

# Implementasi Data Mining Algoritma C5.0 Dalam Sertifikasi Produk Pengguna Tanda SNI Pada Air Minum Dalam Kemasan (Studi Kasus: Balai Riset dan Standardisasi Industri Medan)

Martin Andriano Manurung

Program Studi Teknik Informatika, STMIK Budi Darma, Medan, Indonesia

Email: martinmanurung20@gmail.com

**Abstrak**—Judul skripsi adalah nama Lembaga Sertifikasi Produk yang selanjutnya disebut Lspro adalah lembaga yang melakukan kegiatan Sertifikasi Produk Pengguna Tanda SNI oleh Balai Riset dan Standardisasi Industri Medan. Selama ini Balai Riset dan Standardisasi Industri Medan Sertifikasi Produk Pengguna Tanda SNI yang selanjutnya disebut SPPT – SNI adalah Sertifikasi Produk Pengguna Tanda SNI yang dikeluarkan oleh Lembaga Sertifikasi Produk kepada produsen yang mampu memproduksi barang yang dihasilkan sesuai persyaratan SNI. Namun proses pengklasifikasian data yang dimiliki oleh Balai Riset dan Standardisasi Industri Medan yang terlalu banyak sehingga ketika dilakukan proses klasifikasi data sertifikasi tersebut akan membutuhkan waktu yang sangat lama dan akan berdampak pada pemberian sertifikasi tersebut. Algoritma C5.0 merupakan algoritma yang menghasilkan pohon keputusan yang lebih sederhana dan penggunaan memori yang lebih efisien. Algoritme C5.0 dapat mengklasifikasikan model prediksi berstruktur pohon (tree) dan aturan (rule-based). Penggunaan algoritma C5.0 untuk mendapatkan model keputusan dengan hasil tingkat akurasi yang lebih tinggi Metode ini akan menghitung secara lebih cepat. Tujuan dari penelitian ini untuk dapat membantu mempercepat proses pengambilan keputusan, mengelola informasi terkandung di dalam data transaksi menjadi sebuah pengetahuan (knowledge) yang baru dalam SPPT-SNI oleh Balai Riset dan Standardisasi Industri Medan.

**Kata Kunci:** Data Mining, SPPT – SNI , Algoritma C5.0

**Abstract**—The title of the thesis is the name of the Product Certification Institution, hereinafter referred to as Lspro, is an institution conducting SNI Mark Product User Certification activities by the Medan Industrial Research and Standardization Center. All this time, the Research Center and Medan Industrial Standardization for SNI Mark Product User Certification hereinafter referred to as SPPT - SNI is SNI Mark Product User Certification issued by the Product Certification Institute to producers who are able to produce goods produced in accordance with SNI requirements. However, the process of classifying the data possessed by the Medan Industrial Research and Standardization Center is too much so that when the certification data classification process is carried out it will take a very long time and will have an impact on granting such certification. C5.0 algorithm is an algorithm that produces a simpler decision tree and a more efficient use of memory. C5.0 algorithm can classify tree-based and rule-based prediction models. The use of C5.0 algorithm to get a decision model with a result of a higher level of accuracy This method will calculate faster. The purpose of this research is to be able to help speed up the decision making process, manage information contained in transaction data into new knowledge in SPPT-SNI by the Research and Standardization Center

**Keywords:** Data Mining, SPPT-SNI, C5.0 Algorithm

## 1. PENDAHULUAN

Sertifikasi Produk Pengguna Tanda SNI yang selanjutnya disebut SPPT – SNI adalah Sertifikasi Produk Pengguna Tanda SNI yang dikeluarkan oleh Lembaga Sertifikasi Produk kepada produsen yang mampu memproduksi barang yang dihasilkan sesuai persyaratan SNI. Lembaga Sertifikasi Produk yang selanjutnya disebut Lspro adalah lembaga yang melakukan kegiatan Sertifikasi Produk Pengguna Tanda SNI oleh Balai Riset dan Standardisasi Industri Medan. Selama ini Balai Riset dan Standardisasi Industri Medan sudah memiliki tempat penyimpanan data sertifikasi berbagai bidang industri, salah satunya ialah sertifikasi Air Minum Dalam Kemasan (AMDK). Namun proses pengklasifikasi data yang dimiliki oleh Balai Riset dan Standardisasi Industri Medan yang terlalu banyak ( misalnya data perusahaan, tipe sertifikasi, nomor sertifikasi, nomor dan judul standar), sehingga ketika dilakukan proses pencarian data sertifikasi tersebut akan membutuhkan waktu yang sangat lama dan akan berdampak pada pemberian sertifikasi tersebut.

