

OPTIMASI ALGORITMA C4.5 DENGAN MENGGUNAKAN ALGORITMA GENETIKA UNTUK MEMPREDIKSI MEREK METERAN AIR PAM

Aswan Supriyadi Sunge¹⁾, Risnawati²⁾

Program Studi Teknik Informatika Fakultas Teknik
Universitas Pelita Bangsa
aswan.sunge@pelitabangsa.ac.id

Disetujui 30 Desember 2019

Abstraksi

Perusahaan Air Minum telah mengupayakan memberikan pelayanan untuk memenuhi kebutuhan akan air bersih, namun dalam perjalanannya sering mendapat keluhan dari masyarakat atau pelanggan pam. Keluhan pelanggan bisa di sebabkan dari kerusakan pada meteran Air mengakibatkan air mati kedalam rumah tersebut, adanya kebocoran pada hitungan meteran, kebocoran di badan meteran atau kelebihan penggunaan saat pengecekan berlangsung di setiap rumah, banyaknya meteran yang rusak mengakibatkan stock di Gudang tidak sesuai antara kebutuhan penggunaan pemasangan baru dan penggunaan akibat kerusakan meteran. Karena kurangnya air bersih di wilayah tersebut mengakibatkan adanya permintaan dari masyarakat untuk mengoptimalkan kebutuhan mereka, namun kenyataanya belum bisa di optimalkan karena untuk saat ini jalur pipa air pam belum terpasang kesemua jalur. Penggalan data dalam jumlah besar biasa disebut *data mining*. Karena belum adanya penelitian pada merek meteran menggunakan metode Algoritma C4.5 dan dioptimasi menggunakan Algoritma Genetika. Dalam melakukan pengujian ini tools yang digunakan adalah *Rapidminer*. Hasil yang didapatkan menggunakan algoritma C4.5 tanpa optimasi adalah sebesar 63,33% dan hasil yang didapatkan algoritma C4.5 dan optimasi algoritma genetika adalah sebesar 90,00% atau naik sebesar 26,67% dari algoritma C4.5 tanpa optimasi.

Kata kunci : *Data Mining*, Merek Meteran, Algoritma C4.5, Algoritma Genetika,

Abstract

The Drinking Water Company has endeavored to provide services to meet the need for clean water, but in its journey often received complaints from the community or pam customers. Customer complaints can be caused from damage to the meter Water causes water to die into the house, a leakage in the meter count, leakage in the meter body or excess usage when checking takes place in each house, the number of meters damaged causes the stock in the warehouse to be incompatible with the usage needs new installation and use due to meter damage. Due to the lack of clean water in the area, there is a demand from the community to optimize their needs, but the fact is that it cannot be optimized yet because the water pipeline has not yet been installed. Extracting large amounts of data is usually called data mining. Because there is no research on meter brands using the C4.5 Algorithm method and is optimized using the Genetic Algorithm. In conducting this test the tools used are Rapidminer. The results obtained using the C4.5 algorithm without optimization are 63.33% and the results obtained by the C4.5 algorithm and genetic algorithm optimization are 90.00% or an increase of 26.67% from the C4.5 algorithm without optimization.

Keywords: *Data Mining, Meter Brand, C4.5 Algorithm, Genetic Algorithm.*

1. Pendahuluan

Air merupakan kebutuhan dasar yang paling utama. karena hal tersebut air harus tersedia dalam kuantitas dan kualitas yang memadai. Selain merupakan sumber daya alam, air juga merupakan komponen ekosistem yang sangat penting bagi kehidupan manusia dan makhluk hidup lainnya, yang dikuasai oleh Negara dan dipergunakan untuk sebesar-besarnya kemakmuran rakyat. Hal ini tertuang dalam Pasal 33 ayat (3) Undang Undang Dasar Negara Republik Indonesia Tahun 1945. Mengingat pentingnya kebutuhan akan air bersih, maka sangatlah wajar apabila sektor air bersih mendapatkan prioritas penanganan utama karena menyangkut kehidupan orang banyak.

Perusahaan air minum Tirta Buana Indonesia merupakan salah satu perusahaan air swasta yang mempunyai tugas memberikan pelayanan air bersih untuk masyarakat kabupaten Bekasi khususnya saat ini masih di desa satria jaya, kecamatan tambun utara, pada saat ini pam tirta buana Indonesia mampu melayani sekitar 20% dari seluruh penduduk desa satria jaya, sedangkan target pelayanan sekitar 70 – 80% di desa satria jaya.

