

Penerapan Metode C4.5 Berbasis Particle Swarm Optimization Untuk Memprediksi Penjualan Obat Pada Apotek Bunda Azka

Desi Lestari & Muhammad Nasir

Fakultas Teknik Ilmu Komputer, Universitas Bina Darma

Email: desilestari041298@gmail.com

Abstract

The application of the C4.5 Algorithm based on Particle Swarm Optimization to classify the level of sales of drugs that are often sold at the Bunda Azka Pharmacy, is a strategic thing to reduce the problems experienced by the pharmacy. Classify the level of sales of drugs sold using the C4.5 method. based on particle swarm optimization, to find out whether the C4.5 method based on particle swarm optimization (PSO) can optimize drug sales in the future. This research method uses a descriptive method, namely by conducting case study research by studying activities in the field, observing and interviewing stakeholders. in the initial step of this research is the determination of the attributes that will be processed into data mining with the help of rapidminer tools, this study the author uses the KDD model as a standardization in the data mining process. at the pharmacy. The data will later be processed using the c4.5 algorithm based on Particle Swarm Optimization to find the accuracy results of the prediction of the data. The data sample used is the number of 65 drug transaction records at the Bunda Azka Pharmacy. In the test results, the accuracy of Particle Swarm Optimization was 78.10%, for class recall drug sales was 72.50% and after using Particle Swarm Optimization increased to 78.33%, while precision had an accuracy of 77.92% and after using Particle Swarm Optimization increased to 80.33%. From the results of testing with Particle Swarm Optimization, there is an increase in accuracy of 7.15% from the research application of the C4.5 Method Based on Particle Swarm Optimization to Predict Drug Sales at Bunda Azka Pharmacy.

Keywords : Sales, Bunda Azka Pharmacy, C4.5 Algorithm, PSO

1. Pendahuluan

Penjualan adalah aktivitas atau bisnis dalam menjual produk atau jasa. Aktivitas Penjualan adalah hal yang sangat penting bagi suatu perusahaan, terutama untuk meraih keuntungan. Adapun pengertian penjualan secara umum adalah kegiatan jual beli yang dijalankan oleh kedua belah pihak atau lebih dengan alat pembayaran yang sah. Dalam dunia usaha itu penjualan merupakan salah satu indikator paling penting dalam sebuah perusahaan, bila tingkat penjualan yang dihasilkan oleh perusahaan tersebut besar, maka laba yang dihasilkan perusahaan itu akan besar pula sehingga perusahaan dapat bertahan dalam persaingan bisnis dan bisa mengembangkan usahanya. Apotek Bunda Azka adalah apotek pertama di kecamatan Peninjauan OKU, Apotek Bunda Azka bergerak di bidang Penjualan Obat-obatan. pembeli langsung datang ke Apotek untuk memesan obat yang diperlukan dan melakukan pembayaran secara tunai. Permasalahan yang ada pada Apotek Bunda Azka ini yaitu Apotek tersebut belum bisa mengelompokkan Penjualan Obat-obatan yang sering terjual dan permasalahan tersebut akan diklasifikasikan melalui barangbarang yang sering terjual berdasarkan data dari tahun 2017-2019 setelah itu akan diterapkan pada rincian Penjualan Obat di Apotek Bunda Azka. Tujuan penerapan Algoritma particle swarm Optimization (PSO) untuk mengklasifikasikan obat yang sering terjual. C4.5 Merupakan salah satu dari Decision Tree menyerupai struktur flowchart, yang masing-masing internal nodenya dinyatakan sebagai atribut pengujian, setiap cabang mewakili output dari pengujian, dan setiap node daun (terminal node) menentukan label class. Node paling atas dari sebuah pohon adalah node akar. Salah satu metode klasifikasi yang

menarik melibatkan konstruksi pohon keputusan, koleksi node keputusan, terhubung oleh cabang-cabang, memperpanjang bawah dari simpul akar sampai berakhir di node daun. Dimulai di node root, yang oleh konvensi ditempatkan di bagian atas dari diagram pohon keputusan, atribut diuji pada node keputusan, dengan setiap hasil yang mungkin dihasilkan dalam suatu cabang. Setiap cabang kemudian mengarah baik ke node lain keputusan atau ke node daun untuk mengakhiri. Algoritma Particle Swarm Optimization (PSO) diperkenalkan oleh Kennedy dan Eberhart pada tahun 1995, proses algoritmanya diinspirasi oleh perilaku sosial dari binatang, seperti sekumpulan burung dalam suatu swarm. Particle Swarm Optimization (PSO) adalah salah satu dari teknik komputasi evolusioner, yang mana populasi pada PSO didasarkan pada penelusuran algoritma dan diawali dengan suatu populasi yang random yang disebut dengan particle. Berbeda dengan teknik komputasi evolusionerlainnya, setiap particle di dalam PSO juga berhubungan dengan suatu velocity. Particle-particle tersebut bergerak melalui penelusuran ruang dengan velocity yang dinamis yang disesuaikan menurut perilaku historisnya. Oleh karena itu, particle-particle mempunyai kecenderungan untuk bergerak ke area penelusuran yang lebih baik setelah melewati proses penelusuran. Penerapan Algoritma C4.5 berbasis Particle Swarm Optimization untuk klasifikasi tingkat penjualan Obat yang sering terjual pada Apotek Bunda Azka, merupakan hal yang strategis untuk mengurangi masalah yang di alami oleh Apotek tersebut. Dari uraian di atas maka penulis mengangkat judul untuk Tugas Akhir yaitu, ” Penerapan Metode dan C4.5 Berbasis Particle Swarm Optimization Untuk Memprediksi Penjualan Obat Pada Apotek Bunda Azka ”.

2. Tinjauan Literatur

Algoritma *Particle Swarm Optimization* (PSO) diperkenalkan oleh Kennedy dan Eberhart pada tahun 1995, proses algoritmanya diinspirasi oleh perilaku sosial dari binatang, seperti sekumpulan burung dalam suatu swarm.

