

## PREDIKSI WAKTU TANAM CABAI MENGGUNAKAN ALGORITMA C4.5

Dian Rosdiana<sup>1</sup>, Aris Haris Rismayana<sup>2</sup>

<sup>1,2</sup>Teknik Informatika, Politeknik TEDC

e-mail: <sup>1</sup>dianrosdiana72@gmail.com, <sup>2</sup>rismayana@poltektedc.ac.id

### ABSTRAK

Budidaya cabai memang selalu ciamik untuk diperbincangkan. Saat harga tinggi, petani akan menerima keuntungan yang sangat besar. Tapi sebaliknya saat harga cabai anjlok, petani kelimpungan hingga enggan merawat lagi cabainya. Fluktuasi harga cabai ini sangat dipengaruhi cuaca, panen raya, kebijakan perdagangan dan momentum yang mengiringi. Perilaku ikut-ikutan menanam saat harga tinggi juga turut mengubah struktur harga. Apabila masyarakat menanam hanya karena motif mengikuti tingginya harga cabai hari ini, maka seorang petani hanya akan terpelanting dan tidak mampu bertahan dalam kancah budidaya cabai mengingat fluktuasi harga cabai sangat tinggi. Sesuai dengan latar belakang tujuan dari penelitian ini untuk Memprediksi waktu tanam cabai menggunakan algoritma C4.5. pertama proses pengumpulan data cuaca dan harga dilakukan oleh peneliti yang diambil dari DISPERINDAG, BMKG dan data ketersediaan cabai di pasaran untuk dijadikan data training, data training dikumpulkan di lakukan preprocessing dengan melakukan kategorisasi, melakukan pelatihan untuk menghasilkan model, dari model kemudian dibuat sebuah aplikasi yang mampu melakukan prediksi waktu tanam cabai. Hasil didapat hasil akurasi 72.2 % dengan kriteria akurasi Good classification menggunakan confusion matrix

**Kata Kunci:** waktu tanam cabai, Harga Cabai, Data mining, klasifikasi, C4.5, Prediksi

### 1. PENDAHULUAN

Budidaya cabai memang selalu ciamik untuk diperbincangkan. Saat harga tinggi, petani akan menerima keuntungan yang sangat besar. Tapi sebaliknya saat harga cabai anjlok, petani kelimpungan hingga enggan merawat lagi cabainya. Fluktuasi harga cabai ini sangat dipengaruhi cuaca, panen raya, kebijakan perdagangan dan momentum yang mengiringi. Perilaku ikut-ikutan menanam saat harga tinggi juga turut mengubah struktur harga.

Jika ingin tanam pada saat kemarau atau sekitar bulan Juni hingga Agustus, kemungkinan petani akan mendapatkan harga yang rendah pada awal panen, tetapi masih bisa mendapatkan harga sedang dan harga tinggi di musim penghujan dengan syarat kondisi tanaman masih sehat terawat dan panen masih optimal hingga akhir panen. Agar hasil terus optimal, meskipun tanam pada musim kemarau, tinggi guludan tetap harus 40 cm untuk mengantisipasi masuknya musim penghujan. Selain itu aplikasi pestisida hayati juga perlu diberikan untuk menjaga kondisi tanaman).

Apabila tanam pada musim penghujan, atau berkisar antara bulan Oktober hingga Desember, kemungkinan petani akan mendapatkan harga yang tinggi di awal tanam, kemudian harga sedang dan harga rendah di akhir usia panen. Tanam pada musim ini memang rawan terserang penyakit, tetapi ini dapat diantisipasi apabila menerapkan teknis budidaya yang tepat.

Tujuan dari penelitian ini Membuat model yang mampu memprediksi waktu tanam cabai dan mengetahui tingkat akurasi algoritma C4.5 yang diterapkan untuk prediksi waktu tanam cabai.

Pada penelitian ini digunakan dua metode untuk dua aktivitas utama yaitu dokumentasi dan implementasi.

### 2. TINJAUAN PUSTAKA

Data mining adalah 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.

Algoritma decision tree didasarkan padapendekatan divide-and-conquer untuk klasifikasisuatu masalah. Algoritma tersebut bekerja dari atas ke bawah, mencari pada setiap tahap atribut untuk membaginya ke dalam bagian terbaik class tersebut, dan memproses secara rekursif submasalah yang dihasilkan dari pembagian tersebut. Strategi ini menghasilkan sebuah decision tree yang dapat diubah menjadi satu set classification rules [1]. Decision tree merupakan salah satu metode klasifikasi yang menggunakan representasi struktur pohon (tree) di mana setiap node merepresentasikan atribut, cabangnya merepresentasikan nilai dari atribut, dan daun merepresentasikan kelas.

