

## IMPLEMENTASI ALGORITME C4.5 UNTUK PREDIKSI PENANAMAN CABAI MERAH

Muhammad Syam Firdaus<sup>1</sup>, Aji Primajaya<sup>2</sup>, Asep Jamaludin<sup>3</sup>  
Program Studi Teknik Informatika, Universitas Singaperbangsa Karawang<sup>1,2,3</sup>  
muhammad.syam17155@student.unsika.ac.id

*Submitted August 9, 2021; Revised November 25, 2021; Accepted November 26, 2021*

### Abstrak

Tanaman cabai menjadi tanaman yang paling banyak ditanam di Indonesia, karena tanaman ini dapat berbuah dan tumbuh di dataran tinggi maupun di dataran rendah. Setiap tahunnya permintaan pasar akan cabai terus meningkat, bahkan setiap tahunnya harga cabai cenderung meningkat yang di akibatkan pasokan cabai berkurang dan permintaan pasar kurang terpenuhi. Kegagalan panen yang dialami para petani cabai menyebabkan pertumbuhan inflasi di tahun 2020 sebanyak 0,31% disumbang oleh naiknya harga di sektor pangan, termasuk di dalamnya cabai rawit dan cabai merah. Fluktuasi harga cabai ini sangat dipengaruhi cuaca, panen raya, kebijakan perdagangan dan momentum yang mengiringi. Sesuai dengan latar belakang, tujuan dari penelitian ini untuk memprediksi penanaman cabai menggunakan algoritme C4.5. pertama proses pengumpulan data cuaca dan harga dilakukan oleh peneliti yang diambil dari DISPERINDAG, BMKG dan data harga cabai di pasaran untuk dijadikan *dataset*, *dataset* dikumpulkan dilakukan *preprocessing* untuk menghilangkan *missing value*, *outliers data* dan juga *imbalance data*, lalu setelah itu dilanjutkan dengan tahap pembuatan model, dari model tersebut dapat memprediksi penanaman cabai. Hasil prediksi dengan menggunakan *rule* dari pohon keputusan yang di dapatkan mempunyai tingkat akurasi sebesar 97,8% berdasarkan perhitungan dari prediksi menggunakan data validasi sebanyak 95 data.

**Kata Kunci :** Prediksi Penanaman Cabai, Algoritma C4.5, Klasifikasi, Data Mining

### Abstract

*Chili has been the most widely grown plant in Indonesia because it can bear fruit and grow in the highlands and the lowlands. Every year the market demand for chili continues to increase, even the annual price of chili tends to increase as a result of reduced chili supply and less fulfilled market demand. Besides, the crop failure experienced by chili farmers cause inflation growth in 2020 as much as 0.31%, resulting from rising prices in the food sector, including the price of cayenne pepper and red chili. The fluctuations in chili prices are strongly influenced by weather, harvest season, trade policies and the accompanying momentum. According to the background, the purpose of this study is to predict chili planting using the C4.5 algorithm. First, the researchers collect weather and price data taken from DISPERINDAG, BMKG, use chili price data on the market as a dataset to be later preprocessed to eliminate missing values, data outliers and imbalance data and then make a model that can predict chili planting. The prediction results using the rule from the decision tree have an accuracy rate of 97.8% based on the calculation of the prediction using validation data as much as 95 data*

**Keywords :** Chili Planting Prediction, C4.5 Algorithm, Classification, Data Mining

## 1. PENDAHULUAN

Tanaman cabai menjadi tanaman yang paling banyak ditanam di Indonesia, karena tanaman ini dapat berbuah dan tumbuh di dataran tinggi maupun di dataran rendah. Cabai biasa digunakan oleh orang Indonesia sebagai bumbu

masakan dapur yang bercita rasa pedas. Setiap tahunnya permintaan pasar akan cabai terus meningkat, meningkatnya kebutuhan pasar dapat dilihat dari banyaknya makanan-makanan yang ada di Indonesia yang memerlukan bahan baku cabai sebagai pelengkap. Namun

persediaan cabai yang minim mengakibatkan harga yang dipatok oleh petani atau pengepul menjadi mahal. Jika harga cabai melonjak tinggi dan kebutuhan pasar kurang terpenuhi maka akan menimbulkan inflasi [1].

Cabai yang rusak atau busuk dapat mengakibatkan kegagalan panen yang dialami oleh para petani cabai. Kegagalan panen tersebut menjadi salah satu faktor penyebab dari kurangnya pasokan cabai di Indonesia. Cabai yang rusak dan busuk tersebut diakibatkan iklim cuaca yang tidak mendukung sehingga buah tidak tumbuh dengan baik dan menyebabkan kekurangan pasokan pada akhir 2019 dan terus menunggak hingga akhir tahun 2020 [2].

