

# ANALISA DATA PERTANIAN TANAMAN PANGAN UNTUK MEMPREDIKSI HASIL PANEN DENGAN DATA MINING ALGORITMA C.45 (STUDI KASUS : DINAS TANAMAN PANGAN dan HOLTIKUTURA PROVINSI SUMUT)

Sri Maya Sari Sianturi<sup>1</sup>, Nelly Astuti Hasibuan<sup>2</sup>

Prodi Teknik Informatika STMIK Budi Darma, Medan, Indonesia  
Jln. Sisingamangaraja No. 338 Simp. Limun Medan

## ABSTRAK

Tanaman Pangan adalah segala jenis tanaman yang di dalamnya terdapat karbohidrat dan protein sebagai sumber energi manusia. Tanaman pangan juga dapat dikatakan sebagai tanaman utama yang di konsumsi manusia, sebagai makanan untuk memberikan asupan energi bagi tubuh. Umumnya tanaman pangan yang tumbuh dalam waktu semusim. Penelitian ini bertujuan membentuk model pohon keputusan untuk memprediksi hasil panen dan melihat variabel yang paling berpengaruh. Obyek dalam penelitian ini adalah tanaman padi, data yang digunakan untuk penelitian adalah data mining yang merupakan suatu proses menemukan hubungan yang berarti, pola, dan kecenderungan dengan memeriksa dalam sekumpulan besar data yang tersimpan dalam penyimpanan dengan menggunakan teknik pengenalan pola seperti teknik statistik dan matematika. Pembentukan model prediksi menggunakan metode C4.5. Pada algoritma C4.5 dilakukan perhitungan entropy dan information gain. Sebagai atribut sumber untuk memperoleh node akar dan node lainnya. Berdasarkan hasil klasifikasi menggunakan algoritma C4.5 menunjukkan bahwa algoritma C4.5 cocok digunakan untuk memprediksi hasil panen padi.

**Kata Kunci :** *Tanaman Pangan, Algoritma C4.5, Data Mining.*

## I. PENDAHULUAN

Dinas Tanaman Pangan Dan Holtikultura Provinsi Sumut merupakan sebagai salah satu Departemen yang mempunyai kegiatan Melaksanakan Penyuluhan Pertanian, memberikan saran-saran dalam bidang Pertanian dan pemberian tanah kepada perusahaan-perusahaan bidang Pertanian. Selain itu Dinas Pertanian juga mengadakan penelitian tentang ekonomi masyarakat dan membuat laporan keadaan Pertanian termasuk statistic.

Data statistik Indonesia tahun 2015 menunjukkan Sumatera utara merupakan provinsi penghasil padi terbesar ketiga setelah pulau Jawa (Badan Pusat Statistik, 2016). Ditahun 2015, luas panen Sumut mencapai 71.756 hektare dengan produksi 857.553 ton. Sedangkan ditahun 2016, dengan luas panen yang lebih sedikit, yakni 69.960 hektare justru produksinya tetap diatasnya, yakni mencapai 907.908 ton. Dilihat dari produktivitasnya, tahun 2015 sebesar 119,51 ton per hektare sedangkan di tahun 2016 naik menjadi 129,78 ton per hektare. Sumatera Utara memiliki potensi wilayah yang sangat besar sebagai sentral penghasil padi.

Dengan tingkat kesuburan tanahnya, tak heran jika produksi padi Sumut hingga kini masih berada pada peringkat 3 di Indonesia setelah propinsi lainnya di Jawa. Upaya peningkatan produksi padi telah dilakukan 3 pemerintah dengan berbagai cara, namun lemahnya teknologi pendukung menjadi salah satu kendala peningkatan hasil panen, Kurangnya saran prediksi hasil panen mengakibatkan kurangnya informasi yang dibutuhkan untuk meningkatkan hasil panen.

Selama ini Dinas Tanaman Pangan Dan Holtikultura Provinsi Sumut telah berupaya melakukan prediksi hasil panen berdasarkan hasil panen dari tahun sebelumnya. Namun peramalan yang dilakukan belum mempertimbangkan faktor irigasi dan iklim, dalam hal ini adalah tenaga kerja. Oleh karena itu dibutuhkan model prediksi produksi panen komoditas padi dengan mempertimbangkan faktor luas lahan, irigasi dan tenaga kerja. Berdasarkan kondisi di atas, dibutuhkan model prediksi hasil panen menggunakan teknik data mining dengan algoritma C4.5. Diharapkan dengan mempertimbangkan faktor luas lahan, irigasi dan tenaga kerja, hasil prediksi yang diperoleh dapat memberikan informasi untuk peningkatan hasil panen.

## II. TEORITIS

### A. Kecerdasan Buatan

*Artificial Intelligence* (AI) atau kecerdasan buatan termasuk bidang ilmu yang relatif muda. Pada tahun 1950-an para ilmuwan dan peneliti mulai memikirkan bagaimana caranya agar mesin dapat melakukan pekerjaannya seperti yang bisa dikerjakan oleh manusia. Alan Turing, seorang matematikawan dari Inggris pertama kali mengusulkan adanya pengujian untuk melihat bisa tidaknya sebuah mesin dikatakan cerdas. Hasil pengujian tersebut kemudian dikenal dengan Turing Test, di mana mesin tersebut menyamar seolah-olah sebagai seseorang di dalam suatu permainan yang mampu memberikan respon terhadap serangkaian pertanyaan yang diajukan. Turing beranggapan bahwa, jika mesin dapat membuat seseorang percaya bahwa dirinya mampu

berkomunikasi dengan orang lain, maka dapat dikatakan bahwa mesin tersebut cerdas (seperti layaknya manusia).

