

**PENERAPAN *DATA MINING* UNTUK KLASIFIKASI KUALITAS PIPA PVC MENGGUNAKAN
METODE *ALGORITMA C4.5* , STUDI KASUS
PT CIPTA ANEKA AGUNG**

Asep Muhidin, Ading Bagus Saputra

Program Studi Teknik Informatika STT Pelita Bangsa
asep.muhidin@pelitabangsa.ac.id

Disetujui, 27 Juni 2019

ABSTRAKSI

Penelitian ini ditujukan untuk mengetahui hasil prediksi uji kualitas pipa PVC dengan melihat akurasi *Algoritma C4.5* sehingga memudahkan QC dalam menentukan kualitas pipa PVC. Penelitian dilakukan pada data uji kualitas pipa PVC, dari data tersebut dilakukan pembagian *data training* dan *data testing*. *Data mining* mengekstraksi data untuk menemukan informasi dan *pattern* dalam menentukan kualitas pipa PVC. Metode *classification* dilakukan pada *data training* hingga menemukan suatu *rule* (proses *learning*) yang dapat diterapkan untuk mengklasifikasikan kategori uji kualitas pipa PVC pada data baru. Proses *learning* menggunakan teknik *decision tree C.45* mengubah fakta menjadi pohon keputusan yang merepresentasikan aturan yang lebih mudah dipahami. Variabel yang digunakan dalam metode ini yaitu: Uji Linyak, Uji Kuat Tarik, Uji Kelenturan, Uji Tekanan Hidrostatik, dan Uji Kejatuhan. Berdasarkan Hasil klasifikasi menggunakan *algoritma C4.5* menunjukkan bahwa diperoleh akurasi mencapai 92.92%, yang menunjukkan bahwa *algoritma C4.5* cocok digunakan untuk mengukur uji kualitas pipa PVC.

Kata kunci: Kualitas, *Algoritma C4.5*, *Data mining*, Pohon keputusan

ABSTRACT

The research was aimed to determine the results of the prediction of PVC pipe quality test by looking at the accuracy of the C4.5 Algorithm so as to facilitate QC in determining the quality of PVC pipes. The research was carried out on the PVC pipe quality test data, from the data was carried out the distribution of training data and testing data. Data mining extracts data to find information and patterns in determining the quality of PVC pipes. Classification method is carried out on training data to find a rule that can be applied to classify PVC pipe quality test categories in new data. The learning process using the C.45 decision tree technique transforms facts into decision trees that represent rules that are easier to understand. Variables used in this method are: Flattening Test, Tensile Strength Test, Elongation Test, Hydrostatic Pressure Test, and Impact Test. Based on the classification results using C4.5 algorithm shows that the accuracy reached 92.92%, which shows that the C4.5 algorithm is suitable for measuring the quality of PVC pipe tests.

Keyword: Quality, C4.5 Algorithm, Data mining, Decision Tree

1. PENDAHULUAN

Pipa PVC (*PolyVinyl Chloride*) merupakan polimer termoplastik urutan ketiga dalam hal jumlah pemakaian di dunia, setelah PE (*polietilena*) dan PP (*polipropilena*). Di seluruh dunia, lebih dari 50% PVC yang di produksi di pakai dalam kontruksi. Pipa PVC pada umumnya digunakan sebagai saluran air dalam suatu proyek perumahan, gedung atau jalan. Pipa PVC memiliki sifat keras, ringan, kuat, tahan terhadap hampir semua alkalin atau zat beracun serta mudah dipasang [3] (Hadi, Takwin, & Dani, 2016). Pipa PVC yang baik dan memenuhi "strandart aplikasinya" harus lulus beberapa syarat pengujian atau tes laboratorium. Sehingga Departemen QC (*Quality Control*) dapat mempertimbangkan dan memberikan keputusan urutan prioritas kelayakannya. Namun pada pelaksanaan kriteria-kriteria tersebut tidak mempunyai bobot yang baku sehingga terjadi penyamaan diantara kriteria tersebut. Padahal suatu kriteria dapat menjadi lebih penting dari kriteria yang lain dengan berbagai pertimbangan.

Salah satu metode yang dapat digunakan untuk hal ini adalah dengan metode *data mining algoritma C4.5*. Kelebihan menggunakan model klasifikasi *decision tree* ini adalah hasil pohon keputusannya sederhana dan mudah di mengerti. Proses learning dan klasifikasi sederhana dan cepat. Secara umum model klasifikasi *algoritma decision tree* mempunyai tingkat akurasi tinggi. Oleh karena itu dapat disimpulkan *algoritma* ini cocok untuk diimplementasikan pada penelitian yang melibatkan proses pengujian kualitas pipa PVC.

