

# Analisis Perbandingan Algoritma ID3 Dan C4.5 Untuk Klasifikasi Penerima Hibah Pemasangan Air Minum Pada PDAM Kabupaten Kendal

*Comparative Analysis Of ID3 And C4.5 Algorithm For Classification Of Grant Recipients Of Drinking Water Installation At PDAM Kendal*

Dana Melina Agustina<sup>1</sup>, Wijanarto<sup>2</sup>

<sup>1,2</sup>Jurusan Teknik Informatika, Fakultas Ilmu Komputer, Universitas Dian Nuswantoro, Semarang

e-mail: <sup>1</sup>111201207023@mhs.dinus.ac.id, <sup>2</sup>wijanarto.udinus@gmail.com

## Abstrak

Program hibah air minum bertujuan untuk meningkatkan cakupan pelayanan air minum yang diprioritaskan bagi masyarakat berpenghasilan rendah dalam rangka meningkatkan derajat kualitas kesehatan masyarakat. Pengklasifikasian data masyarakat berperan untuk menentukan pemberian sambungan air minum secara objektif dan akurat. Dalam penelitian ini dilakukan perbandingan metode data mining yaitu algoritma ID3 dan C4.5 yang diterapkan pada data masyarakat berpenghasilan rendah pada PDAM Kabupaten Kendal dengan menggunakan RapidMiner. Hasil pengujian yang menunjukkan bahwa algoritma ID3 nilai akurasi sebesar 98,91%. Sedangkan pada algoritma C4.5 nilai accuracy sebesar 99,14%. Jadi algoritma C4.5 memiliki tingkat akurasi yang lebih besar dari pada algoritma ID3. Sehingga pada kasus penerima hibah pemasangan sambungan air minum diterapkan pada framework php dapat menentukan penerima hibah pemasangan sambungan air minum dengan menggunakan acuan pada algoritma C4.5 yang memiliki akurasi yang lebih baik.

**Kata kunci**—Klasifikasi, Algoritma ID3, Algoritma C4.5, Air Minum, PDAM

## Abstract

The grant program aims to improve the drinking water service coverage of drinking water a priority for low-income people in order to improve the quality of public health. Data classification societies serve to determine the provision of water connection objectively and accurately. In this study, a comparison of data mining methods ie ID3 and C4.5 algorithms are applied to data taps low-income communities in Kendal using RapidMiner. The test results show that the algorithm ID3 accuracy value amounted to 98.91%. While the algorithm C4.5 value amounted to 99.14% accuracy. So C4.5 algorithm has an accuracy rate greater than the ID3 algorithm. So that in the case of a grantee water connections can be applied to determine the PHP framework grantee water connections by using a reference on the C4.5 algorithm has better accuracy.

**Keywords**— Classification, ID3 algorithm, C4.5 algorithm, drinking water, PDAM

## 1. PENDAHULUAN

Berdasarkan proposal Program Hibah Air Minum APBN Februari 2015, pada PDAM ke pemerintah pusat, bahwa pemberian pemasangan sambungan air minum merupakan bantuan dari pemerintah pusat yang terdapat 1473 masyarakat yang akan direkomendasikan untuk mendapatkan pemasangan sambungan air minum. Adapun

yang mendapatkan pemasangan sambungan air minum 1223 masyarakat, sedangkan yang gagal mendapatkan sebanyak 250 masyarakat. Dalam pemberian pemasangan sambungan air minum pada masyarakat dilihat dari daya listrik yang digunakan, fasilitas, bangunan rumah, dan kesediaan menjadi pelanggan.

Pengklasifikasian data masyarakat berperan untuk menentukan pemberian sambungan air minum secara objektif dan akurat. Salah satu metode yang akan digunakan yaitu dengan data mining. Menurut penelitian Yadav dan Pal [1] bahwa klasifikasi *Decision Tree* dilakukan untuk menemukan klasifikasi terbaik untuk prediksi kinerja siswa di tahun pertama ujian. Pada penelitian yang dilakukan Adhatrao [2] bahwa prediksi siswa disemester pertama menggunakan algoritma ID3 menunjukkan tingkat akurasi 75.145%, sedangkan dengan algoritma C4.5 menunjukkan tingkat akurasi 75.145%. Menurut penelitian Sharma dan Sahni [3] dalam klasifikasi data spam email, untuk menentukan apakah email tertentu termasuk spam atau tidak. Pada algoritma J48 memiliki akurasi tertinggi 92,7624%, sedangkan pada algoritma ID3 menunjukkan akurasi 89.111%.

