

Penerapan Algoritma C4.5 Dalam Klasifikasi Kualitas Sayur Kol Di Kabupaten Simalungun

Mhd Wendico Herdian*, Riki Winanjaya, Susiani

Program Studi Sistem Informasi, STIKOM Tunas Bangsa

Jalan Jendral Sudirman Blok A, No.1,2 & 3, Pematangsiantar, Sumatera Utara, Indonesia

Email: ^{1,*}herdianwhendico@gmail.com, ²rikiwinanjaya@amiktunasbangsa.ac.id, ³susiani@amiktunasbangsa.ac.id

Email Penulis Korespondensi: herdianwhendico@gmail.com

Abstrak—Kol merupakan tanaman semusim atau lebih yang berbentuk perdu. Salah satu sentra produksi kol berada di kabupaten Simalungun, kecamatan Purba. Kol memiliki prospek permintaan yang baik. Namun faktor cuaca yang tidak menentu dan hama tanaman yang mengancam menjadikan kualitas sayur kol kurang baik sehingga berdampak pada penghasilan petani menjadi tidak menentu. Dalam menentukan sayur kol yang layak dan tidak layak digunakan Metode Datamining dengan Algoritma C4.5. Algoritma C4.5 dapat memberikan pohon keputusan yang mudah diinterpretasikan dan memiliki akurasi yang baik sehingga dari hasil yang diperoleh diharapkan mampu menjadi masukan dan membantu dinas pertanian maupun para petani agar dapat menentukan kualitas sayur kol yang baik sesuai dengan target pasar. Menjadi referensi bagi penelitian selanjutnya yang berkaitan dengan penggunaan Algoritma C4.5.

Kata Kunci: Data mining; Algoritma C4.5; Klasifikasi; Sayur Kol; Dinas Pertanian

Abstract—Cabbage is an annual plant or more in the form of a shrub. One of the centers for cabbage production is in Simalungun Regency, Ancient District. Cabbage has good demand prospects. However, erratic weather factors and plant pests that threaten to make the quality of cabbage vegetables are not good so that the impact on farmers' income is uncertain. In determining the appropriate and unfeasible cabbage, the Datamining Method with the C4.5 Algorithm is used. The C4.5 algorithm can provide a decision tree that is easy to interpret and has good accuracy so that the results obtained are expected to be input and help the agricultural office and farmers to determine the quality of cabbage vegetables that are good according to the target market. Become a reference for further research related to the users of the C4.5 Algorithm.

Keywords: Data mining; C4.5 Algorithm; Classification; Cabbage; Agriculture Of Departmen

1. PENDAHULUAN

Indonesia mempunyai peluang besar dan sangat prospektif untuk mengembangkan budidaya sayuran yang ditunjukkan dengan tersedianya kekayaan potensi sumber daya pertanian berupa lahan yang sangat luas dan belum dimanfaatkan secara optimal[1]. Total luas daratan Indonesia sekitar 190 juta hektare, terbagi atas 123 juta hektare (64,6%) kawasan budidaya dan sisanya 67 juta hektare (35,4%) merupakan kawasan lindung. Dari total luas kawasan budidaya di daratan yang berpotensi untuk areal pertanian seluas 101 juta hektare. Sampai saat ini yang berpotensi untuk pertanian yang sudah dibudidayakan seluas 47 juta hektare[2]. Kol merupakan tanaman semusim atau lebih yang berbentuk perdu. Tanaman kol atau kubis berbatang pendek dan beruas-ruas. Tanaman ini berbatang tunggal dengan akar samping yang sedikit tapi dangkal. Daunnya berbentuk bulat telur dan lunak. Daun yang muncul terlebih dulu menutup daun yang muncul kemudian, demikian seterusnya, hingga membentuk krop daun bulat seperti telur dan padat berwarna putih[3].