2. LANDASAN TEORI

2.1 Data Mining

Data mining adalah serangkaian proses untuk menggali nilai tambah berupa informasi yang selama ini tidak diketahui secara manual dari suatu basis data. Informasi yang dihasilkan diperoleh dengan cara mengekstraksi dan mengenali pola yang penting atau menarik dari data yang terdapat pada basis data. *Data mining* terutama digunakan untuk mencari pengetahuan yang terdapat dalam basis data yang besar sehingga sering disebut *knowledge discovery database (KDD)* [9] (Vulandari, 2017:1). Menurut Clifton (Vulandari, 2017:1), *data mining* merupakan gabungan sejumlah disiplin ilmu komputer, yang didefinisikan sebagai proses penemuan pola baru dari kumpulan-kumpulan data sangat besar, meliputi metode-metode yang merupakan irisan dari *artificial intelligence, machine learning, statistic*, dan *database system*. Berdasarkan definisi diatas maka dapat penulis simpulkan bahwa *Data Mining* merupakan suatu proses menggali informasi pada sekumpulan besar data dengan tujuan menemukan suatu pengenalan pola tertentu.

2.2 Algoritma C4.5

Algoritma adalah urutan logis langkah-langkah penyelesaian masalah yang disusun secara sistematis.

Algoritma C4.5 adalah suatu deretan *algoritma* untuk permasalahan klasifikasi di dalam sebuah mesin dan himpunan data. *Algoritma C4.5* diperkenalkan oleh Quinlan (1996) pada *Datamining* (Candraningsih & Nurhadiyono, 2014) sebagai versi pembenanahan dari ID3, [2]. Dalam ID3, induksi *decision tree* hanya dapat digunakan pada fitur kategorikal (nimonal), sedangkan tipe numerik (ratio) tidak bisa digunakan. Perubahan yang membedakan *algoritma C4.5* dan ID3 adalah dapat menangani fitur dengan numerik, melakukan pemotongan (*pruning*) *decision tree*, dan penurunan (*deriving*) *rule set*. *Algoritma C4.5* juga menggunakan kriteria *gain* dan menentukan fitur yang menjadi pemecah *node* pada pohon yang diinduksi. Secara umum algoritma C4.5 untuk membangun pohon keputusan adalah sebagai berikut (Vulandari, 2017:26) :

1. Pilih atribut sebagai akar.
2. Buat cabang untuk masing-masing nilai.
3. Bagi kasus dalam cabang.
4. Ulangi proses untuk masing-masing cabang sampai semua kasus pada cabang memiliki klas yang sama.

Untuk memilih atribut sebagai akar, didasarkan pada nilai gain tertinggi dari atribut-atribut yang ada.

—

Sehingga akan diperoleh nilai *gain* dari *atribut* yang paling tinggi. *Gain* adalah salah satu *attribute selection measure* yang digunakan untuk memilih *test attribute* tiap *node* pada *tree*. Atribut dengan *information gain* tertinggi dipilih sebagai *test attribute* dari suatu *node*. Sementara itu, perhitungan *entropy* dapat dilihat pada persamaan berikut di bawah ini :

Rumus untuk menghitung *Accuracy*, *Precision*, dan *Recall*

1. *Accuracy* adalah Tingkat kedekatan antara nilai prediksi dengan nilai aktual.
2. *Precision* adalah Tingkat ketepatan antara informasi yang diminta oleh pengguna dengan jawaban yang diberikan oleh system.
3. *Recall* adalah Tingkat keberhasilan system dalam menentukan kembali sebuah informasi.

Rumus *confusion matrix* :

Tabel 1. Perhitungan *Akurasi*, *Presisi*, dan *Recall*

		Nilai Sebenarnya	
		<i>TRUE</i>	<i>FALSE</i>
Nilai Prediksi	<i>TRUE</i>	TP (True Positive)	FP (False Positive)
	<i>FALSE</i>	FN (False Negative)	TN (True Negative)

2.3 Klasifikasi

Menurut Han dan Kamber dalam bukunya Vulandari, 2017:13, [9] *Classification* adalah sebuah model dalam dimana, classifier dikonstruksi untuk memprediksi categorical label, seperti “aman” atau “beresiko” untuk data aplikasi peminjaman uang;”ya” atau “tidak” untuk data `marketing; atau “treatment A”,”treatment B”, atau “treatment C” untuk data medis. Kategori tersebut dapat direpresentasikan dengan nilai yang sesuai dengan kebutuhannya, dimana pengaturan dari nilai tersebut tidak memiliki aturan tersebut.. Proses data klasifikasi memiliki dua tahapan, yang pertama adalah *Learning*: yaitu training data dianalisa dengan menggunakan sebuah algoritma klasifikasi. Dan yang kedua adalah *Classification*: yaitu pada tahap ini test data digunakan untuk mengestimasi ketepatan dari *classification rules*. Jika keakuratan yang dikondisikan dan yang diperkirakan dapat diterima, *rule* tersebut dapat diaplikasikan pada klasifikasi lainnya dari tipe data yang baru (Vulandari, 2017:14) [9].