Berdasarkan studi literatur diatas, penelitian ini akan mengimplementasikan metode data mining untuk membangun model klasifikasi masyarakat berpenghasilan rendah untuk penerima pemasangan sambungan air minum pada program hibah air minum di PDAM Kabupaten Kendal. Metode yang digunakan adalah metode *Decision Tree* dengan menggunakan algoritma ID3 dan C4.5. Metode ini dipilih karena berdasarkan penelitian sebelumnya memiliki konsep yang mudah diinterpretasikan hasilnya, serta akan membandingkan algoritma ID3 dan C4.5 mana yang memiliki tingkat akurasi yang lebih tinggi pada klasifikasi penerima hibah pemasangan air minum. Data yang akan digunakan dalam klasifikasi ini yaitu data masyarakat penerima hibah pemasangan sambungan air minum pada tahun 2015. Hasil dari *Decision Tree* mengklasifikasikan menerima atau tidak menerima hibah pemasangan sambungan air minum.

## 2. METODE PENELITIAN

### 2.1 Bahan

Dalam penelitian ini bahan yang dibutuhkan adalah data masyarakat berpenghasilan rendah di PDAM Kabupaten Kendal tahun 2015.

### 2.2 Peralatan

Peralatan dalam penelitian ini meliputi kebutuhan perangkat lunak dan kebutuhan perangkat keras .

Kebutuhan perangkat lunak :

1. Microsoft Windows 7 Ultimate sebagai notebook.
2. Microsoft Exel 2007 sebagai media penulisan dataset.
3. RapidMiner versi 5.3, framework yang akan digunakan untuk melihat hasil akurasi.
4. PHP, framework yang akan digunakan untuk prototipe.

Kebutuhan perangkat keras :

1. Processor Intel Core 2 Duo
2. Memori RAM 2GB
3. Hardisk 320GB

### 2.3 Prosedur pengambilan atau pengumpulan data

Data yang diperoleh berupa data softcopy yang tersedia pada PDAM Kabupaten Kendal berupa data masyarakat berpenghasilan rendah tahun 2015 untuk penerima hibah pemasangan air minum sebanyak 1473 data. Dengan data tersebut terdapat atribut nama, alamat, desa, RT, RW, Kecamatan, daya listrik yang digunakan, KTP dan gambar. Atribut tambahan yang didapat dari hasil survai, yaitu berupa fasilitas yang dimiliki, kondisi bangunan, dan kesediaan.

### 2.4 Teknik analisis data (cara pengolahan data awal)

Dalam penelitian ini menggunakan proses dari *Knowledge Discovery in Database (KDD)* [4] dengan teknik pengolahan data awal sebagai berikut :

#### 1. Pembersihan data

Melakukan pembersihan data dengan tujuan untuk menghilangkan data yang tidak valid dan data yang tidak konsisten. Jika terdapat data atribut yang tidak diperlukan maka atributnya dihilangkan selama proses klasifikasi, sehingga tidak mengganggu proses selanjutnya.

#### 2. Integrasi data

Melakukan integrasi data dengan menggabungkan data dari berbagai *database* ke dalam satu *database* baru. Dimana beberapa sumber data dapat dikombinasikan. Dalam data mining terdapat data yang tidak didapat secara langsung, seperti data yang di peroleh dari hasil survai dilakukan analisis. Sehingga data dapat di gabungkan menjadi satu dataset.

Atribut yang akan digunakan dalam pemodelan yaitu sebagai berikut :

Tabel 1. Atribut yang digunakan untuk dalam pemodelan

No	Nama Atribut	Penjelasan
1	Daya Listrik	Daya listrik yang digunakan rumah penerima hibah pemasangan air minum.
2	Fasilitas	Fasilitas yang dimiliki penerima.
3	Bangunan	Kondisi bangunan rumah.
4	Kesediaan	Kesediaan masyarakat menjadi pelanggan.
5	Keterangan Status (atribut label)	Keterangan apakah berhak menerima hibah pemasangan air minum atau tidak.

