

Implementasi Data Mining dengan Algoritma C4.5 pada Penjualan Obat

¹Reza Avrizal, ²Saruni Dwiasnati
Fakultas Informatika, Universitas Indrapasta PGRI¹
Fakultas Ilmu Komputer, Universitas Mercu Buana²

Jl. Raya Tengah No.80, RT.6/RW.1, Gedong, Kec. Ps. Rebo, Kota Jakarta Timur¹
Jl. Meruya Selatan No.1, RT.4/RW.1, Jakarta Barat²

E-mail : Reza.Avrizal@unindra.ac.id¹
Saruni.dwiasnati@mercubuana.ac.id²

Abstract - The rise of various types of drugs issued by BPOM makes hospitals, clinics or pharmacies have to provide drugs with various versions, can be in the standard version or exclusive version. Problems that often occur either in hospitals or pharmacies such as in offering these drug products, especially vitamins for cross selling with products that are commonly purchased by drug users, still do not go directly to what is needed by drug users still have to guess what drugs which is suitable for a disease that is being suffered by the patient. The current research conducted by the author aims to implement a number of existing algorithms in data mining, but the authors focus more on the C4.5 algorithm which functions to be able to produce complex decisions to be simpler such as a prediction of sales of drugs specifically vitamins where the drug is DO or NOT LAKU is to be consumed by consumers so that the accuracy results can be used to provide information on which products DO or DO NOT. The results showed that the tree produced from software created by the decision tree method with C4.5 algorithm has a percentage of 87.90% between the truth trees to 100% that can be obtained from a prediction. A large percentage of the truth tree is strongly influenced by the training data used to build a decision tree model.

Keywords: Drug Sales, Data Mining, Decision Tree, C4.5 Algorithm

Abstrak – Maraknya berbagai jenis obat yang dikeluarkan oleh BPOM membuat rumah sakit-rumah sakit, klinik-klinik ataupun apotek harus menyediakan obat dengan berbagai versi, bisa dalam versi standar ataupun versi exclusive. Permasalahan yang sering terjadi baik pada rumah sakit ataupun apotek seperti dalam menawarkan produk obat tersebut yang khususnya vitamin untuk di cross selling dengan produk yang biasa di beli oleh pengguna obat masih belum langsung menuju ke yang di butuhkan oleh pengguna obat masih harus menerka-nerka obat apa yang cocok untuk sebuah penyakit yang sedang di derita si pasien. Penelitian saat ini yang penulis lakukan bertujuan untuk mengimplementasikan beberapa algoritma yang ada pada data mining, namun penulis lebih memfokuskan pada algoritma C4.5 yang berfungsi untuk mampu menghasilkan keputusan kompleks menjadi lebih sederhana seperti sebuah prediksi penjualan obat yang khususnya vitamin dimana obat tersebut LAKU atau TIDAK LAKU untuk dikonsumsi oleh konsumen sehingga di dapat hasil accuracy nya digunakan untuk pemberian informasi produk mana yang LAKU atau TIDAK LAKU. Hasil penelitian menunjukkan bahwa pohon yang dihasilkan dari perangkat lunak yang dibuat dengan metode decision tree dengan algoritma C4.5 memiliki persentase 87.90% antara pohon kebenaran sampai 100% yang dapat didapatkan dari sebuah prediksi. Persentase besar pohon kebenaran sangat dipengaruhi oleh data training yang digunakan untuk membangun suatu model pohon keputusan.

Kata Kunci : Penjualan Obat, Data Mining, Decision Tree, Algoritma C4.5

I. PENDAHULUAN

Dunia teknologi informasi saat ini sedang sangat berkembang dengan pesat. Perkembangan ini menyebabkan kebutuhan pengguna akan data dan informasi dalam berbagai bidang semakin meningkat. Untuk mengolah data tersebut dibutuhkan sebuah sistem terkomputerisasi dengan mengkombinasikan sebuah database satu dengan database lainnya. Data yang dahulu hanyalah data yang tertumpuk jika sudah selesai di olah dan tidak diolah lebih lanjut, namun zaman sekarang sebuah data dapat diolah kembali sehingga mendapatkan suatu informasi baru. Apotek yang dimaksud pada penelitian ini adalah apotek yang bergerak dalam bidang distribusi penyaluran obat-obat ke para retail baik rumah sakit ataupun apotek-apotek yang menjual langsung ke konsumen.