Algoritma C4.5 merupakan pengembangan dari algoritma ID3. Algoritma C4.5 dan ID3 diciptakan oleh seorang peneliti dibidang kecerdasan buatan bernama J. Ross Quinlan pada akhir tahun 1970-an [2]. Algoritma C4.5 membuat pohon keputusan dari atas ke bawah, dimana atribut paling atas merupakan akar, dan yang paling bawah dinamakan daun. Secara umum, algoritma C4.5 untuk membangun sebuah pohon keputusan adalah sebagai berikut:

- a. Hitung jumlah data, jumlah data berdasarkan anggota attribute hasil dengan syarat tertentu. Untuk proses pertama syaratnya masih kosong.
- b. Pilih atribut sebagai *Node*.
- c. Buat cabang untuk tiap-tiap anggota dari *Node*.
- d. Periksa apakah nilai *entropy* dari anggota *Node* ada yang bernilai nol. Jika ada, tentukan daun yang terbentuk. Jika seluruh nilai *entropy* anggota *Node* adalah nol, maka proses pun berhenti.
- e. Jika ada anggota *Node* yang memiliki nilai *entropy* lebih besar dari nol, ulangi lagi proses dari awal dengan *Node* sebagai syarat sampai semua anggota dari *Node* bernilai nol.

*Node* adalah atribut yang mempunyai nilai *gain* tertinggi dari atribut-aribut yang ada. Untuk menghitung nilai *gain* suatu atribut digunakan rumus seperti yang tertera dalam persamaan berikut:

$$Gain(S, A) = Entropy(S) - \left[ \sum_{i=1}^n \frac{A_i}{S} * Entropy(A_i) \right] \tag{1}$$

Keterangan:

*S* : Kasus.

*A* : Atribut

*n* : Jumlah partisi atribut *A*

*A<sub>i</sub>* : Jumlah kasus pada partisi ke-*i*.

*S* : Jumlah kasus.

Sementara itu, untuk menghitung nilai *Entropy* dapat dilihat pada persamaan berikut ini:

$$Entropy(S) = \sum_{i=1}^n - p_i * \log_2 p_i \tag{2}$$

Keterangan:

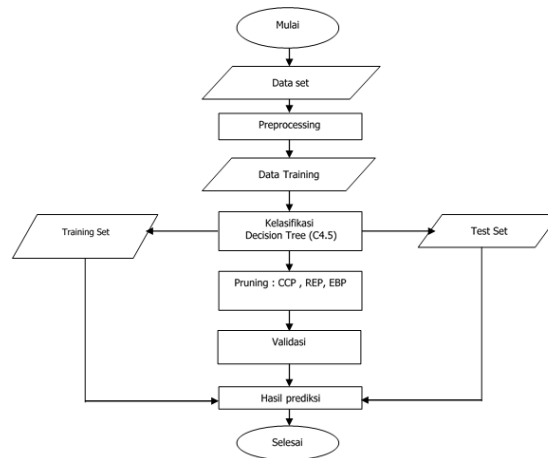
*S* : Himpunan kasus.

*n* : Jumlah partisi *S*

*p<sub>i</sub>* : Proporsi dari *S<sub>i</sub>* ke *S*.

### 3. METODE PENELITIAN

Pada penelitian ini digunakan dua metode untuk dua aktivitas utama yaitu dokumentasi dan implementasi. Dibawah ini adalah alur penelitian :



Gambar 1. Blok diagram sistem

Dari gambar blok diagram sistem dapat di jelaskan sebagai berikut :

#### 1. *Data Mentah*

Data yang digunakan berupa data historis yang didapat dari Dinas Perindustrian dan Perdagangan (DISPERINDAG) serta data curah hujan yang dijadikan data pendukung didapatkan dari Badan Meteorologi, Klimatologi dan Geofisika (BMKG).

#### 2. *Preprocessing*

Tahap ini dilakukan untuk memudahkan algoritma membaca data inputan yang akan diolah. Data harga dan curah hujan merupakan input dari sistem yang akan dibangun. Pada penelitian ini, data akan diolah melalui pembersihan data dan transformasi data. Pembersihan data menghilangkan atribut atau data yang didapatkan tetapi tidak diperlukan dalam penelitian ini, sedangkan transformasi data yaitu dengan melakukan kategorisasi data yang sudah di dapat.