2.4. Pohon Keputusan (Decision Tree)

Bahwa “Pohon dalam analisis pemecahan masalah pengambilan keputusan adalah pemetaan mengenai alternatif- alternatif pemecahan masalah yang dapat diambil dari masalah tersebut. Pohon tersebut juga memperlihatkan faktor-faktor kemungkinan/ probabilitas yang akan mempengaruhi alternatif-alternatif keputusan tersebut, disertai dengan estimasi hasil akhir yang akan didapat bila kita mengambil alternatif keputusan tersebut” [9].

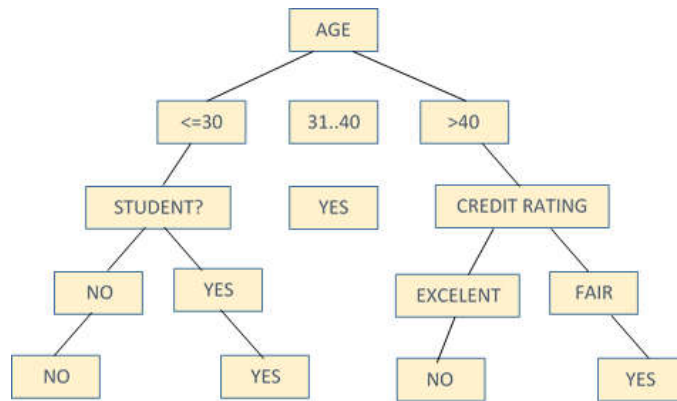
Manfaat utama dari penggunaan pohon keputusan adalah kemampuannya untuk mem-*breakdown* proses pengambilan keputusan yang kompleks menjadi lebih simpel sehingga pengambil keputusan akan lebih menginterpretasikan solusi dari permasalahan.

Pada pohon keputusan terdapat 3 jenis *node*, yaitu :

- a) *Root Node*, merupakan *node* paling atas, pada *node* ini tidak ada *input* dan bisa tidak mempunyai *output* atau mempunyai *output* lebih dari satu.
- b) *Internal Node*, merupakan *node* percabangan, pada *node* ini hanya terdapat satu *input* dan mempunyai *output* minimal dua.
- c) *Leaf node* atau terminal *node*, merupakan *node* akhir, pada *node* ini hanya terdapat satu *input* dan tidak mempunyai *output* (Andriani, 2013).

Pohon keputusan adalah model prediksi menggunakan struktur pohon atau struktur berhirarki.

Contoh dari pohon keputusan dapat dilihat pada gambar berikut:



Gambar 1. Model Pohon Keputusan

Sumber : Pramudiono dalam buku Vulandari, 2017 [10]

Disini setiap percabangan menyatakan kondisi yang harus dipenuhi dan tiap ujung pohon menyatakan kelas data. Pada pohon keputusan setiap simpul daun menandai label kelas. Proses dalam pohon keputusan yaitu mengubah bentuk data (tabel) menjadi model pohon (*tree*) kemudian mengubah model pohon tersebut menjadi aturan (*rule*).

3. METODE PENELITIAN

Pada penelitian ini, data yang digunakan adalah data hasil uji kualitas pipa PT Cipta Aneka Agung. Data tersebut akan diolah menggunakan beberapa metode *Algoritma C4.5*. Tahapan ini menerangkan tentang darimana sumber data dalam penelitian ini didapatkan dan menemukan informasi yang bisa digunakan untuk penelitian. Data yang digunakan dalam penelitian ini adalah data hasil uji laboratorium kualitas pipa PVC ukuran SNI S-10 Ø110 pada departemen QC pada PT Cipta Aneka Agung tahun 2017.

Tabel 2. Data Uji Kualitas

Kode Produksi	Nama Operator	Nama QC	Tebal Dinding	Diameter Luar	Uji Linyak	Uji Kuat Tarik	Uji Kelenturan	Uji Tekanan Hidrostatik	Uji Kejutahan	Hasil
518A051	Dwi	Rinin	5.4	110.1	Tidak Retak	462	110	Tidak Pecah	5	Ng
518A052	Fachrudin	Gunawan	5.3	109.8	Tidak Retak	460	98	Tidak Pecah	7	Ng
518A061	Okem	Ari	5.3	109.8	Retak	455	99	Tidak Pecah	1	Ok
518A062	Alex	Yudi	5.4	110	Tidak Retak	456	86	Tidak Pecah	1	Ok
518A063	Saidin	Rinin	5.4	110	Retak	442	115	Tidak Pecah	4	Ng
518A102	Elmo	Rinin	5.4	110.2	Tidak Retak	452	105	Tidak Pecah	0	Ok
518A103	Sulaemi	Yudi	5.3	109.9	Tidak Retak	480	87	Tidak Pecah	0	Ok
518A111	Burhan	Gunawan	5.6	110.5	Tidak Retak	493	76	Tidak Pecah	1	Ok
518B011	Sulaemi	Yudi	5.4	110.3	Tidak Retak	477	79	Tidak Pecah	2	Ok
518B011	Sulaemi	Yudi	5.4	110.3	Tidak Retak	473	70	Tidak Pecah	4	Ng
418B011	Alex	Yudi	5.5	110.3	Retak	494	71	Tidak Pecah	0	Ng
418B012	Fachrudin	Gunawan	5.3	109.8	Tidak Retak	439	110	Tidak Pecah	2	Ok
418B013	Dwi	Rinin	5.3	109.8	Retak	471	75	Tidak Pecah	7	Ng
418B021	Okem	Ari	5.4	110.2	Retak	491	162	Tidak Pecah	1	Ok
418B021	Okem	Ari	5.3	110	Retak	485	166	Tidak Pecah	2	Ok
418B022	Dwi	Rinin	5.3	110	Retak	461	164	Tidak Pecah	6	Ng
418B062	Fachrudin	Gunawan	5.4	110.1	Retak	448	164	Tidak Pecah	7	Ng
418B063	Deki	Ari	5.3	110.1	Tidak Retak	445	170	Tidak Pecah	0	Ok
418B071	Elmo	Rinin	5.4	110.2	Tidak Retak	483	79	Tidak Pecah	1	Ok

