
Penerapan Data Mining untuk Memprediksi Hasil Panen Pertanian Tanaman Padi Daerah Serdang Bedagai Menggunakan Metode C4.5 Pada Dinas Tanaman Pangan Dan Holtikultura Sumut

Sri Kandi ^{#1}, Muhammad Syahril,S.E.,M.Kom ^{#2}, Milfa Yetri, S.Kom.,M.Kom³

^{#1} Program Studi Sistem Informasi, STMIK Triguna Dharma

^{#2,3} Program Studi Sistem Informasi, STMIK Triguna Dharma

Article Info

Article history:

Received xxxx xxth, 2020

Revised xxxx xxth, 2020

Accepted xxxx xxth, 2020

Keyword:

Hasil Panen

Data Mining

C.45

ABSTRAK

Dinas Tanaman Pangan dan Holtikultura Provinsi Sumut merupakan sebagai salah satu departemen yang mempunyai kegiatan melaksanakan penyuluhan pertanian, memberikan pendapat dalam bidang pertanian dan membantu para petani untuk menyelesaikan masalah tentang tanaman yang bermasalah.

Data Mining merupakan suatu alat dan aplikasi menggunakan analisis statistic, matematika, kecerdasan buatan dan machine learning untuk mengekstrasi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai database besar..

Dalam hal ini Data Mining memiliki beberapa jenis algoritma yang sesuai dengan manfaatnya diantaranya Algoritma C4.5 yang merupakan Metode Decision Tree. Decision Tree adalah mengubah data yang paling besar menjadi pohon-pohon keputusan yang mempresentasikan aturan

Kata Kunci : Hasil Panen, Data Mining, C.45

Copyright © 201x STMIK Triguna Dharma.

All rights reserved.

Nama : Sri Kandi
Kator : STMIK Triguna Dharma
Program Studi : Sistem Informasi
Email : Srikandi@gmail.com

1. PENDAHULUAN

Indonesia merupakan Negara agraris dimana pertanian berperan penting dikarenakan kaya sumber daya alam, seperti tanaman pangan dan hortikultura. Sebagai salah satu Negara Asia beras merupakan makanan produksi terbesar di Indonesia. Hasil perkiraan panen adalah salah satu persoalan paling kritis yang ditemui dalam sektor pertanian, dikarenakan minimnya pengetahuan para petani tentang panen yang melimpah, kondisi cuaca yang tidak pasti dan naik-turunnya harga pasar mempengaruhi hasil produksi tanaman.

Dinas Tanaman Pangan dan Holtikultura Provinsi Sumut merupakan sebagai salah satu departemen yang mempunyai kegiatan melaksanakan penyuluhan pertanian, memberikan pendapat dalam bidang pertanian dan membantu para petani untuk menyelesaikan masalah tentang tanaman yang bermasalah.

Data Mining merupakan suatu alat dan aplikasi menggunakan analisis statistic, matematika, kecerdasan buatan dan machine learning untuk mengekstrasi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai database besar.

Dalam hal ini *Data Mining* memiliki beberapa jenis algoritma yang sesuai dengan manfaatnya diantaranya Algoritma C4.5 yang merupakan Metode *Decision Tree*. *Decision Tree* adalah mengubah data yang paling besar menjadi pohon-pohon keputusan yang mempresentasikan aturan.

%1. Kajian Pustaka

%1.%2. Data Mining

Data Mining adalah proses untuk mencari nilai tambah berupa informasi yang selama ini tidak diketahui secara manual dari suatu basis data. Informasi yang penting ini diperoleh dengan cara mengekstraksi dan mengenali pola yang penting atau berkesan dari data yang terdapat dalam bisnis data. *Data Mining* itu juga dapat diartikan sebagai proses penambangan data dalam jumlah data yang besar dan menghasilkan sebuah *output* (keluaran) berupa pengetahuan..