Penanganan yang selama ini terjadi dalam proses klasifikasi data oleh Balai Riset dan Standardisasi Industri Medan adalah melakukan pengecekan ulang proses yang dibutuhkan dari awal hingga akhir secara manual dan komputerisasi melalui layanan internet website yang dimiliki Balai Riset dan Standardisasi Industri Medan.

Solusi untuk mengatasi masalah proses penggalan data dengan jumlah yang sangat banyak dalam Sertifikasi Produk Pengguna Tanda SNI oleh Balai Riset dan Standardisasi Industri Medan yaitu menggunakan teknik Data Mining, karena penggunaan teknik Data Mining diharapkan dapat membantu mempercepat proses pengambilan keputusan, memungkinkan perusahaan untuk mengelola informasi terkandung di dalam data transaksi menjadi sebuah pengetahuan (knowledge) yang baru dalam Sertifikasi Produk Pengguna Tanda SNI oleh Balai Riset dan Standardisasi Industri Medan.

## 2. METODOLOGI PENELITIAN

### 2.1 Data Mining

Data adalah segala fakta, angka, atau teks yang dapat diproses oleh komputer . Saat ini akumulasi pertumbuhan jumlah data berjalan dengan cepat dalam format basis data yang berbeda. Sementara informasi adalah pola, asosiasi, atau hubungan antara semua data yang dapat memberikan informasi. Sebagai contoh , analisis titik eceran (retail point) data transaksi penjualan dapat menghasilkan informasi mengenai produk apa yang sebaiknya dijual dan kapan menjualnya. Informasi dapat dikonversi menjadi pengetahuan mengenai pola – pola historis dan tren masa depan, misalnya ringkasan informasi tentang penjualan eceran supermarket dapat dianalisis sehubungan dengan upaya promosi untuk memberikan pengetahuan mengenai perilaku konsumen dalam membeli. Dengan demikian, produsen atau pengecer dapat menentukan item yang paling rentan terhadap upaya promosi [1].

### 2.2 Algoritma C5.0

Algoritma C5.0 adalah salah satu algoritma data mining yang khususnya diterapkan pada teknik decision tree. C5.0 merupakan penyempurnaan algoritma sebelumnya yang dibentuk oleh Ross Quinlan pada tahun 1987, yaitu ID3 dan C4.5. Dalam algoritma ini pemilihan atribut yang akan diproses menggunakan information gain. Dalam memilih atribut untuk pemecah obyek dalam beberapa kelas harus dipilih atribut yang menghasilkan information gain paling besar. Atribut dengan nilai information gain tertinggi akan dipilih sebagai parent bagi node selanjutnya. Algoritma C5.0 adalah versi komersial dari Algoritma C4.5 yang secara luas digunakan di banyak pemakatan data mining seperti Clementine and RuleQuest. Tidak seperti Algoritma C4.5, penggunaan algoritma yang tepat untuk Algoritma C5.0 belum terungkap. Hasil menunjukkan bahwa Algoritma C5.0 meningkatkan pada penggunaan memori sekitar 90%, lebih cepat daripada Algoritma C4.5.[4]