Particle Swarm Optimization (PSO) adalah salahsatu dari teknik komputasi *evolusioner*, yang mana populasi pada PSO didasarkan pada penelusuran algoritma dan diawali dengan suatu populasi yang random yang disebut dengan *particle*. Berbeda dengan teknik komputasi *evolusionerlainnya*, setiap particle di dalam PSO juga berhubungan dengan suatu *velocity*. Particle-particle tersebut bergerak melalui penelusuran ruang dengan *velocity* yang dinamis yang disesuaikan menurut perilaku historisnya. Oleh karena itu, particle-particle mempunyai kecenderungan untuk bergerak ke area penelusuran yang lebih baik setelah melewati proses penelusuran.

Particle Swarm Optimization (PSO) mempunyai kesamaan dengan genetic algorithm yang mana dimulai dengan suatu populasi yang random dalam bentuk matriks. Namun PSO tidak memiliki operatorevolusi yaitu crossover dan mutasi seperti yang ada pada *genetic algorithm*. Baris pada matriks disebut particle atau dalam *genetic algorithm* sebagai kromosom yang terdiri dari nilai suatu variable. Setiap particle berpindah dari posisinya semula keposisi yang lebih baik dengan suatu *velocity* . Pada algoritma PSO vektor *velocity* diupdate untuk masing-masing particle kemudian menjumlahkan vektor velocity tersebut ke posisi particle. *Update velocity* dipengaruhi oleh kedua solusi yaitu *global best* yang berhubungan dengan biaya yang paling rendah yang pernah diperoleh dari suatu particle dan solusi *local best* yang berhubungan dengan biaya yang paling rendah pada populasi awal. Jika solusi *local best* mempunyai suatu biaya yang kurang dari biaya solusi global yang ada, maka solusi *local best* menggantikan solusi *global best*. Kesederhanaan algoritma dan performansinya yang baik, menjadikan PSO telah menarik

banyak perhatian di kalangan para peneliti dan telah di aplikasikan dalam berbagai persoalan optimisasi sistem tenaga seperti economic dispatch, design kontrol PID pada sistem AVR, kontrol tegangan dan daya reaktif, *unit commitment* dan lain sebagainya. PSO telah populer menjadi optimisasi global dengan sebagian besar permasalahan dapat diselesaikan dengan baik di mana variabel-variabelnya adalah bilangan riil.

Algoritma PSO meniru perilaku sosial organisme ini. Perilaku sosial terdiri dari tindakan individu dan pengaruh dari individu-individu lain dalam suatu kelompok. Kata partikel menunjukkan, misalnya, seekor burung dalam kawanan burung. Setiap individu atau partikel berperilaku secara terdistribusi dengan cara menggunakan kecerdasannya (intelligence) sendiri dan juga dipengaruhi perilaku kelompok kolektifnya. Dengan demikian, jika satu partikel atau seekor burung menemukan jalan yang tepat atau pendek menuju ke sumber makanan, sisa kelompok yang lain juga akan dapat segera mengikuti jalan tersebut meskipun lokasi mereka jauh di kelompok tersebut.

Metode optimasi yang didasarkan pada swarm intelligence ini disebut algoritma behaviorally inspired sebagai alternatif dari algoritma genetika, yang sering disebut evolution-based procedures. Algoritma PSO ini awalnya diusulkan oleh. Dalam konteks optimasi multivariabel, kawanan diasumsikan mempunyai ukuran tertentu atau tetap dengan setiap partikel posisi awalnya terletak di suatu lokasi yang acak dalam ruang multidimensi. Setiap partikel diasumsikan memiliki dua karakteristik: posisi dan kecepatan. Setiap partikel bergerak dalam ruang/space tertentu dan mengingat posisi terbaik yang pernah dilalui atau ditemukan terhadap sumber makanan atau nilai fungsi objektif. Setiap partikel menyampaikan informasi atau posisi bagusnya kepada partikel yang lain dan menyesuaikan posisi dan kecepatan masing-masing berdasarkan informasi yang diterima mengenai posisi yang bagus.

Sebagai pembuktian, misalnya perilaku burung-burung dalam dalam kawanan burung. Meskipun setiap burung mempunyai keterbatasan dalam hal kecerdasan, biasanya ia akan mengikuti kebiasaan (rule) seperti berikut :

1. Seekor burung tidak berada terlalu dekat dengan burung yang lain
2. Burung tersebut akan mengarahkan terbangnya ke arah rata-rata keseluruhan burung

Akan memposisikan diri dengan rata-rata posisi burung yang lain dengan menjaga sehingga jarak antar burung dalam kawanan itu tidak terlalu jauh.

Dengan demikian perilaku kawanan burung akan didasarkan pada kombinasi dari 3 faktor simpel berikut:

1. Kohesi - terbang bersama
2. Separasi - jangan terlalu dekat
3. Penyesuaian(alignment) - mengikuti arah bersama

Jadi PSO dikembangkan dengan berdasarkan pada model berikut:

1. Ketika seekor burung mendekati target atau makanan (atau bisa minimum atau maximum suatu fungsi tujuan) secara cepat mengirim informasi kepada burung-burung yang lain dalam kawanan tertentu

2. Burung yang lain akan mengikuti arah menuju ke makanan tetapi tidak secara langsung
3. Ada komponen yang tergantung pada pikiran setiap burung, yaitu memorinya tentang apa yang sudah dilewati pada waktu sebelumnya (Juwardi and Andilala, 2012).

2.1 Algoritma C4.5

Algoritma C4.5 merupakan salah satu dari *Decision Tree* menyerupai struktur *flowchart*, yang masing-masing internal *node*-nya dinyatakan sebagai atribut pengujian, setiap cabang mewakili output dari pengujian, dan setiap *node* daun (*terminal node*) menentukan label *class*. *Node* paling atas dari sebuah pohon adalah *node* akar. Salah satu metode klasifikasi yang menarik melibatkan konstruksi pohon keputusan, koleksi *node* keputusan, terhubung oleh cabang-cabang, memperpanjang bawah dari simpul akar sampai berakhir di *node* daun. Dimulai di *node root*, yang oleh konvensi ditempatkan di bagian atas dari diagram pohon keputusan, atribut diuji pada *node* keputusan, dengan setiap hasil yang mungkin dihasilkan dalam suatu cabang. Setiap cabang kemudian mengarah baik ke *node* lain keputusan atau ke *node* daun untuk mengakhiri (Author K.M, 2009).