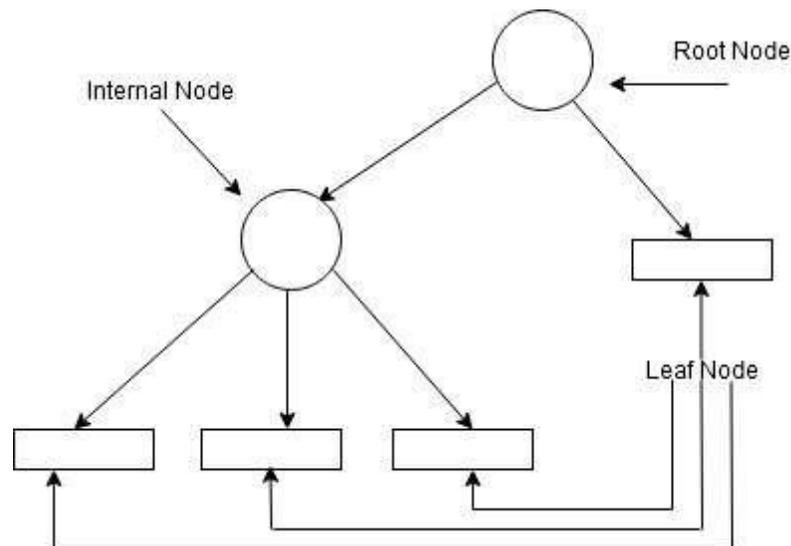
%1.%2. C5

Algoritma C4.5 merupakan algoritma yang digunakan untuk membentuk pohon keputusan. Algoritma C4.5 juga merupakan algoritma dari metode *Decision Tree*. Metode *Decision Tree* merupakan metode klasifikasi yang paling populer digunakan. Selain karena pembangunannya relatif cepat, hasil dari model yang dibangun mudah untuk di pahami. Pada *Decision Tree* terdapat 3 jenis node, yaitu:

- %1. *Root Node*, merupakan node paling atas, pada node ini tidak ada input dan bisa mempunyai output atau mempunyai output lebih dari satu.
- %1. *Internal Node*, merupakan node percabangan, pada node ini hanya terdapat satu input dan mempunyai output minimal dua.
- %1. *Leaf Node* atau *Terminal Node*, merupakan node akhir, pada node ini hanya terdapat satu input dan mempunyai output.

Adapun penjelasan tentang Algoritma C4.5 itu sendiri yaitu salah satu Algoritma C4.5 induksi pohon keputusan yaitu ID3 (*Iterative Dechotomizer 3*).

Input berupa sampel training, label training dan atribut. Algoritma C4.5 merupakan pengembangan dari ID3. *Decision tree* pada (Gambar 2.2) setiap simpul daun menandai label kelas. Simpul yang bukan akhir terdiri dari akar dan simpul *internal* yang terdiri dari kondisi tes atribut pada sebagian *record* yang mempunyai karakteristik yang berbeda. Simpul akar dan simpul internal ditandai dengan bentuk oval dan simpul daun ditandai dengan bentuk segi empat.



Gambar 2.2 Komposisi *Decision Tree*

Untuk penyelesaian kasus di dalam Algoritma C4.5 ada beberapa elemen yang diketahui yaitu:

%1. Entropy

Entropy (S) merupakan jumlah bit yang diperkirakan dibutuhkan untuk dapat mengekstrak suatu kelas (+ atau -) dari sejumlah data acak pada ruang sampel S . Entropy dapat dikatakan sebagai kebutuhan bit untuk menyatakan

suatu kelas. Semakin kecil nilai *Entropy* maka *Entropy* akan semakin digunakan dalam mengekstrak suatu kelas. *Entropy* digunakan untuk mengukur ketidak aslian *S*. Adapun rumus untuk mencari nilai *Entropy*:

$$Entropy(S) \equiv - \sum P \log_2 P - \sum P_0 \log_2 P_0$$

Dimana:

- *S* merupakan ruang (data) sampel yang digunakan untuk pelatihan.
- *P* merupakan jumlah yang bersolusi positif atau mendukung pada data sampel untuk variabel tertentu.
- *P* merupakan jumlah yang bersolusi negatif atau tidak mendukung pada data sampel untuk variabel tertentu.
- $Entropy(S) = 0$, jika semua contoh pada *S* berada dalam kelas yang sama.
- $Entropy(S) = 1$, jika jumlah contoh positif dan negatif dalam *S* adalah sama.
- $0 > Entropy(S) = 1$, jika jumlah contoh positif dan negatif dalam *S* adalah sama.