Dalam penyajian data LAKU atau TIDAK LAKU sebuah apotek sering mengalami kesulitan untuk menentukan barang mana yang LAKU atau TIDAK LAKU yang diminati oleh para konsumennya. Hal ini diakibatkan oleh menumpuknya stok vitamin yang kurang diminati oleh customer di gudang, namun untuk vitamin yang diminati oleh customer sama sekali tidak ada dalam persediaan pada apotek tersebut.

Untuk mengatasi masalah yang dijelaskan di atas, maka dilakukan pengklasifikasian obat yang LAKU dan TIDAK LAKU terjual di Apotek tersebut berdasarkan variabel-variabel yang di dapatkan. Algoritma yang digunakan dalam pengklasifian obat yang LAKU atau TIDAK LAKU menggunakan algoritma C4.5 yang mampu memberikan informasi mengenai vitamin yang LAKU atau TIDAK LAKU pada apotek tersebut untuk meminimalisasikan stok yang menumpuk untuk vitamin yang di kategorikan TIDAK LAKU.

II. LANDASAN TEORI DAN METODE

Data Mining Merupakan proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, *machine learning* untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai *database* besar. Terdapat beberapa istilah lain yang memiliki makna sama dengan data mining, yaitu *Knowledge discovery in databases* (KDD), ekstraksi pengetahuan (*knowledge extraction*), Analisa data/pola (*data/pattern analysis*), kecerdasan bisnis (*business intelligence*) dan data *archaeology* dan data *dredging* [1]. Definisi Data Mining dari ilmuwan lain dapat di definisikan sebuah tahapan yang dapat digunakan untuk analisa terhadap data untuk menemukan hubungan yang jelas serta menyimpulkannya yang belum diketahui sebelumnya dengan cara terkini dipahami dan berguna bagi pemilik data tersebut. Karakteristik data mining dapat dijelaskan sebagai berikut:

- a. Data mining berhubungan dengan penemuan sesuatu yang tersembunyi dan pola data tertentu yang tidak diketahuindari data sebelumnya.
- b. Data mining biasa menggunakan data yang sangat besar untuk dapat diolah menjadi sebuah informasi. Biasanya data yang besar digunakan untuk membuat hasil lebih percaya dari data yang sedikit.
- c. Data mining berguna untuk membuat keputusan yang kritis, terutama dalam strategis (Davies, et al., 2004).

Algoritma klasifikasi merupakan suatu teknik yang dilakukan dengan melihat pada kelakuan dan atribut dari kelompok yang telah didefinisikan sebelumnya. Teknik ini dapat memberikan klasifikasi pada data baru dengan memanipulasi data yang ada yang telah diklasifikasi dan dengan menggunakan hasilnya untuk memberikan sejumlah aturan untuk memberikan sebuah informasi baru. Aturan-aturan tersebut digunakan pada data-data baru untuk diklasifikasi agar menjadi sebuah model yang dapat digunakan untuk penentuan.

Secara konsep, Pohon Keputusan merupakan salah satu teknik dari Analisis Keputusan (Decision Analysis). Terdapat banyak definisi teknis yang bisa ditemukan dalam berbagai sumber mengenai pohon keputusan karena beragamnya aplikasi pohon keputusan ini pada berbagai jenis proses dan industri yang berbeda-beda. Definisi yang paling sederhana mengenai apa itu Pohon Keputusan, adalah diagram analisis yang dapat membantu pengambil keputusan ketika menghadapi beberapa pilihan dengan cara memproyeksikan hasil yang mungkin terjadi di segala aspek. 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 Definisi teknis yang diambil dari beberapa buku Manajemen Operasional, bahwa Pohon Keputusan merupakan “Model alternatif pilihan yang tersedia bagi pengambil keputusan, beserta kemungkinan konsekuensinya.” Pohon keputusan dapat digunakan saat membuat berbagai pilihan, mulai dari keputusan yang paling sederhana sampai yang sangat yang rumit. Konsep dari pohon keputusan adalah mengubah data menjadi pohon keputusan dan aturan-aturan keputusan. Pohon keputusan memadukan antara eksplorasi data dan pemodelan, sehingga sangat bagus sebagai langkah awal dalam proses pemodelan bahkan ketika dijadikan sebagai model akhir dari beberapa teknik lainnya.