Beberapa tahap dalam membuat sebuah pohon keputusan dengan algoritma C4.5, yaitu:

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. Sebelum menghitung gain dari atribut, hitung dahulu nilai entropy yaitu:

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

Keterangan:

S : himpunan kasus A : atribut

N : jumlah partisi S

Pi : proporsi dari Si terhadap S

1. Kemudian hitung nilai gain dengan metode *informasi gain*:

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

Ulangi langkah ke-2

2. Proses partisi pohon keputusan akan berhenti saat:

- a. Semua tupel dalam node N mendapat kelas yang sama.
- b. Tidak ada atribut di dalam tupel yang dipartisi lagi.
- c. Tidak ada tupel di dalam cabang yang kosong.

Langkah-langkah diatas digunakan untuk menangani atribut nominal. Perhitungan dengan metode *Entropy Based Discretization* di gunakan untuk menangani atribut yang bersifat kontinu. Metode ini menggunakan entropy sebagai bagian dari proses pemisahan selang data selang kontinu.

2.2 Rapidminer

RapidMiner adalah aplikasi data mining berbasis open-source yang terkemuka dan ternama. Didalamnya terdapat aplikasi yang berdiri sendiri untuk analisis data dan sebagai mesin data mining seperti untuk loading data, transformasi data, pemodelan data, dan metode visualisasi data. *RapidMiner* pertama kali dinamai Yet Another Learning Environment atau singkat YALE. Pada tahun 2007 akhirnya diganti namanya menjadi *RapidMiner*. *Rapidminer* merupakan perangkat lunak yang bersifat terbuka (*open source*). *Rapidminer* adalah sebuah solusi untuk melakukan analisis terhadap data mining, text mining dan analisis prediksi. *Rapidminer* menggunakan berbagai teknik deskriptif dan prediksi dalam memberikan wawasan kepada pengguna sehingga dapat membuat keputusan yang paling baik.

Rapidminer memiliki kurang lebih 500 operator data mining, termasuk operator untuk input, output, data *preprocessing* dan visualisasi. *Rapidminer* merupakan software yang berdiri sendiri untuk analisis data dan sebagai mesin data mining yang dapat diintegrasikan pada produknya sendiri. *Rapidminer* ditulis dengan menggunakan bahasa Java sehingga dapat bekerja di semua sistem operasi. *RapidMiner* menyediakan *GUI* (*Graphic User Interface*) untuk merancang sebuah pipeline analitis. *GUI* ini akan menghasilkan file *XML* (*Extensible Markup Language*) yang mendefinisikan proses analitis keinginan pengguna untuk diterapkan ke data. File ini kemudian dibaca oleh *RapidMiner* untuk menjalankan analisis secara otomatis (Rika Nofitri, 2019).

Dari hasil analisis optimasi model algoritma C4.5 berbasis PSO memberikan nilai akurasi yang lebih tinggi yaitu 78.16% dibandingkan dengan model algoritma C4.5 yaitu 73.88%. Dari hasil tersebut didapatkan selisih antara kedua model yaitu 4,28%. Sementara untuk evaluasi menggunakan ROC curve untuk kedua model yaitu, untuk model algoritma C4.5 nilai AUC adalah 0.764 dengan tingkat diagnosa Fair classification, dan untuk model algoritma C4.5 berbasis PSO nilai AUC adalah 0.780 dengan tingkat diagnosa Fair classification. Dari evaluasi ROC curve tersebut terlihat bahwa model algoritma C4.5 berbasis PSO lebih besar. Sehingga dapat disimpulkan bahwa algoritma C4.5 berbasis particle swarm optimization lebih akurat dalam memprediksi minat beli produk sepatu (Ari Puspita, Jefa, 2019).

Pada penelitian ini dilakukan pemodelan menggunakan algoritma C4.5 dan C4.5 yang dikombinasi dengan PSO, data yang digunakan adalah data nasabah bank yang mana pada tujuan ini untuk mengetahui nasabah mana yang nantinya berpotensi melakukan deposito. Penelitian ini difokuskan pada penerapan algoritma PSO sebagai pembobotan atribut teknik klasifikasi data mining C4.5. Validasi model menggunakan 10fold cross-validation dan evaluasi model menggunakan confusion matrix dan kurva ROC. Dari penelitian ini didapat algoritma C4.5 yang dikombinasi dengan algoritma PSO

mempunyai akurasi yang lebih baik dibanding penggunaan algoritma C4.5 saja, yaitu 88.83% berbanding 89.26% akan tetapi untuk waktu eksekusi algoritma C4.5 yang dikombinasi PSO memakan waktu lebih lama yaitu 6 menit 44 detik untuk C4.5 dan menjadi 4 jam 48 menit 44 detik untuk C4.5 berbasis PSO (Syaeful Mujab, 2016).

3. Metode Penelitian

Metode Penelitian ini menggunakan metode deskriptif yakni dengan cara melakukan penelitian studi kasus dengan mempelajari aktivitas di lapangan, mengamati dan wawancara kepada pihak yang terkait. Metode deskriptif adalah suatu metode penelitian yang ditunjukkan untuk membuat gambaran atau lukisan secara sistematis, aktual, dan akurat melalui data sampel atau populasi sebagaimana adanya.

3.1 Data Mining

Data mining adalah proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu. Teknik-teknik, metode-metode, atau algoritma dalam data mining sangat bervariasi. Pemilihan metode atau algoritma yang tepat sangat bergantung pada tujuan dan proses Knowledge Discovery in Database (KDD) secara keseluruhan. Data Mining merupakan teknologi baru yang sangat berguna untuk membantu menemukan informasi yang sangat penting dari data mereka. Data mining memperkenalkan tren dan sifat-sifat perilaku bisnis yang sangat berguna untuk mendukung pengambilan keputusan penting. Analisis yang Komponen data mining pada proses KDD seringkali merupakan aplikasi iteratif yang berulang dari metodologi data mining tertentu. Pada pembahasan di sini akan digunakan istilah pola dan model. Pola dapat diartikan sebagai instansiasi dari model.