Tabel diatas menjelaskan atribut yang digunakan untuk pemodelan data yang mana terdapat atribut daya listrik, fasilitas, bangunan, kesediaan, dan keterangan status. Yang terakhir atribut yang dijadikan target adalah keterangan status apakah masyarakat yang berpenghasilan rendah berhak menerima hibah pemasangan air minum atau tidak.

### 2.5 Model atau metode yang digunakan

Metode yang akan digunakan pada penelitian ini dengan 2 metode yaitu ID3 dan C4.5.

#### • Tahapan Algoritma ID3

Tahapan dari proses algoritma ID3 yaitu sebagai berikut [5]:

1. Menyiapkan data training

2. Hitung nilai Entropy dengan rumus :

$$Entropy(S) = \sum_i^c - P_i \log_2 P_i$$

$$Entropy(S) = (- P_+ \log_2 P_+) - (P_- \log_2 P_-)$$

3. Setelah mendapatkan nilai Entropy akan mencari Information Gain dari setiap atribut untuk mendapatkan nilai Information Gain yang paling tinggi.
4. Rumus dari Information Gain yaitu :
 
$$Gain(S,A) = Entropy(S) - \sum_{v \in \text{nilai}(A)} \frac{|S_v|}{|S|} Entropy(S_v)$$
5. Nilai Information Gain yang tertinggi akan terbentuk menjadi simpul yang pertama dan akan menempati paling atas.
6. Proses perhitungan Information Gain diulangi sampai semua data yang termasuk dalam kelas yang sama.
7. Sehingga akan terbentuk pohon keputusan.
8. Maka akan terbentuk *Rule-Rule*.

- Tahapan Algoritma C4.5

Tahapan dari proses algoritma C4.5 adalah [6]:

1. Mempersiapkan data training.
2. Hitung nilai entropy dengan rumus :
 
$$Entropy(S) = - \sum_{i=1}^n p_i + \log_2 p_i$$
3. Setelah mendapatkan nilai dari Entropy maka akan digunakan untuk mencari nilai Gain.
4. Rumus yang digunakan untuk menghitung Gain sebagai berikut :
 
$$Gain(S,A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} Entropy(S_i)$$
5. Kemudian mencari nilai Split Info dengan rumus :
 
$$Split\ Information(S,A) = - \sum_{i=1}^n \frac{|S_i|}{|S|} \log_2 \frac{|S_i|}{S}$$
6. Setelah mendapatkan nilai Gain dan Split Info, lalu mencari nilai Gain Ratio dengan rumus sebagai berikut :
 
$$Gain\ Ratio(S,A) = \frac{Gain\ Ratio(S,A)}{Split\ Information(S,A)}$$
7. Nilai Gain Ratio tertinggi akan digunakan sebagai atribut akar. Dengan itu akan terbentuk pohon keputusan sebagai node 1.
8. Ulangi proses ke-2 sampai semua cabang memiliki kelas yang sama.
9. Maka akan terbentuk pohon keputusan.
10. Dari pohon keputusan yang terbentuk maka dapat ditentukan *Rule-Rule*.

## 2.6 Evaluasi dan validasi

*Confusion Matrix* merupakan evaluasi kinerja dari model klasifikasi berdasarkan objek dengan memperkirakan yang benar atau salah. *Confusion Matrix* memberikan keputusan yang diperoleh dalam data *training* dan *testing* [7].

Tabel 2. Confusion Matrix 2 Kelas

CLASSIFICATION	PREDICTED CLASS		
	Class = YES	Class = NO	
OBSERVED CLASS	Class = YES	<p style="text-align: center;"><b>A</b> (<i>true positive-TP</i>)</p>	<p style="text-align: center;"><b>b</b> (<i>false positive-FP</i>)</p>
	Class = NO	<p style="text-align: center;"><b>C</b> (<i>false negative-FN</i>)</p>	<p style="text-align: center;"><b>d</b> (<i>true negative-TN</i>)</p>

Keterangan :

- True Positive (TP) : proporsi positif yang terdapat dalam data set yang diklasifikasikan positif.
- False Negative (FN) : proporsi negatif yang terdapat dalam data set yang diklasifikasikan negatif.
- False Positive (FP) : proporsi negatif yang terdapat dalam data set yang diklasifikasikan positif.
- True Negatif (TN) : proporsi positif yang terdapat dalam data set yang diklasifikasikan negatif.