### 2.3 Decision Tree

*Decision Tree* atau Pohon Keputusan merupakan metode klasifikasi dan prediksi yang sangat kuat dan terkenal. Metode *Decision Tree* mengubah fakta yang sangat besar menjadi Pohon Keputusan yang mempersentasikan aturan. Aturan dengan mudah dipahami dengan bahasa alami. Dan mereka juga dapat diekspresikan ke dalam bentuk bahasa basis data *Structured Query Language* untuk mencari *record* pada kategori tertentu. Pohon keputusan juga berguna untuk mengeksplorasi data, menemukan hubungan tersembunyi antara sejumlah calon variable input dengan sebuah variable target. Karena pohon keputusan memadukan antara eksplorasi data dan permodelan, sangat bagus sebagai langkah awal dalam proses pemodelan bahkan dijadikan sebagai model akhir dari beberapa teknik lain. Sebuah pohon keputusan adalah sebuah struktur yang dapat digunakan untuk membagi kumpulan data yang besar menjadi himpunan – himpunan record yang lebih kecil dengan menerapkan serangkaian aturan keputusan. Dengan masing – masing rangkaian pembagian, anggota himpunan hasil menjadi mirip satu dengan yang lain. Sebuah pohon keputusan mungkin dibangun dengan seksama secara manual atau dapat tumbuh secara otomatis dengan menerapkan salah satu atau beberapa algoritma pohon keputusan untuk memodelkan himpunan data yang belum terklasifikasi. Variabel tujuan biasanya dikelompokkan dengan pasti dan model pohon keputusan lebih mengarah kepada perhitungan probabilitas dari tiap – tiap *record* terhadap kategori – kategori tersebut atau untuk mengklasifikasi dengan mengelompokkannya dalam satu kelas. Data dalam pohon keputusan biasanya dinyatakan dalam bentuk table dengan atribut dan *record*. Atribut menyatakan suatu parameter yang sebagai kriteria dalam pembentukan pohon. Misalkan untuk menentukan main tenis, kriteria yang diperhatikan adalah cuaca, angin, dan temperatur . Salah satu atribut merupakan atribut yang menyatakan data solusi per *item* data yang disebut target atribut. Atribut memiliki nilai – nilai yang dinamakan *instance* berupa cerah, berawan, dan hujan Proses pada pohon keputusan adalah mengubah bentuk data (*table*), menjadi model pohon, mengubah model pohon menjadi *rule*, dan menyederhanakan *rule* [1].

## 3. HASIL DAN PEMBAHASAN

### 3.1 Analisa Masalah

Penelitian dilakukan di Balai Riset dan Standardisasi Industri Medan yaitu di bagian Lembaga Sertifikasi Produk (LSPPro). Langkah pertama adalah melakukan penelitian pengamatan secara langsung di Balai Riset dan Standardisasi, dalam hal ini proses sertifikasi produk pengguna tanda SNI Air Minum dalam Kemasan. Pengamatan dilakukan dengan cara observasi langsung ke instansi untuk melihat proses yang ada, dan dengan mengetahui proses yang ada diharapkan dapat membuat data yang sudah ada menjadi kumpulan data yang sangat banyak sehingga dalam melakukan klasifikasi data sertifikasi produk pengguna tanda SNI dengan menggunakan teknik Data Mining

Tujuan klasifikasi data dalam Sertifikasi Produk Pengguna Tanda SNI oleh Balai Riset dan Standardisasi Industri Medan menggunakan teknik *Data Mining*, yaitu membantu mempercepat proses pengambilan keputusan, memungkinkan perusahaan untuk mengelola informasi terkandung di dalam data transaksi menjadi sebuah

pengetahuan (knowledge) yang baru dalam Sertifikasi Produk Pengguna Tanda SNI oleh Balai Riset dan Standardisasi Industri Medan.

### 3.2 Penerapan Algoritma C5.0

Pengolahan *Data Mining* dengan menggunakan algoritma C5.0 dilakukan dengan perhitungan *Entropy* dan *Gain* untuk menentukan akar (*root*) dari pohon keputusan dalam menentukan klasifikasi anggota berdasarkan tingkat potensi anggota dalam memberikan kontribusi pada perusahaan. Data dikelompokkan terlebih dahulu dengan menghitung jumlah kasus, jumlah kasus untuk anggota yang berpotensi tinggi, sedang dan rendah. Selanjutnya dihitung nilai *Entropy* dan *Gain* dari semua atribut. Untuk menentukan akar ditentukan dari nilai *Gain* tertinggi. Hasil pengelompokan data anggota dapat dilihat pada tabel 1. berikut