Kecerdasan buatan atau *Artificial Intelligence* (AI) itu sendiri dimunculkan oleh seorang professor dari Massachusetts Institute of Technology yang bernama John McCarthy pada tahun 1956 pada Dartmouth Conference yang dihadiri oleh para peneliti AI. Pada konferensi tersebut juga didefinisikan tujuan utama dari kecerdasan buatan, yaitu mengetahui dan memodelkan proses-proses berpikir manusia dan mendesain mesin agar dapat menirukan kelakuan manusia tersebut. [1]

## B. Data Mining

Data Mining adalah suatu istilah yang digunakan untuk menguraikan penemuan pengetahuan didalam database. Data mining adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan *meachine learning* untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terakit dari berbagai database besar [2]

Istilah data mining dan *knowledge discovery in database* (KDD) seringkali digunakan secara bergantian untuk menjelaskan proses penggalian inormasi tersembunyi dalam suatu basis data yang besar. Sebenarnya kedua istilah tersebut memiliki konsep yang berbeda, tetapi berkaitan satu sama lain. Dan salah satu tahapan dalam keseluruhan proses KDD adalah data mining.

## C. Algoritma C4.5

Algoritma C4.5 adalah algoritma klasifikasi data dengan teknik pohon keputusan yang memiliki kelebihan-kelebihan. Kelebihan ini misalnya dapat mengolah data numerik (kontinyu) dan diskret, dapat menangani nilai atribut yang hilang, menghasilkan aturan - aturan yang mudah diintrepetasikan dan tercepat diantara algoritma-algoritma yang lain.[3]

Keakuratan prediksi yaitu kemampuan model untuk dapat memprediksi label kelas terhadap data baru atau yang belum diketahui sebelumnya dengan baik. Dalam hal kecepatan atau efisiensi waktu komputasi yang diperlukan untuk membuat dan menggunakan model. Kemampuan model untuk memprediksi dengan benar walaupun data ada nilai dari atribut yang hilang. Dan juga skalabilitas yaitu kemampuan untuk membangun model secara efisien untuk data berjumlah besar (aspek ini akan mendapatkan penekanan).

Terakhir interpretabilitas yaitu model yang dihasilkan mudah dipahami. Dalam algoritma C4.5 untuk membangun pohon keputusan hal pertama yang dilakukan yaitu memilih atribut sebagai akar. Kemudian dibuat cabang untuk tiaptiap nilai didalam akar tersebut. Langkah berikutnya yaitu membagi kasus dalam cabang. Kemudian ulangi proses untuk setiap cabang sampai semua kasus pada cabang memiliki kelas yang sama.

Untuk memilih atribut dengan akar, didasarkan pada nilai gain tertinggi dari atribut - tribut yang ada. Untuk menghitung gain digunakan rumus seperti tertera dalam persamaan 1 berikut:[3]

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

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.

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

Keterangan:

S : Himpunan kasus

A : Atribut

N : Jumlah partisi S

Pi : Proporsi dari Si terhadap S

## D. Tanaman Pangan

Tanaman pangan adalah segala jenis tanaman yang di dalamnya terdapat karbohidrat dan protein sebagai sumber energi manusia. Tanaman pangan juga dapat dikatakan sebagai tanaman utama yang dikonsumsi manusia sebagai makanan untuk memberikan asupan energi bagi tubuh. Umumnya tanaman pangan adalah tanaman yang tumbuh dalam waktu semusim.[4]

## III. ANALISA

### A. Analisa Masalah

Dinas Tanaman Pangan Dan Holtikutura Provinsi Sumut merupakan sebagai salah satu Departemen yang mempunyai kegiatan Melaksanakan Penyuluhan Pertanian, memberikan saran-saran dalam bidang Pertanian dan pemberian tanah kepada perusahaan-perusahaan bidang Pertanian. Prediksi hasil panen mengacu kepada memprediksi jumlah hasil panen dimasa mendatang berdasarkan data histori.

Berkaitan dengan kompetensi dan globalisasi, prediksi hasil panen padi memainkan peran penting pada system siklus hidup para petani. Faktor-faktor yang mempengaruhi hasil panen padi adalah jumlah tanaman (Ha), lokasi, panen (ha) , Produktivitas (Kw/ha) dan Produksi (ton). Hubungan antara faktor-faktor yang mempengaruhi hasil panen padi dengan hasil panen

merupakan pendukung kelancaran hasil panen padi di satu priode. Hasil panen padi yang tidak dapat memenuhi kebutuhan pasar mengakibatkan kelangkaan beras yang akan mempengaruhi kecukupan konsumsi gizi penduduk Indonesia.