Perusahaan air minum Tirta Buana Indonesia memiliki kewajiban untuk meningkatkan keuntungan. Salah satu cara dalam meningkatkan keuntungan tersebut adalah dengan cara marketing untuk mendapatkan pelanggan baru yang berpotensi besar,

Salah satu permasalahan yang di hadapi saat ini adalah adanya perbedaan merek meteran yang di gunakan pada setiap rumah. Perbedaan merek meteran mengakibatkan terjadinya air mati kedalam rumah tersebut, adanya kebocoran pada hitungan meteran atau bahkan kebocoran di badan meteran atau kelebihan penggunaan pada saat pengecekan meteran berlangsung di setiap rumah.

2. Tinjauan Studi

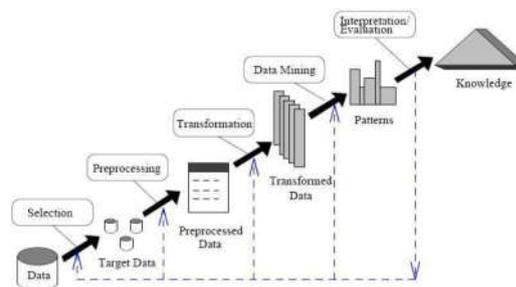
2.1 Analisis Penelitian Terkait

1. *Optimasi Algoritma Naïve Bayes Dengan Menggunakan Algoritma Genetika Untuk Prediksi Kesuburan (Fertility)*. [1] Tujuan dari penelitian ini adalah untuk mengetahui tingkat kesuburan (fertilitas) karena dalam dua dekade mengalami penurunan, dalam penelitian ini total data yang dikumpulkan berjumlah 160 data dan dibagi menjadi 2 bagian yaitu data training 90% dan data testing 10% proses pengujian ini menggunakan naïve bayes dan algoritma genetika. Hasil dari penelitian ini sebelumnya menggunakan naïve bayes menunjukkan tingkat akurasi 97,66% dan setelah di optimasi menggunakan data yang sama, untuk mengoptimalkan naïve bayes dengan algoritma genetika result meningkat menjadi akurasi 99.33%.
2. Review : Data mining menggunakan Algoritma Genetik [2]. Algoritma genetika dapat diterapkan ke dalam berbagai sistem data mining, algoritma genetika dapat digunakan untuk meningkatkan suatu akurasi data. Berdasarkan hasil penelitian yang telah dilakukan oleh berbagai peneliti, tingkat keakuratan menggunakan algoritma genetika pada data mining berada pada rentang 70% hingga 90%

2.2 Data mining

Data mining merupakan salah satu bagian dalam KDD. KDD adalah singkatan dari *Knowledge Data Discovery*. Banyak orang yang menganggap *data mining* dan KDD itu sama, karena sebagian besar pekerjaan dalam KDD difokuskan pada *data mining*. Ada dua jenis metode pada *data mining*, metode *predictive* dan metode *descriptive*. Metode *predictive* adalah proses untuk menemukan pola dari data yang menggunakan beberapa variable untuk memprediksi variabel lain yang tidak diketahui jenis atau nilainya. Teknik yang termasuk dalam *predictive mining* antara lain Klasifikasi, Regresi dan Deviasi. Sedangkan metode *descriptive* adalah proses untuk menemukan suatu karakteristik penting dari data dalam suatu basis data. Teknik data mining yang termasuk dalam *descriptive mining* adalah *Clustering*, *Association* dan *Sequential Mining*. [3]

2.3 Tahapan Data Mining



Gambar 1. Tahapan Data Mining

Proses *data mining* memiliki beberapa tahapan, berikut adalah tahapan pada proses *data mining* :

- 1) *Data Selection*, data akan diseleksi berdasarkan kecocokan data yang akan diambil keputusan.
- 2) *Data Preprocessing/Cleaning*, pada tahap ini dilakukan pembersihan data yang kosong, *duplicate* atau tidak sesuai dari yang ingin diputuskan.
- 3) *Transformation*, melihat data yang sudah dipilih dan dipresentasikan dari hasil yang diinginkan.
- 4) *Data Mining*, melihat pola yang ingin ditampilkan dari metode maupun teknik yang dipilih sebelumnya misalnya klasifikasi, *clustering*, regresi, CART dan lain sebagainya.
- 5) *Interpretation/Evaluation*, pada tahap ini sebagai penerjemah dari data yang telah ditampilkan dan melihat hasil dari teknik atau metode yang digunakan. [4]