Manfaat utama dari penggunaan pohon keputusan adalah kemampuannya untuk menyederhanakan proses pengambilan keputusan yang rumit menjadi lebih simpel sehingga pengambil keputusan akan lebih mudah menginterpretasikan solusi dari permasalahan yang ada.

Terdapat beberapa keuntungan menggunakan pohon keputusan, di antaranya adalah:

- Menghilangkan perhitungan-perhitungan yang tidak dibutuhkan. Sample yang diuji hanya berdasarkan kriteria atau kelas tertentu.
- Daerah pengambilan keputusan yang sebelumnya kompleks dan sangat global, dapat diubah menjadi lebih simpel dan spesifik.
- Dalam analisis multivariat, dengan kriteria dan kelas yang jumlahnya sangat banyak, seorang penguji biasanya perlu untuk mengestimasi baik itu distribusi dimensi tinggi ataupun parameter tertentu dari distribusi kelas tersebut. Metode pohon keputusan menghindari munculnya permasalahan dengan menggunakan kriteria yang jumlahnya lebih sedikit pada setiap node internal tanpa banyak mengurangi kualitas keputusan yang dihasilkan.
- Bersifat fleksibel, memilih fitur dari internal node yang berbeda, fitur yang terpilih akan membedakan suatu kriteria dibandingkan kriteria yang lain dalam node yang sama. Kefleksibelan metode pohon keputusan ini meningkatkan kualitas keputusan yang dihasilkan jika dibandingkan ketika menggunakan metode penghitungan satu tahap yang lebih konvensional

Seperti halnya tools pada umumnya, pohon keputusan juga mempunyai kekurangan, antara lain:

- Terjadi overlap dalam menggunakan kelas-kelas dan kriteria yang digunakan dalam jumlah besar. Hal tersebut juga dapat menyebabkan meningkatnya waktu pengambilan keputusan dan jumlah memori yang diperlukan.
- Pengakumulasian jumlah eror dari setiap tingkat dalam sebuah pohon keputusan yang besar.
- Kesulitan dalam mendesain pohon keputusan yang optimal.

- Hasil kualitas keputusan yang didapatkan dari metode pohon keputusan sangat tergantung pada bagaimana pohon tersebut didesain.

Decision tree menggunakan struktur hierarki untuk pembelajaran supervised. Proses dari decision tree dimulai dari root node hingga leaf node yang dilakukan secara rekursif. Di mana setiap percabangan menyatakan suatu kondisi yang harus dipenuhi dan pada setiap ujung pohon menyatakan kelas dari suatu data. Proses dalam pohon keputusan yaitu mengubah bentuk data (tabel) menjadi model pohon (tree) kemudian mengubah model pohon tersebut menjadi aturan (rule). Metode pohon keputusan digunakan untuk memperkirakan nilai diskret dari fungsi target yang mana fungsi pembelajaran direpresentasikan oleh sebuah pohon keputusan (decision tree). Pohon keputusan terdiri dari himpunan IF...THEN. Setiap path dalam tree dihubungkan dengan sebuah aturan, dimana premis terdiri atas sekumpulan node-node yang ditemui dan kesimpulannya dari aturan atas kelas yang terhubung dengan leaf node dari path.

Arsitektur pohon keputusan dibuat menyerupai bentuk pohon, dimana pada umumnya sebuah pohon terdapat akar (root), cabang dan daun (leaf). Pada pohon keputusan juga terdiri dari tiga bagian sebagai berikut :

- a. Root node Root node atau node akar merupakan node yang terletak paling atas dari suatu pohon.
- b. Internal node Internal Node ini merupakan node percabangan, dimana pada node ini hanya terdapat satu input dan mempunyai minimal dua output.
- c. Leaf node Node ini merupakan node akhir, hanya memiliki satu input, dan tidak memiliki output.

Pada pohon keputusan setiap leaf node menandai label kelas. Pada pohon keputusan di setiap percabangan menyatakan kondisi yang harus dipenuhi dan tiap ujung pohon menyatakan nilai kelas data.

Algoritma C4.5 adalah algoritma yang membentuk sebuah pohon keputusan yang sangat kuat dan klasifikasi tebal serta metode prediksi. Algoritma pohon keputusan mengubah fakta yang sangat besar menjadi pohon keputusan yang mewakili aturan. Pohon keputusan juga berguna untuk mengeksplorasi data, menemukan hubungan tersembunyi antara sejumlah variabel input prospektif dengan variabel target. Karena pohon keputusan menggabungkan eksplorasi data dan pemodelan, itu sangat bagus sebagai langkah pertama dalam proses pemodelan [2].