Pengujian pada algoritma ID3 dan C4.5 dilakukan dengan *confusion matrix* untuk mengetahui tingkat *Sensitivity (recall)*, PPV (*positive predictive value*) atau *precision*, dan akurasi dari metode klasifikasi yang dibuat [8] :

Untuk menghitung *recall* dengan rumus :

$$Recall = \frac{TP}{TP+FN}$$

Recall bertujuan untuk mengukur proporsi *true positive* (TP) terhadap tupelo positif yang diidentifikasi secara benar.

Untuk perhitungan *precision* menggunakan rumus :

$$Precision = \frac{TP}{TP+FP}$$

Precision bertujuan untuk mengukur proporsi jumlah kasus yang diprediksi positif yang juga positif benar pada data yang sebenarnya.

Akurasi dihitung dengan rumus :

$$Accuracy = \frac{a+d}{a+b+c+d} = \frac{TP+TN}{TP+TN+FP+FN}$$

Akurasi bertujuan untuk menjumlah prediksi penerima hibah pemasangan air minum yang benar.

### 3. HASIL DAN PEMBAHASAN

Didalam bab ini penulis akan membahas mengenai data yang akan digunakan dalam penelitian, data tersebut akan dihitung menggunakan algoritma ID3 dan C4.5 yang kemudian akan diuji menggunakan *Cross Validation*. Pada penelitian ini, data yang digunakan adalah data Masyarakat Berpenghasilan Rendah tahun 2015 pada PDAM Kabupaten Kendal dengan jumlah data 1473. Pembagian pengujian dengan data training dan testing yang berbeda.

Pengujian data sebagai berikut :

- 1 Data training dan testing 50%
- 2 Data training 60% dan data testing 40%
- 3 Data training 70% dan data testing 30%
- 4 Data training 80% dan data testing 20%
- 5 Data training 90% dan data testing 10%

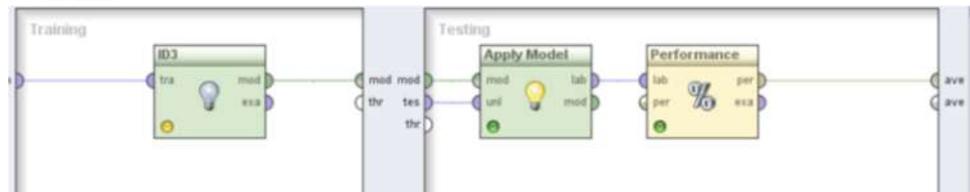
#### 3.1 Metode yang digunakan

- Model Algoritma ID3



Gambar 1. Model Cross Validation pada Algoritma ID3

Pada gambar diatas dataset Masyarakat Berpenghasilan Rendah tahun 2015 dihubungkan dengan operator cross validation yang didalamnya terdapat proses seperti gambar dibawah ini.



Gambar 2. Pengujian Cross Validation pada Algoritma ID3

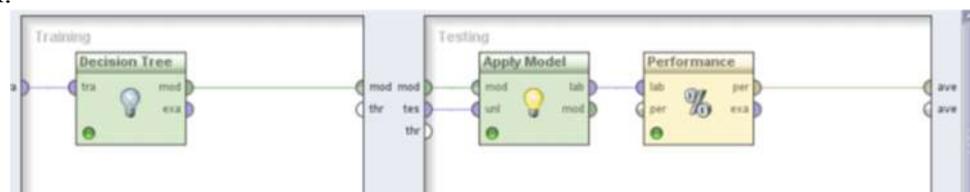
Gambar diatas merupakan model dari algoritma ID3 yang diterapkan dalam rapidminer, dimana operator dan parameter yang digunakan akan berpengaruh terhadap akurasi dan model yang terbentuk.