%1. *Gain*

Gain (*S,A*) merupakan perolehan informasi dari atribut *A* relatif terhadap *output* data *S*. Perolehan informasi didapat dari *output* data atau variabel *dependent S* yang dikelompokkan berdasarkan atribut *A*, dinotasikan dengan *Gain* (*S,A*). Adapun rumus untuk mencari nilai *Gain* yaitu:

$$Gain(S,A) = Entropy(S) -$$

Dimana :

A : Atribut

S : Sampel

N : Jumlah partisis himpunan atribut *A*

$|S_i|$: Jumlah sampel pada pertisi ke – *i*

$|S|$: Jumlah sampel dalam *S*

Adapun langkah – langkah untuk penyelesaian Algoritma C4.5 yaitu:

- %1. Mencari nilai *Entropy* dari Variabel
 - %1. Mencari nilai *Gain* dari setiap atribut
 - %1. Pembentukan atribut sebagai akar berdasarkan *Gain* tertinggi
 - %1. Pembentukan cabang berdasarkan masing – masing nilai
- Ulangi proses untuk masing – masing cabang

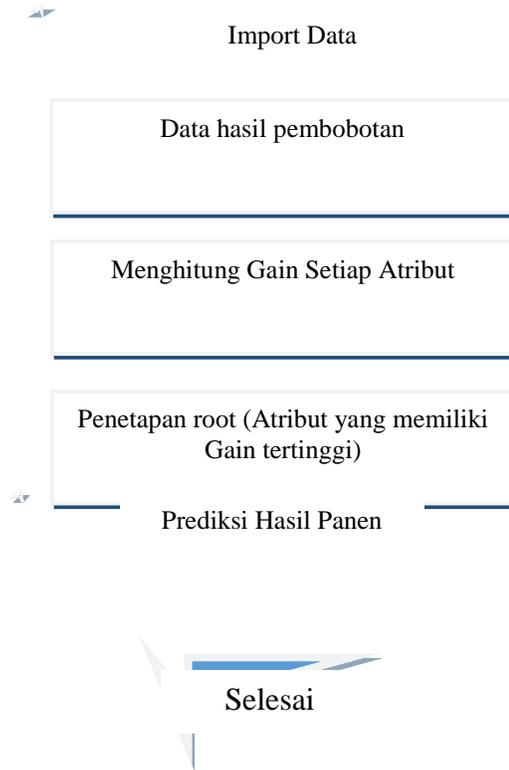
%1. Metodologi Penelitian

%1.%2. Algoritma Sistem

Algoritma system ini menjelaskan langkah-langkah penyelesaian masalah dalam perancangan sistem penerapan data mining dalam memprediksi hasil panen pertanian dengan menggunakan algoritma C.45. Hal ini dilakukan untuk meningkatkan pendapatan instansi pemerintah serta menguntungkan instansi pemerintah dalam hal persaingan dengan instansi pemerintah lain.

Berikut ini adalah *flowchart* dari algoritma C.45 yaitu sebagai berikut:

Mulai



Gambar 3.1 Flowchart C.45

Algoritma C4.5 digunakan untuk mengetahui tingkat akurasi prediksi hasil panen pertanian tanaman padi daerah Serdang berdagai. Langkah untuk dalam proses algoritma C4.5 dengan membuat cabang untuk setiap nilai akar adalah sebagai berikut.