Algoritma C4.5 mengkonstruksi pohon keputusan dari data pelatihan, yang berupa kasus-kasus atau record (tupel) dalam basis data. Setiap kasus berisikan nilai dari atribut-atribut untuk sebuah kelas. Setiap atribut dapat berisi data diskret atau kontinyu (numerik). C4.5 juga menangani kasus yang tidak memiliki nilai untuk sebuah atau lebih atribut. Akan tetapi, atribut kelas hanya bertipe diskret dan tidak boleh kosong. Ada tiga prinsip kerja algoritma C4.5 pada tahap belajar dari data, yaitu sebagai berikut :

1. Pembuatan Pohon Keputusan Obyektif dari algoritma pohon keputusan adalah mengkonstruksi struktur data pohon (dinamakan pohon keputusan) yang dapat digunakan untuk memprediksi kelas dari sebuah kasus atau record baru yang belum memiliki kelas. Algoritma ini memilih pemecahan kasus-kasus yang terbaik dengan menghitung dan membandingkan gain ratio, kemudian pada node-node yang terbentuk di level berikutnya. Demikian seterusnya sampai terbentuk daun-daun.
2. Pemangkasan Pohon Keputusan dan Evaluasi (Opsional) Karena pohon yang dikonstruksi dapat berukuran besar dan tidak mudah dibaca, C4.5 dapat menyederhanakan pohon dengan melakukan pemangkasan berdasarkan nilai tingkat kepercayaan (confidence level). Selain untuk pengurangan ukuran pohon, pemangkasan juga bertujuan untuk mengurangi tingkat kesalahan prediksi pada kasus (rekord) baru.
3. Pembuatan Aturan-Aturan dari Pohon Keputusan (Opsional) Aturan-aturan dalam bentuk if-then diturunkan dari pohon keputusan dengan melakukan penelusuran dari akar sampai ke daun. Setiap node dan syarat pencabangannya akan diberikan di if, sedangkan nilai pada daun akan menjadi ditulis di then. Setelah semua aturan dibuat, maka aturan akan disederhanakan (digabung atau diperumum)

Pada konstruksi pohon C4.5, di setiap simpul pohon, atribut dengan nilai gain ratio tertinggi dipilih sebagai atribut split untuk simpul. Rumus dari gain ratio adalah sebagai berikut :

Dimana $gain(a)$ adalah information gain dari atribut a untuk himpunan sampel X dan $split\ info(a)$ menyatakan entropi atau informasi potensial yang didapat pada pembagian X menjadi n sub himpunan berdasarkan telaahan pada atribut a . Dalam teori informasi, entropi mengukur ketidakpastian antar variabel acak dalam file data. Claude E. Shannon telah mengembangkan gagasan tentang entropi dari variabel acak. Entropi dan informasi terkait menyediakan perilaku jangka panjang dari proses acak yang sangat berguna untuk menganalisis data. Perilaku dalam proses acak juga merupakan faktor kunci untuk mengembangkan pengkodean untuk teori informasi. Entropi merupakan pengukuran ketidakpastian rata-rata kumpulan data ketika kita tidak tahu hasil dari sumber informasi. Itu berarti bahwa seberapa banyak pengukuran informasi yang kita tidak punya. Ini juga menunjukkan jumlah rata-rata informasi yang kami akan menerima dari hasil sumber informasi. Untuk mendapatkan nilai gain ratio dalam pembentukan pohon keputusan, perlu menghitung dulu nilai informasi dalam satuan bits dari suatu kumpulan objek Bentuk perhitungan untuk entropi.

Pada dasarnya konsep dari algoritma C4.5 adalah mengubah data yang sudah di olah menjadi pohon keputusan dan aturan-aturan keputusan (*rule*). C4.5 adalah algoritma yang cocok untuk masalah klasifikasi dan prediksi. C4.5 memetakan nilai atribut menjadi kelas yang dapat diterapkan untuk klasifikasi baru [3].

Ada beberapa tahapan dalam membangun sebuah pohon keputusan dengan algoritma C4.5 yaitu [4].