No.	Komoditas	Inflasi/Deflasi (% mtm)	Sumbangan (%)
<b>INFLASI</b>			
1	Cabai Merah	36.39	0.13
2	Cabai Rawit	26.43	0.05
3	Minyak Goreng	4.85	0.04
4	Beras	0.85	0.03
5	Bawang Merah	5.35	0.02
6	Bawang Putih	6.39	0.01
7	Ikan Kembung/Ikan Gembung/Ikan Banyar/Ikan Gembelo/Ikan Aso-aso	4.14	0.01
8	Kentang	6.96	0.01
9	Tomat	5.76	0.01
<b>DEFLASI</b>			
1	Daging Ayam Ras	-1.99	-0.03
2	Telur Ayam Ras	-1.88	-0.01

**Gambar 1. Inflasi di Sektor Pangan**

Sumber : Tim Pengendalian Inflasi Pusat (TPIP)

Kegagalan panen yang dialami para petani cabai menyebabkan pertumbuhan inflasi di tahun 2020 sebanyak 0,31% disumbang oleh naiknya harga di sektor pangan, termasuk di dalamnya cabai rawit dan cabai merah. Kenaikan inflasi disumbang oleh cabai rawit yaitu menyumbang kenaikan sebesar 0,05 persen di satu sisi cabai merah menyumbang kenaikan sebesar 0,13 persen [3]. Perubahan cuaca ataupun iklim dapat mempengaruhi produksi cabai setiap tahunnya seperti penelitian yang telah dilakukan oleh Ridho (2020) dalam penelitiannya pada petani cabai di Kota Malang, Kecamatan Pakis. Penelitian tersebut menyebutkan di Kecamatan Pakis mengalami penurunan produksi buah cabai pada tahun 2015 sebesar 319,6 kilogram, yang sebelumnya pada tahun 2014 petani cabai dapat menghasilkan 505,5 kilogram buah cabai.

Tanaman cabai dapat menghasilkan buah cabai yang baik dan bagus apabila cara menanam tanaman cabai tersebut dapat disesuaikan dengan keadaan lingkungan.

Petani dapat membuat rencana sendiri untuk budidaya tanaman cabainya, agar dapat membuat perencanaan tersebut petani cabai memerlukan informasi cuaca. Cuaca menjadi salah satu penyebab besar yang mempengaruhi keberhasilan dalam budidaya tanaman cabai, pertumbuhan tanaman cabai akan baik jika petani cabai dapat mempersiapkan ladangnya seperti pengairan, bedengan dan jenis cabai sesuai dengan cuacanya. Cuaca sifatnya tidak dapat dikontrol oleh manusia, oleh karena itu petani hanya dapat mengantisipasinya dan menyesuaikan cara budidaya tanaman cabai berdasarkan dengan kondisi cuaca [5]. Waktu yang diperlukan untuk budidaya tanaman cabai cukup panjang, yaitu sekitar enam bulan dari pembibitan sampai masa panen [6]. Oleh karena itu informasi terhadap kondisi cuaca yang akan datang sangat diperlukan.

Algoritme C4.5 digunakan untuk mengklasifikasi keadaan cuaca dan harga yang nantinya akan menghasilkan berupa pohon keputusan, dari pohon keputusan tersebut akan menghasilkan sebuah *rule* atau aturan. Aturan yang didapatkan digunakan untuk memprediksi data cuaca dan harga sehingga dapat digunakan pengambilan keputusan untuk menanam cabai. Data cuaca yang digunakan menggunakan data penelitian sebelumnya oleh Rosdiana & Rismayana (2018) yaitu data tahun 2011-2018 dan data terbaru dari tahun 2019 sebagai data tambahan untuk wilayah Bandung. Manfaat yang diharapkan dalam penulisan ini adalah dapat memprediksi keputusan penanaman cabai secara akurat dengan menggunakan algoritme C4.5. Penelitian ini mengacu pada penelitian sebelumnya oleh Rosdiana & Rismayana (2018) yang memprediksi tanam cabai menggunakan algoritme C4.5

dengan hasil akurasi yaitu 72.2 % dengan kriteria akurasi baik menggunakan *confusion matrix*. Berdasarkan penelitian tersebut peneliti ingin melakukan penelitian serupa dengan menggunakan algoritme C4.5 dengan penanganan data seperti *missing value*, *outliers data* dan *imbalance data* serta pembagian data menggunakan *k-fold cross validation* agar menghindari terjadinya *overfitting*, lalu evaluasi dari model menggunakan *confusion matrix* dan mengharapkan dapat memprediksi penanaman cabai lebih baik lagi.

Berdasarkan permasalahan tersebut, penelitian untuk memprediksi penanaman cabai dibutuhkan sehingga diambillah judul penelitian “Implementasi Algoritme C4.5 dalam Memprediksi Penanaman Cabai Merah”.