%1. Keterangan I (Luas Lahan Berdasarkan Rencana Panen) :

%1. Jika Luas < 50 maka keterangan "SEMPIT"

%1. Jika Luas > 50 maka keterangan "SEDANG"

%1. Jika Luas > 100 maka keterangan "LUAS"

%1. Keterangan II (Hasil Panen / Ton April-Sept 2017) :

%1. keterangan "MENURUN"

%1. keterangan "MENINGKAT"

%1. Keterangan III (Hasil Panen / Ton April-Sept 2018) :

%1. "MENURUN"

%1. "MENINGKAT"

%1. Keterangan IV (Hama) :

%1. Jika Terserang maka : - 20 %

%1. Jika Tidak maka : 20 %

%1. Keterangan V (Curah Hujan) :

%1. Jika Curah Hujan Tinggi maka : -20%

%1. Jika Curah Hujan Sedang maka : 20 %

%1. Jika Curah Hujan Rendah maka : 22,5 %

Sebelum menghitung nilai *gain* dari atribut, hitung dahulu nilai *entropy*. Untuk menghitung nilai *entropy* digunakan rumus:

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

Keterangan:

S = Himpunan kasus

n = Jumlah partisi S

p_i = Proporsi S_i terhadap S

%1. Entropy Total = Entropy (S) = $\sum_{i=1}^n -p_i \log_2 p_i$

$$\begin{aligned} Entropy \text{ Total} &= ((-90/100 * \log_2(90/100)) + (-10/100 * \log_2(10/100))) \\ &= 0.1368027841 + 0.3321928095 \\ &= 0.8276225419 \end{aligned}$$

%1. Entropy Luas Lahan

$$\begin{aligned} \text{Luas Lahan "SEMPIT"} &= ((-43/47 * \log_2(43/47)) + (-2/47 * \log_2(2/47))) \\ &= 0.1174028972 + 0.2401640932 \\ &= 0.3575669904 \end{aligned}$$

$$\begin{aligned} \text{Luas Lahan "SEDANG"} &= ((-34/41 * \log_2(34/41)) + (-7/41 * \log_2(7/41))) \\ &= 0.2239763794 + 0.4353995019 \\ &= 0.6593758813 \end{aligned}$$

$$\begin{aligned} \text{Luas Lahan "LUAS"} &= ((-10/12 * \log_2(10/12)) + (-1/12 * \log_2(1/12))) \\ &= 0.2191953382 + 0.2987468751 \\ &= 0.5179422133 \end{aligned}$$

%1. Entropy Hasil Panen 2017

$$\begin{aligned} \text{Hasil Panen "MENURUN"} &= ((-2/8 * \log_2(2/8)) + (-0/6 * \log_2(0/6))) \\ &= 0.5 + 0 \\ &= 0.5 \end{aligned}$$

$$\begin{aligned} \text{Hasil Panen "MENINGKAT"} &= ((-1/89 * \log_2(1/89)) + (-89/92 * \log_2(89/92))) \\ &= 0.0727610498 + 0.0462688993 \\ &= 0.1190299491 \end{aligned}$$

%1. Entropy Hama

$$\begin{aligned} \text{Hama "TERSERANG"} &= ((-33/35 * \log_2(33/35)) + (-2/35 * \log_2(2/35))) \\ &= 0.0800381034 + 0.2359590295 \\ &= 0.3159971329 \end{aligned}$$

$$\begin{aligned} \text{Hama "TIDAK"} &= ((-56/65 * \log_2(56/65)) + (-9/65 * \log_2(9/65))) \\ &= 0.1852418753 + 0.3949536201 \\ &= 0.5801954954 \end{aligned}$$

%1. Entropy Curah Hujan

$$\begin{aligned} \text{"TINGGI"} &= ((-33/36 * \log_2(33/36)) + (-2/36 * \log_2(2/36))) = 0.1150699752 + 0.2316625001 = 0.3467324753 \\ \text{"SEDANG"} &= ((-27/30 * \log_2(27/30)) + (-3/30 * \log_2(3/30))) = 0.1368027841 + 0.3321928095 = 0.4689955936 \\ \text{"RENDAH"} &= ((-28/34 * \log_2(28/34)) + (-3/34 * \log_2(3/34))) = 0.2306771099 + 0.3090441477 = 0.5397212576 \end{aligned}$$

Langkar selanjutnya menghitung nilai *gain* dari atribut. Untuk menghitung nilai *Gain* digunakan rumus:

$$Gain(S, A) = entropy(S) -$$

$$\sum_{i=1}^n \frac{|S_i|}{|S|} Entropy (S_i)$$

Keterangan :