Berdasarkan permasalahan maka muncul gagasan untuk membuat model baru luas panen padi yang melibatkan indikator / variabel area tambah tanam dan curah hujan. Agar para petani dapat memenuhi kebutuhan penduduk Indonesia, Sehingga permasalahan akan berkurang dan para petani pun dapat melakukan prediksi untuk tahun yang akan datang. Dimana para petani akan memprediksi daerah –daerah untuk mendapatkan hasil panen yang terbaik, Agar tidak terjadi kerusakan pada beras yang akan di pasarkan.

Hal ini menjadi latar belakang penggunaan metode prediksi menggunakan algoritma C4.5 Untuk mempermudah hasil prediksi tanaman padi oleh para petani yang akan di pasarkan, Sehingga dilakukan prediksi menggunakan algoritma C4.5 tersebut.

### B. Penyelesaian Dengan Algoritma C4.5

Berikut ini adalah hasil panen padi yang akan dijadikan sampel untuk analisa dan untuk pengujian diambil bulan priode tahun 2015 -2016.

Tabel 1 Data Hasil Panen Priode 2015 -2016

N	Kabupaten/ Kota	Sasaran Tanam (ha)	Hasil Panen (Ton) Okt'15-Maret'16	Hasil panen Ton April-Sept'16	Curah Hujan	Ham	Kategori
1	Medan	3,0	12,970	13,493	Sedang	Tidak	Meningkat
2	Langkat	89,	484,040	503,547	Sedang	Tidak	Meningkat
3	Deli Serdang	106	584,057	607,595	Sedang	Tidak	Meningkat
4	Simalungun	99,	580,793	604,199	Tinggi	Tidak	Meningkat
5	Tanah Karo	25,	145,416	151,276	Tinggi	Tidak	Meningkat
6	Asahan	19,	106,222	110,503	Tinggi	Tidak	Meningkat
7	Lab. Batu	30,	144,289	150,104	Tinggi	Tidak	Meningkat
8	Tap. Utara	34,	169,969	176,819	Tinggi	Tidak	Meningkat

N	Kabupaten/ Kota	Sasaran Tanam (ha)	Hasil Panen (Ton) Okt'15-Maret'16	Hasil panen Ton April-Sept'16	Curah Hujan	Ham	Kategori
9	Tap. Tengah	35,	151,569	157,677	Tinggi	Tidak	Stabil
10	Tap. Selatan	46,	225,567	234,657	Tinggi	Tidak	Stabil
11	Nias	14,	53,907	56,079	Rendah	Tidak	Stabil
12	Dairi	22,	129,268	134,477	Tinggi	Tidak	Meningkat
13	Teb. Tinggi	824	4,241	4,412	Sedang	Tidak	Stabil
14	Tanj. Balai	480	2,186	2,274	Rendah	Tidak	Rendah
15	Binjai	3,5	17,625	18,335	Rendah	Tidak	Stabil
16	Pem. Siantar	5,7	33,646	35,002	Sedang	Tidak	Stabil
17	Toba Samosir	34,	186,404	193,916	Tinggi	Tidak	Stabil
18	Mand. Natal	53,	264,969	275,647	Tinggi	Tidak	Stabil
19	P. Sidempun	10,	50,178	52,200	Tinggi	Tidak	Stabil
20	Nias Selatan	33,	156,123	162,414	Rendah	Tidak	Stabil
21	H. Hasundutan	27,	128,487	133,665	Tinggi	Tidak	Stabil
22	S. Bedagai	82,	453,400	471,672	Sedang	Tidak	Meningkat
23	Pakpak Bharat	2,8	12,025	12,509	Tinggi	Tidak	Rendah
24	Samosir	12,	59,323	61,714	Tinggi	Tidak	Stabil
25	Batubara	40,	210,373	218,851	Rendah	Tidak	Stabil
26	Padang Lawas	21,	92,837	96,579	Rendah	Tidak	Stabil
27	P. Lawas Utara	42,	174,436	181,466	Rendah	Tidak	Stabil

N	Kabupaten/ Kota	Sasaran Tanam (ha)	Hasil Panen (Ton) Okt'15- Maret'16	Hasil panen Ton April- Sept'16	Cu rah Hu jan	Ha ma	Kategori
28	L. Batu Utara	42,	176,325	183,431	Se dan g	Tid ak	Stabil
29	L. Batu Selatan	1,0	5,092	5,298	Re nda h	Tid ak	Stabil
30	Nias Utara	13,	46,144	48,004	Re nda h	Tid ak	Stabil
31	Nias Barat	5,9	23,638	24,591	Re nda h	Tid ak	Stabil
32	Gunung Sitoli	5,4	26,845	27,927	Re nda h	Tid ak	Stabil

Data yang digunakan dalam penelitian ini adalah data hasil panen tahun 2015-2016 yang memberikan informasi hasil panen dengan status meningkat, stabil dan menurun,. Setelah data terkumpul dilakukan analisa data untuk menghasilkan kesimpulan berdasarkan *rule*. Analisis data tersebut dilakukan berdasarkan klasifikasi dan decision tree dengan algoritma C4.5. Data keseluruhan yang digunakan sebanyak.

### Penentuan Atribut

Pada tahap ini dilakukan penentuan atribut untuk membentuk pohon keputusan dari kasus. Adapun atribut prediksi hasil panen adalah sebagai berikut :

#### 1. Sasaran tanam

Dari data sasaran tanam yang didapat Dinas Tanaman Pangan Dan Holtikutura Provinsi Sumut seperti tertera pada table 2. berikut.