- Model Algoritma C4.5



Gambar 3. Model Cross Validation pada Algoritma C4.5

Pada gambar diatas sama halnya pada Algoritma sebelumnya yaitu algoritma ID3, dimana dataset Masyarakat Berpenghasilan Rendah tahun 2015 dihubungkan dengan operator cross validation yang didalamnya terdapat proses seperti gambar dibawah ini.



Gambar 4. Pengujian Cross Validation pada Algoritma C4.5

Gambar diatas merupakan model dari algoritma C4.5, dimana proses dalam membangun model C4.5 penggunaan operator dan parameter sama dengan algoritma ID3. Model C4.5 juga akan berpengaruh terhadap akurasi yang dihasilkan.

### 3.2 Hasil Pengujian

Berdasarkan pengujian yang telah dilakukan sebanyak 9 kali yang terdiri dari pengujian data training dan data testing. Berikut ini adalah hasil perbandingan antara kedua algoritma tersebut :

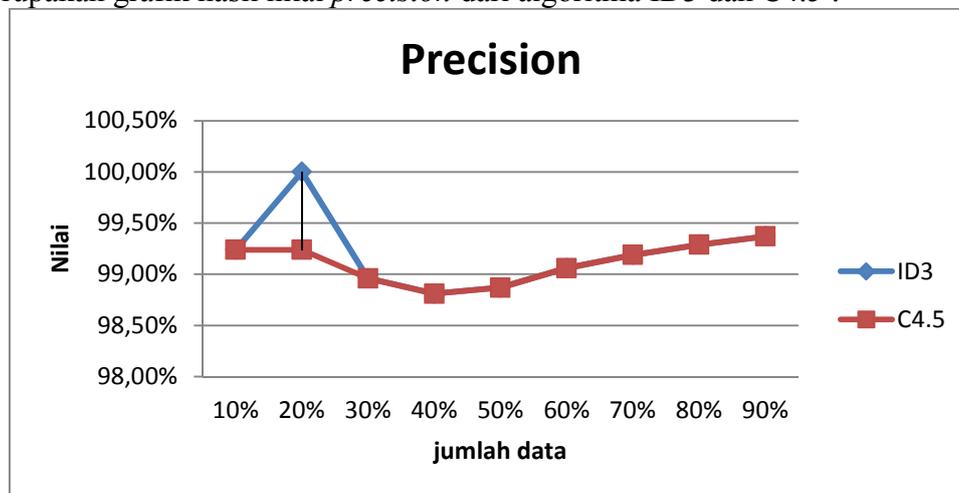
Tabel 3: Perbandingan hasil klasifikasi dari data training

Metode		Pengujian Data Training (%)				
		50	60	70	80	90
ID3	Accuracy	98,64	98,87	99,03	99,15	99,25
	Precision	98,87	99,06	99,19	99,29	99,37
	Recall	97,41	97,90	98,24	98,45	98,63
C4.5	Accuracy	98,92	99,10	99,22	99,32	99,40
	Precision	98,87	99,06	99,19	99,29	99,37
	Recall	99,14	99,30	99,41	99,48	99,54