**Tabel 1.** Pengelompokan Data Anggota

Node		Jumlah Kasus	Potensi Anggota	
			Layak	Dipertimbangkan
1	Total Keadaan Warna	50	30	20
	Rendah	15	5	10
	Sedang	15	9	6
	Tinggi	20	16	4
	pH			
	Rendah	15	7	8
	Sedang	15	10	5
	Tinggi	20	13	7
	Kekeruhan			
	Rendah	20	5	15
	Sedang	5	5	0
	Tinggi	25	20	5
	Jumlah zat terlarut			
	Rendah	15	2	13
	Sedang	15	11	4
	Tinggi	20	17	3
	Zat Organik			
	Rendah	15	5	10
	Sedang	15	9	6
	Tinggi	20	16	4

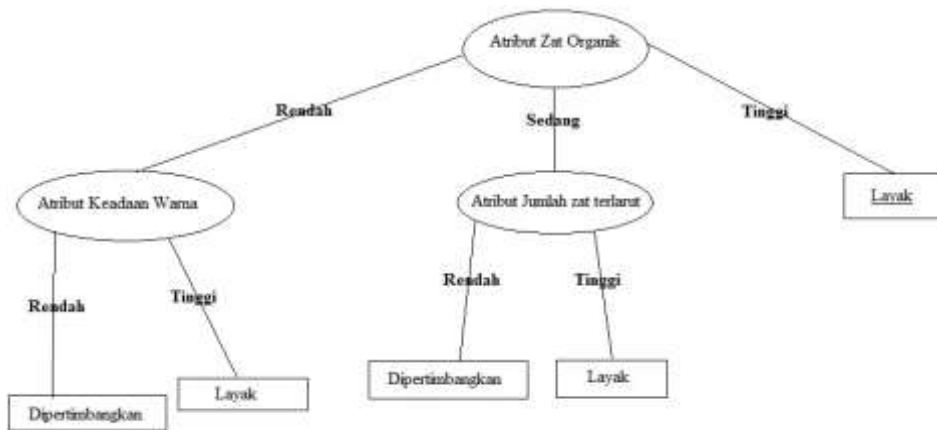
Setelah data kita kelompokkan seperti pada tabel 8, langkah selanjutnya adalah mencari nilai *Entropy* total dan juga nilai semua atribut yang digunakan. Kemudian dihitung *Gain* dari masing – masing atribut yang digunakan. Berikut perhitungan nilai *Entropy* dan *Gain* pada node 1

**Tabel 2.** Hasil Perhitungan Node 1

Node 1	Jumlah Kasus	Potensi Anggota		<i>Entropy</i>	<i>Gain</i>
		Layak	Dipertimbangkan		
Total Keadaan Warna	50	30	20	0,349	0,076
Rendah	15	5	10	0,166	
Sedang	15	9	6	0,156	
Tinggi	20	16	4	0,204	
pH					0,231
Rendah	15	7	8	0,15	
Sedang	15	10	5	0,166	
Tinggi	20	13	7	0,163	
Kekeruhan					0,143
Rendah	20	5	15	0,187	
Sedang	5	5	0	0,301	
Tinggi	25	20	5	0,204	
Jumlah zat terlarut					0,198

Node 1	Jumlah Kasus	Potensi Anggota		Entropy	Gain
		Layak	Dipertimbangkan		
Zat organik	Rendah	15	2	13	0,23
	Sedang	15	11	4	0,182
	Tinggi	20	17	3	0,223
	Rendah	15	5	10	0,166
	Sedang	15	9	6	0,156
	Tinggi	20	16	4	0,204

Dari hasil perhitungan pada Node 1, nilai Entropy dan Gain seperti terlihat pada tabel 4.9 dapat diketahui bahwa atribut yang memiliki Gain tertinggi adalah atribut pH yaitu sebesar 0.231. Ada tiga nilai atribut dari pH, yaitu rendah, sedang, tinggi. Dari ketiga nilai atribut tersebut, nilai atribut tinggi sudah mengklasifikasikan kasus menjadi 1, yaitu keputusan Layak sehingga tidak perlu dilakukan perhitungan lebih lanjut, tetapi untuk nilai atribut Layak masih perlu diperlukan perhitungan lagi.