1. Menyiapkan data training. Data training biasanya diambil dari data histori yang pernah terjadi sebelumnya dan sudah dikelompokkan ke dalam kelas-kelas tertentu.
2. Menentukan akar dari pohon. Akar akan diambil dari atribut yang terpilih, dengan cara menghitung nilai gain dari masing-masing atribut, nilai gain yang paling tinggi yang akan menjadi akar pertama.
3. Sebelum menghitung nilai gain dari atribut, hitung dahulu nilai entropi. Untuk menghitung nilai entropi digunakan rumus:

$$Entropy(S) = \sum_{i=1}^N - p_i \log_2 p_i \quad (1)$$

Di mana:

- S = Himpunan Kasus n = Jumlah partisi S
- P_i = Proporsi S_i terhadap S

4. Kemudian hitung nilai gain yang menggunakan rumus:

$$Gain(S, A) = entropy(S) - \sum_{i=1}^n \frac{|S_i|}{S} * entropy(S_i) \quad (2)$$

Di mana:

- S = Himpunan Kasus
- A = Fitur
- n = Jumlah Partisi Atribut A
- $|S_i|$ = Proporsi S_i terhadap S
- $|S|$ = Jumlah Kasus dalam S

5. Ulangi langkah ke-2 hingga semua record terpartisi.
6. Proses partisi pohon keputusan akan berhenti saat:
 - a. Semua record dalam simpul N mendapat kelas yang sama
 - b. Tidak ada atribut di dalam record yang dipartisi lagi.
 - c. Tidak ada record di dalam cabang yang kosong.

Salah satu algoritma induksi pohon keputusan yaitu ID3 (Iterative Dichotomiser 3). ID3 dikembangkan oleh J. Ross Quinlan. Dalam prosedur algoritma ID3, input berupa sampel training, label training dan atribut. Algoritma Decision Tree C4.5 merupakan pengembangan dari ID3. Sedangkan pada perangkat lunak open source WEKA mempunyai versi sendiri dari C4.5 yang dikenal sebagai J48.

Penjualan merupakan sebuah proses dimana kebutuhan pembeli dan kebutuhan penjual dipenuhi, melalui antar pertukaran informasi dan kepentingan. Dalam proses penjualan, penjual atau penyedia barang dan jasa memberikan kepemilikan suatu komoditas kepada pembeli untuk suatu harga tertentu. Secara umum pengertian penjualan adalah sejumlah total yang dikenakan kepada pelanggan untuk barang atau jasa yang di jual[3].

Rapidminer adalah salah satu software yang digunakan untuk pengolahan data mining. Pekerjaan yang dilakukan oleh RapidMiner text mining adalah berkisar dengan analisis teks, mengekstrak pola-pola dari data set yang besar dan mengkombinasikannya dengan metode statistika, kecerdasan buatan, dan database. *RapidMiner* menyediakan prosedur data mining dan machine learning, di dalamnya termasuk: ETL (extraction, transformation, loading), data preprocessing, visualisasi, modelling dan evaluasi. Proses data mining tersusun atas operator-operator yang nestable, dideskripsikan dengan XML, dan dibuat dengan GUI. Penyajiannya dituliskan dalam bahasa pemrograman Java.

Lambang *Rapidminer* dapat di lihat Gambar di bawah ini:



Gambar 1. Rapidminer

III. PEKERJAAN DAN DISKUSI HASIL

Analisis Kebutuhan

Yang berusaha dicapai dalam penelitian ini melalui implementasi data *mining* ini adalah mengetahui berapa nilai *accuracy* yang akan di hasilkan dari algoritma C4.5, dengan beberapa *indicator* yang dapat digunakan dan dengan data yang di dapat. *Tools* yang digunakan untuk mengolah data ini adalah *rapidminer* yang dibuat model nya dahulu lalu menghasilkan nilai. Nilai yang dapat di lihat dari *Rapidminer* dapat dijelaskan pada di bawah ini:

Confusion Matrix algoritma C4.5

Gambar 2 merupakan perhitungan *accuracy* yang dihitung dari data *training* menggunakan algoritma C4.5 yang menghasilkan *accuracy* 87.90%. Diketahui data *training* terdiri dari 149 *record* data, 6 data yang diprediksi Tidak ternyata Ya dan dari 12 data diprediksi Ya namun Tidak, tidak ada data yang di prediksi Tidak dan benar-benar Tidak dan 131 data secara benar diprediksi Ya.