Tabel 4 : Perbandingan hasil klasifikasi dari data testing

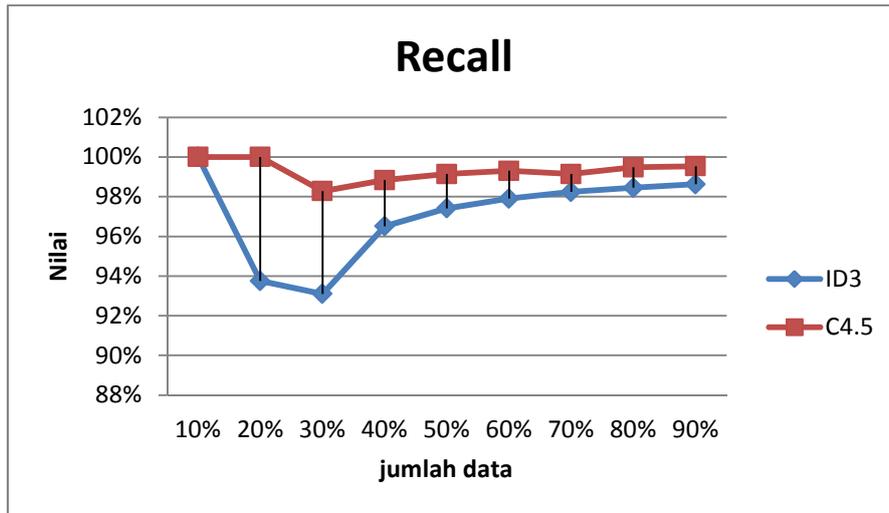
Metode		Pengujian Data Testing (%)				
		10	20	30	40	50
ID3	Accuracy	99,33	99,31	98,19	98,47	98,64
	Precision	99,24	100	98,96	98,81	98,87
	Recall	100	93,75	93,10	96,51	97,41
C4.5	Accuracy	99,33	99,32	98,87	98,81	98,92
	Precision	99,24	99,24	98,96	98,81	98,87
	Recall	100	100	98,28	98,84	99,14

Pengujian dilakukan berdasarkan jumlah data yang diuji berdasarkan data training dan data testing. Hasil dari Tabel 4.9 dan Tabel 4.10 menunjukkan bahwa pengujian data dilakukan mulai dari 10% hingga 90% dari data keseluruhan 1473 data. Pengujian data terbagi sebagai berikut ini 147 (10%), 295 (20%), 442 (30%), 589 (40%), 736 (50%), 884 (60%), 1031 (70%), 1178 (80%), dan 1326 (90%) data. Berikut ini merupakan grafik hasil nilai *precision* dari algoritma ID3 dan C4.5 :



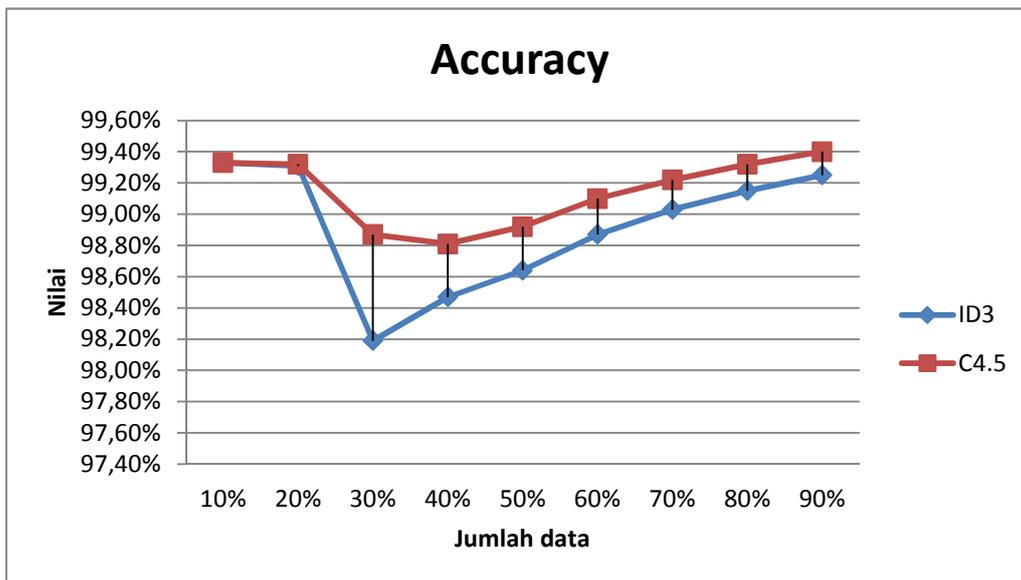
Gambar 5. Grafik Precision

Hasil pengujian menunjukkan bahwa Algoritma ID3 memiliki nilai *precision* 100% dengan jumlah data 295 (20%). Sedangkan pada Algoritma C4.5 nilai *precision* tertinggi dicapai pada data 1326 (90%) sebesar 99,37%. Sehingga algoritma ID3 menunjukkan nilai *precision* lebih tinggi dari pada algoritma C4.5, meskipun pada jumlah data yang lain nilai *precision* yang dihasilkan cenderung sama, tetapi nilai rata-rata yang dihasilkan pada algoritma ID3 lebih tinggi. Sedangkan untuk grafik dari hasil *recall* algoritma ID3 dan C4.5 dapat dilihat dibawah ini :