Gambar 1. Pohon Keputusan Hasil Perhitungan Node 1

Berdasarkan pohon keputusan diatas dapat diambil aturan final yaitu:

1. Atribut Zat organik = rendah
2. Atribut Keadaan warna = rendah: Dipertimbangkan {Dipertimbangkan=4, Layak=0}
3. Atribut Keadaan warna = tinggi: Layak {Dipertimbangkan=0, Layak=2}
4. Atribut Zat organik = sedang
5. Atribut Jumlah zat terlarut = rendah: Dipertimbangkan {Dipertimbangkan=4, Layak=0}
6. Atribut Jumlah zat terlarut = sedang: Layak {Dipertimbangkan=0, Layak=2}
7. Atribut Zat organik = tinggi: Layak {Dipertimbangkan=0, Layak=8}

### 3.3 Pengujian

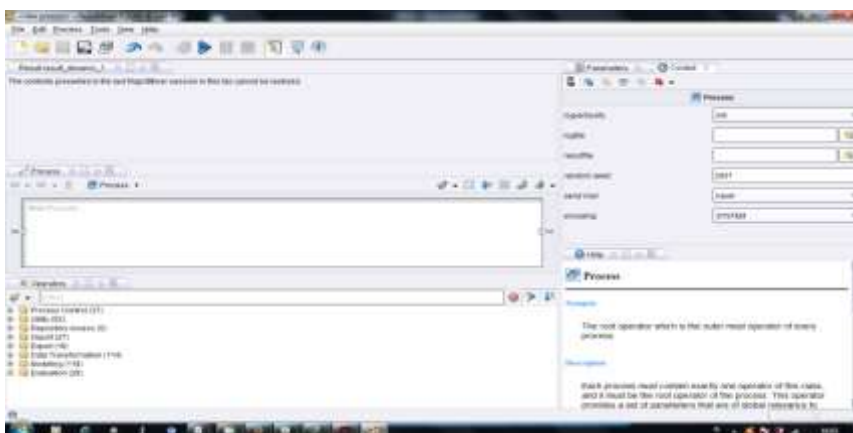
Pengujian perangkat lunak merupakan suatu investigasi yang dilakukan untuk mendapatkan informasi mengenai kualitas dari produk atau layanan yang sedang diuji. Pengujian perangkat lunak juga memberikan pandangan mengenai perangkat lunak secara obyektif dan independen, yang bermanfaat dalam operasional bisnis untuk memahami tingkat resiko pada implementasinya. Teknik-teknik pengujian mencakup, namun tidak terbatas pada proses mengeksekusi suatu bagian program atau keseluruhan aplikasi dengan tujuan untuk menemukan kesalahan atau cacat lainnya. Adapun pengujian dalam perangkat lunak RapidMiner ini dapat dilakukan dalam perangkat komputer. Pada tahap pengujian dilakukan pengolahan Data Mining menggunakan aplikasi RapidMiner berdasarkan data anggota sebanyak 50 buah data. Berikut terlihat dataset yang terdiri 50 record data dan berisi variabel-variabel atribut kondisi dan atribut keputusan untuk menganalisa dan klasifikasi data Sertifikasi Produk Pengguna tanda SNI Air minum dalam kemasan disimpan dalam aplikasi microsoft excel dengan nama file datasptsniamdk.xls

Adapun tampilan perangkat lunak RapidMiner yang masih belum diinputkan data atau kosong dapat dilihat pada tampilan gambar berikut:



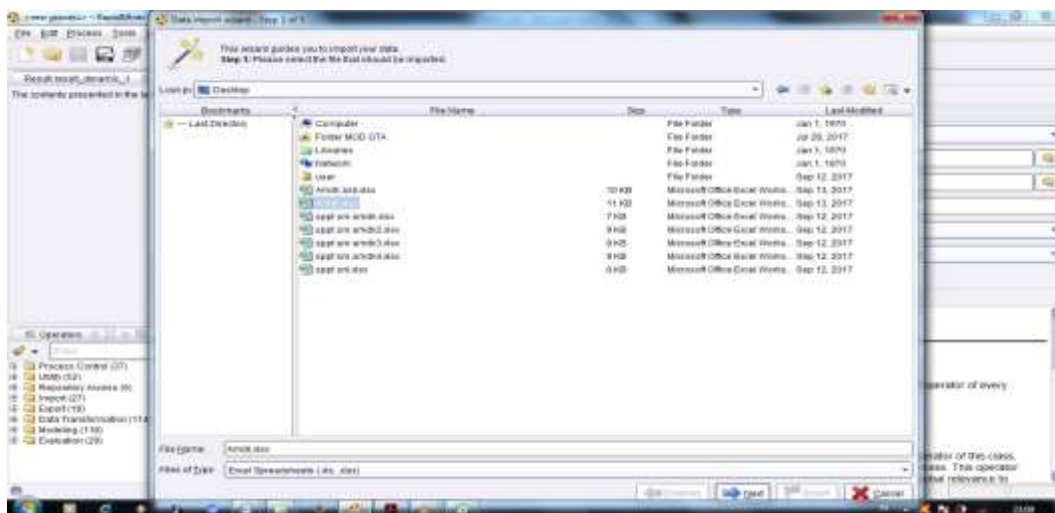
Gambar 2. Tampilan awal RapidMiner 5.3

Sebelum data diolah, maka data harus terlebih dahulu dipilih seperti terlihat pada gambar 3.



Gambar 3. Tampilan Menu New Process

Pilih Menu *File*, pilih *Import Data* dan pilih *Import Excel Sheet*, sehingga akan muncul tampilan seperti berikut :



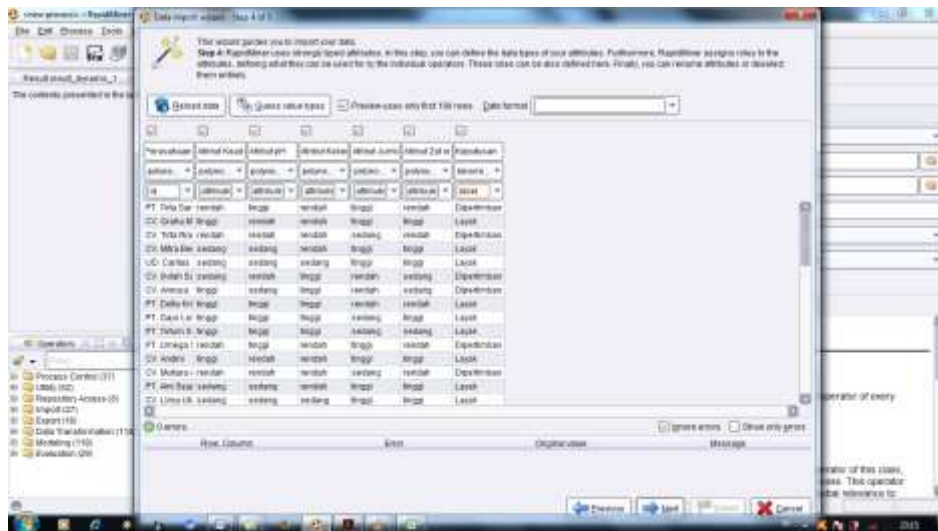
Gambar 4. Tampilan Import Data Excel Sheet

Arahkan ke folder tempat file *data.pptsniamdk.xls* disimpan, kemudian klik pada file *data.pptsniamdk.xls*. Klik pada baris pertama di kolom *Annotation*, pilih *name* dan klik *Next* seperti tampilan berikut :