3. Data training

Setelah tahap preprocessing selesai, data ini siap di jadikan data training, data training tersebut diambil untuk diproses ke algoritma C4.5 agar menemukan model prediksi waktu tanam cabai yang tepat.

4. Proses Decision Tree (C4.5)

Metode ini merupakan salah satu metode yang ada pada teknik klasifikasi dalam data mining. Metode pohon keputusan mengubah fakta yang sangat besar menjadi pohon keputusan yang merepresentasikan aturan.

5. Proses Pruning : CCP, REP, EBP

Pruning merupakan bagian dari proses pembentukan decision tree. Saat pembentukan decision tree, beberapa node merupakan outlier maupun hasil dari noisedata. Penerapan pruning pada decision tree, dapat mengurangi outlier maupun noise data pada decision tree awal sehingga dapat meningkatkan akurasi pada klasifikasi data [3]. Oleh sebab itu pemilihan algoritma pruning yang tepat perlu dilakukan untuk mendapat hasil klasifikasi yang maksimal. Pada penelitian yang dilakukan oleh Esposito et al. 1997 [4], algoritma Reduced Error Pruning (REP) disimpulkan sebagai algoritma yang menghasilkan subtree terkecil dengan error rate minimum. Penelitian Esposito et al. (1997) [5], menggunakan algoritma C4.5 untuk membangun decision tree yang di-pruning. Algoritma CART biasa menggunakan Cost Complexity Pruning (CCP) sebagai algoritma pruning-nya. Pada penelitian ini algoritme pruning CCP dibandingkan dengan dua algortime pruning lain yaitu REP dan Error Based Pruning (EBP). Penelitian ini bertujuan untuk menerapkan teknik CCP, REP dan EBP pada metode klasifikasi decision tree dengan algoritma CART. Selain itu penelitian ini juga membandingkan nilai akurasi dari decision tree yang terbentuk, serta waktu proses yang dihasilkan oleh algoritma pruning CCP, REP dan EBP.

6. Validasi

Pengujian dilakukan menggunakan k-flip lintas validasi teknik. Cross-validasi metode digunakan untuk memprediksi keakuratan data pengujian [6].

7. Hasil Prediksi

Output yang dihasilkan oleh algoritma C4.5

4. HASIL DAN PEMBAHASAN

1. Proses Pengolahan Data

Data didapat dari Dinas Perindustrian dan Perdagangan (DISPERINDAG) serta data curah hujan yang dijadikan data pendukung didapatkan dari Badan Meteorologi, Klimatologi dan Geofisika (BMKG). Data yang dikumpulkan yaitu data harga komoditi cabai provinsi Jawa barat dan cuaca sebanyak 858 dari tahun 2011 sd 2018 dengan attribute Suhu rata-rata, kelembaban rata-rata, Curah hujan, harga.