Gambar 6. Grafik Recall

Dari grafik diatas tertera tingkat *recall* pada jumlah data. Nilai recall tertinggi pada algoritma C4.5 dicapai pada jumlah data 147 (10%) dan 295 (20%) dengan nilai recall yang dihasilkan sama besar yaitu 100%. Sedangkan pada algoritma ID3 nilai recall yang tertinggi yaitu pada jumlah data 147 (10%) dengan nilai recall 100%. Berikut ini merupakan grafik perbedaan tingkat akurasi algoritma ID3 dan algoritma C4.5 :



Gambar 7. Grafik Accuracy

Pada gambar diatas menunjukkan bahwa pada algoritma ID3 nilai Accuracy tertinggi pada jumlah data 147 (data testing 10%) dengan nilai Accuracy 99,33%. Sedangkan pada algoritma C4.5 nilai Accuracy tertinggi adalah 99,40% yang dicapai pada jumlah data 1326 (data training 90%). Sehingga dapat disimpulkan bahwa nilai accuracy algoritma C4.5 lebih baik dari pada algoritma ID3.

Dari pengujian yang telah dilakukan pada algoritma ID3 dan C4.5 dapat disimpulkan bahwa algoritma C4.5 memiliki tingkat akurasi yang lebih tinggi

berdasarkan pengujian jumlah data. Sehingga untuk menentukan penerima hibah pemasangan sambungan air minum dapat dilakukan atau dapat berpedoman pada algoritma C4.5 yang lebih memiliki tingkat akurasi lebih tinggi. Pemilihan penerima sambungan air minum dapat berdasarkan pada pohon keputusan yang terbentuk pada algoritma C4.5. Rule yang terbentuk diambil pada algoritma C4.5 yang memiliki tingkat akurasi tertinggi yaitu pada jumlah data 1326 (90%).

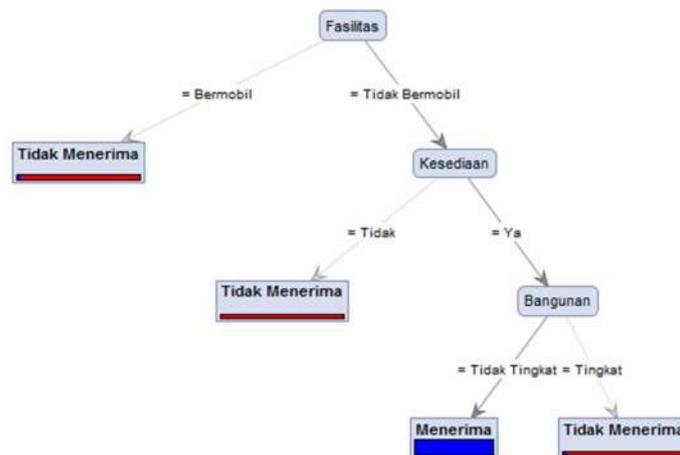
“JIKA fasilitas = bermobil MAKA = Tidak Menerima”

“JIKA fasilitas = tidak bermobil DAN kesediaan = tidak MAKA = tidak menerima”

“JIKA fasilitas = tidak bermobil DAN kesediaan = ya DAN bangunan = tidak tingkat MAKA = Menerima”

“JIKA fasilitas = tidak bermobil DAN kesediaan = ya DAN bangunan = tingkat MAKA = Tidak Menerima”

Berikut ini merupakan pohon keputusan yang dapat dijadikan pedoman :



Gambar 8. Pohon Keputusan Dari Algoritma Tertinggi

### 3.3 Prototype

Berikut ini merupakan prototype dari *decision tree* dengan menggunakan framework PHP :

Gambar 9. Prototype

Gambar diatas merupakan penerapan dari decision tree yang dihasilkan dari nilai akurasi yang tertinggi, sehingga proses yang dilakukan merupakan proses dari rule-rule yang terakhir yaitu dari algoritma C4.5 yang memiliki akurasi tinggi. Prototype digunakan dengan memasukkan data baru yang akan digunakan untuk mengetahui hasil keputusannya.