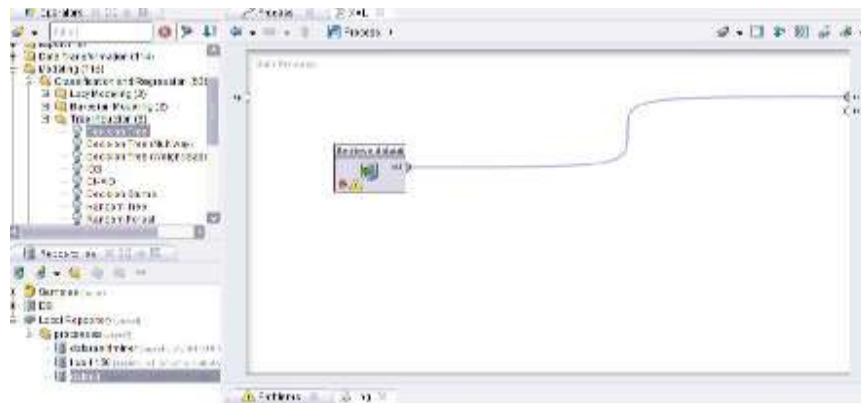
Gambar 5. Tampilan pilihan Annotation

Selanjutnya ubah properties atribut norek menjadi *id* dan atribut Status menjadi *label*, dan sesuaikan tipe tiap field atribut lalu klik *Next* seperti tampilan berikut :



Gambar 6. Tampilan pilihan Properties atribut dan field

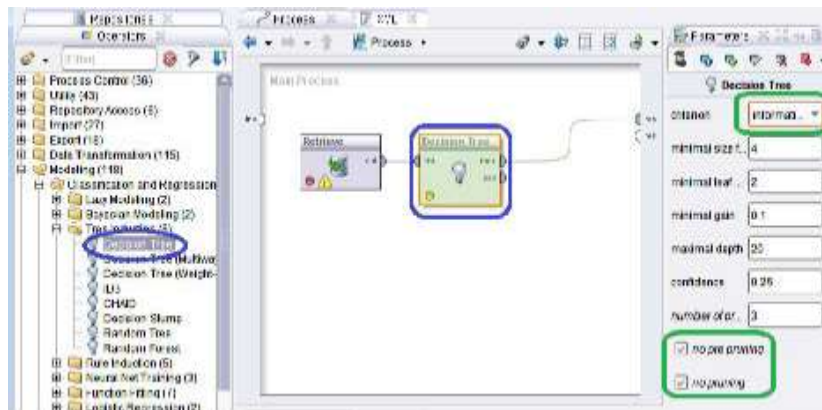
Ketikkan *DataSptSniAmdk* di *textbox* Name, kemudian klik tombol *Finish* untuk menyimpan *dataset* yang akan kita pakai. Untuk menghasilkan pohon keputusan dengan algoritma C5.0, pilih *dataset* yang kita sudah simpan sebelumnya di *Tab Repositories*, kemudian *drag dataset* datauji ke *Tab Process*, seperti tampilan berikut :



Gambar 7. Tampilan Repositoris

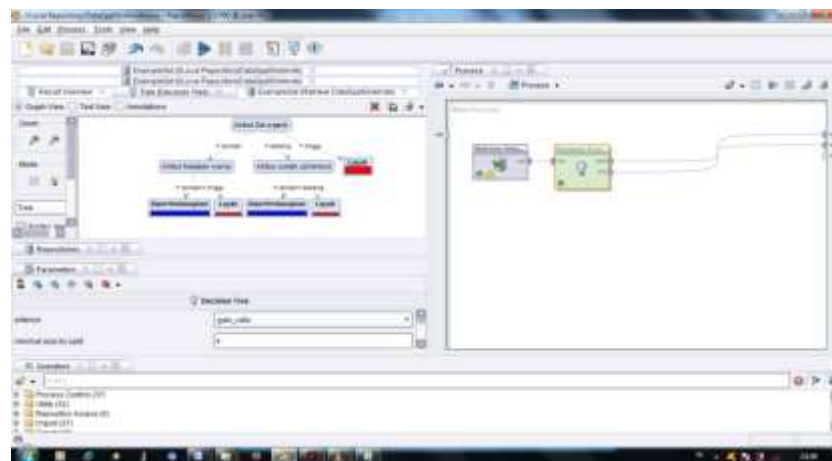
Selanjutnya tambahkan algoritma decision tree dari *Tab Operators*, pilih *Modelling(118)*, pilih *Classification and Regression (53)*, pilih *Tree Induction(8)*, kemudian *drag Decision Tree* ke *Tab Process*, pada