Tabel 2. Jumlah daerah berdasarkan rancangan panen (ha)

Rencana Panen (ha)	Kategori
< 50.000	Sempit
=50.000	Sedang
>50.000	Luas

#### 2. Hasil Panen

Dari data hasil panen yang didapat Dinas Tanaman Pangan Dan Holtikutura Provinsi Sumut seperti tertera pada table 3 berikut.

Tabel 3. Jumlah Daerah berdasarkan Hasil produksi (Ton)

Hasil Panen (/Ton)	Kategori
< 50.000	Menurun
=50.000	Stabil
>50.000	Meningkat

#### 3. Curah Hujan

Curah hujan dibagi menjadi tiga golongan :

Tinggi : -20%

Hasil panen dengan curah hujan yang tinggi berkurang 20% dari rata-rata hasil panen maksimal

Sedang : 20%

Hasil panen dengan curah hujan yang sedang bertambah 20% dari rata-rata hasil panen maksimal

Rendah : 22,5%

Hasil panen dengan curah hujan rendah bertambah 20% dari rata-rata hasil panen maksimal

#### 4. Hama

Hama dibagi menjadi dua golongan :

Ya : -20%

Hasil panen yang terserang hama akan mengalami penurunan sebesar 20% Tidak : 20%

Pembentukan pohon keputusan pada dasarnya memiliki tahap awal menentukan *node* awal sebagai akar pohon keputusan. Langkah-langkah penentuan akar (*node* awal) adalah :

1. Dalam metode *decision tree* dengan algoritma C4.5, penentuan simpul awal dilakukan melalui perhitungan jumlah kasus. Jumlah kasus dengan hasil panen meningkat, stabil dan menurun untuk setiap atribut, *entropy* dari semua kasus dan *gain* dari setiap atribut. Berikut tabel pembagian kasus dari atribut sasaran tanam (luas, sedang, sempit), hasil panen (meningkat, stabil, menurun), curah hujan (tinggi, sedang rendah), hama (ya, tidak).

Table 4. Pembagian Khusus

Atribut	Nilai	Jumlah kasus	Meningkat	Stabil	Menurun
Sasaran tanam	<50	32	10	20	2
	50-100	25	6	18	1
	>100	4	3	1	0
Hasil Panen Priode Okt'15-Maret'16	<50	3	1	1	1
	50-100	10	1	7	2
	>100	4	0	4	0
Hasil Panen Priode April-Sept'16	<50	18	9	9	0
	50-100	10	1	7	2
	>100	4	0	4	0
Curah hujan	Tinggi	18	9	9	0
	Sedang	14	6	7	1
	Rendah	7	4	3	0
Hama	Ya	11	0	10	1
	Tidak	0	0	0	0
		32	10	20	2

2. Menghitung nilai *entropy* total dari semua kasus dengan menggunakan persamaan

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

$$\begin{aligned} Entropy(\text{Total}) &= (-\frac{10}{32} * \log_2(\frac{10}{32})) + (-\frac{20}{32} * \log_2(\frac{20}{32})) + (-\frac{2}{32} * \log_2(\frac{2}{32})) \\ &= (-0.3125 * \log_2 0.3125) + (-0.625 * \log_2 0.625) \\ &\quad + (-0.1 * \log_2 0.1) \\ &= -0.1225 + (-0.1176) + (-0.0030) \\ &= -0.2431 \end{aligned}$$

3. Menghitung nilai *entropy* tiap atribut (*instance*) dan *gain* tiap atribut. Berikut adalah perhitungan nilai *entropy* dari atribut sasaran panen yaitu <50, 50-100 dan >100.

$$\begin{aligned} Entropy(\text{sasaran panen} < 50) &= (-\frac{6}{25} * \log_2(\frac{6}{25})) + (-\frac{18}{25} * \log_2(\frac{18}{25})) + (-\frac{1}{25} * \log_2(\frac{1}{25})) \\ &= (-0.24 * \log_2 0.24) + (-0.72 * \log_2 0.72) + (-0.04 * \log_2 0.04) \\ &= -0.0173 + -0.1560 + -0.1871 \\ &= -0.3604 \end{aligned}$$

$$\begin{aligned} Entropy(\text{sasaran panen } 50-100) &= (-\frac{3}{4} * \log_2(\frac{3}{4})) + (-\frac{1}{4} * \log_2(\frac{1}{4})) \\ &= (-0.75 * \log_2 0.75) + (-0.25 * \log_2 0.25) \\ &= -0.1694 + -0.1882 \\ &= -0.3576 \end{aligned}$$

$$\begin{aligned} Entropy(\text{sasaran panen } > 100) &= (-\frac{1}{3} * \log_2(\frac{1}{3})) + (-\frac{1}{3} * \log_2(\frac{1}{3})) + (-\frac{1}{3} * \log_2(\frac{1}{3})) \\ &= (-0.3333 * \log_2 0.3333) + (-0.3333 * \log_2 0.3333) + (-0.3333 * \log_2 0.3333) \\ &= -0.0334 + -0.0334 + -0.0334 \\ &= -0.1002 \end{aligned}$$