Pemilihan Kol merupakan ungkapan selera pribadi konsumen, ditentukan oleh faktor subjektif dan dipengaruhi oleh lokasi, suku bangsa, lingkungan, dan tingkat pendapatan. Salah satu sentra produksi kol Kabupaten Simalungun, yaitu Kecamatan Purba. Meskipun kol memiliki prospek permintaan yang baik tetapi petani masih menghadapi masalah atau kendala. Tingkat keuntungan petani juga tidak pasti tergantung harga jual hasil panen kol. Faktor cuaca yang tidak menentu dan hama tanaman yang mengancam sewaktu-waktu merupakan alasan pokok yang berakibat produksi kol cepat rusak dan tingkat kehilangan hasil tinggi. *Data Mining* disebut juga *Knowledge Discovery in Database* (KDD) didefinisikan sebagai ekstraksi informasi potensial, implisit dan tidak dikenal dari sekumpulan data [4], [5]. Proses *Knowledge Discovery in Database* melibatkan hasil proses *Data Mining* (proses pengeksktrak kecenderungan suatu pola data), kemudian mengubah hasilnya secara akurat menjadi informasi yang mudah dipahami[6]. Teknik *Data Mining* secara garis besar dapat dibagi dalam dua kelompok: verifikasi dan *discovery*.

Metode verifikasi umumnya meliputi teknik-teknik statistik seperti *goodness of fit* dan *analisis variansi*. Metode *discovery* lebih lanjut dapat dibagi atas model prediktif dan model deskriptif. Teknik prediktif melakukan prediksi terhadap data dengan menggunakan hasil-hasil yang telah diketahui dari data yang berbeda[7]. Model ini dapat dibuat berdasarkan penggunaan data historis lain. Sementara itu, model deskriptif bertujuan mengidentifikasi pola-pola atau hubungan antar data dan memberikan cara untuk mengeksplorasi karakteristik data yang diselidiki [8].

2. METODOLOGI PENELITIAN

2.1 Metode Penelitian

Pada bagian ini penulis akan menjelaskan bagaimana *procedure* pengumpulan data serta proses analisa data penelitian yang akan digunakan. Pada penelitian ini juga akan dijelaskan lokasi dan waktu dilakukan penelitian hingga memperoleh data yang akurat untuk dilakukan pengolahan data menggunakan Algoritma C4.5[9].

2.1.1 Lokasi dan Waktu Penelitian

Penelitian ini dilakukan di Dinas Pertanian Simalungun. Lamanya waktu yang dibutuhkan untuk menyelesaikan penelitian ini selama 1 Minggu dari tanggal 31 Agustus s/d 05 September 2020.

2.1.2 Rancangan Penelitian

Rancangan penelitian ini akan dijelaskan proses penelitian dan pengumpulan data, selanjutnya data diolah melalui proses perhitungan dan mengikuti langkah-langkah penyelesaian Algoritma C4.5[10]. Selanjutnya hasil perhitungan tersebut akan diaplikasikan menggunakan tools RapidMiner untuk melihat keakuratan hasil yang diperoleh. Rancangan penelitian dapat dilihat pada gambar 1[11].



Gambar 1. Rancangan Penelitian

Gambar 1. menjelaskan rancangan penelitian yang dilakukan untuk menentukan kualitas sayur kol di Kabupaten Simalungun menggunakan Algoritma C4.5 yang terdiri dari[12],[13]:

- a) Analisis Masalah
Menganalisis masalah yang terkait dengan penentuan bagaimana menentukan sayur kol yang layak jual dan tidak layak jual.
- b) Mempelajari Literatur
Penelitian ini harus didasari rujukan yang digunakan untuk mendapatkan rujukan yang digunakan untuk mendapatkan informasi dalam penelitian.
- c) Menetapkan Metode
Menetapkan metode untuk memecahkan masalah yang mudah dan akurat dalam menentukan sayur kol yang layak dan tidak layak menggunakan Algoritma C4.5.
- d) Mengumpulkan Data
Data yang digunakan sebagai sampel adalah data kuisisioner dari para Stakeholder sayur kol di Kabupaten Simalungun.
- e) Mengelolah Data
Melakukan pengolahan data dengan menggunakan Algoritma C4.5. Dalam proses pengolahan data digunakan *Microsoft Excel* 2010 untuk mempermudah proses pengolahan data menggunakan Algoritma C4.5.
- f) Menguji data
Pengujian data dilakukan dengan menggunakan menggunakan tools RapidMiner. Proses pengujian data dilakukan guna menguji keakuratan hasil dari pengolahan data secara manual maupun menggunakan sistem.
- g) Kesimpulan
Kesimpulan yang didapatkan dalam menentukan sayur kol yang layak dan tidak layak menggunakan Algoritma C4.5 akan menjadi masukan bagi para petani dalam memilih sayur kol yang layak dan tidak layak jual agar dapat menghasilkan sayur kol yang berkualitas dan layak edar sehingga dapat membantu meningkatkan penghasilan petani di Kabupaten Simalungun.