accuracy: 87.90% +/- 4.04% (mikro: 87.92%)

	true Tidak	true Ya	class precision
pred. Tidak	0	6	0.00%
pred. Ya	12	131	91.61%
class recall	0.00%	95.62%	

Gambar 2 *Confusion Matrix* algoritma C4.5

PerformanceVector

```

PerformanceVector:
accuracy: 87.90% +/- 4.04% (mikro: 87.92%)
ConfusionMatrix:
True:  Tidak Ya
Tidak:  0    6
Ya:    12   131
precision: 91.65% +/- 2.52% (mikro: 91.61%) (positive class: Ya)
ConfusionMatrix:
True:  Tidak Ya
Tidak:  0    6
Ya:    12   131
recall: 95.66% +/- 4.77% (mikro: 95.62%) (positive class: Ya)
ConfusionMatrix:
True:  Tidak Ya
Tidak:  0    6
Ya:    12   131
AUC (optimistic): 0.710 +/- 0.171 (mikro: 0.710) (positive class: Ya)
AUC: 0.270 +/- 0.413 (mikro: 0.270) (positive class: Ya)
AUC (pessimistic): 0.263 +/- 0.402 (mikro: 0.263) (positive class: Ya)
    
```

Gambar 3. Text view Confusion Matrix algoritma C4.5

Dari hasil *confusion matrix* diatas, selanjutnya dilakukan perhitungan nilai *accuracy*, *precision*, dan *recall*. Perbandingan nilai *accuracy*, *precision*, dan *recall* yang telah dihitung untuk algoritma C4.5 dapat dilihat pada Tabel 1.

Tabel 1 Perbandingan Nilai *Accuracy*, *Precision*, dan *Recall*

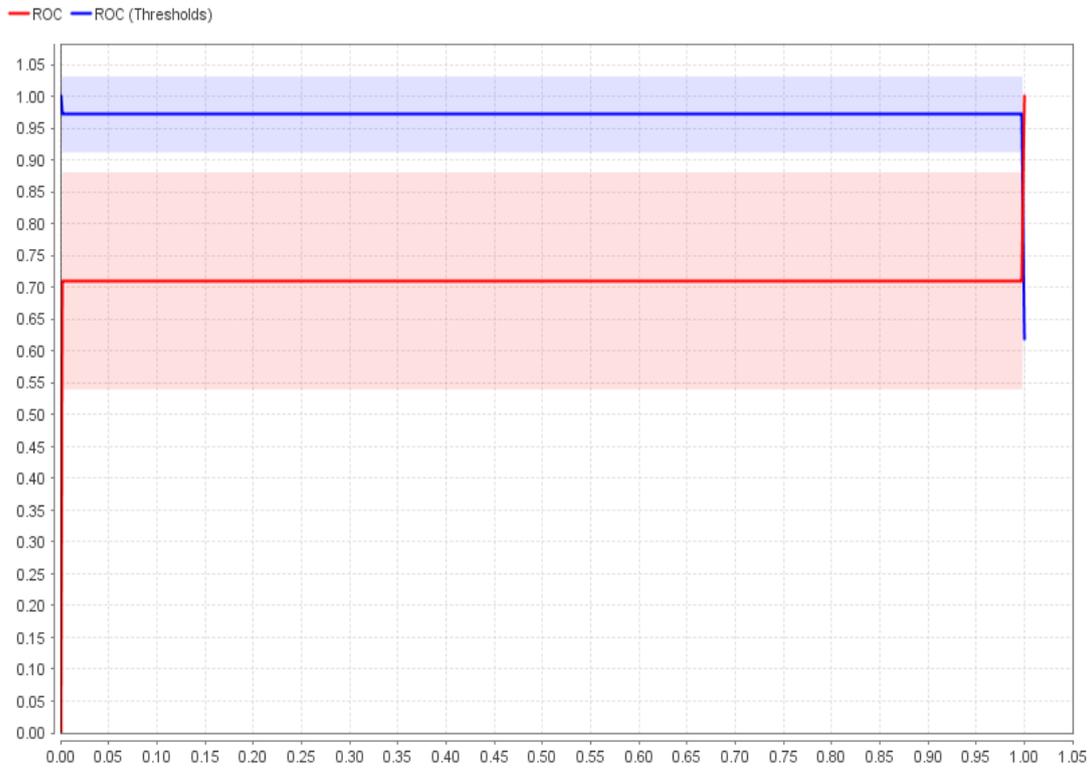
	C4.5
	Training
<i>Accuracy</i>	87.90%
<i>Precision</i>	91.65%
<i>Recall</i>	95.66%