2.4 C4.5

Algoritma C4.5 merupakan pengembangan dari ID3 yang digunakan untuk mengambil keputusan dengan membentuk pohon keputusan yang memiliki simpul cabang secara optimal sampai menghasilkan cabang akhir. [5]

2.5 Keuntungan dan kekurangan Algoritma C4.5

Algoritma C4.5 memungkinkan pemangkasan pohon keputusan yang dihasilkan. Hal ini meningkatkan tingkat kesalahan pada data pelatihan, tetapi yang lebih penting, menurunkan tingkat kesalahan pada data pengujian yang tidak terlihat. Algoritma C4.5 juga dapat menangani atribut numerik, nilai yang hilang, data *noise*. Dan memiliki kelebihan dan kerugian, [6] sebagai berikut :

1. Keuntungan, yaitu :
 - a. C4.5 dapat menangani atribut kontinyu dan diskrit
 - b. C4.5 memungkinkan nilai atribut yang hilang diubah menjadi tanda “?”, Nilai atribut yang hilang tidak digunakan dalam penghitungan *gain* dan *entropy*.
 - c. C4.5 melalui pohon yang sudah dibuat dan dicoba untuk menghapus cabang yang tidak membantu dengan menggantinya dengan simpul daun.
2. Kerugian, yaitu :
 - a. C4.5 dapat membangun cabang kosong yang nilainya tidak berkontribusi untuk menghasilkan aturan.
 - b. Terjadi *overfitting* akibat dari *noise* data.
 - c. *Noise* data yang rentan terjadi pada C4.5

2.6 Kerangka Pemikiran

Dalam suatu penelitian perlu adanya kerangka pemikiran yang digunakan sebagai landasan dan pedoman agar penelitian ini berjalan sesuai dengan yang telah direncanakan. Permasalahan pada penelitian ini adalah belum adanya metode yang digunakan untuk memprediksi merek meteran, kemudian dilakukan studi literatur yang berguna untuk mendapatkan referensi tentang permasalahan yang sedang diteliti. Metode yang digunakan pada penelitian ini adalah algoritma C4.5 dan juga algoritma genetika untuk dilakukan pengujian. Pengujian dari metode yang telah di terapkan

menggunakan cara *Confusion Matrix* dan Kurva ROC. Untuk tool yang digunakan untuk melakukan pengujian metode adalah Rapidminer. Adapun kerangka pemikiran yang digunakan sebagai berikut :

MASALAH		
Belum ada peneliti yang prediksi atau melakukan pengujian dengan metode algoritma C4.5.		
STUDI LITERATUR		
Mencari referensi teori yang sesuai dengan permasalahan yang akan dilakukan peneliti.		
PERCOBAAN		
DATA	METODE	TOOL
Dataset penggunaan merek meteran pelanggan air PAM	Algoritma C4.5 dan Algoritma Genetika	Rapid Miner

PENGUKURAN
Confusion Matrix dan Curva ROC
HASIL
teknik optimasi menggunakan algoritma genetika berhasil meningkatkan akurasi dari algoritma C4.5 dalam memprediksi merek meteran air pam. Hasil akurasi dari algoritma C4.5 tanpa optimasi adalah sebesar 63,33% dan setelah dioptimasi menggunakan algoritma genetika hasil akurasinya menjadi sebesar 90.00% atau naik sebesar 26,67% dari algoritma C4.5 tanpa optimasi.

3. Desain Penelitian/Metodologi

3.1 Data yang Digunakan

Data yang di gunakan pada penelitian ini menggunakan data primer karena dataset yang diperoleh langsung dari PT. Tirta Buana Indonesia sebanyak 300 data.

3.2 Pengolahan Data Awal

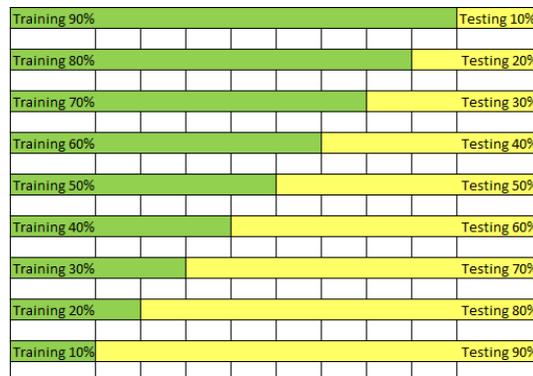
Pada data merek meteran air PAM PT. Tirta Buana Indonesia yang telah dikumpulkan, maka atribut yang diperlukan seperti tarif air, status, active, lokasi pemasangan, jenis dan merek. hal ini dilakukan karena atribut tersebut berpengaruh terhadap pengolahan data pada proses selanjutnya.