## 2. METODE PENELITIAN

### Teknik Pengumpulan Data

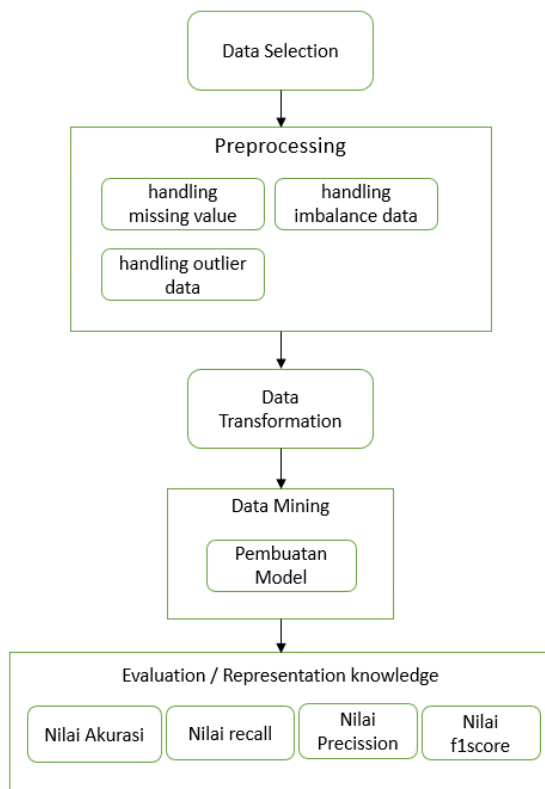
Data diperoleh menggunakan teknik pengambilan data sekunder yaitu mengambil data dari situs Badan Meteorologi Klimatologi dan Geofisika (BMKG) (<https://dataonline.bmkg.go.id/>) dan Dinas Perindustrian dan Perdagangan (Disperindag) (<http://disperindag.jabarprov.go.id/>) di wilayah Bandung, Jawa Barat. Data diambil dari tahun 2011 - 2019. Data yang diambil dari BMKG adalah data cuaca yang terdiri dari suhu rata-rata, kelembapan rata-rata, dan curah hujan, sedangkan data yang di ambil dari Disperindag berupa data harga lokal cabai merah di pasar tradisional. Deskripsi data tersebut dapat dilihat pada **Tabel 1**

**Tabel 1. Deskripsi Data**

No	Atribut	Tipe Data	Keterangan
1	Nama Stasiun	Kategorik	Nama stasiun Meteorologi
2	WMO ID	Numerik	ID stasiun
3	Tanggal	Numerik	Data per hari menggunakan tanggal
4	Suhu rata-rata (°C)	Numerik	Suhu rata-rata yang di ukur dan di hitung rata-rata berdasarkan harian
5	Kelembapan rata-rata (%)	Numerik	Kelembapan rata-rata yang di ukur dan di hitung rata-rata berdasarkan harian
6	Curah hujan (mm)	Numerik	Curah hujan yang di ukur berdasarkan harian
7	Lamanya penyinaran matahari (jam)	Numerik	Lamanya matahari bersinar yang di ukur berdasarkan harian
8	Harga Jabar	Numerik	Data Harga cabai merah di Jawa Barat
9	Harga Lokal	Numerik	Data Harga cabai merah di pasar tradisional di wilayah Bandung
10	Label (tanam/tidak)	Kategorik	Label penanaman cabai perhariannya

### Teknik Analisis dan Pemrosesan Data

Proses analisis dan pemrosesan data menggunakan *tools jupyter notebook* dan juga menggunakan bahasa pemrograman *python*. Penelitian ini menggunakan metode Knowledge Discovery in Database (KDD) untuk mengolah data. *Knowledge Discovery in Database* (KDD) merupakan metode cara memperoleh pengetahuan dari data yang ada. Data tersebut memiliki tabel yang berhubungan / terkoneksi. Pengetahuan yang diperoleh dalam proses ini dapat digunakan sebagai sumber pengetahuan untuk pengambilan keputusan [8]. Berikut adalah tahapan atau proses dalam penelitian ini pada **Gambar 2**.



**Gambar 2. Proses KDD**

### 1. *Data Selection*

Penentuan basis data yang akan digunakan untuk prediksi merupakan tahapan pertama dalam *data mining*. Data yang akan digunakan adalah data penelitian sebelumnya yang berasal dari data BMKG dari tahun 2011 – 2019 di wilayah Bandung, Jawa Barat. Pada tahap seleksi data ini, data yang berasal dari basis data akan diseleksi, sehingga data yang sudah diseleksi tersebut akan dianalisis dan masuk ke tahap selanjutnya.

### 2. *Preprocessing*

*Preprocessing* merupakan salah satu tahap pada *data mining* yang akan menganalisis data dan melakukan pembersihan *noise* pada *dataset*, pembersihan data pada tahap data mining ini perlu dilakukan guna menyingkirkan data-data yang tidak diperlukan untuk proses prediksi nantinya, seperti data *outlier*, *imbalance data* ataupun *missing value*. Penanganan yang akan dilakukan terhadap *missing value* dan *outliers data* yaitu dengan menghapus

sejumlah *record* yang terdapat *missing value* dan juga menghapus *outliers data* (data yang sebarannya terlalu jauh) lalu penanganan terhadap *imbalance data* akan digunakan pendekatan *re-sampling*

### 3. *Data transformation*

Menurut referensi [9] transformasi data adalah proses mentransformasi atau mengubah data ke dalam bentuk yang sesuai, agar dapat diproses. Dalam data *transformation*, terdapat beberapa operasi/teknik untuk melakukan transformasi data, yaitu *smoothing*, *attribute construction*, *normalization*, *aggregation*, dan *discretization* dan yang akan digunakan adalah *discretization* karena pada *dataset* penelitian data harus diubah dari data numerik menjadi kategorik.