No	Nama Stasiun	WMO ID	Tanggal	Suhu Rata-rata (°C)	Kelembaban Rata-rata (%)	Curah Hujan (mm)	Curah Hujan (mm/tahun)	Lama Penyinaran (jam)	Harga Jabar	Harga Lokal	Label (tanam/tidak)
1											
2	1 Stasiun Geofisika Bandung	96789	1/15/2011	23	85	20	2000	4.1	34252	33000	tidak
3	2 Stasiun Geofisika Bandung	96789	2/12/2011	23.6	74	21	2100	7.8	31428	14000	tidak
4	3 Stasiun Geofisika Bandung	96789	3/12/2011	24.4	75	0	0	8	21704	16000	tanam
5	4 Stasiun Geofisika Bandung	96789	4/9/2011	25.9	72	0	0	6.7	20076	18000	tidak
6	5 Stasiun Geofisika Bandung	96789	5/7/2011	24.5	79	0	0	3.8	15643	18000	tidak
7	6 Stasiun Geofisika Bandung	96789	6/4/2011	24.3	89	0	0	1.7	13329	7500	tidak
8	7 Stasiun Geofisika Bandung	96789	6/12/2011	24.8	88	0	0	5.5	11722	16000	tanam
9	8 Stasiun Geofisika Bandung	96789	7/10/2011	25	79	0	0	5.5	10894	16000	tidak
10	9 Stasiun Geofisika Bandung	96789	8/7/2011	24.9	82	0.2	20	6.3	11333	18000	tidak
11	10 Stasiun Geofisika Bandung	96789	9/4/2011	24.9	77	1	100	6.9	13885	58000	tidak
12	11 Stasiun Geofisika Bandung	96789	10/2/2011	22.7	82	27.4	2740	1.3	13739	32000	tanam
13	12 Stasiun Geofisika Bandung	96789	10/30/2011	21.6	92	24	2400	0	19847	38400	tidak
14	13 Stasiun Geofisika Bandung	96789	11/27/2011	23.2	77	0	0	5.1	32991	38000	tidak
15	14 Stasiun Geofisika Bandung	96789	12/25/2011	23.1	91	8.9	890	4.6	31850	38400	tidak
16	15 Stasiun Geofisika Bandung	96789	1/8/2012	22.2	89	1.6	160	0.8	38400	34800	tanam
17	16 Stasiun Geofisika Bandung	96789	1/10/2012	22.5	88	3.6	360	4.9	38400	28000	tidak
18	17 Stasiun Geofisika Bandung	96789	1/14/2012	23.1	88	4.2	420	2.9	35400	28000	tidak
19	18 Stasiun Geofisika Bandung	96789	1/18/2012	23.5	89	2.3	230	6	30000	27000	tidak
20	19 Stasiun Geofisika Bandung	96789	1/22/2012	24.2	84	0	0	3.6	25786	20600	tanam
21	20 Stasiun Geofisika Bandung	96789	1/26/2012	22	93	9	900	3.6	28000	19600	tidak
22	21 Stasiun Geofisika Bandung	96789	1/30/2012	22.4	90	29.5	2950	4	28000	22600	tidak
23	22 Stasiun Geofisika Bandung	96789	2/3/2012	22	90	1.5	150	3.9	19600	22600	tidak

Gambar 2. Dataset mentah untuk proses

Setelah data dimasukan atau diolah menggunakan Microsoft Office Excel, maka dilakukan pengolahan data dengan cara mengambil data rata-rata x1 (Suhu), rata-rata x2 (Kelembaban), rata-rata x3 (Curah Hujan), rata-rata x4 (Harga) dan Y (Hasil) dan dilakukan pemilihan atribut pada proses pengolahan data maka hasil tabel seperti tabel dibawah ini :

No	Tanggal	Suhu (X1)	Kelembaban (X2)	Curah Hujan	Harga Lokal (X4)	Hasil (Y)	Label (tanam/tidak)
1	1/15/2011	23	85	20	33000	Baik	tidak
2	2/12/2011	23.6	74	21	14000	Tidak	tidak
3	3/12/2011	24.4	75	0	16000	Tidak	tanam
4	4/9/2011	25.9	72	0	18000	Tidak	tidak
5	5/7/2011	24.5	79	0	18000	Tidak	tidak
6	6/4/2011	24.3	83	0	7500	Tidak	tidak
7	6/12/2011	24.8	88	0	16000	Tidak	tanam
8	7/10/2011	25	79	0	16000	Tidak	tidak
9	8/7/2011	24.9	82	0.2	18000	Baik	tidak
10	9/4/2011	24.9	77	1	58000	Baik	tidak
11	10/2/2011	22.7	82	27.4	32000	Baik	tanam
12	10/30/2011	21.6	92	24	38400	Baik	tidak
13	11/27/2011	23.2	77	0	38000	Baik	tidak
14	12/25/2011	23.1	91	8.9	38400	Baik	tidak
15	1/6/2012	22.2	89	1.6	34600	Baik	tanam
16	1/10/2012	22.5	88	3.6	28000	Baik	tidak
17	1/14/2012	23.3	88	4.2	28000	Baik	tidak
18	1/18/2012	23.5	89	2.3	27000	Baik	tidak
19	1/22/2012	24.2	84	0	20600	Baik	tanam
20	1/26/2012	22	93	9	19600	Tidak	tidak
21	1/30/2012	22.4	90	29.5	22600	Baik	tidak
22	2/3/2012	22	90	1.5	22600	Baik	tidak

Gambar 3. Dataset setelah preprocessing

2. Proses Data Mining

Pada tahap selanjutnya data akan di proses menggunakan bantuan Tools Rapid Miner dan untuk pengujiannya menggunakan cross validation yang ada di dalam Rapid Miner, maka data akan terlihat nilai akurasi, pohon keputusan, dan rulenya. Berikut adalah

Data yang akan di import ke dalam Rapid Miner:

Tanggal	Suhu	Kelembaban	Curah Hujan	Harga Lokal	Hasil	Label (tanam/tidak)
Polynomial	Numeric	Numeric	Numeric	Numeric	Bionomial	Bionomial
Id	Attribute	Attribute	Attribute	Attribute	Label	Label
1/15/2011	23	85	20	33000	Baik	tidak
2/12/2011	23.6	74	21	14000	Tidak	tidak
3/12/2011	24.4	75	0	16000	Tidak	tanam
4/9/2011	25.9	72	0	18000	Tidak	tidak
5/7/2011	24.5	79	0	18000	Tidak	tidak
6/4/2011	24.3	83	0	7500	Tidak	tidak
6/12/2011	24.8	88	0	16000	Tidak	tanam
7/10/2011	25	79	0	16000	Tidak	tidak
8/7/2011	24.9	82	0.2	18000	Baik	tidak
9/4/2011	24.9	77	1	58000	Baik	tidak
10/2/2011	22.7	82	27.4	32000	Baik	tanam
10/30/2011	21.6	92	24	38400	Baik	tidak
11/27/2011	23.2	77	0	38000	Baik	tidak
12/25/2011	23.1	91	8.9	38400	Baik	tidak
1/6/2012	22.2	89	1.6	34600	Baik	tanam
1/10/2012	22.5	88	3.6	28000	Baik	tidak
1/14/2012	23.3	88	4.2	28000	Baik	tidak
1/18/2012	23.5	89	2.3	27000	Baik	tidak
1/22/2012	24.2	84	0	20600	Baik	tanam
1/26/2012	22	93	9	19600	Tidak	tidak

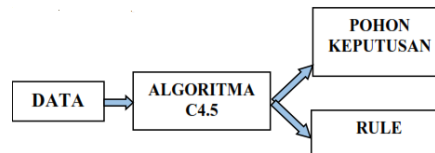
Gambar 4. Data yang di import ke rapid miner

Keterangan :

	:	Nama Variabel
	:	Jenis atau Tipe data
	:	attribute / label

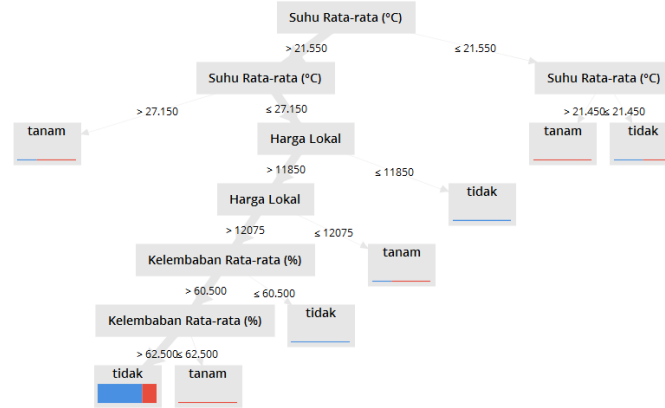
Gambar 5. Keterangan data yang di import

Berikutnya proses pengolahan data menggunakan algoritma C4.5 untuk melihat pohon keputusan dan rule.



Gambar 6. Alur Pohon keputusan C4.5

Setelah dilakukan perhitungan metode algoritma C4.5 pada Rapid Miner, maka terbentuk pohon keputusan seperti gambar di bawah ini :



Gambar 7. Pohon keputusan hasil rapid miner

3. Pengujian

Pengujian ini dilakukan untuk mengetahui kinerja dari metode algoritma C4.5 dalam melakukan klasifikasi terhadap kelas yang telah ditentukan dalam uji coba ini.

Algoritma C4.5 melakukan training terhadap data-data yang telah dibagi oleh cross validation menjadi dua kotak yaitu training dan testing. Training terdiri dari decision tree, sedangkan testing terdiri dari apply model dan performance [7].

Berikut dijelaskan parameter dan operator yang digunakan pada model algoritma C4.5:

- a. Retrieving Data adalah operator yang digunakan untuk mengimport dataset yang akan digunakan, pada penelitian ini data diimport dari file excel.
- b. Validation operator yang digunakan dalam penelitian ini adalah split validation, validation ini hanya membagi total dari keseluruhan dataset menjadi data training dan data testing.
- c. Decision tree adalah metode klasifikasi yang digunakan dalam penelitian ini.
- d. Apply model adalah operator yang digunakan dalam penelitian ini yaitu algoritma C4.5.
- e. Performance adalah operator yang digunakan untuk mengukur performa akurasi dari model