3.2 Metode (PSO) Particle Swarm Optimazition

Algoritma Particle Swarm Optimization (PSO) diperkenalkan oleh Kennedy dan Eberhart pada tahun 1995, proses algoritmanya diinspirasi oleh perilaku sosial dari binatang, seperti sekumpulan burung dalam suatu swarm. Particle Swarm Optimization (PSO) adalah salahsatu dari teknik komputasi evolusioner, yang mana populasi pada PSO didasarkan pada penelusuran algoritma dan diawali dengan suatu populasi yang random yang disebut dengan particle. Berbeda dengan teknik komputasi evolusionerlainnya, setiap particle di dalam PSO juga berhubungan dengan suatu velocity. Particle Swarm Optimization (PSO) mempunyai kesamaan dengan genetic algorithm yang mana dimulai dengan suatu populasi yang random dalam bentuk matriks. Pada algoritma PSO vektor velocity diupdate untuk masingmasing particle kemudian menjumlahkan vektor velocity tersebut ke posisi particle. Update velocity dipengaruhi oleh kedua solusi yaitu global best yang berhubungan dengan biaya yang paling rendah yang pernah diperoleh dari suatu particle dan solusi local best yang berhubungan dengan biaya yang paling rendah pada populasi awal. Jika solusi local best mempunyai suatu biaya yang kurang dari biaya solusi global yang ada, maka solusi local best menggantikan solusi global best. Kesederhanaan algoritma dan performansinya yang baik, menjadikan PSO telah menarik banyak perhatian di kalangan para peneliti dan telah di aplikasikan dalam berbagai persoalan optimisasi sistem tenaga seperti economic dispatch, design kontrol PID pada sistem AVR, kontrol tegangan dan daya reaktif, unit commitment dan lain sebagainya.

2.3 Algoritma C4.5

Algoritma C4.5 merupakan salah satu dari Desicion Tree menyerupai struktur flowchart,

yang masing-masing internal node-nya dinyatakan sebagai atribut pengujian, setiap cabang mewakili output dari pengujian, dan setiap node daun (terminal node) menentukan label class. Node paling atas dari sebuah pohon adalah node akar. Salah satu metode klasifikasi yang menarik melibatkan konstruksi pohon keputusan, koleksi node keputusan, terhubung oleh cabangcabang, memperpanjang bawah dari simpul akar sampai berakhir di node daun. Dimulai di node root, yang oleh konvensi ditempatkan di bagian atas dari diagram pohon keputusan, atribut diuji pada node keputusan, dengan setiap hasil yang mungkin dihasilkan dalam suatu cabang. Setiap cabang kemudian mengarah baik ke node lain keputusan atau ke node daun untuk mengakhiri.

2.4 Rapidminer

RapidMiner adalah aplikasi data mining berbasis open-source yang terkemuka dan ternama. Didalamnya terdapat aplikasi yang berdiri sendiri untuk analisis data dan sebagai mesin data mining seperti untuk loading data, transformasi data, pemodelan data, dan metode visualisasi data. RapidMiner pertama kali dinamai Yet Another Learning Environment atau singkat YALE. Pada tahun 2007 akhirnya diganti namanya menjadi RapidMiner. Rapidminer.

merupakan perangkat lunak yang bersifat terbuka (open source). Rapidminer adalah sebuah solusi untuk melakukan analisis terhadap data mining, text mining dan analisis prediksi. Rapidminer menggunakan berbagai teknik deskriptif dan prediksi dalam memberikan wawasan kepada pengguna sehingga dapat membuat keputusan yang paling baik..

4. Hasil dan Pembahasan

4.1 Analisis Data

Pada tahapan ini merupakan tahapan analisis data dari penelitian yang berjudul Penerapan Metode C4.5 Berbasis Particle Swarm Optimization Untuk Memprediksi Penjualan Obat Pada Apotek Bunda Azka, pada langkah awal penelitian ini adalah penentuan atribut yang akan diolah kedalam data mining dengan bantuan tools rapidminer, pada penelitian ini akan dilakukan penyusunan atribut sebelum diolah kedalam proses minning dengan menggunakan rapidminer.

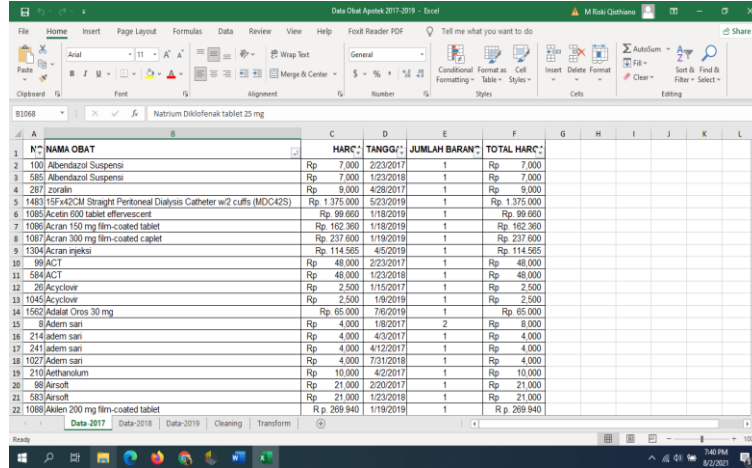
4.2 Desain Penelitian

Dalam penelitian ini penulis menggunakan model KDD sebagai standarisasi dalam proses data mining dengan langkah-langkah sebagai berikut ini:

4.2.1 Data Cleaning

Data Cleaning, tahap pembersihan data merupakan tahap awal dari proses KDD. Seluruh atribut yang ada pada dataset di atas selanjutnya akan diseleksi untuk mendapatkan atribut-atribut yang berisi nilai yang relevan. Tidak redundant dan tidak missing value, dimana syarat tersebut merupakan syarat awal yang harus dikerjakan dalam data mining sehingga akan diperoleh dataset yang bersih untuk digunakan pada tahap mining data. Data dikatakan missing value bila atributatribut dalam dataset tidak berisi nilai atau kosong, sedangkan data dikatakan redundant jika dalam satu dataset yang sama terdapat lebih dari satu record yang berisi nilai yang sama. Data yang berkualitas sangat mempengaruhi kualitas keputusan yang akan diperoleh. Data yang tidak berkualitas adalah data yang nilai atributnya hilang, error dan data yang tidak konsisten dalam pengisian atributnya. Untuk mendapatkan data yang berkualitas dilakukan tahapan dalam data Cleaning sebagai berikut:

1. Menghilangkan record yang tidak lengkap,
2. Menghilangkan data yang atribut ataupun label yang tidak di pakai.
3. Mengisi nilai-nilai yang missing atau tidak lengkap.



NAMA OBAT	HARC	TANGGAL	JUMLAH BARANG	TOTAL HARC
100 Alendazol Suspensi	Rp. 7.000	2/23/2017	1	Rp. 7.000
585 Alendazol Suspensi	Rp. 7.000	1/23/2018	1	Rp. 7.000
2871 ezale	Rp. 9.000	4/28/2017	1	Rp. 9.000
1483 15FxD3M Straight Peritoneal Dialysis Catheter w/2 cuffs (MDC425)	Rp. 1.375.000	5/23/2019	1	Rp. 1.375.000
1085 Acetin 600 tablet effervescent	Rp. 99.660	1/18/2019	1	Rp. 99.660
1086 Acran 150 mg film-coated tablet	Rp. 162.360	1/18/2019	1	Rp. 162.360
1087 Acran 300 mg film-coated caplet	Rp. 237.600	1/18/2019	1	Rp. 237.600
1304 Acran injeksi	Rp. 114.565	4/5/2019	1	Rp. 114.565
99 ACT	Rp. 48.000	2/23/2017	1	Rp. 48.000
11 584 ACT	Rp. 48.000	1/23/2018	1	Rp. 48.000
26 Acyclovir	Rp. 2.500	1/15/2017	1	Rp. 2.500
1045 Acyclovir	Rp. 2.500	1/9/2019	1	Rp. 2.500
14 1562 Adalat Oros 30 mg	Rp. 65.000	7/6/2019	1	Rp. 65.000
8 Adem sari	Rp. 4.000	1/8/2017	2	Rp. 8.000
214 Adem sari	Rp. 4.000	4/3/2017	1	Rp. 4.000
241 Adem sari	Rp. 4.000	4/12/2017	1	Rp. 4.000
1027 Adem sari	Rp. 4.000	7/31/2018	1	Rp. 4.000
210 Arhanolun	Rp. 10.000	4/2/2017	1	Rp. 10.000
88 Arsoft	Rp. 21.000	2/20/2017	1	Rp. 21.000
583 Arsoft	Rp. 21.000	1/23/2018	1	Rp. 21.000
1088 Aklon 200 mg film-coated tablet	Rp. 269.940	1/19/2019	1	Rp. 269.940

Gambar 1. Data Awal setelah *Cleaning*

Data Integration and Transformation

Data Integration atau integrasi data merupakan proses menggabungkan atau menyatukan dua atau lebih sebuah data dari berbagai sumber database yang berbeda ke dalam sebuah penyimpanan seperti gudang data (data warehouse). Teknik yang digunakan untuk menganalisis data kolerasi, atribut yang redundan disebut *integration*,

Tabel 1. Sampel Data setelah Transformasi

No	NamaObat	Harga	Terjual	JenisHarg a	JenisObat
1	Alleron	3000	38	Murah	Non-Antibiotik
2	Allopurinol	8000	72	Murah	Non-Antibiotik
3	Ambroxol	7000	213	Murah	Non-Antibiotik
4	Aminofilin	65000	67	Mahal	Non-Antibiotik
5	Amlodipin	40000	211	Mahal	Non-Antibiotik
6	Amoxilin	100000	370	Mahal	Antibiotik
7	Ampicilin	40000	110	Mahal	Antibiotik
8	Antidepresen	8000	58	Murah	Non-Antibiotik
9	Asam mefemat	8000	66	Murah	Antibiotik
10	Caviplex	4000	214	Murah	Non-Antibiotik
11	Cefazolin	6000	58	Murah	Antibiotik
12	Cetirizen	3000	126	Murah	Antibiotik
13	Cisapride	8000	61	Murah	Non-Antibiotik
14	Diazepam	9000	55	Murah	Non-Antibiotik
15	Entecapir	4000	78	Murah	Non-Antibiotik
16	Eritromisin	70000	112	Mahal	Non-Antibiotik
17	Fenobarbital	10000	22	Normal	Non-Antibiotik
18	Flamar	91000	81	Mahal	Antibiotik
19	Folavit	8000	81	Murah	Non-Antibiotik
20	Gentaviolet	10000	38	Normal	Non-Antibiotik
21	Glimepiride	30000	28	Mahal	Non-Antibiotik
22	Haloperidol	13000	124	Normal	Non-Antibiotik
23	Ibuprofen	29000	33	Mahal	Non-Antibiotik
24	Kloramfenikol	5000	110	Murah	Antibiotik
25	Kuinidin	35000	62	Mahal	Non-Antibiotik
....					
.	..	.	.	.	

sedangkan *transformation* berguna untuk meningkatkan akurasi dan efisiensi algoritma *Algoritma C4.5* memiliki kelebihan memproses data yang bernilai nominal, ordinal maupun kontinyu. Sehingga nilai – nilai setiap atribut yang terdapat pada dataset tidak perlu ditransformasikan.