418B072	Nardi	Rinin	5.4	110.2	Tidak Retak	459	79	Tidak Pecah	1	Ok
518B111	Deki	Ari	5.4	110.2	Tidak Retak	443	70	Tidak Pecah	5	Ng
518B112	Sarjan	Gunawan	5.3	109.9	Retak	454	79	Tidak Pecah	4	Ng
518B121	Alex	Yudi	5.3	110	Tidak Retak	493	77	Tidak Pecah	0	Ok
518B131	Muchlis	Gunawan	5.5	110.4	Tidak Retak	454	67	Tidak Pecah	5	Ng
518B141	Soleh	Ari	5.3	110	Tidak Retak	446	65	Tidak Pecah	6	Ng
518B151	Alex	Yudi	5.3	109.9	Tidak Retak	443	66	Tidak Pecah	0	Ng
518B161	Didit	Rinin	5.4	110.2	Retak	440	70	Tidak Pecah	2	Ng
518B171	Didit	Rinin	5.4	110.2	Retak	442	70	Tidak Pecah	6	Ng
518B211	Sulaemi	Yudi	5.4	110.1	Retak	461	76	Tidak Pecah	6	Ng
518B221	Deki	Ari	5.5	110.4	Retak	475	77	Tidak Pecah	4	Ng

3.2 Variabel Data

Variabel yang terdapat pada data penelitian ini meliputi variabel *input* dan variabel *output* atau target :

1. Variabel *input*

Variabel *input* dinyatakan sebagai data sampel pipa serta kriteria-kriteria dalam pengujian kualitas pipa. Berikut ini adalah variabel *input* yang terdapat pada data uji kualitas yang akan digunakan untuk mengklasifikasi yaitu Kode Produksi, Nama Operator, Nama QC, Diameter Luar (*Outside Diameter*), Tebal Dinding (*Wall Thickness*), Uji Linyak (*Flattening Test*), Uji Kuat Tarik (*Tensile Strength Test*), Uji Kelenturan (*Elongation Test*), Uji Tekanan Hidrostatik (*Hidrostatic Pressure test*), dan Uji Kejutuhan (*Impact Test*).

2. Variabel Target

Variabel target dinyatakan sebagai hasil dari penelitian yang dilakukan terhadap variabel *input*. Sedangkan variabel target dari penelitian ini adalah atribut "Hasil" yang berisikan label "Ok" dan "NG".

3.3 Pengelolaan Data Awal

Pengelolaan awal pada penelitian ini akan mencakup semua kegiatan untuk mempersiapkan data sebelum masuk proses pemodelan. Dalam melakukan pengolahan data awal, akan dilakukan beberapa tahapan agar pada akhirnya akan didapatkan data yang bisa digunakan pada tahap berikutnya. Tahapan tersebut antara lain: *select data*, *cleaning*, *transformation* serta akan dilakukan *split validation*.

3.4 Pemilihan Data (Data Select)

Data yang sudah ada selanjutnya dilakukan pemilihan terhadap parameter yang akan dianalisis. Parameter yang diambil adalah atribut dari data uji kualitas yang telah didapatkan sebelumnya. Pemilihan parameter ini dilakukan dengan mempelajari lebih lanjut setiap atribut dengan mempertimbangkan tujuan penulisan, maka akan didapatkan atribut-atribut yang akan digunakan untuk menjadi masukan atau variabel *input*.

Tabel 3. Pemilihan Data

No	Atribut	Indikator	Datail Penggunaan
1	Kode Produksi	x	-
2	Nama Operator	x	-
3	Nama QC	x	-
4	Diameter Luar	√	Nilai Model

5	Tebal Dinding	√	Nilai Model
6	Uji Linyak	√	Nilai Model
7	Uji Kuat Tarik	√	Nilai Model
8	Uji Kelenturan	√	Nilai Model
9	Uji Tekanan Hidrostatik	√	Nilai Model
10	Uji Kejatuhan	√	Nilai Model
11	Hasil	√	Label Target

Tabel diatas menerangkan atribut yang akan dipakai dalam penelitian ini. Indikator “√” menandakan atribut akan digunakan, sedangkan indicator “x” menandakan atribut tersebut akan di eliminasi pada tahap pengolahan data awal.