### 4. *Data Mining*

Pada tahapan ini data yang sudah di-*preprocessing* akan diolah menggunakan metode dari *data mining*. Pada penelitian ini metode yang digunakan adalah C4.5 dengan proses yang telah dijelaskan sebelumnya. Nantinya data cuaca dan harga cabai tersebut akan dianalisis oleh sistem sehingga membentuk pohon keputusan dan dapat memprediksi penanaman cabai berdasarkan rule dari pohon keputusan tersebut.

### 5. *Evaluation/Representation Knowledge*

Tujuan akhir dari tahap ini yaitu memproyeksikan informasi yang berguna bagi pengguna yang memakainya. Hasil akhir berupa model yang dibuat menggunakan algoritma C4.5, evaluasi yang dihasilkan oleh model yang di-*training* yaitu berupa nilai akurasi, *precision*, *recall* dan *f1score* dari pengujian menggunakan *decision tree classifier* dan pembagian data menggunakan *k-fold cross validation*.

### 3. HASIL DAN PEMBAHASAN

#### Seleksi Data

Tidak semua atribut pada data yang telah dikumpulkan digunakan untuk proses pengolahan *data mining*. Maka dilakukan proses *data selection* atau seleksi data yang bertujuan untuk menganalisis data-data yang relevan dan memilih atribut atau variabel yang berpengaruh pada data cuaca dan harga cabai dalam memprediksi penanaman tanaman cabai. Atribut pada data yang dipilih akan diolah terlihat pada **Tabel 2**.

**Tabel 2. Atribut Terpilih**

No	Atribut Terpilih	Detail Penggunaan
1	Suhu rata-rata (°C)	Atribut Penjelas
2	Kelembapan rata-rata (%)	Atribut Penjelas
3	Curah hujan (mm)	Atribut Penjelas
4	Harga	Atribut Penjelas
5	Label (tanam/tidak)	Atribut Target / Kelas Target

Atribut suhu, kelembapan, dan curah hujan dipilih karena sangat berpengaruh terhadap pertumbuhan tanaman cabai, serta berpengaruh terhadap waktu penanaman cabai. Atribut yang tidak digunakan akan dihapus dan data yang terpilih akan diolah menggunakan algoritma klasifikasi yaitu algoritma C4.5 yang nantinya akan menghasilkan sebuah pohon keputusan, dan pohon tersebut akan menjadi *rule*, yang nantinya dapat memprediksi penanaman cabai.

#### Data Preprocessing

*Preprocessing* merupakan salah satu tahapan penting pada proses *mining* dikarenakan data yang digunakan dalam proses *mining* tidak selamanya dalam kondisi yang ideal untuk diproses terkadang pada data tersebut terdapat berbagai permasalahan yang dapat mengganggu hasil dari proses *mining* itu

sendiri diantaranya seperti *missing value*, *outliers*, dan *imbalance data*. Setelah melihat dan menganalisis data yang ada pada tahapan *data preprocessing* dengan data yang ada yaitu data penelitian sebelumnya yang telah diseleksi dan ternyata harus dilakukan penanganan terhadap *outliers* data, *missing value* dan juga *penanganan imbalance data*.

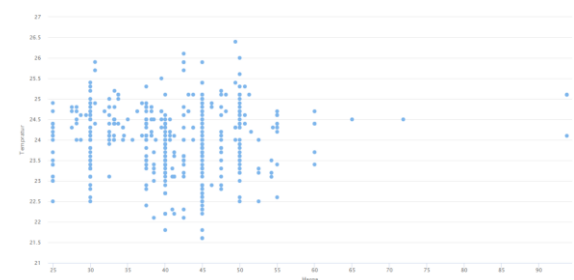
#### a. Penanganan Outliers Data

Pada *dataset* yang sudah diseleksi, data memiliki outliers yang perlu ditangani, *outliers* pada *dataset* dapat dilihat pada atribut pada **Tabel 3**.

**Tabel 3. Outliers Data**

Nama Atribut	Min	Max	Average
Suhu	21	26	23.5
Kelembapan	20	8888	4454
Curah Hujan	0	8888	4444
Harga	21000	94000	57500

Data yang outliers dapat divisualisasikan melalui grafik, salah satunya adalah scatter plot, sebaran data harga terhadap suhu dapat dilihat pada **Gambar 3**.



**Gambar 3. Sebaran Data Atribut Harga**

Pada **Gambar 3** terlihat harga memiliki sebaran data yang padat, akan tetapi data tersebut memiliki *outliers* data, yang sebarannya terlalu jauh, data tersebutlah yang harus ditangani agar data tidak memiliki *noise* yang nantinya akan berpengaruh terhadap model *data mining*. Penanganan pada *outliers* data tersebut adalah, dengan menghapus *range* yang terlalu jauh, seperti *max value* pada atribut kelembapan dan curah hujan yang bernilai

8888, lalu pada atribut harga dengan *range* antara 25000 – 72000 saja sisanya yang melebihi *range* tersebut akan dihapus. Setelah melalui tahap *outliers* data, jumlah *dataset* menjadi 785.