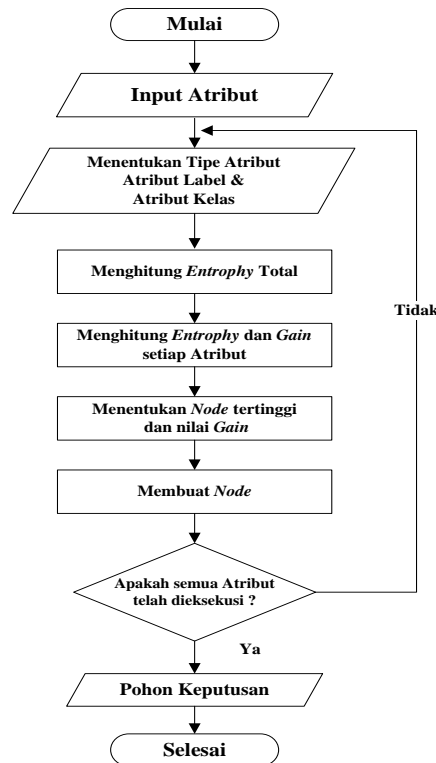
2.2 Analisis Data

Analisis data dapat dilakukan setelah adanya pengumpulan data dan setelah data dikumpulkan dan diolah ke *Microsoft Excel*. Setelah hasil yang telah dapat di *Microsoft Excel* akan diaplikasikan menggunakan tools RapidMiner untuk menyesuaikan hasil yang didapatkan. Jenis data yang digunakan dalam penelitian ini merupakan data primer, Data

primer adalah data yang diperoleh langsung dari tangan pertama atau responden atau informan[14]. Dalam melakukan penelitian ini penulis menggunakan *Microsoft Excel 2010* dan aplikasi *RapidMiner* sebagai alat untuk memproses data. *Microsoft Excel 2010* digunakan untuk membantu dalam proses perhitungan yang akan dilakukan setelah proses perhitungan selesai maka selanjutnya diaplikasikan menggunakan tools *rapidminer* untuk melihat keakuratan hasil yang didapat dari perhitungan di *Microsoft Excel 2010* serta akan dicocokkan dengan hasil yang sudah diaplikasikan menggunakan tools *RapidMiner*[15].

2.3 Pemodelan Metode

Dalam penelitian ini penulis menggunakan teknik klasifikasi *Data Mining* dengan Algoritma *C4.5* dan melakukan pengujian di *RapidMiner*. Berikut ini pemodelan yang penulis gunakan dalam penelitian ini[16]. Berikut ini pemodelan Algoritma *C4.5* yang digambarkan dalam *flowchart* pada Gambar 2.



Gambar 2. Pemodelan Algoritma *C4.5*

Berdasarkan Pemodelan Gambar 2 dapat diketahui alur Algoritma *C4.5* yang digunakan. Pada persiapan awal ditentukan penginputan atribut yang digunakan kemudian melakukan penentuan tipe atribut dengan mencari atribut label dan atribut kelas berdasarkan perhitungan *entropy* total dari masing-masing atribut. Apabila ditemukan gain disetiap atribut maka gain tersebut akan menjadi node tertinggi dengan nilai gain. Selanjutnya dilakukan penentuan cabang dengan cara yang sama dengan melihat gain tertinggi dari tiap hasil partisi.