No	Atribut	Tipe Data	Keterangan
1	Tarif Air	Text	-
2	Status	Text	-
3	Active	Text	-
4	Lokasi pemasangan	Text	-
5	Jenis	Text	-
6	Merek	Text	Class

Data yang sudah disiapkan untuk klasifikasi dibagi menjadi dua yaitu untuk data *training* 80% dan data *testing* 20%. Pembagian data menjadi data training dan data testing menggunakan tools *Split Validation*. *Split Validation* adalah Teknik validasi yang membagi data menjadi dua bagian secara acak. Dengan menggunakan *Split Validation* akan dilakukan percobaan training berdasarkan split ratio yang telah di tentukan sebelumnya, dan kemudian sisa dari split ratio data training akan dianggap sebagai data testing.

Data training adalah data yang akan dipakai dalam melakukan pembelajaran sedangkan data testing adalah data yang belum pernah dipakai sebagai pembelajaran dan akan berfungsi sebagai data pengujian kebenaran atau keakurasian hasil pembelajaran. [7]

Berikut adalah ilustrasi *Split Validation*:



Gambar 2. Ilustrasi *Split Validation*

1. Choosing the appropriate Data Mining task

Pada tahap ini jenis data mining yang digunakan pada penelitian untuk memprediksi penentuan merek meteran air PAM yang lebih banyak digunakan maka jenis data mining yang akan digunakan adalah Klasifikasi.

2. Choosing the Data Mining Algorithm

Setelah memilih jenis data mining yang akan digunakan yaitu Klasifikasi, maka selanjutnya menentukan algoritma klasifikasi yang akan digunakan yaitu Algoritma C4.5 setelah itu dioptimasi menggunakan Algoritma genetika.

3. Employing the Data Mining Algorithm

Tahap ini di lakukan untuk pengolahan data dengan algoritma yang telah di tentukan untuk mendapatkan algoritma terbaik dan tingkat akurasi yang tinggi dalam klasifikasi prediksi merek meteran air PAM.

4. Evaluation

Dalam tahap ini dilakukan evaluasi dan menafsirkan pola yang di dapatkan dari hasil algoritma yang dipakai. Evaluasi dilakukan dengan menerapkan pola yang didapat dari proses sebelumnya terhadap data testing yang disediakan. Evaluasi di lakukan dengan *confusion matrix* dan Kurva ROC. Hasil dari *confusion matrix* akan digunakan untuk menampilkan *Accuracy*, *Precision*, dan *Recall*. *Accuracy* merupakan presentase antara nilai prediksi dengan nilai sebenarnya. *Recall* merupakan kinerja keberhasilan algoritma yang digunakan. *Precision* merupakan nilai akurasi dengan *class* yang telah di prediksi. Dan Kurva ROC digunakan untuk menghasilkan nilai *Area Under Curve* (AUC). [8] Berikut adalah table *Confusion matrix*:

Tabel 3. 1 Model *Confusion Matrix*

Confusion Matrix		Nilai Prediksi	
		Positif	Negatif
Nilai Sebenarnya	Positif	(a) TP	(b) FP
	Negatif	(c) FN	(d) TN

Keterangan ;

a: jika nilai prediksi positif dan kelas sebenarnya positif

b: jika nilai prediksi negatif dan kelas sebenarnya positif

c: jika nilai prediksi positif dan kelas sebenarnya negatif

d: jika nilai prediksi negatif dan kelas sebenarnya negatif

berikut adalah rumus untuk menghitung *Accuracy*:

$$Accuracy = \frac{a + d}{a + b + c + d}$$

Berikut adalah rumus untuk menghitung

Precision :

$$Precision = \frac{a}{a + b}$$

$$Recall = \frac{a}{a + c}$$

Berikut adalah rumus untuk menghitung *Recall*:

a + c Keakurasian nilai AUC dapat dikategorikan

menjadi 5 kategori antara lain:

1. Akurasi bernilai 0.90 – 1.00 = *Excellent Clasification*
2. Akurasi bernilai 0.80 – 0.90 = *Good Clasification*
3. Akurasi bernilai 0.70 – 0.80 = *Fair Clasification*
4. Akurasi bernilai 0.60 – 0.70 = *Poor Clasification*
5. Akurasi bernilai 0.50 – 0.60 = *Failure Clasification*

2. Using the discovered knowledge

Pada tahap ini menggunakan pengetahuan yang diperoleh dari proses data mining untuk penerapan pada aplikasi atau lainnya. Pengetahuan klasifikasi merek meteran pada air PAM diterapkan pada data baru untuk membuat klasifikasi merek meteran air PAM mana yang lebih banyak digunakan.

4. Hasil Dan Pembahasan

4.1 Pengujian Decision Tree

Tahap awal yang dilakukan dalam penelitian ini yaitu mempersiapkan data yang telah di peroleh, Data yang di ambil adalah dataset Pelanggan PT. Tirta Buana Indonesia yang merupakan data penggunaan merk meteran air PAM data tersebut berjumlah 300 data dan memiliki 6 atribut.

Tabel 4. 1 Data awal

Data diatas dapat dikatakan bahwa ini adalah data awal yang harus dikelola terlebih dahulu, setelah data dianggap sudah cukup maka tahap selanjutnya adalah pengujian pada aplikasi *rapidminer*.

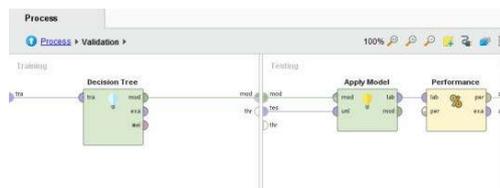
tahap selanjutnya adalah menggunakan operator *Split Validation*. Operator *Split Validation* disini berguna untuk membagi data menjadi dua bagian berdasarkan *split ratio* yang telah ditentukan

Tarif Air	Status	Active	Lokasi	Jenis	Merk
K3	SWADAYA	AKTIF	Griya satria pesona	CICILAN	Arita
K3	KONTRAK	SAMB. BARU	Taman Edelwis	CICILAN	Arita
K2	SWADAYA	AKTIF	Taman Edelwis	CICILAN	Meglio
K1	SWADAYA	AKTIF	Graha Prima	CASH	Meglio
K3	SWADAYA	AKTIF	Taman Satria indah	CICILAN	Arita
K3	SWADAYA	AKTIF	Graha Prima	CICILAN	Arita
K2	SWADAYA	SAMB. BARU	Graha Prima	CICILAN	Meglio
K3	KONTRAK	SAMB. BARU	Griya satria pesona	CASH	Arita
K2	SWADAYA	AKTIF	Taman Satria indah	CICILAN	Meglio
K2	SWADAYA	AKTIF	Taman Satria indah	CASH	Meglio
K3	SWADAYA	SAMB. BARU	Taman Edelwis	CASH	Arita
K3	SWADAYA	AKTIF	Taman Edelwis	CASH	Arita
K2	SWADAYA	AKTIF	Taman Edelwis	CICILAN	Meglio
K3	KONTRAK	SAMB. BARU	Taman Satria indah	CASH	Arita
K1	SWADAYA	SAMB. BARU	Griya satria pesona	CICILAN	Arita
K2	SWADAYA	AKTIF	Griya satria pesona	CICILAN	Meglio
K3	SWADAYA	SAMB. BARU	Taman Satria indah	CICILAN	Arita
K3	SWADAYA	SAMB. BARU	Taman Edelwis	CASH	Arita
K2	SWADAYA	AKTIF	Graha Prima	CASH	Meglio
K2	KONTRAK	AKTIF	Graha Prima	CICILAN	Meglio
....

