ini setelah mendapatkan klasifikasi dari aturan algoritme C4.5 setelah itu kita hitung keakuratan nilainya..

2. METODE

Metode Pada penelitian ini menggunakan data sekunder yang diperoleh dari Badan Pusat Statistik (BPS[4]) dan Dinas Pertanian [7]. Data sekunder yang digunakan mengenai produksi ubi jalar dan variabel-variabel yang mempengaruhi produksi ubi jalar di kabupaten atau kota di pulau Jawa pada tahun 2015. Produksi ubi jalar di kabupaten atau kota di pulau Jawa sebagai variabel dependen. Luas lahan panen ubi jalar, suhu udara, curah hujan, dan ketinggian wilayah di setiap kabupaten atau kota di pulau Jawa sebagai variabel independen. Menurut Larose [8] terdapat beberapa langkah dalam membangun pohon keputusan algoritme C4.5. Perhitungan dalam penelitian ini menggunakan Microsoft Excel dan Software WEKA.

Berikut langkah-langkah yang dilakukan untuk mencapai tujuan dalam penelitian.

- a. Mengumpulkan data Produksi dan Faktor - faktor yang mempengaruhi produksi Ubi Jalar di Pulau Jawa sejumlah 98 kabupaten/kota dengan 4 atribut pada masing-masing kota/kabupaten.
- b. Mendeskripsikan data faktor-faktor yang mempengaruhi produksi (S).
- c. Menentukan atribut faktor-faktor produksi (A) yang akan digunakan.
- d. Menghitung jumlah kasus untuk keputusan nilai produksi tinggi dan jumlah kasus untuk keputusan nilai produksi rendah.
- e. Menghitung entropy dari semua kasus dan kasus yang dibagi berdasarkan atribut nilai produksi untuk menunjukkan keberagaman data sampel menggunakan persamaan dimana S adalah himpunan kasus, n adalah banyaknya kelas dan p_i adalah proporsi S_i terhadap S.
- f. Menghitung nilai gain dari tiap-tiap atribut menggunakan persamaan
- g. Menghitung nilai gain ratio dari tiap-tiap atribut menggunakan persamaan (Wei Ji, et al. [9]).
- h. Menetapkan salah satu atribut faktor-faktor produksi dengan nilai gain ratio tertinggi sebagai root.
- i. Membuat cabang untuk tiap-tiap nilai atribut faktor produksi dengan menghapus atribut yang telah terpilih sebagai akar.
- j. Mengulangi langkah 5 hingga 8 sampai semua kasus pada cabang memiliki kelas yang sama.
- k. Menganalisis dan mengambil kesimpulan

3. HASIL DAN PEMBAHASAN

3.1. Deskripsi atribut data

Data hasil produksi ubi jalar di pulau Jawa tahun 2015 terdiri dari 98 kabupaten atau kota. dapat dilihat pada Tabel 1.

Tabel 1. Jumlah Produksi Ubi Jalar di Pulau Jawa Produksi Ubi Jalar

Produksi Ubi Jalar	Jumlah
< 1426 ton	42
≥ 1426 ton	56
Total	98

Berdasarkan Tabel 1. untuk variabel produksi ubi jalar digunakan nilai median hasil produksi ubi jalar di kabupaten atau kota di pulau Jawa. Adapun rincian kategori dari masing-masing variable dapat dilihat pada tabel 2

Tabel 2. Tabulasi kategori setiap variabel

Variable	Kategori	Keterangan
Produksi ubi jalar (Ton)	1. < 1426 ton	Kurang dari 1426 ton
	2. ≥ 1426 ton	Lebih dari 1426 ton
Luas lahan (Ha)	1. < 400 Ha	Sempit (R)
	2. ≥ 400 Ha	Luas (T)
Suhu udara (°C)	1. < 23°C	Rendah (R)
	2. 23°C – 25°C	Sedang (S)
	3. > 25°C	Tinggi (T)
Curah hujan (mm)	1. < 75	Rendah (R)
	2. 75 – 150	Sedang (S)
	3. > 150	Tinggi (T)
Tinggi wilayah (mdpl)	1. < 100	Rendah (R)
	2. 100 – 600	Sedang (S)
	3. > 600	Tinggi (T)

3.2. Analisis Algoritme C4.5

a. Menentukan Node Akar

Langkah pertama yang dilakukan adalah menghitung entropy, gain, split info dan gain ratio dari empat atribut data sebagai nilai awal dari pohon (tree). Hasil perhitungan ditunjukkan pada tabel 3.