$$\begin{aligned} Entropy(\text{Hasil panen Ton Okt'15-Maret'16 } < 50) &= (-\frac{1}{10} * \log_2(\frac{1}{10})) + (-\frac{7}{10} * \log_2(\frac{7}{10})) + (-\frac{2}{10} * \log_2(\frac{2}{10})) \\ &= (-0.1 * \log_2 0.1) + (-0.7 * \log_2 0.7) + (-0.2 * \log_2 0.2) \\ &= -0.0030 + -0.1475 + -0.0120 \\ &= -0.1625 \end{aligned}$$

$$\begin{aligned} Entropy(\text{Hasil panen Ton Okt'15-Maret'16 } 50-100) &= (-\frac{4}{4} * \log_2(\frac{4}{4})) \\ &= (-1 * \log_2 1) \\ &= -0.301 \end{aligned}$$

$$\begin{aligned} Entropy(\text{Hasil panen/ Ton Okt'15-Maret'16 } > 100) &= (-\frac{9}{18} * \log_2(\frac{9}{18})) + (-\frac{9}{18} * \log_2(\frac{9}{18})) \\ &= (-0.5 * \log_2 0.5) + (-0.5 * \log_2 0.5) \\ &= -0.0753 + -0.0753 \\ &= -0.1506 \end{aligned}$$

$$\begin{aligned} Entropy(\text{Hasil panen /Ton April-Sept'16 } < 50) &= (-\frac{1}{10} * \log_2(\frac{1}{10})) + (-\frac{7}{10} * \log_2(\frac{7}{10})) + (-\frac{2}{10} * \log_2(\frac{2}{10})) \\ &= (-0.1 * \log_2 0.1) + (-0.7 * \log_2 0.7) + (-0.2 * \log_2 0.2) \end{aligned}$$

$$\begin{aligned} &= -0.0030 + -0.1475 + -0.0120 \\ &= -0.1625 \end{aligned}$$

$$\begin{aligned} Entropy(\text{Hasil panen /Ton April-Sept'16 } 50-100) &= (-\frac{4}{4} * \log_2(\frac{4}{4})) \\ &= (-1 * \log_2 1) \\ &= -0.301 \end{aligned}$$

$$\begin{aligned} Entropy(\text{Hasil panen/ Ton April-Sept'16 } > 100) &= (-\frac{9}{18} * \log_2(\frac{9}{18})) + (-\frac{9}{18} * \log_2(\frac{9}{18})) \\ &= (-0.5 * \log_2 0.5) + (-0.5 * \log_2 0.5) \\ &= -0.0753 + -0.0753 \\ &= -0.1506 \end{aligned}$$

$$\begin{aligned} Entropy(\text{curah hujan tinggi}) &= (-\frac{6}{14} * \log_2(\frac{6}{14})) + (-\frac{7}{14} * \log_2(\frac{7}{14})) + (-\frac{1}{14} * \log_2(\frac{1}{14})) \\ &= (-0.4283 * \log_2 0.4283) + (-0.5 * \log_2 0.5) + (-0.0714 * \log_2 0.0714) \\ &= -0.0097 + -0.0753 + -0.0015 \\ &= -0.0865 \end{aligned}$$

$$\begin{aligned} Entropy(\text{curah hujan sedang}) &= (-\frac{4}{7} * \log_2(\frac{4}{7})) + (-\frac{3}{7} * \log_2(\frac{3}{7})) \\ &= (-0.5714 * \log_2 0.5714) + (-0.4286 * \log_2 0.4286) \\ &= -0.0987 + -0.0320 \\ &= -0.1307 \end{aligned}$$

$$\begin{aligned} Entropy(\text{curah hujan rendah}) &= (-\frac{10}{11} * \log_2(\frac{10}{11})) + (-\frac{1}{11} * \log_2(\frac{1}{11})) \\ &= (-0.9091 * \log_2 0.9091) + (-0.0909 * \log_2 0.0909) \\ &= -0.2485 + -0.0025 \\ &= -0.251 \end{aligned}$$

$$\begin{aligned} Entropy(\text{hama tidak}) &= (-\frac{10}{32} * \log_2(\frac{10}{32})) + (-\frac{20}{32} * \log_2(\frac{20}{32})) + (-\frac{2}{32} * \log_2(\frac{2}{32})) \\ &= (-0.3125 * \log_2 0.3125) + (-0.625 * \log_2 0.625) + (-0.1 * \log_2 0.1) \\ &= -0.1225 + (-0.1176) + (-0.0030) \\ &= -0.2431 \end{aligned}$$

Berikut adalah hasil perhitungan *entropy* yang diletakkan pada Tabel 5.