Tabel 4. 1 Data awal

Data diatas dapat dikatakan bahwa ini adalah data awal yang harus dikelola terlebih dahulu, setelah data dianggap sudah cukup maka tahap selanjutnya adalah pengujian pada aplikasi *rapidminer*.

tahap selanjutnya adalah menggunakan operator *Split Validation*. Operator *Split Validation* disini berguna untuk membagi data menjadi dua bagian berdasarkan *split ratio* yang telah ditentukan sebelumnya, *split ratio* yang telah dipilih akan digunakan sebagai data *training* dan sisanya akan digunakan sebagai data *testing*. Pada *parameters Split Validation* pilih *split ratio* 0.8 karena peneliti ingin datanya dibagi menjadi 80% untuk data *training* dan 20% untuk data *testing*. Namun jika diklik dua kali maka *Split Validation* akan memiliki dua bagian yaitu *Training* dan *Testing*, atau bisa dilihat seperti gambar berikut :

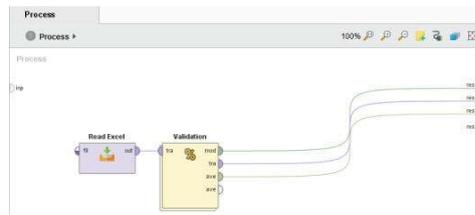


Gambar 3. Split Validation

Pada Gambar 3 dapat dilihat pada bagian *Training* digunakan algoritma klasifikasi *Decision Tree* dan pada bagian *Testing* menggunakan operator *Apply Model* untuk mengaplikasikan model pada data *testing* dan operator *Performance* yang digunakan untuk menampilkan *accuracy*.

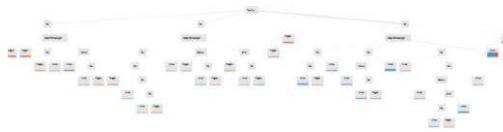
Tahap selanjutnya adalah menghubungkan seluruh operator yang akan digunakan, dapat kita

lihat pada gambar berikut :



Gambar 4. Model pengujian Decision tree

Dari proses yang telah dijalankan dalam **Gambar 4** maka didapatkan hasil sebagai berikut :



Gambar 5. Hasil pengujian Decision tree

Deskripsi diatas menunjukkan aturan-aturan yang berlaku untuk digunakan sebagai landasan memprediksi penggunaan merek meteran air PAM. Berikut adalah hasil accuracy, recall, precision, dan nilai AUC.

accuracy: 63,33%			
	true Airfa	true Meglo	class precision
pred Airfa	37	14	72,55%
pred Meglo	19	20	51,28%
class recall	66,67%	59,82%	

Gambar 6. Hasil Accuracy Decision Tree

Hitungan accuracy :

$$\text{Accuracy} = \frac{a + d}{a + b + c + d} = \frac{37 + 20}{37 + 14 + 19 + 20} = \frac{57}{90} = 0,6333 = 63,33\%$$

precision: 51,28% (positive class: Meglo)			
	true Airfa	true Meglo	class precision
pred Airfa	37	14	72,55%
pred Meglo	19	20	51,28%
class recall	66,67%	59,82%	

Dari **Gambar 7**, maka dapat dilihat nilai *accuracy* yang didapat dari pengujian algoritma *Decision Tree C4.5* yaitu sebesar 63,33%.

Hitungan Precision :

$$\text{Precision} = \frac{a}{a + b}$$

$$= \frac{37}{37+14} = \frac{37}{51} = 0,725 = 72,55\%$$

Dari **Gambar 4.5** maka dapat dilihat nilai *precision* yang didapat dari pengujian algoritma *Decision Tree C4.5* yaitu sebesar 51,28%

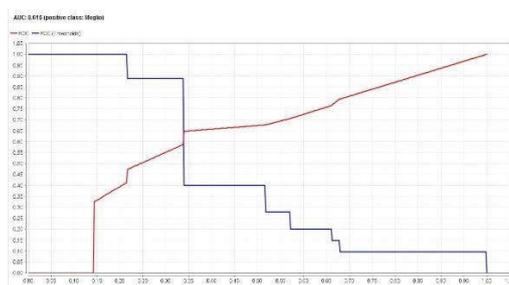
recall: 58,82% (positive class: Megis)			
	true Aita	true Megis	class precision
pred Aita	37	14	72,55%
pred Megis	19	20	51,28%
class recall	59,07%	58,82%	

Gambar 8. Hasil Recall Decision tree

Hitungan Recall :

$$\text{Recall} = \frac{a}{a+c} = \frac{37}{37+19} = \frac{37}{56} = 0,6607 = 66,07\%$$

Dari **Gambar 4.6** maka dapat dilihat nilai *recall* yang didapat dari pengujian algoritma *Decision Tree C4.5* yaitu sebesar 58,82%.