3. HASIL DAN PEMBAHASAN

3.1 Hasil

Pada tahap ini disajikan sesuai penelitian yang dilakukan. Pada penelitian ini digunakan *tools RapidMiner* sebagai pengujian data yang telah di analisa. Berikut adalah bagian - bagian dari *RapidMiner* beserta isi dari data yang dibuat.

3.1.1 Perhitungan Menggunakan Algoritma *C4.5*

Data penelitian terdiri dari 40 responden yang diperoleh dari hasil kuisioner dan terdiri dari 5 kriteria. Berikut adalah data penelitian yang digunakan:

Tabel 1. Data Penelitian

No	Warna	Tekstur	Berat	Lapisan Daun	Ukuran	Pekerjaan	Output
1	S	S	S	S	S	Pedagang Kol	Layak
2	S	S	TS	S	TS	Pedagang Kol	Layak
3	TS	S	TS	S	S	Pedagang Kol	Layak
.
.

No	Warna	Tekstur	Berat	Lapisan Daun	Ukuran	Pekerjaan	Output
36	TS	S	S	S	S	Pedagang Kol	Layak
37	TS	S	TS	TS	S	Pedagang Kol	Tidak Layak
38	S	TS	S	S	S	Petani Kol	Layak
39	S	S	S	S	S	Pedagang Kol	Layak
40	TS	TS	S	TS	S	Agen Kol	Tidak Layak

Data penelitian diatas menggunakan 5 kriteria yang terdiri dari warna, tekstur, berat, lapisan dan ukuran. Setiap kriteria terdiri dari beberapa sub kriteria yang menjadi pernyataan dalam kuisioner yang diberikan kepada responden. Setiap pernyataan terdiri dari jawaban setuju dan tidak setuju.

Selanjutnya masuk kedalam tahap pengolahan data menggunakan Algoritma C4.5 yaitu:

- 1) Menghitung jumlah kasus seluruhnya, jumlah dengan keputusan setuju dan tidak setuju. Pada penelitian ini jumlah kasus 8, jumlah setuju 6, dan tidak setuju 2.
- 2) Menghitung entropy dari semua kasus yang terbagi berdasarkan kriteria Warna, Tekstur, Berat, Lapisan Dan Ukuran menggunakan persamaan berikut.

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

Keterangan:

S = himpunan kasus,

A = fitur,

n = jumlah partisi S,

Pi = proporsi dari Si terhadap S.

$$Entropy (Total) = (- 27/40 * \log_2(27/40))+(-13/40 * \log_2(13/40))= 0,90973$$

$$Entropy (Warna, Setuju) = (-16/18 * \log_2(16/18))+(-2/18 * \log_2(2/18)) = 0,503258$$

$$Entropy (Warna, Tidak Setuju) = (-11/22*\log_2(11/22))+(-11/22 * \log_2(11/22)) = 1$$

Dan begitu seterusnya hingga diperoleh nilai *entropy* dari setiap kriteria. Setelah diperoleh nilai *entropy* dari setiap kriteria, selanjutnya menghitung nilai *Gain* untuk setiap kriteria menggunakan persamaan berikut.

$$Gain (S,A) = Entropy (S) - \sum_{i=1}^n * Entropy (Si) \tag{2}$$

S = himpunan

A = atribut

n = jumlah partisi antribut A

|Si| = jumlah kasus pada partisi ke-i

|S| = jumlah kasus dalam S

$$Gain (Total, Warna) = 0,909736 - ((18/40*0,503258) + (22/40*1)) = 0,133269$$

$$Gain (Total, Tekstur) = 0,909736 - ((30/40*0,650022) + (10/40*0,721928)) = 0,241737$$

$$Gain (Total, Berat) = 0,909736 - ((25/40*0,795040) + (25/40*1,040254)) = -0,237323$$

$$Gain (Total, Lapisan Daun) = 0,909736 - ((25/40*0,529360) + (15/40*0,0918295)) = 0,234524$$

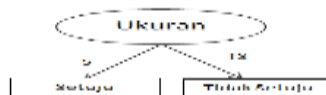
$$Gain (Total, Ukuran) = 0,909736 - ((16/40*0,009480)+(14/40*0,98522)) = 0,81128$$

Berikut adalah hasil dari perhitungan nilai *entropy* dan *gain* pada node 1.