Tabel 5. *Sample* perhitungan *entropy* Condition

Atribut	Nilai	Jumlah kasus	Meningkat	Stabil	Menu run	entropy
		32	10	20	2	-0.2431
Sasaran tana m	<50	25	6	18	1	-0.3604
	50-100	4	3	1	0	-0.3576
	>100	3	1	1	1	-0.1002
Hasil Pane n	<50	10	1	7	2	-0.1625
	50-100	4	0	4	0	-0.301

Atribut	Nilai	Jumlah kasus	Meningkat	Stabil	Menurun	entropy
Priode Okt'15-Maret'16	>100	18	9	9	0	-0.1506
Hasil Panen Priode April-Sept'16	<50	10	1	7	2	-0.1625
	50-100	4	0	4	0	-0.301
	>100	18	9	9	0	-0.1506
Curah hujan	Tinggi	14	6	7	1	-0.0865
	Sedang	7	4	3	0	0.1307
	Rendah	11	0	10	1	-0.251
Hama	Ya	0	0	0	0	0
	Tidak	32	10	20	2	-0.2431

4. Menghitung nilai *gain* dari atribut, nilai *gain* pada atribut *condition* dihitung dengan persamaan sebagai berikut.

$$Gain(S,A)=Entropy(S) - \sum_{i=1}^n \frac{S_i}{S} \times Entropy(S_i) \dots \dots \dots (3.3)$$

$$Gain(\text{Total Nilai sasaran panen}) = -0.2431 - ((\frac{25}{32} \times -0.3604) + (\frac{4}{32} \times -0.3576) + (\frac{3}{32} \times -0.1002)) = -0.2431 - (-0.2816) + (-0.0447) + (-0.0094) = -0.0156$$

$$Gain(\text{Total Nilai Hasil Panen Priode Okt'15-Maret'16}) = -0.2431 - ((\frac{10}{32} \times -0.0508) + (\frac{4}{32} \times -0.301) + (\frac{18}{32} \times -0.1506)) = -0.2431 - (-0.2816) + (-0.0376) + (-0.08471) = -0.0838$$

$$Gain(\text{Total Nilai Hasil Panen Priode April-Sept'16}) = -0.2431 - ((\frac{10}{32} \times -0.0508) + (\frac{4}{32} \times -0.301) + (\frac{18}{32} \times -0.1506)) = -0.2431 - (-0.2816) + (-0.0376) + (-0.08471) = -0.0838$$

$$Gain(\text{Total Nilai curah hujan}) = -0.2431 - ((\frac{14}{32} \times -0.0865) + (\frac{7}{32} \times -0.1307) + (\frac{11}{32} \times -0.251)) = -0.2431 - (-0.0378) + (-0.0286) + (-0.0862) = -0.3201$$

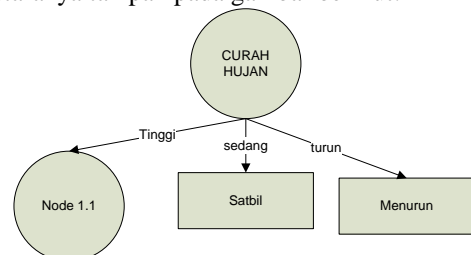
$$Gain(\text{Total Nilai hama})$$

$$= -0.2431 - ((\frac{32}{32} \times -0.2431)) = -0.2431 - (-0.2431) = 0$$

Tabel 7. Perhitungan *note Akar*

Atribut	Nilai	Jumlah kasus	Meningkat	Stabil	Menurun	entropy	Gain
		32	10	20	2	-0.2431	
Sasaran tanam	<50	25	6	18	1	-0.3604	-0.0156
	50-100	4	3	1	0	-0.3576	
	>100	3	1	1	1	-0.1002	
Hasil Panen Priode Okt'15-Maret'16	<50	10	1	7	2	-0.1625	-0.0838
	50-100	4	0	4	0	-0.301	
	>100	18	9	9	0	-0.1506	
Hasil Panen Priode April-Sept'16	<50	10	1	7	2	-0.1625	-0.0838
	50-100	4	0	4	0	-0.301	
	>100	18	9	9	0	-0.1506	
Curah hujan	Tinggi	14	6	7	1	-0.0865	-0.3201
	Sedang	7	4	3	0	0.1307	
	Rendah	11	0	10	1	-0.251	
Hama	Ya	0	0	0	0	0	0
	Tidak	32	10	20	2	-0.2431	

Dari hasil Tabel dapat diketahui bahwa atribut dengan *gain* tertinggi adalah curah hujan yaitu sebesar -0.3201. dengan demikian, Perilaku menjadi *Node* akar. Ada tiga atribut pada curah hujan, yaitu tinggi, sedang dan rendah. Dari tiga nilai atribut tersebut, nilai atribut baik sedang dan turun sudah mengklasifikasikan menjadi satu, yaitu stabil dan menurun, sehingga tidak perlu dilakukan perhitungan lebih lanjut, tetapi untuk nilai atribut tertinggi masih perlu dilakukan perhitungan lagi. Dari hasil tersebut dapat digambarkan Pohon Keputusan semestarnya tampak pada gambar berikut:



Gambar 1. Pohon keputusan hasil perhitungan *node* 1

*Node* berikutnya dapat dipilih pada bagian yang mempunyai atau memiliki nilai antara meningkat, stabil

dan dan menurun, pada *node* 1 diatas atribut sedang dan turun sudah memiliki keputusan stabil dan menurun,sehinga tidak perlu dijelaskan lagi.

Berikut merupakan Hasil perhitungan untuk *node* 1 dengan menggunakan persamaan (3.2) dan persamaan (3.3) terlihat pada Tabel 8.