Hasil perhitungan divisualisasikan dengan kurva ROC (*Receiver Operating Characteristic*) atau AUC (*Area Under Curve*). ROC memiliki tingkat nilai diagnosa yaitu (Gorunescu, 2011):

- Accuracy* bernilai 0.90 – 1.00 = *excellent classification*
- Accuracy* bernilai 0.80 – 0.90 = *good classification*
- Accuracy* bernilai 0.70 – 0.80 = *fair classification*
- Accuracy* bernilai 0.60 – 0.70 = *poor classification*
- Accuracy* bernilai 0.50 – 0.60 = *failure*

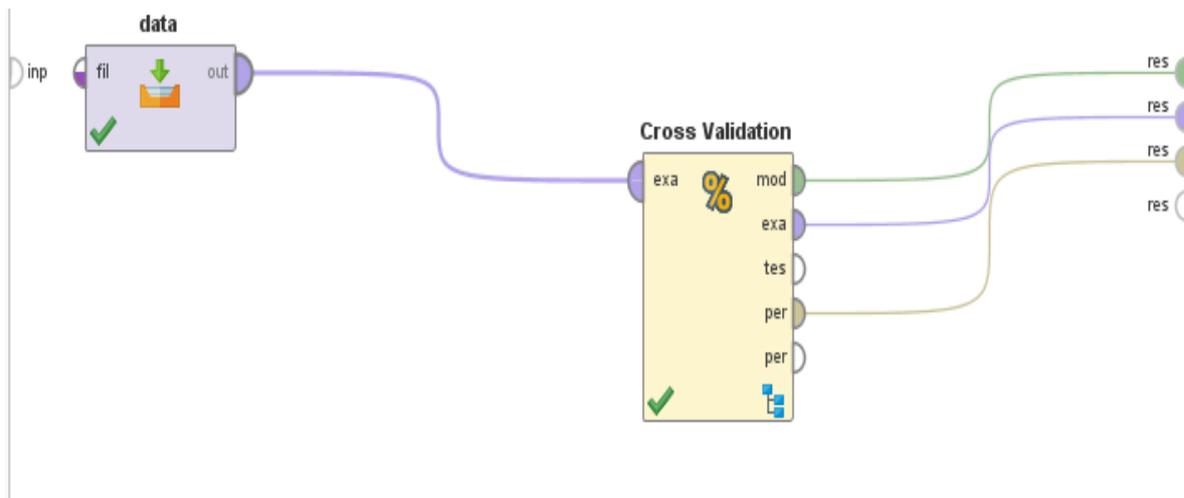
Hasil yang didapat dari pengolahan ROC menggunakan data *training* untuk algoritma C4.5 sebesar 0.710 dapat dilihat pada gambar 3 dengan tingkat diagnosa *fair classification*.

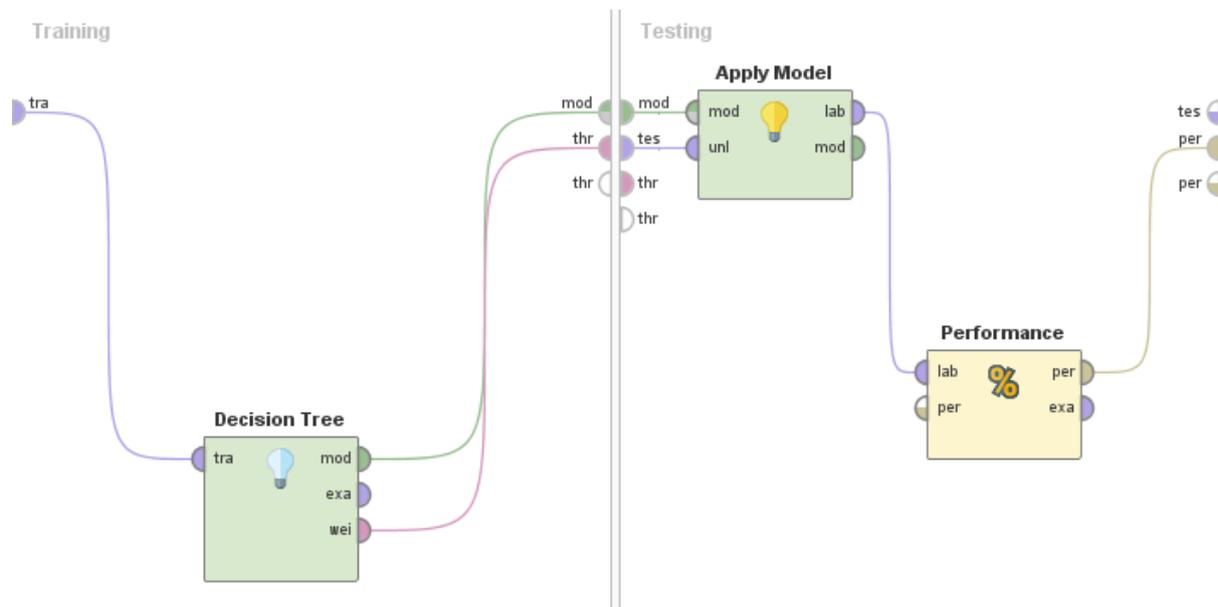
AUC (optimistic): 0.710 +/- 0.171 (mikro: 0.710) (positive class: Ya)