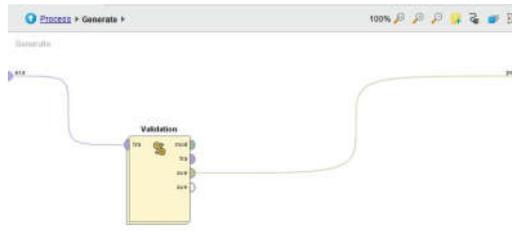
Gambar 5. Hasil AUC Decision tree

Dari **Gambar 4.7** maka dapat dilihat nilai AUC yang didapat dari pengujian algoritma *Decision Tree C4.5* yaitu sebesar 0,616 dan termasuk dalam kategori *Poor Clasifikasi* karena berada pada range 0,60 - 0,70 yang artinya menunjukkan hasil yang baik dalam akurasi.

4.2 Pengujian Decision tree + Algoritma Genetika

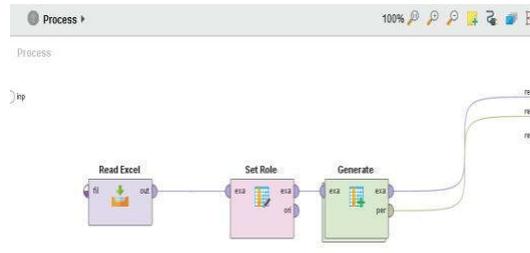
Model awal pengujian Decision Tree + Algoritma Genetika sama seperti pengujian Decision Tree, hanya saja pada pengujian Decision Tree + Algoritma Genetika diperlukan operator tambahan yaitu Set Role dan Generate. Operator Set Role digunakan untuk menerapkan Algoritma Genetika dan Generate digunakan untuk metode Algoritma Genetika, didalam Generate terdapat Split Validation

untuk membagi data menjadi dua bagian seperti gambar berikut :



Gambar 6. Konfigurasi operator *Generate*

Dari **Gambar 10** dapat dilihat bahwa didalam operator *Generate* terdapat *Split validation* yang sama dengan *Split Validation* yang ada di pengujian *Decision Tree*.



Gambar 7. Model Decision tree + Algoritma Genetika

Dari proses yang telah di jalankan pada **gambar 11** maka didapatkan hasil sebagai berikut :

accuracy 90,00%			
	true Arita	true Meglio	class precision
pred. Arita	53	6	89,82%
pred. Meglio	3	28	90,32%
class recall	94,64%	82,35%	

Gambar 8. Hasil Accuracy *Decision tree* + *Algoritma Genetika*

Hasil accuracy :

$$\begin{aligned}
 \text{Accuracy} &= \frac{a + d}{a + b + c + d} \\
 &= \frac{53 + 28}{53 + 6 + 3 + 28} = \frac{81}{90} \\
 &= 0,9 \\
 &= 90,00 \%
 \end{aligned}$$

recall: 82,35% (positive class: Meglio)			
	true Arita	true Meglio	class precision
pred. Arita	53	6	89,82%
pred. Meglio	3	28	90,32%
class recall	94,64%	82,35%	

Dari **Gambar 9** maka dapat dilihat nilai *accuracy* yang didapat dari pengujian *Decision Tree* + Algoritma Genetika yaitu sebesar 90,00%.

Dari **Gambar 13** maka dapat dilihat nilai *Precision* yang didapat dari pengujian *Decision Tree* + Algoritma Genetika yaitu sebesar 90,32%.

Dari **Gambar 14** maka dapat dilihat nilai *Recall* yang didapat dari pengujian *Decision Tree* + Algoritma Genetika yaitu sebesar 82,35%.

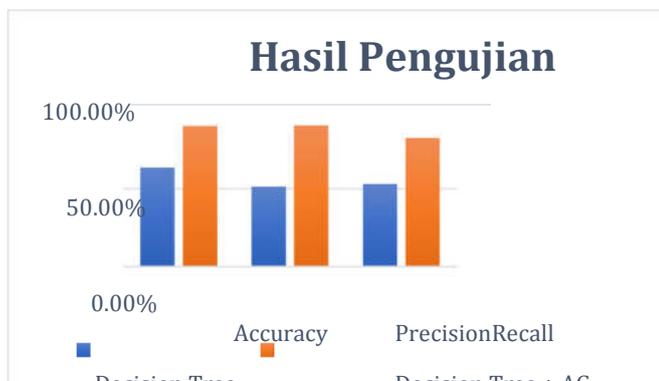


Gambar 14. Hasil AUC *Decision Tree* + Algoritma genetika

Dari **Gambar 14** maka dapat dilihat nilai AUC yang didapat dari pengujian *Decision Tree* + Algoritma Genetika yaitu sebesar 0,911.