3.5 Transformasi

Setelah data sudah dipilih maka akan dilakukan tahapan untuk melakukan transformasi terhadap atribut, transformasi akan dilakukan untuk memodifikasi sumber data ke format berbeda yang dapat diterima oleh proses data *mining* pada tahap selanjutnya. Transformasi nilai-nilai dari atribut juga perlu dilakukan karena jangkauan nilai terlalu lebar, sehingga dapat mengakibatkan proses pengenalan pola data dan pembentukan pohon keputusan menjadi lama.

Jenis atribut yang ada pada data awal penelitian ini berupa atribut nominal (Kode Produksi, Nama Operator, Nama QC), ordinal (Diameter Luar, Tebal Dinding, Uji Kuat Tarik, Uji Kelenturan, Uji Kejatuhan), biner (Uji Linyak, Uji Kejatuhan). Selanjutnya atribut akan ditransformasi ke dalam bentuk atribut numerik agar memudahkan proses pemodelan. Berikut adalah hasil dari transformasi atribut :

Tabel 4. Hasil Transformasi

NO	ATRIBUT	NILAI
1	Uji Linyak	Tidak Retak
		Retak
2	Uji Kuat Tarik	< 450
		450-470
		471-490
		> 490
		> 490
3	Uji Kelenturan	< 80
		80-120
		121-160
		> 160
4	Uji Tekanan Hidrostatik	Tidak Pecah
		Pecah
5	Uji Kejatuhan	0
		1
		2
		3

		4
		5
		6
		7

3.6 Split Validation

Split Validation merupakan teknik validasi yang membagi data menjadi dua bagian secara acak, sebagian data *training* dan sebagian data *testing*. Data yang sudah disiapkan untuk klasifikasi Data uji kualitas dibagi menjadi 2 bagian. Dari seluruh populasi, data dipisah menjadi 80% *data training* dan 20% *data testing*. Dalam hal pembagian data ini digunakan sebuah teknik yang bernama *sampling random sistematis (Systematic Random Sampling)*. Tata cara menggunakan teknik *sampling random sistematis* ini dengan cara perandoman atau pengacakan yang dilakukan hanya satu kali, yaitu saat menentukan sebuah unsur pertama dari *sampling* yang akan diambil. Penentuan unsur *sampling* selanjutnya ditempuh dengan cara memanfaatkan interval sampel. Interval sampel adalah angka yang menunjukkan jarak antara nomor-nomor urut yang terdapat dalam kerangka *sampling* yang akan dijadikan patokan dalam menentukan atau memilih unsur-unsur *sampling* kedua dan seterusnya hingga unsur ke-n. Interval sampel biasanya dilambangkan dengan huruf k.

Interval sampel atau juga disebut *sampling rasio* diperoleh dengan cara membagi ukuran populasi dengan ukuran sampel yang dikehendaki (N/n). Hasil perhitungan untuk mengambil *data testing* adalah sebagai berikut :

$$\text{Jumlah popilasi (N)} = 300$$

$$\text{Jumlah data testing} = 20\% \times 300 = 60$$

$$\text{Jumlah sampel (n)} = 60$$

$$\text{Interval Sampling (k)} = N/n = 300 / 60 = 5$$

Unsur pertama yang diambil untuk data testing (s) = 1 Unsur kedua

$$= s + k$$

Unsur ketiga

$$= s + 2k$$

Unsur keempat

$$= s + 3k, \text{ dan seterusnya hingga unsur ke-n.}$$

Dari hasil diatas diperoleh *data training* sebanyak 240 data uji kualitas (lampiran 1), maka sisanya *data testing* sebanyak 60 dijadikan data uji kualitas (lampiran 2).

Tabel 5. Cuplikan *Data Training* (Lampiran 1)

No	Uji Linyak	Uji Kuat Tarik	Uji Kelenturan	Uji Tekanan Hidrostatik	Uji Kejatuhan	Hasil
1	Tidak Retak	450-470	80-120	Tidak Pecah	7	Ng
2	Retak	450-470	80-120	Tidak Pecah	1	Ok
3	Tidak Retak	450-470	80-120	Tidak Pecah	1	Ok
4	Retak	< 450	80-120	Tidak Pecah	4	Ng
5	Tidak Retak	471-490	80-120	Tidak Pecah	0	Ok
6	Tidak Retak	> 490	< 80	Tidak Pecah	1	Ok
7	Tidak Retak	471-490	< 80	Tidak Pecah	2	Ok
8	Tidak Retak	471-490	< 80	Tidak Pecah	4	Ng
9	Tidak Retak	< 450	80-120	Tidak Pecah	2	Ok
10	Retak	471-490	< 80	Tidak Pecah	7	Ng
11	Retak	> 490	> 160	Tidak Pecah	1	Ok
12	Retak	471-490	> 160	Tidak Pecah	2	Ok
13	Retak	< 450	> 160	Tidak Pecah	7	Ng
14	Tidak Retak	< 450	> 160	Tidak Pecah	0	Ok
15	Tidak Retak	471-490	< 80	Tidak Pecah	1	Ok
16	Tidak Retak	450-470	< 80	Tidak Pecah	1	Ok