#### b. Penanganan Missing Value

Pada *dataset* yang sudah diseleksi, ternyata masih memiliki *missing value* pada atribut suhu, kelembapan dan juga curah hujan, dapat dilihat keterangan *missing value* pada atribut pada **Gambar 4**

```
In [5]: ds.isnull().any()
#Mengecek adakah nilai null/missing

Out[5]: Suhu                True
         Kelembapan         True
         Curah Hujan         True
         Harga                False
         Label                False
         dtype: bool
```

**Gambar 4. Missing Value**

Penanganan pada data *missing value* tersebut yaitu peneliti menggunakan nilai rata-rata setiap bulannya, misalkan ada data *missing* pada atribut suhu pada bulan *juni* tahun 2014, berarti tinggal dirata-ratakan nilai suhu pada bulan *juni* tahun 2014 lalu isi kolom yang *missing* tersebut pada atribut suhu.

#### c. Penanganan Imbalance Data

*Dataset* setelah diteliti kembali ternyata memiliki ketidakseimbangan data, terdeteksi bahwa data yang memiliki label “tanam” memiliki jumlah data yaitu sebanyak 360, sedangkan label “tidak” memiliki jumlah data lebih banyak yaitu sebanyak 421. Penanganan yang peneliti lakukan adalah dengan melakukan pendekatan *re-sampling* yaitu *under-sampling*. Cara *under-sampling* dalam menangani *imbalance* data ini yaitu dengan cara mengurangi data pada *class* yang besar agar seimbang dengan *class* yang kecil, akan tetapi memiliki ketentuan yaitu hanya dapat dilakukan apabila jumlah data antar *class* tidak terlalu jauh agar cara tersebut efektif. Pada penelitian ini *dataset* antar *class*nya memiliki selisih yang tidak terlalu jauh sehingga bisa menggunakan

cara *under-sampling* ini dengan mengurangi *class* yang besar yaitu *class* “tidak”, sehingga saat ini *dataset* yang digunakan sama-sama memiliki 360 jumlah data setiap *class*-nya.

#### Data Transformation

Tahapan *transformation* merupakan tahapan perubahan data ke dalam bentuk sesuai yang diinginkan dengan kebutuhan yang ada. Pada *dataset* data yang akan dilakukan transformasi yaitu pada atribut label dikarenakan pada proses evaluasi pada *python* label harus menggunakan numerik.

**Tabel 4. Perubahan Data pada Atribut Label**

No	Label	Perubahan
1	Tanam	1
2	Tidak	0

Perubahan tersebut akan diterapkan pada *dataset* yang memiliki jumlah 360 *class* “tanam” diganti menjadi nilai 1, dan 360 lainnya dengan *class* “tidak” diganti menjadi nilai 0.

#### Data Mining

Data telah melewati beberapa tahap diantaranya tahap *selection*, *preprocessing*, dan *transformation* kemudian dilakukan permodelan. Teknik pemodelan data *mining* yang dipilih adalah klasifikasi prediksi dengan menggunakan algoritma C4.5. Klasifikasi algoritma C4.5 digunakan untuk mencapai tujuan awal penelitian yaitu memprediksi penanaman cabai.

Implementasi algoritma C4.5 menggunakan bahasa pemrograman *python* dengan bantuan IDE atau *text editor jupyter notebook* guna membuat model klasifikasi dengan metode C4.5 dan juga dapat memprediksi penanaman cabai. Model yang dibuat menggunakan *Decision tree classifier*, sebuah library dari *sklearn* yang dapat memproses data

menggunakan algoritma decision tree, lalu data akan dibagi menjadi data uji dan data latih. Pembagian data uji dan data latih tersebut menggunakan k-folds cross validation dengan nilai k=10, sehingga nantinya akan terdapat 10 skenario pengujian dengan data uji dan data latih yang berbeda-beda. Proses split data menggunakan k-folds cross validation dengan nilai k=10 yang berarti terdapat 10 skenario pengujian. Terdapat 10 skenario pengujian dengan data uji dan data latih yang berbeda-beda. Berikut adalah penjelasan mengenai skenario yang dilakukan. Berikut adalah contoh untuk skenario pertama:

1. Skenario 1

**Tabel 5. Hasil Prediksi Skenario 1**

	Actual Tanam	Actual Tidak
Pred. Tanam	36	0
Pred. Tidak	0	36

Berdasarkan hasil pada **Tabel 5** terlihat bahwa sebanyak 36 data diprediksi “Tanam” dengan valid kemudian 36 data diklasifikasikan “Tidak” dengan valid. Setelah diketahui hasil dari prediksi tersebut, maka dapat diketahui nilai *accuracy*, *recall*, *precision* dan *f1score*.