4.3 Pembahasan Hasil Pengujian

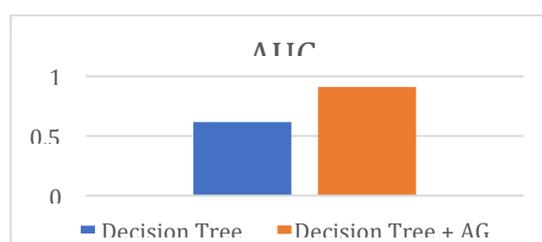
Dari beberapa pengujian yang telah dilakukan diatas maka didapatkan hasil sebagai berikut



Gambar 15. Grafik Hasil Pengujian

Dari **Gambar 15** dapat dilihat nilai *accuracy* yang diperoleh dari pengujian *Decision Tree* sebesar 63.33%, sedangkan pada pengujian *Decision Tree* + Algoritma Genetika sebesar 90.00%, yang artinya nilai *accuracy* meningkat 26.67%. Pada nilai *precision*, hasil yang diperoleh dari pengujian *Decision Tree* sebesar 51.28%, sedangkan pada pengujian *Decision Tree* + Algoritma Genetika sebesar 90.32%, nilainya meningkat sebesar 39.04%. Selanjutnya nilai *recall*, pada pengujian *Decision Tree* diperoleh hasil sebesar 58.82%, sedangkan pengujian *Decision Tree* + Algoritma Genetika menghasilkan nilai *recall* sebesar 82.35%, yang artinya nilai *recall* meningkat 23.07%

Gambar 15 Grafik Hasil AUC



Dari **Gambar 15** dapat dilihat bahwa nilai AUC yang diperoleh dari pengujian *Decision Tree* adalah sebesar 0,616 yang termasuk dalam kategori *Poor Classification*. Sedangkan pada pengujian *Decision Tree + Algoritma Genetika* nilai AUC yang diperoleh sebesar 0,911. Dari hasil pengujian Nilai AUC meningkat sebesar 0,295 dan termasuk dalam kategori *Excellent Classification*.

5. Kesimpulan

Berdasarkan dari hasil penelitian yang telah dilakukan maka dapat disimpulkan bahwa teknik optimasi menggunakan Algoritma Genetika berhasil meningkatkan akurasi dari algoritma C4.5 dalam memprediksi merek meteran air pam. Hasil akurasi dari algoritma C4.5 sebesar 63,33%, dan setelah dioptimasi menggunakan algoritma genetika hasil akurasinya menjadi sebesar 90,00% atau naik sebesar 26,67% dari algoritma C4.5.

Daftar Pustaka

- D. - STMIK Nusa Mandiri Jakarta, "Optimasi Algoritma Naïve Bayes dengan Menggunakan Algoritma Genetika untuk Prediksi Kesuburan (Fertility)," *Evolusi*, 2016.
- N. M. E. Putri dan Ermatita, "Review : Data Mining menggunakan Algoritma Genetika," *Pros. Annu. Res. Semin. 2017*, 2017.
- S. Lorena, W. Zarman, dan I. Hamidah, "Analisis Dan Penerapan Algoritma C4.5 Dalam Data Mining Untuk Memprediksi Masa Studi Mahasiswa Berdasarkan Data Nilai Akademik," *Pros. Semin. Nas. Apl. Sains dan Teknol.*, 2014.
- A. S. Sunge, "Optimasi Algoritma C4.5 Dalam Prediksi Web Phishing Menggunakan Seleksi Fitur Genetic Algoritma," *Paradigma*, 2018.
- I. Rahmayuni, "Perbandingan performansi algoritma c4.5 dan cart dalam klasifikasi data nilai mahasiswa prodi teknik komputer politeknik negeri padang," *Teknoif*, 2014.
- S. Singh dan P. Gupta, "Comparative Study Id3, Cart and C4.5 Decision Tree Algorithm: a Survey," *Int. J. Adv. Inf. Sci. Technol.*, 2014.
- I. H. Witten, E. Frank, M. A. Hall, dan C. J. Pal, *Data Mining: Practical Machine Learning Tools and Techniques*. 2016.
- J. Han, M. Kamber, dan J. Pei, "Chapter 9: Graph Mining, Social Network Analysis, and Multirelational Data Mining," *Data Min. concepts Tech.*, 2006.