Tab Parameters pilih *information\_gain* pada *criterion*, dan cek pada pilihan *no pre pruning* dan *no pruning*, seperti tampilan berikut :



Gambar 8. Tampilan *Process Decision tree*

Pilih menu *Process* dan pilih *Run*, pilih *No* dan pilih *Yes*, secara otomatis Rapidminer akan memproses untuk menghasilkan pohon keputusan seperti gambar dibawah :



Gambar 9. Tampilan hasil akhir *Process Decision Tree*

#### 4. KESIMPULAN

Berdasarkan riset dan pembahasan yang telah dilakukan, dapat diberikan beberapa kesimpulan sebagai berikut :

1. Mengetahui perusahaan industri Air Minum dalam Kemasan dalam menyajikan data serta data pra proses yang telah dilakukan Balai Riset dan Standardisasi Industri Medan
2. Penerapan teknik *Data Mining* Algoritma C5.0 untuk mengetahui proses Sertifikasi Produk Pengguna Tanda SNI Air Minum dalam Kemasan pada Balai Riset dan Standardisasi Medan.
3. Menggunakan *tools RapidMiner 5.3* didalam mengolah data pada proses Sertifikasi Produk Pengguna Tanda SNI pada Balai Riset dan Standardisasi Industri Medan
4. Setelah dilakukan pengimplementasian teknik *Data Mining* Algoritma C5.0 pada *RapidMiner 5.3*, maka hasil pengolahan data Sertifikasi Produk Pengguna Tanda SNI pada Air Minum dalam Kemasan, setelah dilakukan perhitungan, menyatakan bahwa nilai *Gain* tertinggi adalah atribut pH dan nilai *Gain* terendah adalah atribut Keadaan warna.

#### REFERENCES

- [1] Eko Prasetyo, DATA MINING KONSEP DAN APLIKASI MENGGUNAKAN MATLAB, I ed, Indonesia: Andi, 2009.
- [2] Kusriani and Emha Taufiq Luthfi, ALGORITMA DATA MINING, I ed, Theresia Ari Prabawati, Ed. Yogyakarta, Indonesia: Andi, 2009.
- [3] I Kadek Juni Arta, Gede Indrawan. Gede Rasben Dantes, "DATA MINING REKOMENDASI CALON MAHASISWA MENGGUNAKAN METODE TECHNIQUE FOR OTHERS REFERENCE BY SIMILARITY TO IDEAL SOLUTION," Jurnal Sains dan Teknologi, vol. 5, no. 5, Oktober 2016.
- [4] Kurniati Putri Wiradhaningsih, Dian Eka Ratnawati. Marji, "PENERAPAN ALGORITMA DECISION TREE C5.0 UNTUK PERAMALAN FOREX.,"
- [5] Abdul Kadir, PENGENALAN ALGORITMA, I ed, Indonesia: Andi, 2009.

- [6] [http://bsn.go.id/main/berita/berita\\_det/7008/Infografis---Alur-ProsesSertifikasi-SNI-padaProduk#.WZFO-jWJjCM](http://bsn.go.id/main/berita/berita_det/7008/Infografis---Alur-ProsesSertifikasi-SNI-padaProduk#.WZFO-jWJjCM).
- [7] Retna G, BELAJAR CEPAT MICROSOFT EXCEL 2007, I ed, Indonesia: Andi, 2009.
- [8] D. P. Utomo and B. Purba, "Penerapan Datamining pada Data Gempa Bumi Terhadap Potensi Tsunami di Indonesia," Prosiding Seminar Nasional Riset Information Science (SENARIS), vol. 1, pp. 846-853, 2019.
- [9] D. P. Utomo and M. Mesran, "Analisis Komparasi Metode Klasifikasi Data Mining dan Reduksi Atribut Pada Data Set Penyakit Jantung," JURNAL MEDIA INFORMATIKA BUDIDARMA, vol. 4, no. 2, pp. 437-444, 2020.