Tabel 2. Nilai *Entropy* dan *Gain* Node 1

Faktor		Jumlah	Setuju	Tidak Setuju	Entropy	Gain
Warna	S	40	27	13	0,909736	0,133269
	TS	18	16	2	0,503258	
Tekstur	S	22	11	11	1	0,241737
	TS	30	25	5	0,650022	
Berat	S	10	2	8	0,721928	-0,237323
	TS	25	19	6	0,795040	
Lapisan Daun	S	25	8	7	1,040254	0,234524
	TS	25	22	3	0,529360	
Ukuran	S	15	5	10	0,918295	0,561113
	TS	16	21	5	0,009480	
	TS	14	6	8	0,985228	

Dari hasil diatas dapat dilihat bahwa faktor dengan nilai *gain* tertinggi adalah ukuran dengan nilai 0,909736. Untuk ukuran dengan pernyataan setuju dan pernyataan tidak setuju sudah diperoleh hasil maka proses selesai. Adapun pohon keputusan yang terbentuk dari pencarian Node 1 adalah seperti pada gambar 3:



Gambar 3. Node

Pada gambar diatas dapat dilihat bahwa hasil dari pengolahan data secara manual dengan sistem diperoleh hasil yang sama. Dari penjelasan diatas dapat disimpulkan bahwa data yang digunakan adalah data *valid* dan dapat dibuktikan dengan *tools RapidMiner* serta dapat menampilkan hasil yang sama dari perhitungan manual maupun dengan *tools RapidMiner*. Berikut adalah tampilan *rule model* berupa teks yang ada dalam *tools RapidMiner* dimana tampilan ini adalah penjelasan dari pohon keputusan yang terbentuk :



Gambar 7. Rule Model Decision Tree

Dan berikut adalah hasil *accuracy* menggunakan *tools RapidMiner*:

accuracy: 100.00%

	true Setuju	true Tidak Setuju	class precision
pred. Setuju	8	0	100.00%
pred. Tidak Setuju	0	2	100.00%
class recall	100.00%	100.00%	

Gambar 8. Hasil Accuracy

Berdasarkan gambar diatas dapat dilihat bahwa *accuracy* dari *RapidMiner* dipeoleh sebesar 100%.

4. KESIMPULAN

Berdasarkan hasil penelitian yang dilakukan bahwa menerapkan *Data Mining* dengan menggunakan Algoritma C4.5 dapat diterapkan. Berdasarkan hasil pengujian software RapidMiner diperoleh hasil yang sama dengan analisis perhitungan Algoritma C4.5 dimana diperoleh tingkat *accuracy* yaitu sebesar 100%. Penelitian ini penulis melakukan beberapa proses pengumpulan data dengan menggunakan Algoritma C4.5 sehingga ditemukan hasil klasifikasi yang membentuk dua kelas yaitu kelas layak dan tidak layak.