S = Himpunan kasus

A = Fitur

n = Jumlah partisi atribut A

$|S_i|$ = Proporsi S_i terhadap S

$|S|$ = jumlah kasus dalam S

Berikut ini adalah nilai *gain* dari setiap criteria :

- Nilai *Gain* Luas Lahan

$$\begin{aligned} &= 0.8276225419 - ((47/100) * 0.3575669904 + ((41/100) * 0.6593758813)) + \\ &((12/100) * 0.5179422133)) \\ &= 0.4513750107 \end{aligned}$$

- Nilai *Gain* Hasil Panen 2017

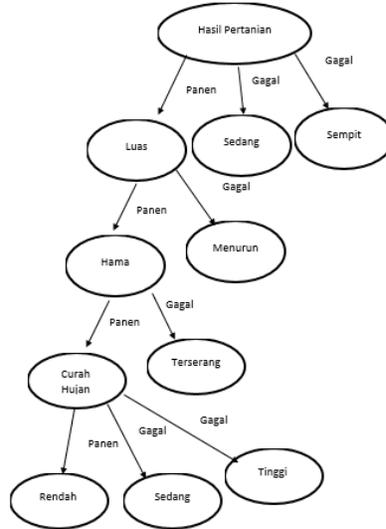
$$= 0.8276225419 - ((8/100) * 0.5) + ((89/100) * 0.1190299491))$$

- = 0.8935591966
- Nilai *Gain* Hasl Panen 2018
= $0.8276225419 - ((3/100) * 0.3895750005 + (97/100) * 0.45639866591)$
= 0.373216586
 - Hama
= $0.8276225419 - ((35/100) * 0.3159971329 + (65/100) * 0.5801954954)$
= 0.3398964734
 - Curah Hujan
= $0.8276225419 - ((36/100) * 0.3467324753 + (30/100) * 0.4689955936 + ((34/100) * 0.4689955936 + ((34/100) * 0.5397212576))$
= 0.3785949451

Setelah mendapatkan data sampel, kemudian melakukan proses perhitungan jumlah data, *entropy* dan *gain*. Hasil tersebut terdapat pada tabel berikut ini

		Keterangan	Jumlah Kasus	Meningkat	Stabil	Entropy	Gain
Node							
1	Total		100	90	10	0.827623	
	Luas Lahan	Sempit	47	43	2	0.357567	
			41	34	7	0.659376	
		Sedang					
			12	10		0.517942	
		Luas			1		
							0.451375
	Hasil Panen 2017	Menurun	8	2	0	0.5	
			89	1	89	0.11903	
		Meningkat					
							0.893559
	Hasil Panen 2018	Menurun	3	2	0	0.389975	
			97	90	11	0.456399	
		Meningkat					
							0.456399
		Terserang	35	33	2	0.315997	
	Hama						
		Tidak	65	56	9	0.580195	
							0.339896
	Curah Hujan	Tinggi	36	33	2		
						0.346732	
		Sedang	30		3	0.468996	
				27			
		Rendah	34	28	3	0.539721	
							0.378595

Penjelasan dari tabel diatas bahwa tabel node berikutnya adalah table keputusan dari 5 varibel adalah sebagai berikut .

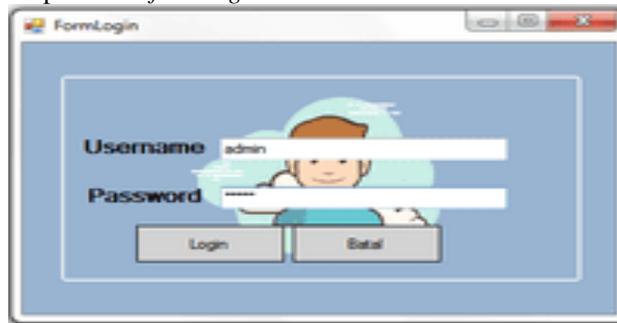


%1. Pengujian dan implementasi

Implementasi sistem merupakan kegiatan akhir dari proses penerapan sistem baru dimana proses yang baru ini akan dioperasikan secara menyeluruh. Sebelum sistem benar-benar bias digunakan dengan baik, sistem harus melalui tahap pengujian terlebih dahulu untuk menjamin tidak ada kendala yang muncul pada saat sistem digunakan. Implementasi yang dilakukan terdapat beberapa tahap prosedur untuk menyelesaikan analisa yaitu aplikasi yang disetujui, melakukan penginstalan, pengujian data, dan mulai menggunakan sistem yang diperbaiki atau system baru. Berikut merupakan implementasi dari system.