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} = \frac{36 + 36}{72} = 1$$

Precision, recall dan f1score class tanam pada skenario 1

$$precision = \frac{TP}{TP + FP} = \frac{36}{36 + 0} = 1$$

$$recall = \frac{TP}{TP + FN} = \frac{36}{36 + 0} = 1$$

$$f1score = \frac{2 * recall * precision}{recall + precision} = \frac{2 * 1 * 1}{1 + 1} = 1$$

*Precision*, *recall* dan *f1score* class tidak pada skenario 1

$$precision = \frac{TP}{TP + FP} = \frac{36}{36 + 0} = 1$$

$$recall = \frac{TP}{TP + FN} = \frac{36}{36 + 0} = 1$$

$$f1score = \frac{2 * recall * precision}{recall + precision} = \frac{2 * 1 * 1}{1 + 1} = 1$$

Dalam skenario pertama ini mendapatkan akurasi sempurna, jika dapat memprediksi semuanya maka *precision*, *recall* dan *f1score* pun akan bernilai sama. Terdapat 10 skenario pengujian yang dilakukan pada penelitian ini, penulis hanya menjelaskan detail pada skenario pertama saja, selanjutnya skenario 2 sampai skenario 10 cara dan rumusnya sama persis dengan skenario pertama namun hasil akurasi, *precision*, *recall* dan *f1score* berbeda karena setiap skenario memiliki data uji dan data latih yang berbeda-beda. Berikut adalah hasil evaluasi dari 10 skenario pengujian pada **Tabel 6**.

**Tabel 6. Hasil Evaluasi**

Sk	Akurasi	Precision	Recall	F1score
1	1	1	1	1
2	1	1	1	1
3	1	1	1	1
4	1	1	1	1
5	0.98611111	0.97297297	1	0.98630137
6	1	1	1	1
7	0.97222222	0.94736842	1	0.97297297
8	0.98611111	0.97297297	1	0.98630137
9	1	1	1	1
10	1	1	1	1

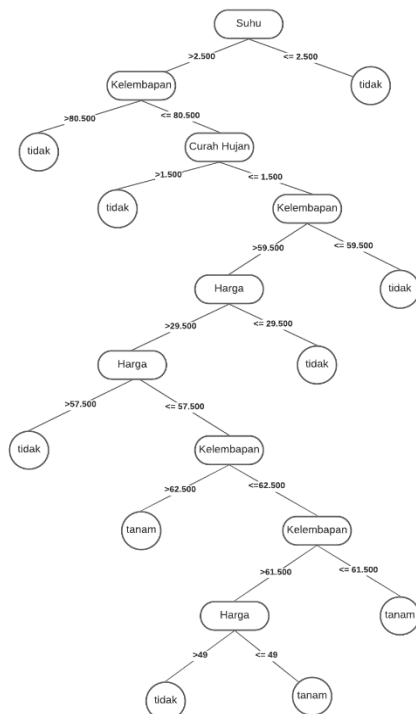
Keterangan:  
SK = Skenario



Terlihat pada **Tabel 6**, selain skenario 5, 7, 8 memiliki nilai akurasi, *precision*, *recall*, dan *f1score* yang sempurna, akan tetapi model yang dibuat belum tentu sempurna, setelah melakukan skenario terhadap *cross validation*, maka seluruh proses dari skenario pertama sampai akhir dirata-ratakan tingkat akurasinya dari proses *cross validation* tersebut, dan hasil akurasi rata-ratanya adalah 0.9944444, *precision* mendapatkan skor rata-rata 0,989331436, *recall* mendapatkan nilai rata-rata 1 dan *f1score* mendapatkan *score* rata-rata 0,994557571.

### 1. Evaluasi

Penelitian ini dilakukan menggunakan metodologi KDD (*Knowledge Discovery in Database*) yang memiliki tahapan lima yaitu, *data selection*, *data processing*, *transformation*, *data mining*, dan *evaluation*. Dari skenario yang sudah dilakukan, algoritma C4.5 menghasilkan suatu pemodelan pohon keputusan yang berupa beberapa *rule* atau aturan untuk



penelitian ini.

**Gambar 5. Pohon Keputusan**

Berdasarkan pohon keputusan tersebut yang dibangun oleh algoritma C4.5 mengenai prediksi tanam cabai di wilayah Bandung, Jawa Barat, menghasilkan sejumlah aturan atau rule sebagai berikut:

- JIKA Suhu  $\leq 2.500$  Maka tidak
- JIKA Suhu  $> 2.500$  dan Kelembapan  $> 80.500$  MAKA tidak
- JIKA Suhu  $> 2.500$  dan Kelembapan  $\leq 80.500$  dan Curah Hujan  $> 1.500$  MAKA tidak
- JIKA Suhu  $> 2.500$  dan Kelembapan  $\leq 80.500$  dan Curah Hujan  $\leq 1.500$  dan Kelembapan  $\leq 59.500$  MAKA tidak
- JIKA Suhu  $> 2.500$  dan Kelembapan  $\leq 80.500$  dan Curah Hujan  $\leq 1.500$  dan Kelembapan  $> 59.500$  dan Harga  $\leq 29.500$  MAKA tidak
- JIKA Suhu  $> 2.500$  dan Kelembapan  $\leq 80.500$  dan Curah Hujan  $\leq 1.500$  dan Kelembapan  $> 59.500$  dan Harga  $> 29.500$  dan Harga  $> 57.500$  MAKA tidak
- JIKA Suhu  $> 2.500$  dan Kelembapan  $\leq 80.500$  dan Curah Hujan  $\leq 1.500$  dan Kelembapan  $> 59.500$  dan Harga  $> 29.500$  dan Harga  $\leq 57.500$  dan Kelembapan  $> 62.500$  MAKA tanam
- JIKA Suhu  $> 2.500$  dan Kelembapan  $\leq 80.500$  dan Curah Hujan  $\leq 1.500$  dan Kelembapan  $> 59.500$  dan Harga  $> 29.500$  dan Harga  $\leq 57.500$  dan Kelembapan  $\leq 62.500$  dan Kelembapan  $\leq 61.500$  MAKA tidak
- JIKA Suhu  $> 2.500$  dan Kelembapan  $\leq 80.500$  dan Curah Hujan  $\leq 1.500$  dan Kelembapan  $> 59.500$  dan Harga  $> 29.500$  dan Harga  $\leq 57.500$  dan Kelembapan  $\leq 62.500$  dan Kelembapan  $> 61.500$  dan Harga  $> 49$  MAKA tidak
- JIKA Suhu  $> 2.500$  dan Kelembapan  $\leq 80.500$  dan Curah Hujan  $\leq 1.500$  dan Kelembapan  $> 59.500$  dan Harga  $> 29.500$  dan Harga  $\leq 57.500$  dan Kelembapan  $\leq 62.500$  dan Kelembapan  $> 61.500$  dan Harga  $\leq 49$  MAKA tanam





45	23.2	83	2	44.15	tidak	tidak
46	24	75	20	44.15	tidak	tidak
47	24.2	76	0	44.15	tanam	tanam
48	25.6	67	5.8	44.15	tanam	tanam
49	24.3	72	0	44.15	tanam	tanam
50	22.8	84	0	40	tidak	tidak
51	23.5	85	7.5	38.25	tidak	tidak
52	24.1	80	19.8	37.75	tanam	tanam
53	23.2	85	34	37.9	tidak	tidak
54	24.5	75	9.7	37.9	tanam	tanam
55	24.2	75	0.3	37.9	tanam	tanam
56	23.2	79	0.5	37.9	tidak	tidak
57	22.3	80	26	35.9	tidak	tidak
58	24	78	36	35.9	tidak	tidak
59	23.7	80	12.8	34.5	tidak	tidak
60	22.8	89	7.4	34.5	tidak	tidak
61	23.3	84	13	34.5	tidak	tidak
62	22.2	90	34.7	34.5	tidak	tidak
63	23.4	86	8.8	34.5	tidak	tidak
64	23.9	77	0.3	34.15	tidak	tidak
65	25.3	69	0	35	tanam	tidak
66	24.1	56	0	35	tidak	tidak
67	24	71	0	35	tanam	tanam
68	24.6	74	0	35	tanam	tanam
69	23.4	87	0.4	35	tidak	tidak
70	23.2	86	13.7	35	tidak	tidak
71	23.1	89	26.5	36.25	tidak	tidak
72	22.1	92	38.5	36.25	tidak	tidak
73	22.8	83	15.8	36.25	tidak	tidak
74	22.5	88	2.3	36.25	tidak	tidak
75	21.6	90	65	36.25	tidak	tidak
76	23.7	83	50.6	42.25	tidak	tidak
77	23.8	80	32.7	42.25	tidak	tidak
78	21.9	91	16.2	42.25	tidak	tidak
79	23.1	83	17.3	42.25	tidak	tidak
80	23.5	85	23.6	42.25	tidak	tidak
81	23.2	84	1.6	44.65	tidak	tidak
82	22.2	88	42	46.9	tidak	tidak
83	23.1	84	8.5	46.9	tidak	tidak
84	22.7	86	5.7	46.9	tidak	tidak
85	22.8	82	10.9	46.65	tidak	tidak
86	23.2	79	5	46.65	tidak	tidak
87	23.5	87	13.2	46.65	tidak	tidak
88	24.7	75	5.8	46.65	tanam	tanam
89	24.7	69	0	41.25	tanam	tanam
90	24	71	7	48.25	tanam	tanam

91	23.5	78	0	48.25	tidak	tidak
92	22.8	86	7.2	45.9	tidak	tidak
93	23.1	81	34.8	45.9	tidak	tidak
94	24	78	0.5	44	tanam	tanam
95	21.7	90	9.2	44	tidak	tidak

**Keterangan:**

- Klmbp = Kelembapan
- CH = Curah Hujan
- Hrg = Harga
- LA = Label Aktual
- LP = Label Prediksi

Setelah disandingkan dengan label aktualnya lihat dengan seksama label yang diprediksi oleh sistem hampir sama persis, namun dalam prediksi tersebut memiliki 2 kesalahan yaitu pada nomor 65 dan 43 yang tidak sama dengan label sebenarnya dan 93 data sisanya terprediksi dengan baik yang artinya sistem dapat memprediksi penanaman cabai dengan baik. Pada tes menggunakan *validation* data ini untuk menunjukkan bahwa model yang dibuat merupakan *good fitting*.