Dan dalam pengujian ini akan menghasilkan nilai akurasi, *precision*, *recall* dan *AUC(Area Under Curve)*. Hasil akurasi *confusion matrix* pada pengujian sebagai berikut:

accuracy: 74.29% +/- 1.23% (micro average: 74.30%)

	true tidak	true tanam	class precision
pred. tidak	393	130	75.14%
pred. tanam	7	3	30.00%
class recall	98.25%	2.26%	

Gambar 8. Hasil akurasi *confusion matrix*

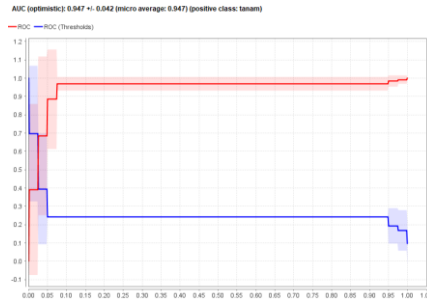
Berikut adalah perhitungan akurasi :

$$\text{Akurasi} = \frac{tp+tn}{tp+tn+fp+fn} \times 100\% \tag{3}$$

$$= \frac{130+7}{130+7+3+393} \times 100\% \tag{4}$$

$$= 74.2\%$$

Hasil AUC (Area Under Curve) pada pengujian ini adalah sebagai berikut:



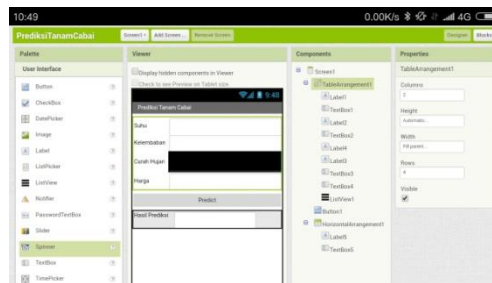
Gambar 9. Hasil AUC (Area Under Curve)

Dari Grafik pengujian AUC didapat bahwa hasil 0.947 berarti pengujian tersebut termasuk good classification.

4. Implementasi

Berdasarkan proses pengolahan data menggunakan Algoritma C4.5 yang telah menghasilkan pohon keputusan dan rule yang telah terbentuk, selanjutnya diimplementasikan rule tersebut dengan membuat program prediksi data pada aplikasi mobile android, dengan menggunakan tools MIT AppInventor.

Berikut ini tampilan desain tempat muka dan juga desain blok-blok logika pemrogramannya.



Gambar 10. Desain tampilan aplikasi



Gambar 11. Blok logika aplikasi

Hasil dari desain dan pengerjaan aplikasi mobile android tersebut adalah sebagai berikut



Gambar 12. Aplikasi Prediksi taman cabai berbasis android



Gambar 13. Prediksi aplikasi

## 5. KESIMPULAN

Dari hasil penelitian yang telah dilakukan oleh peneliti, maka dapat disimpulkan bahwa prediksi tanam cabai dengan memanfaatkan teknik data mining menggunakan algoritma decision tree C4.5 untuk memprediksi (menentukan kelas) dengan data training yang telah diperoleh hasil akurasi 72.2 % dengan criteria akurasi *Good classification* menggunakan confusion matrix.

## 6. SARAN

Untuk meningkatkan kinerja dan menyempurnakan penelitian yang telah dibuat, peneliti memberikan saran sebagai berikut :

- a. Penelitian ini dapat dikembangkan dengan menggabungkan atau membandingkan dengan algoritma klasifikasi lain untuk mendapatkan hasil prediksi yang lebih baik.
- b. Sebaiknya jumlah data ditambah, sehingga dapat diperoleh hasil akurasi

## DAFTAR PUSTAKA

- [1] Witten, Frank, & Hall, 2011, *Data Mining: Practical Machine Learning Tools and Techniques*, second edition, Elsevier.
- [2] J. Rose quinlan, 1986, *INDUCTION OF DECISION TREES, MACHINE LEARNING I*, Elsevier.
- [3] Han & Kamber 2006, *Data Mining; Concepts and Techniques*, 3<sup>rd</sup> Edition, Elsevier.
- [4] Esposito et al. 1997, *Pattern analysis and machine intelligenc*, *Journal IEEE* vol.19
- [5] Hastie & Tibshirani, 2009, *The Elements of Statistical Learning*, Springer
- [6] Mochamad Rizki Ilham S, Purwanto, 2015, *Implementasi Data Mining Menggunakan Algoritma C4.5 Untuk Prediksi Kepuasan Pelanggan Taksi Kosti*, Universitas Dian Nusantoro, Semarang.