Tabel 3. Perhitungan Node 1

Node	Atribut	Jumlah kasus	Produksi rendah	Produksi tinggi	Entropy	Gain	Split info	Gain ratio
1		total	42	56	98	0.985		
	Suhu					0.001	0.795	0.001
	R	3	2	1	0.918			
	S	15	5	10	0.918			
	T	80	35	45	0.985			
	Curah hujan					0.017	0.997	0.017
	R	2	0	2	0			
	S	28	13	15	0.996			
	T	68	29	39	0.984			
	Tinggi wilayah					0.077	1.373	0.056
	R	45	27	18	0.970			
	S	43	13	30	0.884			
	T	10	2	8	0.721			
	Luas lahan					0.252	0.819	0.308
	R	73	42	31	0.983			
	T	25	0	23	0			

Dari Tabel 3 terlihat bahwa atribut dengan nilai gain ratio tertinggi adalah luas lahan yaitu dengan nilai gain ratio sebesar 0,308. Sesuai dengan ketentuan dalam algoritme C4.5, atribut luas lahan terpilih menjadi node akar. Setelah diperoleh node akar dilihat nilai entropy untuk masing-masing rentang skor. Ada dua nilai rentang skor pada atribut Luas lahan yaitu Rendah, dan Tinggi. Kedua nilai rentang skor tersebut perlu dilakukan perhitungan lebih lanjut karena belum mengklasifikasikan produksi ubi jalar.

b. Menentukan Node Cabang

Pada proses penentuan node cabang dipilih berdasarkan gain ratio terbesar setelah menghapus atribut yang telah terpilih sebagai node akar. Setelah mendapatkan atribut luas lahan sebagai node akar, kemudian dilakukan perhitungan untuk mencari node daun sehingga didapatkan suatu kelas. Node daun ini dihitung dari sub kriteria luas lahan yang sama-sama bernilai rendah dan tinggi tetapi kategori hasil masih berbeda. Pada penerapan ini, peneliti mengambil contoh penerapan untuk perhitungan hasil node 2 Luas lahan Rendah.

Tabel 4. Perhitungan node 2 Luas lahan Rendah

Node	Atribut	Jumlah Kasus	Produksi Rendah	Produksi Tinggi	Entropy	Gain	Split Info	Gain Ratio
2	Total	73	42	31		0.983		
	Suhu					0.082	0.848	0.009
	R	3	2	1	0.918			
	S	11	5	6	0.994			
	T	59	35	24	0.974			
	Curah hujan					0.006	0.847	0.007
	R	0	0	0	0			
	S	20	13	7	0.934			
	T	53	29	24	0.993			
	Tinggi wilayah					0.013	1.294	0.01
	R	41	26	15	0.947			
	S	26	13	13	1			
	T	6	3	3	1			

Pada Tabel 4 terlihat bahwa atribut Tinggi wilayah memiliki nilai gain ratio tertinggi yaitu 0,01 sehingga atribut Tinggi wilayah dapat menjadi node cabang dari atribut Luas lahan rendah. Ada tiga nilai rentang skor pada atribut Tinggi wilayah yaitu Rendah, Sedang dan Tinggi. Dari ketiga sub kriteria, sehingga nilai Rendah, Sedang dan Tinggi perlu dilakukan perhitungan lebih lanjut Karena belum mengklasifikasikan Produksi Ubi jalar. Dalam penerapan contoh di sini peneliti menetapkan Tinggi wilayah Sedang sebagai perhitungan node 3. Perhitungan node 3 Luas lahan Rendah Tinggi wilayah Sedang ditunjukkan pada Tabel 5.

Tabel 5. Perhitungan node 3 Luas lahan Rendah Tinggi wilayah Sedang

Node	Atribut	Jumlah Kasus	Produksi Rendah	Produksi Tinggi	Entropy	Gain	Split Info	Gain Ratio
3	Total	26	13	13		1		
	Suhu					0.193	0.845	0.229
	R	1	1	0	0			
	S	4	0	4	0			
	T	21	11	10	0.998			
	Curah hujan					0.024	0.779	0.031
	R	0	0	0	0			
	S	6	4	2	0.918			
	T	20	9	11	0.992			

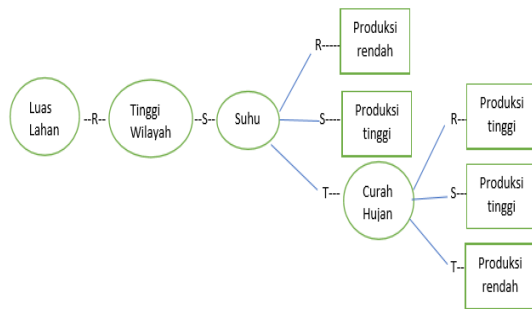
Berdasarkan Tabel 5 dapat diketahui bahwa atribut dengan gain ratio tertinggi adalah 0,229 sehingga atribut suhu dapat menjadi cabang dari node Luas lahan

Rendah Tinggi wilayah Sedang. Pada Tabel 5 atribut suhu mempunyai tiga sub kriteria yaitu Rendah, Sedang, dan Tinggi. Nilai Rendah mengklasifikasikan kasus ke dalam produksi rendah dan nilai Sedang mengklasifikasikan kasus ke dalam produksi tinggi sehingga tidak perlu dilakukan perhitungan lagi. Sedangkan nilai Tinggi perlu dilakukan perhitungan lebih lanjut. Karena belum mengklasifikasikan Produksi Ubi jalar. Dalam penerapan contoh di sini peneliti menetapkan Suhu Tinggi sebagai perhitungan node 4. Perhitungan node 4 Luas lahan Rendah Tinggi wilayah Sedang Suhu Tinggi ditunjukkan pada Tabel 6

Tabel 6. Perhitungan node 3 Luas lahan Rendah Tinggi wilayah Sedang dan Suhu Tinggi.