4.2.3 Data Reduction

Dataset dapat direduksi dengan mengurangi jumlah atribut dan *record* supaya menjadi lebih sedikit tetapi tetap bersifat informatif. Memperoleh representasi dalam bentuk volume data yang telah berkurang jumlahnya namun tetap mendapatkan hasil analisis yang sama. Deskretasi data merupakan bagian dari reduksi data bagian penting untuk data numerik. Berikut adalah keseluruhan atribut sebelum dilakukan *reduction data*. Maka dapat di peroleh parameter data tersebut antara lain sebagai berikut.

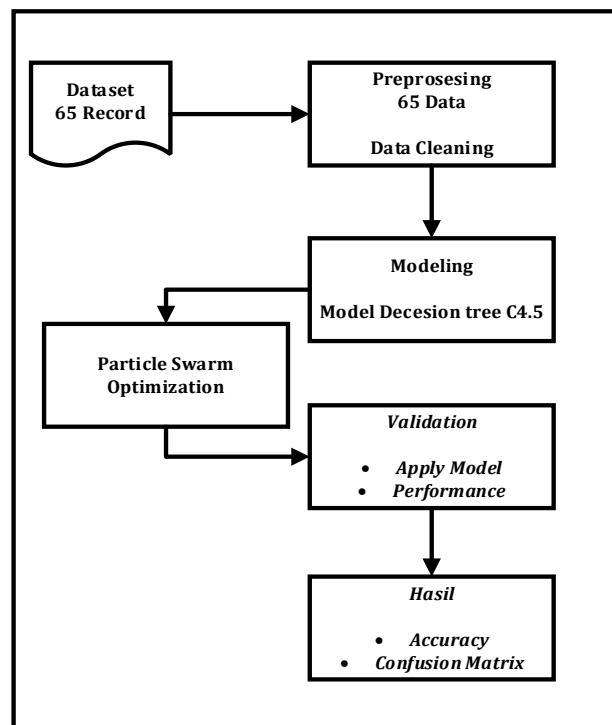
Tabel 2 Parameter Atribut Data

No	Parameter	Keterangan
1	NamaObat	Atribut
2	Harga	Atribut
3	Terjual	Atribut
4	JenisHarga	Atribut
5	JenisObat	Atribut
6	Kategori	Label

Dari hasil pengolahan parameter data diatas maka terpilih lah 5atribut data, serta 1 Label data yang akan di olah menggunakan proses Data Mining.

4.3 Pemodelan

Metode yang akan digunakan pada penelitian adalah Algoritma C4.5. Dalam permodelan ini Algoritma C4.5 akan dicari performanya yaitu *Performance Vector (accuracy)* dan *confusion matrix*. Data yang digunakan sudah melalui proses cleaning data, untuk melakukan pengukuran dalam penelitian ini menggunakan *toolRapidminer*.

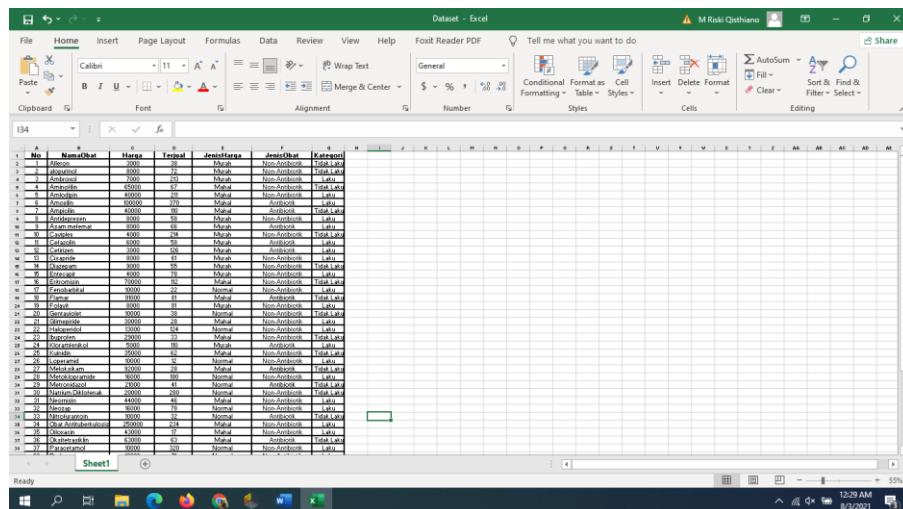


Gambar 2. Modeling

Keterangan:

1. Peneliti melakukan pengumpulan data dan mengelola data penjualan obat tersebut, dengan melakukan proses *selection*, *cleaning* hingga *transform data*.
2. Setelah data siap, peneliti menyiapkan *tools rapidminer* untuk mengelola data penjualan obat untuk mendapatkan hasil akurasi prediksi.
3. Ujicoba pertama peneliti menggunakan alorimat *decision tree C4.5* dengan sampel 50 *record data* penjualan obat.
4. Ujicoba berikutnya peneliti memasukan *particle swarm optimization* untuk meningkatkan akurasi penelitian.
5. Setelah didapat hasil *accruray*, *recall*, dan *precision*, maka didapat hasil akhir dari peneltian.
6. Penutup kesimpulan dan saran.

Dalam pengujian data set pertama dilakukan adalah melakukan proses preprosesing yaitu Data *cleaning*, dari 50 data setelah dilakukan *data cleaning*. Berikut merupakan sampel record data yang akan diolah data mining dengan bantuan *tools rapidminer* yang berjumlah 50 dataset dari data penjualan obat pada apotek bunda azka:



No	NamaObat	Harga	Jumlah	JenisObat	Kategori
1	Albion	1000	25	Khusus	Tidak Laku
2	Albion	1000	25	Khusus	Tidak Laku
3	Albion	1000	25	Khusus	Tidak Laku
4	Albion	1000	25	Khusus	Tidak Laku
5	Albion	1000	25	Khusus	Tidak Laku
6	Albion	1000	25	Khusus	Tidak Laku
7	Albion	1000	25	Khusus	Tidak Laku
8	Albion	1000	25	Khusus	Tidak Laku
9	Albion	1000	25	Khusus	Tidak Laku
10	Albion	1000	25	Khusus	Tidak Laku
11	Albion	1000	25	Khusus	Tidak Laku
12	Albion	1000	25	Khusus	Tidak Laku
13	Albion	1000	25	Khusus	Tidak Laku
14	Albion	1000	25	Khusus	Tidak Laku
15	Albion	1000	25	Khusus	Tidak Laku
16	Albion	1000	25	Khusus	Tidak Laku
17	Albion	1000	25	Khusus	Tidak Laku
18	Albion	1000	25	Khusus	Tidak Laku
19	Albion	1000	25	Khusus	Tidak Laku
20	Albion	1000	25	Khusus	Tidak Laku
21	Albion	1000	25	Khusus	Tidak Laku
22	Albion	1000	25	Khusus	Tidak Laku
23	Albion	1000	25	Khusus	Tidak Laku
24	Albion	1000	25	Khusus	Tidak Laku
25	Albion	1000	25	Khusus	Tidak Laku
26	Albion	1000	25	Khusus	Tidak Laku
27	Albion	1000	25	Khusus	Tidak Laku
28	Albion	1000	25	Khusus	Tidak Laku
29	Albion	1000	25	Khusus	Tidak Laku
30	Albion	1000	25	Khusus	Tidak Laku
31	Albion	1000	25	Khusus	Tidak Laku
32	Albion	1000	25	Khusus	Tidak Laku
33	Albion	1000	25	Khusus	Tidak Laku
34	Albion	1000	25	Khusus	Tidak Laku
35	Albion	1000	25	Khusus	Tidak Laku
36	Albion	1000	25	Khusus	Tidak Laku
37	Albion	1000	25	Khusus	Tidak Laku
38	Albion	1000	25	Khusus	Tidak Laku
39	Albion	1000	25	Khusus	Tidak Laku
40	Albion	1000	25	Khusus	Tidak Laku
41	Albion	1000	25	Khusus	Tidak Laku
42	Albion	1000	25	Khusus	Tidak Laku
43	Albion	1000	25	Khusus	Tidak Laku
44	Albion	1000	25	Khusus	Tidak Laku
45	Albion	1000	25	Khusus	Tidak Laku
46	Albion	1000	25	Khusus	Tidak Laku
47	Albion	1000	25	Khusus	Tidak Laku
48	Albion	1000	25	Khusus	Tidak Laku
49	Albion	1000	25	Khusus	Tidak Laku
50	Albion	1000	25	Khusus	Tidak Laku

Gambar 3. Dataset

4.4 Hasil

Berikut merupakan hasil akurasi dari hasil prediksi terhadap Penerapan Metode C4.5 Berbasis *Particle Swarm Optimization* Untuk Memprediksi Penjualan Obat Pada Apotek Bunda Azkadengan menggunakan *tools Rapidminer*.

accuracy: 70.95% +/- 15.28% (mikro: 70.77%)

	true Tidak Laku	true Laku	class precision
pred. Tidak Laku	23	14	62.16%
pred. Laku	5	23	82.14%
class recall	82.14%	62.16%	

Gambar 4 Hasil Akurasi

Pada hasil diatas terlihat hasil dari akurasi sebesar 70.95% dari 65 *record* data penjualan pada Penerapan Metode C4.5 Berbasis *Particle Swarm Optimization* Untuk Memprediksi Penjualan Obat Pada Apotek Bunda Azka. Berikut juga hasil *plot view* pada *Confusion Matrix*.

accuracy: 78.10% +/- 8.51% (mikro: 78.46%)

	true Tidak Laku	true Laku	class precision
pred. Tidak Laku	23	9	71.88%
pred. Laku	5	28	84.85%
class recall	82.14%	75.68%	

Gambar 5 Hasil Akurasi *Cross Validation* dalam PSO

3.5 Pembahasan

Berikut adalah hasil dari pengujian yang telah dilakukan pada RapidMiner dan *Algoritma C4.5* dengan *Particle Swarm Optimization* sebagai berikut:

1. Pada penelitian ini penulis menggunakan 65 sampel data penjualan obat sebagai data pengujian dari *record excel* pada pengujian dengan menggunakan *software RapidMiner*.
2. Atribut data yang digunakan mulai dari nama obat, harga obat, jenis obat, jenis harga dan label data adalah kategori obat yang merupakan Laku dan Tidak Laku.
3. Untuk mencari performance dari akurasi data digunakan operator *cross validation* yang ada pada *tools rapidminer*.
4. Hasil pengujian di dapat akurasi sebesar 70.95% dan setelah menggunakan *Particle Swarm Optimization* meningkat menjadi 78.10%, untuk *class recall* penjualan obat sebesar 72.50% dan setelah menggunakan *Particle Swarm Optimization* meningkat menjadi 78.33%, sedangkan *precision* memiliki akurasi sebesar 77.92% dan setelah menggunakan *Particle Swarm Optimization* meningkat menjadi 80.33%.
5. Dari hasil pengujian dengan *Particle Swarm Optimization*, ada peningkatan akurasi sebesar 7.15% dari penelitian Penerapan Metode C4.5 Berbasis *Particle Swarm Optimization* Untuk Memprediksi Penjualan Obat Pada Apotek Bunda Azka.

Tabel 3. Hasil Perbandingan

Keterangan	<i>Accuracy</i>	<i>Recall</i>	<i>Precision</i>
Uji Performa Sebelum PSO	70.95%	72.50%	77.92%
Uji Performa dengan PSO	78.10%	78.33%,	80.33%.

Hasil perbandingan diatas menunjukkan ada peningkatan dari hasil akurasi ketika algoritma *Decision Tree C4.5* digabungkan dengan *Particle Swarm Optimization*, dari sebelumnya 70.95% menjadi 78.10%, ada peningkatan akurasi sebesar 7.15% dari penelitian Penerapan Metode C4.5 Berbasis *Particle Swarm Optimization* Untuk Memprediksi Penjualan Obat Pada Apotek Bunda Azka.