17	Retak	450-470	< 80	Tidak Pecah	4	Ng
18	Tidak Retak	> 490	< 80	Tidak Pecah	0	Ok
19	Tidak Retak	450-470	< 80	Tidak Pecah	5	Ng
.....
240	Tidak Retak	> 490	121-160	Tidak Pecah	0	Ok

Tabel 6. Cuplikan *Data Testing* (Lampiran 2)

No	Uji Linyak	Uji Kuat Tarik	Uji Kelenturan	Uji Tekanan Hidrostatik	Uji Kejatuhan	Hasil
1	Tidak Retak	450-470	80-120	Tidak Pecah	5	Ng
2	Tidak Retak	450-470	80-120	Tidak Pecah	0	Ok
3	Retak	> 490	< 80	Tidak Pecah	0	Ng
4	Retak	450-470	> 160	Tidak Pecah	6	Ng
5	Tidak Retak	< 450	< 80	Tidak Pecah	5	Ng
6	Tidak Retak	< 450	< 80	Tidak Pecah	0	Ng
7	Retak	> 490	< 80	Tidak Pecah	2	Ng
8	Retak	> 490	80-120	Tidak Pecah	2	Ok
9	Tidak Retak	< 450	> 160	Tidak Pecah	2	Ok
10	Tidak Retak	471-490	< 80	Tidak Pecah	0	Ok
11	Tidak Retak	450-470	> 160	Tidak Pecah	0	Ok
12	Retak	> 490	80-120	Pecah	4	Ng
13	Tidak Retak	< 450	> 160	Pecah	0	Ng
14	Retak	> 490	> 160	Pecah	6	Ng
15	Tidak Retak	> 490	121-160	Pecah	5	Ng
16	Retak	471-490	80-120	Pecah	1	Ng
17	Retak	> 490	121-160	Pecah	1	Ng
18	Tidak Retak	> 490	> 160	Tidak Pecah	0	Ok
19	Retak	450-470	80-120	Pecah	3	Ng
.....
60	Retak	450-470	121-160	Tidak Pecah	1	Ok

4 HASIL DAN PEMBAHASAN

4.1 Pemodelan Menggunakan *Algoritma C4.5*

Dalam penelitian klasifikasi uji kualitas menggunakan *Algoritma C4.5*, pohon keputusan dibuat berdasarkan hasil perhitungan *Entropy* dan *Gain*, setelah pohon keputusan terbentuk, langkah selanjutnya mencari *Rule* berdasarkan cabang pohon keputusan. Berikut akan dibahas langkah-langkah dalam klasifikasi uji kualitas menggunakan *Algoritma C4.5* secara manual dan analisa memakai *tools RapidMiner*.

Berdasarkan Tabel di atas jumlah kasus untuk keputusan NG berjumlah 160 dan keputusan OK berjumlah 80, dan jumlah TOTAL kasus adalah 240. Setelah diketahui semua kasus yang perlu dihitung, langkah selanjutnya menghitung *Entropy* dan *Gain* dari semua nilai kasus yang dibagi berdasarkan atribut.

1) *Node 1 (Root)* Atribut Total

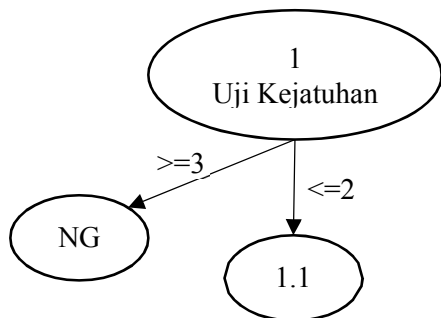
Menghitung jumlah kasus, jumlah kasus untuk keputusan NG, jumlah kasus untuk keputusan OK, dan *entropy* dari semua kasus. Setelah itu, lakukan perhitungan *gain* untuk setiap atribut.

Tabel 7. Perhitungan Node-1

NO	ATTRIBUTE	VALUE	JUMLAH KASUS (S)	NG (S1)	OK (S2)	ENTROPY	GAIN
1	TOTAL		240	160	80	0.918295834	
2	Uji Linyak	Tidak Retak	117	54	63	0.995727452	0.135866836
		Retak	123	106	17	0.579535347	
3	Uji Kuat Tarik	< 450	53	40	13	0.803715924	0.020170217
		450-470	66	47	19	0.86596535	
		471-490	70	46	24	0.927526588	
		> 490	51	27	24	0.997502546	
4	Uji Kelenturan	< 80	51	42	9	0.672294817	0.023974195
		80-120	63	38	25	0.969063253	
		121-160	63	40	23	0.946818832	
		> 160	63	40	23	0.946818832	
5	Uji Tekanan Hidrostatik	Tidak Pecah	128	74	54	0.982316608	0.029587186
		Pecah	112	86	26	0.781728122	
6	Uji Kejatuhan	0	40	17	23	0.983708263	0.364794637
		1	46	18	28	0.965636133	
		2	50	21	29	0.981453895	
		3	26	26	0	0	
		4	17	17	0	0	
		5	16	16	0	0	
		6	23	23	0	0	
		7	22	22	0	0	

Dari Tabel 4.4 memperlihatkan bahwa pada baris total diketahui jumlah kasus (S) adalah 240, jumlah keputusan NG (S1) adalah 160, dan jumlah keputusan OK (S2) adalah 80.