**4. SIMPULAN**

Berdasarkan penelitian yang telah dilakukan, maka dapat disimpulkan bahwa algoritme C4.5 dapat digunakan untuk memprediksi penanaman cabai dengan menggunakan beberapa teknik seperti *cross validation* untuk *splitting* data dan juga *confussion matrix* untuk melihat hasil prediksi.

Hasil penelitian yang telah dilaksanakan menggunakan algoritma C4.5 diterapkan dengan eksperimen menggunakan *k-fold cross validation* dengan jumlah k = 10, maka dari itu pengujian ini memiliki 10 skenario pengujian dan 10 skenario pembagian *dataset*. Pembagian *dataset* menghasilkan nilai akurasi, *precision*, *recall*, *f1score* yang berbeda-beda tiap pembagian datanya dengan kata lain pembagian data berpengaruh terhadap hasil dari algoritma C4.5. Performansi algoritma C4.5 diukur menggunakan nilai akurasi,

*precision, recall, f1score*. Pengujian 10 skenario tersebut menghasilkan tingkat akurasi rata-rata yaitu 0.9944444, *precision* mendapatkan skor rata-rata 0,989331436, *recall* mendapatkan nilai rata-rata 1 dan *f1score* mendapatkan nilai rata-rata 0,994557571. Berdasarkan model yang sudah di buat maka dilakukan pengujian prediksi menggunakan *rule* yang sudah tercipta dari pohon keputusan dan juga pengujian menggunakan data validasi (data yang tidak dilatih oleh model) sebanyak 95 data dan menghasilkan prediksi yang baik dengan 93 data terprediksi benar dan 2 yang salah dan jika di hitung akurasi dari prediksi tersebut menghasilkan akurasi sebesar 0,978947368

#### DAFTAR PUSTAKA

- [1] Hamid, A. (2020, November 25). AACI sebut kurangnya pasokan jadi penyebab kenaikan harga cabai merah. *Industri Kontan*. Available: <https://industri.kontan.co.id/news/aaci-sebut-kurangnya-pasokan-jadi-penyebab-kenaikan-harga-cabai-merah>.
- [2] Suhanto. (2020, November 25). Harga cabai yang melambung diprediksi bertahan hingga akhir february. *Nasional Kontan*. Available: <https://nasional.kontan.co.id/news/harga-cabai-yang-melambung-diprediksi-bertahan-hingga-akhir-february>
- [3] Suhariyanto. (2020, Februari 3) BPS: Inflasi Januari 0,39 persen dipengaruhi kenaikan harga cabai. *Bisnis Tempo*. Available: <https://bisnis.tempo.co/read/1302831/bps-inflasi-januari-039-persen-dipengaruhi-kenaikan-harga-cabai>.
- [4] M. Nukman Ridho, & Nur Edy S., "Pengaruh perubahan iklim terhadap produktivitas tanaman cabai rawit (*Capsicum frutescens* L.) di Kabupaten Malang the effect of the climate change on cayenne pepper (*Capsicum frutescens* L.) productivities in Malang Regency," *Jurnal Produksi Tanaman*, vol. 8, no. 3, pp. 304-314, 2020.
- [5] H. Imtiyaz, B. H. Prasetyo, and N. Hidayat, "Sistem Pendukung Keputusan Budidaya Tanaman Cabai Berdasarkan Prediksi Curah Hujan," *J-PTIHK*, vol. 1, no. 9, pp. 733-738, 2017.
- [6] Sumbar. (2018, April 24). Budidaya Tanaman cabai. Sumbar Litbang. Available: <https://sumbar.litbang.pertanian.go.id/index.php/info-tek/1267-budidaya-tanaman-cabai>  
<http://sumbar.litbang.pertanian.go.id>
- [7] D. Rosdiana and A. H. Rismayana, "Prediksi waktu tanam cabai menggunakan algoritma c4.5," *SINTAK*, pp. 436-442, 2018.
- [8] Y. Mardi, "Data Mining : Klasifikasi Menggunakan Algoritma C4.5," *J. Edik Inform.*, vol. 2, no. 2, pp. 213-219, 2017.
- [9] E. P. Cynthia and E. Ismanto, "Metode Decision Tree Algoritma C.45 Dalam Mengklasifikasi Data Penjualan Bisnis Gerai Makanan Cepat Saji," *Jurasik (Jurnal Ris. Sist. Inf. dan Tek. Inform.*, vol. 3, no. July, pp. 1-13, 2018, doi: 10.30645/jurasik.v3i0.60.
- [10] Tim Pengendalian Inflasi Pusat (TPIP). (2020, Januari 20) Koordinasi pengendalian inflasi tahun 2020. Bank Indonesia. Available: <https://www.bi.go.id>  
<https://www.bi.go.id/id/moneter/koordinasi-pengendalian-inflasi/highlight-news/Pages/Analisis-Inflasi-Januari-2020.aspx>.