1. Tampilan *Form Login*

Berikut ini merupakan tampilan tampilan dari *form login* :



Gambar 5.1 Tampilan *Form Login*

Adapun *form login* berfungsi untuk melakukan proses validasi *username* dan *password* pengguna ketika sebelum masuk ke dalam tampilan *form* menu utama.

2. Tampilan *Form Menu Utama*

Berikut ini merupakan tampilan *form* menu utama setelah pengguna berhasil *login*, Adapun tampilan dari *form* menu utama yaitu :



- SUMUT),” vol. 18, no. April, pp. 212–219, 2019.
- [2] P. Algoritma, C. Untuk, and D. Dini, “Penerapan algoritma c4.5 untuk deteksi dini gangguan autisme pada anak,” vol. 2018, no. Sentika, pp. 23–24, 2018.
- [3] D. Setiyadi, D. Tetap, S. Eresha, S. Laboratorium, and U. Trisakti, “Data Mining Dengan Metode Decision Tree Algoritma C4 . 5 Untuk Memprediksi Permintaan Jenis Produk Barang,” vol. VI, no. 02, pp. 13–34, 2011.
- [4] M. A. Puspa, “Implementasi Data Mining Klasifikasi Algoritma C4 . 5 Dalam Perekrutan Perangkat Desa,” vol. 1, no. 2, pp. 92–97, 2019.
- [5] C. Algoritma *et al.*, “KREDIT (STUDI KASUS DI KOPERASI PEGAWAI REPUBLIK,” vol. 1, no. 2, pp. 6–10, 2014.
- [6] P. Bidang, K. Sains, Y. Mardi, J. Gajah, M. No, and S. Barat, “Jurnal Edik Informatika Data Mining : Klasifikasi Menggunakan Algoritma C4 . 5 Data mining merupakan bagian dari tahapan proses Knowledge Discovery in Database (KDD) . Jurnal Edik Informatika.”
- [7] M. F. Arifin and D. Fitriannah, “Rekomendasi Penerimaan Mitra Penjualan Studi Kasus : PT Atria Artha Persada,” no. January 2018.
- [8] C. U. Klasifikasi, “IMPLEMENTASI METODE DECISION TREE DAN ALGORITMA,” 2016.
- [9] G. Syahputra, M. Kom, P. Studi, and M. Informatika, “Penerapan Algoritma C4 . 5 Dalam Analisa Kelayakan Penerima Bonus Tahunan Pegawai (Studi Kasus : PT . Multi Pratama Nauli Medan),” vol. 16, no. 2, 2015.
- [10] P. P. Putra and A. S. Chan, “Pengembangan Aplikasi Perhitungan Prediksi Stock Motor Menggunakan Algoritma C 4 . 5 Sebagai Bagian dari Sistem Pengambilan Keputusan (Studi Kasus di Saudara Motor),” 2018.
- [11] R. H. Pambudi and B. D. Setiawan, “Penerapan Algoritma C4 . 5 Untuk Memprediksi Nilai Kelulusan Siswa Sekolah Menengah Berdasarkan Faktor Eksternal,” vol. 2, no. 7, pp. 2637–2643, 2018.

UCAPAN TERIMA KASIH

Terima kasih diucapkan kepada pihak-pihak yang telah mendukung dalam proses pembuatan jurnal ini yang tidak dapat disebutkan satu persatu. Kiranya bisa memberi manfaat bagi pembaca dan dapat meningkatkan kualitas jurnal selanjutnya.

BIOGRAFI PENULIS**SRI KANDI****Muhammad Syahril, S.E., M.Kom****Milfa Yetri, S.Kom., M.Kom**