Node	Atribut	Jumlah Kasus	Produksi Rendah	Produksi Tinggi	Entropy	Gain	Split Info	Gain Ratio
4	Total	21	11	10		0.998		
	Curah hujan					0.0006	0.863	0.0007
	R	0	0	0	0			
	S	6	3	3	1			
	T	15	8	7	0.996			

Pada table 6 nilai gain ratio yaitu 0,0007 pada atribut Curah hujan dengan tiga sub kriteria. Dari ketiga sub kriteria tersebut nilai Tinggi mengklasifikasikan kasus ke dalam produksi rendah sedangkan nilai Rendah dan Sedang mengklasifikasikan kasus ke dalam Produksi tinggi sehingga tidak perlu dilakukan perhitungan lebih lanjut. Pohon keputusan yang terbentuk dari kasus Luas lahan Rendah Tinggi wilayah Sedang Suhu Tinggi dapat dilihat pada Gambar 1.



Gambar 1. Pohon Keputusan Perhitungan Luas lahan Rendah Tinggi wilayah Sedang

3.3. Interpretasi Algoritme C4.5

Dalam penelitian ini pembentukan aturan klasifikasi dari pemodelan algoritme C4.5 dengan pembobot empat atribut dari 98 data terdapat 10 aturan. Eksekusi waktu berdasarkan software WEKA yaitu 0.03 second.

Berikut pembentukan aturan klasifikasi untuk Luas lahan rendah Tinggi wilayah sedang..

- Jika Luas lahan Rendah dan Tinggi wilayah Sedang dan Suhu Rendah, maka Produksi rendah
- Jika Luas lahan Rendah dan Tinggi wilayah Sedang dan Suhu Sedang maka Produksi tinggi
- Jika Luas lahan Rendah dan Tinggi wilayah Sedang dan Suhu Tinggi dan Curah hujan Rendah maka Produksi tinggi
- Jika Luas lahan Rendah dan Tinggi wilayah S dan Suhu Tinggi dan Curah hujan Sedang maka Produksi tinggi
- Jika Luas lahan Rendah dan Tinggi wilayah Sedang dan Suhu Tinggi dan Curah hujan Tinggi maka Produksi rendah

3.4. Pengujian dan Evaluasi

Setelah model pohon keputusan dari data training didapatkan, tahap selanjutnya yaitu dilakukan pengujian dan evaluasi terhadap hasil model pohon keputusan data

training tersebut. Dari hasil software WEKA terlihat bahwa dari 98 data training secara keseluruhan terdapat 10 data dengan klasifikasi benar sehingga diperoleh presentase 92,85%. Hal ini menunjukkan bahwa algoritme C4.5 dapat digunakan untuk memberikan informasi dalam penentuan klasifikasi produksi ubi jalar

4. KESIMPULAN

Berdasarkan hasil pembahasan diperoleh kesimpulan bahwa pohon keputusan dengan algoritme C4.5 dapat digunakan dalam penentuan klasifikasi Produksi ubi jalar di Pulau Jawa. Hal ini didukung oleh hasil pengujian yang menggunakan seluruh data training dengan 10 aturan klasifikasi diperoleh presentase akurasi sebesar 92,85%.

REFERENSI

- Zuraida, N. dan Supriati, Y. Usahatani Ubi Jalar sebagai Bahan Pangan alternatif dan Diversifikasi Sumber Karbohidrat. Buletin AgroBio 4(1):13-2. Balai Penelitian Bioteknologi Tanaman Pangan. Bogor, 2001
- Lingga, P. 2001. Pertanaman Ubi-ubian. Penebar Swadaya. Jakarta.
- Juanda, D. dan Cahyono, B. Budidaya dan Analisa Usaha Tani Ubi Jalar. Kanisius, 2000.
- Badan Pusat Statistik, Production of Paddy Maize and Soybeans, www.bps.go.id/release/Production of Paddy Maize and Soybeans, 2015.
- Witten, Ian. H., and Frank, E., and Hall, M. A., Data Mining: Practical Machine Learning Tools and Techniques, 2nd ed. San Francisco: Morgan Kaufmann Publishers is an imprint of Elsevier, 2011
- Khoonsari, P. E. and Motie, A. R., A Comparison of Efficiency and Robustness of ID3 and C4.5 Algorithms Using Dynamic Test and Training Data Sets, International

- Journal of Machine Learning and Computing, (2012) vol 2 no 5.
- [7] Dinas pertanian, Suhu dan Curah Hujan dalam Wilayah Regional Pulau Jawa 2015, [cited Februari 2017]
- [8] www.distan.go.id/download/outlooksuhudancurahhujan.pdf,
- [9] Larose, D.T.. Discovering Knowledge in Data: An Introduction to Data Mining. John Willey & Sons, Inc. 2005.
- [10] Wei, D. and Wei Ji, A Map Reduce Implementation of C4.5 Decision Tree Algorithm, International Journal of Database Theory and Application, (2014) vol 7 no 1,49