Tabel 8. Perhitungan node 1

Atribut	Nilai	Jumlah kasus	Meningkat	Stabil	Menurun	entropy	Gain
Node 1		14	6	7	1	-0.1501	
Sasaran tanam	<50	12	4	7	1	-0.1379	-0.1035
	50-100	2	1	1	0	-0.1506	
	>100	0	0	0	0	0	
Hasil Panen Priode Okt'15-Maret'16	<50	1	0	0	1	-0.3010	-0.291
	50-100	2	0	2	0	-0.3010	
	>100	11	6	5	0	-0.1518	
Hasil Panen Priode April-Sept'16	<50	1	0	0	1	-0.3010	-0.291
	50-100	2	0	2	0	-0.3010	
	>100	11	6	5	0	-0.1518	
Hama	Ya	0	0	0	0	0	0
	Tidak	14	6	7	1	-0.1501	

Entropy(total node 1)

$$\begin{aligned}
 &= (-\frac{6}{14} * \log_2(\frac{6}{14})) + (-\frac{7}{14} * \log_2(\frac{7}{14})) + (-\frac{1}{14} * \log_2(\frac{1}{14})) \\
 &= (-0.4286 * \text{Log}2 0.4286) + (-0.5 * \text{log}2 0.5) + (-0.0714 * \text{Log}2 0.0714) \\
 &= -0.0553 + -0.0753 + -0.0215 \\
 &= -0.1501
 \end{aligned}$$

Entropy(sasaran panen<50)

$$\begin{aligned}
 &= (-\frac{4}{12} * \log_2(\frac{4}{12})) + (-\frac{7}{12} * \log_2(\frac{7}{12})) + (-\frac{1}{12} * \log_2(\frac{1}{12})) \\
 &= (-0.3333 * \text{Log}2 0.3333) + (-0.5833 * \text{log}2 0.5833) + (-0.0833 * \text{Log}2 0.0833) \\
 &= -0.0334 + -0.1024 + -0.0021 \\
 &= -0.1379
 \end{aligned}$$

Entropy(sasaran panen 50-100)

$$\begin{aligned}
 &= (-\frac{1}{2} * \log_2(\frac{1}{2})) + (-\frac{1}{2} * \log_2(\frac{1}{2})) \\
 &= (-0.5 * \text{Log}2 0.5) + (-0.5 * \text{Log}2 0.5) \\
 &= -0.0753 + -0.0753 \\
 &= -0.1506
 \end{aligned}$$

Entropy(Hasil panen Ton Okt'15-Maret'16 <50)

$$\begin{aligned}
 &= (-\frac{1}{1} * \log_2(\frac{1}{1})) \\
 &= (-1 * \text{Log}2 1) \\
 &= -0.0030 + -0.1475 + -0.0120 \\
 &= -0.3010
 \end{aligned}$$

Entropy(Hasil panen Ton Okt'15-Maret'16 50-100)

$$\begin{aligned}
 &= (-\frac{2}{2} * \log_2(\frac{2}{2})) \\
 &= (-1 * \text{Log}2 1) \\
 &= -0.3010
 \end{aligned}$$

Entropy(Hasil panen/ Ton Okt'15-Maret'16 >100)

$$\begin{aligned}
 &= (-\frac{6}{11} * \log_2(\frac{6}{11})) + (-\frac{5}{11} * \log_2(\frac{5}{11})) \\
 &= (-0.5455 * \text{Log}2 0.5455) + (-0.4545 * \text{log}2 0.4545) \\
 &= -0.0896 + -0.0622 \\
 &= -0.1518
 \end{aligned}$$

Entropy(Hasil panen /Ton April-Sept'16 <50)

$$\begin{aligned}
 &= (-\frac{1}{1} * \log_2(\frac{1}{1})) \\
 &= (-1 * \text{Log}2 1) \\
 &= -0.0030 + -0.1475 + -0.0120 \\
 &= -0.3010
 \end{aligned}$$

Entropy(Hasil panen /Ton April-Sept'16 50-100)

$$\begin{aligned}
 &= (-\frac{2}{2} * \log_2(\frac{2}{2})) \\
 &= (-1 * \text{Log}2 1) \\
 &= -0.3010
 \end{aligned}$$

Entropy(Hasil panen/ Ton April-Sept'16 >100)

$$\begin{aligned}
 &= (-\frac{6}{11} * \log_2(\frac{6}{11})) + (-\frac{5}{11} * \log_2(\frac{5}{11})) \\
 &= (-0.5455 * \text{Log}2 0.5455) + (-0.4545 * \text{log}2 0.4545) \\
 &= -0.0896 + -0.0622 \\
 &= -0.1518
 \end{aligned}$$

Entropy(hama tidak)

$$\begin{aligned}
 &= (-\frac{6}{14} * \log_2(\frac{6}{14})) + (-\frac{7}{14} * \log_2(\frac{7}{14})) + (-\frac{1}{14} * \log_2(\frac{1}{14})) \\
 &= (-0.4286 * \text{Log}2 0.4286) + (-0.5 * \text{log}2 0.5) + (-0.0714 * \text{Log}2 0.0714) \\
 &= -0.0553 + -0.0753 + -0.0215 \\
 &= -0.1501
 \end{aligned}$$

Menghitung nilai *gain* dari atribut, nilai *gain* pada *atribut condition* dihitung dengan persamaan (3.3) sebagai berikut.