#### 4. KESIMPULAN

Berdasarkan hasil pengujian dan pembahasan dapat disimpulkan sebagai berikut :

1. Dalam penelitian ini dilakukan perbandingan model algoritma ID3 dengan algoritma C4.5. Keakuratannya diuji dengan mengambil data sebesar 10% hingga 90% dari jumlah data 1473. Pengujian dilakukan sebanyak 9 kali dengan menghitung nilai precision, recall, dan accuracy. Pada algoritma ID3 nilai akurasi sebesar 99,33% pada pengujian 10%, dimana 130 data diklasifikasikan menerima dan 16 data diklasifikasi tidak menerima. Sedangkan pada algoritma C4.5 nilai akurasi sebesar 99,40% pada pengujian 90%, dimana 110 data diklasifikasi menerima dan 218 diklasifikasi tidak menerima.
2. Klasifikasi data penerima hibah pemasangan air minum pada PDAM kabupaten Kendal untuk algoritma C4.5 mempunyai tingkat akurasi yang lebih tinggi dari pada algoritma ID3. Sehingga dengan adanya analisis terhadap algoritma ID3 dan algoritma C4.5 pada kasus penerima hibah pemasangan sambungan air minum PDAM Kabupaten Kendal dapat menggunakan keputusan pada algoritma C4.5 yang memiliki akurasi yang lebih baik, dan sehingga dapat menentukan penerima hibah pemasangan sambungan air minum secara objektif dan tepat sasaran.

## 5. SARAN

Beberapa saran yang penulis sampaikan bagi pengembang penelitian ini antara lain adalah :

1. Dapat dilakukan perbandingan dengan menggunakan metode data mining klasifikasi lainnya yang mendukung pengujian data yang ada.
2. Penelitian ini dapat dikembangkan dengan menambahkan fungsi lainnya pada implementasi decision tree.
3. Penelitian ini dapat digunakan oleh pihak instansi sebagai pedoman pengambil keputusan secara objektif

## DAFTAR PUSTAKA

- [1] S.K.Yadav and S.Pal, "Data Mining : A Prediction for Performance Improvement of Engineering Students using Classification," *World of Computer Science and Information Technology Journal (WCSIT)*, vol. 2, no. 2, ISSN : 2221-0741, pp. 51-56, 2012.
- [2] A. Gaykar, A. Dhawan, R. Jha, and V. Honrao K. Adhatrao, "PREDICTING STUDENTS' PERFORMANCE USING," *International Journal of Data Mining & Knowledge Management Process (IJDMP)*, vol. 3, no. 5, September 2013.
- [3] A.K.Sharma and S.Sahni, "A Comparative Study of Classification Algorithms for Spam Email Data Analysis," *International Journal on Computer Science and Engineering (IJCSE)*, vol. 3, no. 5, ISSN : 0975-3397, May 2011.
- [4] Fajar Astuti Hermawati, *Data Mining*. Surabaya: Andi Offset, 2009.
- [5] Y.Elmande dan P.P.Widodo, "Pemilihan Criteria Splitting Dalam Algoritma Iterative Dichotomiser 3 (ID3) Untuk Penentuan Kualitas Beras : Studi Kasus Pada Peru Bulog Divre Lampung," *Jurnal TELEMATIKA MKOM*, vol. 4, no. 1, ISSN : 2085-725X, Maret 2012.
- [6] R.M.Begum,Dr.V.P.Venkatesan T.M.Lakshmi A.Martin, "An Analysis on Performance of Decision Tree Algorithms using Student's Qualitative Data," *I.J Modern Education and Computer Science*, vol. 5, pp. 18-27, 2013.
- [7] F.Gorunescu, *Data Mining Concept, Model and Technique*. Craiova, Romania: Spinger, 2011.
- [8] Donny.A.B, Lia.A,Wayan.S.W Dennis.A.C, *Belajar Data Mining Dengan Rapid Miner*. Jakarta, 2013.