Gambar 4. Kurva ROC data *training* untuk metode C4.5

Pembentukan rule C4.5 dengan menggunakan *RapidMiner* terdapat pada gambar 4 dibawah ini:





Gambar 5. Pemodelan Proses Algoritma C4.5 dengan RapidMiner

IV. KESIMPULAN

Kesimpulan

Berdasarkan hasil uraian dari penelitian dan pembahasan yang telah dilakukan pada sebelumnya dapat diambil kesimpulan, antara lain:

1. Data yang dapat diambil hanya file yang berekstensi *.csv (*Comma Delimited*).
2. Penelitian menggunakan dataset yang dibentuk dari informasi yang dihasilkan pada data penjualan vitamin di sebuah apotek.
3. *Accuracy* dari algoritma klasifikasi C4.5 menghasilkan nilai sebesar 87.90% menggunakan data testing sebanyak 149 data.

Saran

Mengingat keterbatasan baik data maupun pengetahuan tentang algoritma-algoritma yang dimiliki oleh penulis, maka penulis dapat memberikan beberapa gambaran sebagai saran yang dapat dipakai sebagai acuan dalam pengembangan penelitian ini di penelitian yang ingin dilakukan pada masa yang akan datang, antara lain:

- a. Algoritma C4.5 ini dapat diterapkan di objek-objek data set lainnya, baik dataset yang dikelola oleh pemerintah, BUMN ataupun Swasta agar masyarakat dapat mengetahui berita berdasarkan data yang ada bukan lagi dari prediksi orang perorang yang mengakibatkan banyak berita *HOAX*.
- b. Untuk objek penjualan vitamin di sebuah apotek ini bisa menerapkan beberapa metode yang ada pada data mining dan algoritma yang ada pada beberapa metode yang ada pada data mining, contoh nya menggunakan algoritma apriori yaitu mempelajari pola dari pembelian vitamin dengan produk lainnya sehingga penempatan obat/vitamin bisa lebih maksimal dalam penjualan.
- c. Setelah *accuracy* nilai diketahui bisa membuat aplikasi nya agar bisa diterapkan di objek yang ditujukan, dengan beberapa aplikasi yang ada dalam penyelesaiannya.

REFERENSI

- [1] Kotler Philip, Amstronng Gary. (2013). “Prinsip-prinsip Pemasaran, Edisi ke-12.” Erlangga: Jakarta.
- [2] Larose, Daniel T. (2005). “Discovering Knowledge in Data: An Introduction to Data Mining.” John Willey & Sons, Inc.
- [3] Reza Avrizal, Arief Wibowo, Angger Styo Yuniarti, Deassy Ari Sandy, dan Kamal Prihandani. Analysis Comparison of The Classification Data Mining Method to Predictthe Decisions of Potential Customer Insurance.
- [4] International Journal of Computer Techniques ,ISSN: 2394-2231 Volume 5 Issue 5, Sep- Oct 2018.
- [5] E. T. L. Kusriani, Algoritma Data Mining. Yogyakarta: Penerbit Andi, 2009.
- [6] W. K. V. Xindong, The Top Ten Algorithms in Data Mining, vol. XXXIII, no. 2. USA: Taylor & Francis Group, LLC, 2009
- [7] Davies, et al., 2004.
- [8] Dunham, 2003