**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 must provide drugs with various versions, either standard or exclusive. Problems that often occur either hospitals or pharmacies such as in offering these medicinal products which are especially vitamins for cross selling with products commonly purchased by clients. The research that the author is doing aims to implement several algorithms that exist in data mining, but the authors focus more on the C4.5 algorithm which serves to predict the sale of drugs, especially vitamins, where the drug is used or not consumed by consumers so that the accuracy results can be obtained used to provide information on which products are DUE or NOT. The results showed that trees produced from software made using the decision tree method with the C4.5 algorithm had a percentage of 99.83% between the truth trees to 100%. The large percentage of truth trees 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 exclusive. Permasalahan yang sering terjadi baik rumah sakit ataupun apotek seperti dalam menawarkan produk obat tersebut yang khususnya vitamin untuk di cross selling dengan produk yang biasa di beli oleh klien. Penelitian yang penulis lakukan bertujuan untuk mengimplementasikan beberapa algoritma yang ada pada data mining, namun penulis lebih fokuskan pada algoritma C4.5 yang berfungsi untuk memprediksi 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 99.83% antara pohon kebenaran sampai 100%. 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 di gudang, namun untuk vitamin yang diminati sama sekali tidak ada dalam persedian 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]. Definis 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.

*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.*

*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].*

 *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:*

 *(1)*

*Di mana:*

*S = Himpunan Kasus n = Jumlah partisi S*

*Pi = Proporsi Si terhadap S*

*4. Kemudian hitung nilai gain yang menggunakan rumus:*

 *(2)*

*Di mana:*

*S = Himpunan Kasus*

*A = Fitur*

*n = Jumlah Partisi Atribut A*

*|Si| = Proporsi Si 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.*

*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].*

***Rapidminner****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 Rapidminner dapat di lihat Gambar di bawah ini:*

**

*Gambar 1. Rapidminner*

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 rapidminner yang dibuat model nya dahulu lalu menghasilkan nilai Nilai yang dapat di lihat dari Rapidminner dapat dijelaskan pada di bawah ini:

*Confusion Matrix* algoritma C4.5

 Gambar 1 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.



Gambar 2 *Confusion Matrix* algoritma C4.5



Gambar 2 *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.

|  |  |
| --- | --- |
|   | **C4.5** |
| **Training** |
| *Accuracy* | 87.90% |
| *Precision* | 91.65% |
| *Recall* | 95.66% |

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

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

1. Accuracy bernilai 0.90 – 1.00 = *excellent classification*
2. Accuracy bernilai 0.80 – 0.90 = *good classification*
3. Accuracy bernilai 0.70 – 0.80 = *fair classification*
4. Accuracy bernilai 0.60 – 0.70 = *poor classification*
5. 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.*



Gambar 3 Kurva ROC data *training* untuk metode C4.5

Pembentukan rule C4.5 dengan menggunakan RapidMiner terdapat pada gambar 4 di bawah ini:





Gambar 4. Pemodelan Proses Algoritma C4.5 dengan RapidMiner

# IV. SIMPULAN DAN SARAN

Kesimpulan

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

a. Data yang dapat diambil hanya file yang berekstensi \*.csv ( Comma Delimited).

b. Penelitian menggunakan dataset yang dibentuk dari informasi yang dihasilkan pada data penjualan vitamin di sebuah

 apotek.

c. 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:

1. 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.
2. 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 .
3. Setalah accuracy nilai di ketahui bisa membuat aplikasi nya agar bisa di terapkan di objek yang ditujukan, dengan beberapa aplikasi yang ada dalam penyelesaiiannya.

# REFERENSI

1. Kotler Philip, Amstrong 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.

 International Journal of Computer Techniques ,ISSN: 2394-2231Volume 5 Issue 5, Sep- Oct 2018.

1. E. T. L. Kusrini, Algoritma Data Mining. Yogyakarta: Penerbit Andi, 2009.
2. W. K. V. Xindong, The Top Ten Algorithms in Data Mining, vol. XXXIII, no. 2. USA: Taylor & Francis Group, LLC, 2009