Gain(Total Nilai sasaran panen)

$$\begin{aligned}
 &= -0.1501 - ((\frac{12}{14} * x - 0.1379) + (\frac{2}{14} * x - 0.1506)) \\
 &= -0.2431 - (-0.1181) + (-0.0215) \\
 &= -0.1035
 \end{aligned}$$

Gain(Total Nilai Hasil Panen Priode Okt'15-Maret'16)

$$\begin{aligned}
 &= -0.1501 - ((\frac{1}{14} * x - 0.3010) + (\frac{2}{14} * x - 0.3010) + (\frac{11}{14} * x - 0.1518)) \\
 &= -0.1501 - (-0.0214) + (-0.0430) + (-0.1193) \\
 &= -0.291
 \end{aligned}$$

Gain(Total Nilai Hasil Panen Priode April-Sept'16)

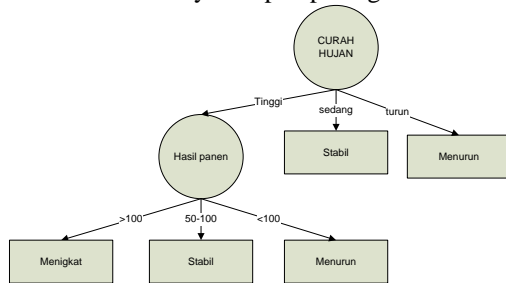
$$\begin{aligned}
 &= -0.1501 - ((\frac{1}{14} * x - 0.3010) + (\frac{2}{14} * x - 0.3010) + (\frac{11}{14} * x - 0.1518)) \\
 &= -0.1501 - (-0.0214) + (-0.0430) + (-0.1193) \\
 &= -0.291
 \end{aligned}$$

Gain(Total Nilai hama)

$$\begin{aligned}
 &= -0.2431 - ((\frac{32}{32} * x - 0.2431)) \\
 &= -0.2431 - (-0.2431) \\
 &= 0
 \end{aligned}$$

Dari hasil Tabel dapat diketahui bahwa atribut dengan *gain* tertinggi adalah hasil panen yaitu sebesar 0.291 dengan demikian, Nilai total hasil panen menjadi Node akar. Ada tiga atribut atribut pada Nilai Akhir, yaitu <50, 50-100 dan >50. Dari tiga nilai atribut tersebut, ketiga atribut tersebut sudah mengklasifikasikan menjadi satu, yaitu meningkat

Dari hasil tersebut dapat digambarkan Pohon Keputusan semestarnya tampak pada gambar berikut:



Gambar 2. Pohon keputusan hasil perhitungan *node* 1.1

#### IV. Kesimpulan

Dari hasil penelitian yang penulis lakukan pada DINAS TANAMAN PANGAN dan HOLTIKUTURA PROVINSI SUMUT, maka penulis mengambil kesimpulan penentuan untuk memprediksi hasil panen padi menggunakan metode data mining khususnya algoritma C4.5 akan bermanfaat sekali dalam memprediksi data tanaman pangan pada DINAS TANAMAN PANGAN dan HOLTIKUTURA PROVINSI SUMUT. Maka dapat disimpulkan bahwa:

1. Dengan algoritma C4.5 dan pengujian dengan Tanagra menghasilkan pola kombinasi *itemssets* dan *rules* sehingga ilmu pengetahuan dan informasi penting dari data prediksi dapat mempengaruhi di Dinas Tanaman Pangan dan Holtikutura Provinsi Sumut .
2. Proses untuk menghasilkan kombinasi item menggunakan algoritma C4.5 dilakukan dengan menetapkan nilai *Entropy* dan nilai *Gain*. Untuk pembentukan pohon keputusan.
3. Pengujian yang telah dilakukan dapat disimpulkan bahwa *data mining classification* dengan menggunakan metode pohon keputusan algoritma C4.5 untuk membentuk pohon keputusan dalam memprediksi hasil panen padi dapat diterapkan menggunakan tanagra dan mendapatkan hasil yang maksimal.

#### REFERENCES

- [1] Edi Wijaya, *Analisis Penggunaan Algoritma Breadth First Search Dalam Konsep Artificial Intellegencia* , 2013, Vol. II No 2 : 18-26,
- [2] Kusriani, luthfi taufiq Emha, *Algoritma Data Mining*, 2009, Penerbit Andi, Yogyakarta.
- [3] Larose D, T., *Data Mining Methods and Models*, 2006, Jhon Wiley & Sons, Inc. Hoboken New Jersey
- [4] Poerwadarminta, W.J.S., *Kamus Besar Bahasa Indonesia*. 1991, Jakarta: P.N Balai Pustaka
- [5] Prabowo Pudjo Widodo, Herlawati. *Menggunakan UML*. 2011. Informatika. Bandung wo
- [6] Julsam, Handryawan Adnan Mooduto, Alexyusandri. *Penambangan Data Format Text Excel dengan Software Tanagra*. Elektron Vol.1 No.1, Edisi : juni 2009. ISSN:2085-6989
- [7] Adi Kusrianto, *Memfaatkan Formula dan Fungsi Microsoft excel 2007*, 2007, Jakarta: PT. Elex Media Komputindo