Untuk perhitungan pada atribut-atribut berikutnya sama seperti perhitungan pada atribut Uji linyak. Sehingga diperoleh hasil perhitungan yang ditunjukkan pada Tabel 4.4. Dari tabel tersebut dapat diketahui bahwa atribut dengan *gain* tertinggi adalah Uji Kejatuhan, yaitu sebesar 0.364794637 dengan demikian uji kejatuhan bisa menjadi *node* akar. Dari nilai ≤ 2 tersebut masih memerlukan perhitungan lebih lanjut untuk mencari *node* cabang. Hasil perhitungan tersebut dapat digambarkan pohon keputusan seperti Gambar 4.1.



Gambar 2. Pohon Keputusan Node 1

Demikian seterusnya untuk Node 1.1 Uji Kejatuhan Nilai ≤ 2 , Node 1.1.1 Uji Linyak Nilai Retak sampai Node 1.1.2.2. 1 Uji Uji Kelenturan Nilai < 450, lihat dihalaman Lampiran.

7) Node 1.1.2.2.1 Uji Kelenturan Nilai <450

Menghitung jumlah kasus, jumlah kasus untuk keputusan NG, jumlah kasus untuk keputusan OK, dan *entropy* dari semua kasus. Setelah itu, lakukan perhitungan *gain* untuk setiap atribut.

Tabel 8. Pehitungan Node 1.1.2.2.1

NO	ATRIBUTE	VALUE	JUMLAH KASUS (S)	NG (S1)	OK (S2)	ENTROPY	GAIN
1	TOTAL		6	2	4	0.918295834	
2	Uji Tekanan Hidrostatik	Tidak Pecah	3	0	3	0	0.459147917
		Pecah	3	2	1	0.918295834	

Dari Tabel 4.11 dapat diketahui bahwa atribut dengan *gain* tertinggi adalah uji Tekanan Hidrostatik, yaitu sebesar 0,459147917.

4.4 Hasil Proses Validasi

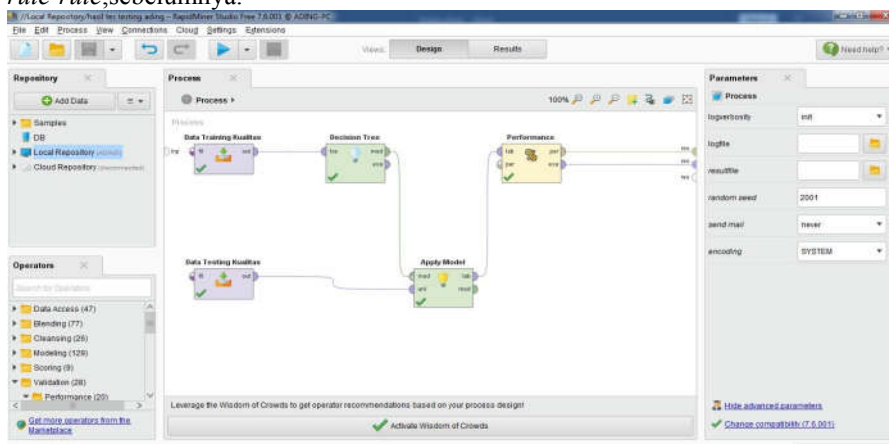
accuracy: 92.92% +/- 4.58% (mikro: 92.92%)

	true Ng	true Ok	class precision
pred. Ng	148	5	96.73%
pred. Ok	12	75	86.21%
class recall	92.50%	93.75%	

Gambar 3. Hasil pengukuran *Data Accuracy Training*

Hasil pengukuran *data accuracy* yang diperoleh dari *data training* mencapai 92.92% dengan +/-4.58% (mikro: 92.92%). Dari data tersebut diketahui prediksi NG dengan *true NG* mencapai 148 pipa dan *true OK* sebanyak 12 pipa, dengan pencapaian *class precision* 96.73%. Sedangkan untuk prediksi OK dengan *true NG* sebanyak 5 pipa dan untuk *true OK* mencapai 75 pipa dengan pencapaian *class precision* 86.21%. Untuk *class recall* dengan *true NG* mencapai 92.50% sedangkan untuk *class recall* dengan *true OK* mencapai 93.75%.