REFERENCES

- [1] T. R. I. B. Tusarwenda, "Penerapan data mining dengan algoritma c4.5 dalam prediksi penjualan botol pada cv. seribukilo," 2018.
- [2] L. Swastina, "Penerapan Algoritma C4.5 untuk Penentuan Jurusan Mahasiswa," *J. Gema Aktual.*, vol. 2, no. 1, pp. 93–98, 2013.
- [3] A. D. I. Suradi, "Penerapan data mining untuk menentukan rekomendasi beasiswa dengan metode algoritma c4.5," 2018.
- [4] Yulia and A. D. Putri, "Data Mining Menggunakan Algoritma C4.5 Untuk Memprediksi Kepuasan Mahasiswa Terhadap Terhadap Kinerja Dosen Di Kota Batam" *Comput. Based Informaon Syst. J.*, vol. 07, pp. 56–66, 2019.
- [5] N. Fajira, "Analisis Kepuasan Pelanggan Terhadap Kualitas Pelayanan Pada Kentucky Fried Chicken Cabang Palembang Square Mall," *Polsri*, pp. 14–20, 2014.
- [6] P. Alkhairi, I. S. Damanik, and A. P. Windarto, "Penerapan Jaringan Saraf Tiruan untuk Mengukur Korelasi Beban Kerja Dosen Terhadap Peningkatan Jumlah Publikasi," *Pros. Semin. Nas. Ris. Inf. Sci.*, vol. 1, no. September, p. 581, 2019, doi: 10.30645/senaris.v1i0.65.
- [7] L. N. Rani, "Klasifikasi Nasabah Menggunakan Algoritma C4.5 Sebagai Dasar Pemberian Kredit," *J. KomTekInfo Fak. Ilmu Komput.*, vol. 2, no. 2, pp. 33–38, 2015.
- [8] M. Zarlis, R. W. Sembiring, T. B. Pematangsiantar, U. Methodist, and P. N. Medan, "Analisa Terhadap Perbandingan Algoritma Decision Tree Dengan Algoritma Random Tree Untuk Pre-Processing Data," *J-SAKTI*, vol. 1, no. 2, pp. 180–185, 2017.
- [9] A. Nugraha, "Pemanfaatan Modul Gsm dan Modul Gps Pada Sistem Keamanan Sepedamotormenggunakan Smartphone Berbasis Arduinouno," *Fatmah Riski Dinniah*, vol. 2, no. 1, pp. 1–16, 2017.
- [10] S. Takalapeta, "Penerapan Data Mining Untuk Menganalisis Kepuasan Konsumen Menggunakan Metode Algoritma C4.5," *J I M P - J. Inform. Merdeka Pasuruan*, vol. 3, no. 3, pp. 34–38, 2018, doi: 10.37438/jimp.v3i3.186.
- [11] I. Romli and A. T. Zy, "Penentuan Jadwal Overtime Dengan Klasifikasi Data Karyawan Menggunakan Algoritma C4.5," *J. Sains Komput. Inform. (J-SAKTI)*, vol. 4, no. 2, pp. 694–702, 2020.
- [12] Hendrianto, "Manajemen Strategi Pengelolaan Pasar Dalam Meningkatkan Pendapatan Pedagang Perspektif Ekonomi Islam (Studi di Pasar Segamas Purbalingga)," pp. 1–94, 2018.
- [13] R. I. P. Jimmy Pratama, "Perancangan Dan Implementasi Animated Sticker Sebagai Media Edukasi Mengenai COVID-19 dengan Menggunakan Metode MDLC," *Pap. Knowl. Towar. a Media Hist. Doc.*, vol. 7, no. 2, pp. 1–12, 2021.
- [14] N. Azwanti and E. Elisa, "Analisa Kepuasan Konsumen Menggunakan Algoritma C4.5," *Pros. Semin. Nas. Ilmu Sos. dan Teknol.*, no. 3, pp. 126–131, 2020.
- [15] D. K. Widiyati, M. Wati, and H. S. Pakpahan, "Penerapan Algoritma ID3 Decision Tree Pada Penentuan Penerima Program Bantuan Pemerintah Daerah di Kabupaten Kutai Kartanegara," *J. Rekayasa Teknol. Inf.*, vol. 2, no. 2, p. 125, 2018, doi: 10.30872/jurti.v2i2.1864.
- [16] Y. S. Luvia, A. P. Windarto, S. Solikhun, and D. Hartama, "Penerapan Algoritma C4.5 Untuk Klasifikasi Predikat Keberhasilan Mahasiswa Di Amik Tunas Bangsa," *Jurasik (Jurnal Ris. Sist. Inf. dan Tek. Inform.)*, vol. 1, no. 1, p. 75, 2017, doi: 10.30645/jurasik.v1i1.12.