5. Kesimpulan

Pada akhir penelitian ini merupakan pembahasan dari kesimpulan dan saran dari hasil peneltian yang berjudul Penerapan Metode C4.5 Berbasis Particle Swarm Optimization Untuk Memprediksi Penjualan Obat Pada Apotek Bunda Azka, setelah melewati hasil dan pembahasan pada bab sebelumnya maka sampailah ditahap akhir penelitian ini, ada beberapa hal yang dapat dijadikan kesimpulan pada penelitian ini, berikut kesimpulannya.

1. Data yang digunakan pada penelitian Penerapan Metode C4.5 Berbasis Particle Swarm Optimization Untuk Memprediksi Penjualan Obat Pada Apotek Bunda Azka merupakan data penjualan obat.
2. Sampel data yang digunakan adalah jumlah 65 record transaksi obat di Apotek Bunda Azka.
3. Pada Hasil pengujian di dapat akurasi sebesar 70.95% dan setelah menggunakan *Particle Swarm Optimization* meningkat menjadi 78.10%, untuk *class recall* penjualan obat sebesar 72.50% dan setelah menggunakan *Particle Swarm Optimization* meningkat menjadi 78.33%, sedangkan *precision* memiliki akurasi sebesar 77.92% dan setelah menggunakan *Particle Swarm Optimization* meningkat menjadi 80.33%. Dari hasil pengujian dengan *Particle Swarm Optimization*, ada peningkatan akurasi sebesar 7.15% dari penelitian Penerapan Metode C4.5 Berbasis *Particle Swarm Optimization* Untuk Memprediksi Penjualan Obat Pada Apotek Bunda Azka

Referensi

- Antoni, D., Herdiansyah, M. I., Akbar, M., & Sumitro, A. (2021). Pengembangan Infrastruktur Jaringan Untuk Meningkatkan Pelayanan Publik di Kota Palembang. *JURNAL MEDIA INFORMATIKA BUDIDARMA*, 5(4), 1652-1659.
- Antoni, D., Jie, F., & Abareshi, A. (2020). Critical factors in information technology capability for enhancing firm's environmental performance: case of Indonesian ICT sector. *International Journal of Agile Systems and Management*, 13(2), 159-181.
- Antoni, D., & Akbar, M. (2019). E-supply chain management value concept for the palm oil industry. *Jurnal Sistem Informasi*, 15(2), 15-29.
- Antoni, D., Fikari, D., & Akbar, M. (2018). The readiness of palm oil industry in enterprise resource planning. *Telkomnika*, 16(6), 2692-2702.
- Fauzi, F., Dencik, A. B., & Asiati, D. I. (2019). Metodologi Penelitian untuk manajemen dan akuntansi. *Jakarta: Salemba Empat*.
- Alhapizi, M. R., Nasir, M., & Effendy, I. (2020). Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Untuk Menentukan Strategi Promosi Mahasiswa Baru Universitas Bina Darma Palembang. *Journal of Software Engineering Ampera*, 1(1), 1-14.

Anisa (2021) 'Sejarah apotek bunda azka', *Bunder-Patuk-Desa.Id*, (434), pp. 1–37. Available at: Visi Misi - Website Bunder (bunder-patuk.desa.id).

Ari puspita dkk (2019) 'Prediksi Peminatan Pelanggan dalam Penjualan Produk Sepatu Menggunakan Metode Decision Tree Berbasis Particle Swarm Optimization pada PT . Baskara Cipta Pratama', V(1), pp. 10–17.

Juhardi dkk (2019) 'Optimalisasi Penjualan Motor Menggunakan Algoritma Particle Swarm Optimization (PSO)', *Jurnal Media Infotama*, 15(2), pp. 71–78.

Mardi, Y. (2017) 'Data Mining : Klasifikasi Menggunakan Algoritma C4.5', *Edik Informatika*, 2(2), pp. 213–219. doi: 10.22202/ei.2016.v2i2.1465.

Moertini, V. (2012) 'Data Mining Sebagai Solusi Bisnis', *Data Mining Sebagai SolusiBisnis*, 7(1), pp. 44–56. Available at: <https://eric.ed.gov/?id=ED539082%0Ahttp://www.win.tue.nl/~mpechen/research/edu.html>

Muarif, A. khoirul (2009) 'Komparasi pemodelan data menggunakan c4.5 dan c4.5 berbasis particle swarm optimization untuk memprediksi kelulusan mahasiswa'.

Mubarokh, M. F., Nasir, M., & Komalasari, D. (2020). Jaringan Syaraf Tiruan Untuk Memprediksi Penjualan Pakaian Menggunakan Algoritma Backpropagation. *Journal of Computer and Information Systems Ampera*, 1(1), 29-43.

Rica Novitri, N. I. (2019) 'INTEGRASI METODE NEIVE BAYES DAN SOFTWARE RAPIDMINER DALAM ANALISIS HASIL USAHA PERUSAHAAN DAGANG Rika Nofitri , Novica Irawati Sekolah Tinggi Manajemen Infromatika dan Komputer Royal PENDAHULUAN Keuntungan menjadi target utama di sebuah Setiap perusahaan', VI(1), pp. 35–42.

Syaeful Mujab (2016) 'PENCARIAN MODEL TERBAIK ANTARA ALGORITMA C4.5 DAN C4.5 BERBASIS PARTICLE SWARM OPTIMIZATION UNTUK PREDIKSI PROMOSI DEPOSITO'.

Tanjung Dkk (2016) 'Pengaruh Penggunaan Metode Pembelajaran Bermain Terhadap Hasil Belajar Matematika Siswa Materi Pokok Pecahan Di Kelas Iii SD Negeri 200407 Hutapadang', *Bina Gogik: Jurnal Ilmiah Pendidikan GuruSekolahDasar*, 3(1), p. 39. Available at: <http://www.ejournal.stkipbbm.ac.id/index.php/pgsd/article/view/26>.

Windarto, kiki fatmawati agus perdana (2018) 'DATA MINING : PENERAPAN RAPIDMINER DENGAN K-MEANS CLUSTER PADA DAERAH TERJANGKIT DEMAM BERDARAH DENGUE (DBD) BERDASARKAN PROVINSI', 3(2), pp. 173–178.

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>)