Proses klasifikasi dengan RapidMiner menggunakan metode *algoritma decision tree C4.5* pada data uji kualitas pipa PVC ini untuk membandingkan *data testing* dengan *data training* yang sudah diketahui *rule-rule*, sebelumnya.



Gambar 4. Proses *Training* dan *Testing*

Dengan mengetahui jumlah data yang diklasifikasikan secara benar maka dapat diketahui akurasi hasil prediksi yaitu 91.67% dari hasil *data testing*.

	true Ng	true Ok	class precision
pred. Ng	30	1	96.77%
pred. Ok	4	25	86.21%
class recall	88.24%	96.15%	

Gambar 5. Hasil Akurasi *data Testing*

Hasil pengukuran *data accuracy* yang diperoleh dari *data testing* mencapai 91.67%. Dari data tersebut diketahui prediksi NG dengan *true NG* mencapai 30 pipa dan *true OK* sebanyak 1 pipa, dengan pencapaian *class precision* 96.77%. Sedangkan untuk prediksi OK dengan *true NG* sebanyak 4 pipa dan untuk *true OK* mencapai 25 pipa dengan pencapaian *class precision* 86.21%. Untuk *class recall* dengan *true NG* mencapai 88.24% sedangkan untuk *class recall* dengan *true OK* mencapai 96.15%.

4. Kesimpulan

- 1) Dari hasil perhitungan *data training* data uji kualitas pipa PVC dengan algoritma C4.5 menggunakan *data training* dengan *confusion matrix* memiliki nilai yaitu *accuracy* 92.92%, *precision* 87.64%, dan *recall* 93.75% serta kurva ROC *optimistic* dengan akurasi *excellent classification* sebesar 0.984. Hal ini menunjukkan bahwa hasil prediksi ini dapat digunakan untuk data uji kualitas yang baru.
- 2) Dari analisis *data training* di peroleh sebuah pohon keputusan yang memiliki 15 *rule model* yang dapat digunakan sebagai acuan dalam pengambilan keputusan kualitas pipa PVC.

DAFTAR PUSTAKA

- [1] Andriani, A. (2013). Sistem Pendukung Keputusan Berbasis Decision Tree Dalam Pemberian Beasiswa Studi Kasus : Amik “ Bsi Yogyakarta .” *Seminar Nasional Teknologi Informasi Dan Komunikasi 2013 (SENTIKA 2013)*, 2013(Sentika), 163–168.
- [2] Candraningsih, & Nurhadiyono, B. (2014). Algoritma C4 . 5 Untuk Klasifikasi Calon Peserta Lomba. *Universitas Dian Nuswantoro, Ilmu Komputer, Teknik Informatika*, (024), 1–6.
- [3] Hadi, S., Takwin, R. N. A., & Dani, A. (2016). Uji Kekuatan Tekan Dan Kekuatan Lentur Pipa Air Pvc. *Maret*, 16(1), 7.
- [4] Hidayatullah, A. F., Prasetyo, A. D., Sari, D. P., & Pratiwi, I. (2014). Analisis Kualitas Data dan Klasifikasi Data Pasien Kanker. *Seminar Nasional Informatika Medis*, 38–47.
- [5] Patoni, K. S. (2016). *Analisa Pengaruh Kualitas Produk, Harga, dan Kepercayaan Terhadap Loyalitas Pelanggan XL Pada Mahasiswa Jurusan Manajemen Angkatan 2014*. STIE Pelita Bangsa.
- [6] Santoso, T. B. (2011). Anaisa dan Penerapan Metode C4.5 Untuk Prediksi Loyalitas Pelanggan. *Jurnal Ilmiah Fakultas Teknik LIMIT'S*, 10(1).
- [7] Shella, P. (2015). *Sistem Pendukung Keputusan Dengan Menggunakan Decission Tree Di Sekolah Menengah Pertama (Studi Kasus Di SMPN 2 Rembang) Sistem Pendukung Keputusan Dengan Menggunakan Decission Tree Di Sekolah Menengah Pertama (Studi Kasus di SMPN 2 Rembang)*. UNIVERSITAS NEGERI SEMARANG.
- [8] Standart, N. B. (n.d.). SNI 06-0084-2002 Pipa PVC untuk saluran air minum. Retrieved August 18, 2015, from <https://www.slideshare.net/dewonugroho/definisi-kualitas-37324948>
- [10] Vuldari, R. T. (2017). *Data Mining Teori dan Aplikasi Rapidminer*. Yogyakarta: Penerbit Gava Media.
- [11] Wiratama, F. R., & Astuti, S. (2017). Implementasi Algoritma C4.5 untuk Analisa Performa Pelayanan Bank Terhadap Nasabah. *Eksplora Informatika*, 6, 127– 135.
- [12] Zega, S. A. (2014). Penggunaan Pohon Keputusan untuk Klasifikasi Tingkat Kualitas Mahasiswa Berdasarkan Jalur Masuk Kuliah. *Seminar Nasional Aplikasi Teknologi Informasi (SNATI)*, 7–13.