

IMPLEMENTASI DATA MINING UNTUK MEMPREDIKSI BARANG RUSAK DENGAN MENGGUNAKAN ALGORITMA C4.5 PADA PERUSAHAAN PT HOME CENTER INDONESIA

Ismasari Nawangsih¹⁾, Agus Setiawan²⁾

Program Studi Teknik Informatika Sekolah Tinggi Teknologi Pelita Bangsa
ismasari.n@pelitabangsa.ac.id

Disetujui, 25 Maret 2019

Abstraksi

Pemanfaatan data yang ada di dalam sistem informasi untuk menunjang kegiatan pengambilan keputusan, tidak cukup hanya mengandalkan data operasional saja, diperlukan suatu analisis data untuk menggali potensi-potensi informasi yang ada. Para pengambil keputusan berusaha untuk memanfaatkan gudang data yang sudah dimiliki untuk menggali informasi yang berguna untuk membantu mengambil keputusan, hal ini mendorong munculnya cabang ilmu baru untuk mengatasi masalah penggalian informasi atau pola yang penting atau menarik dari data dalam jumlah besar, yang disebut dengan data mining. Dalam kasus ini telah terjadi kerusakan barang di PT Home Center Indonesia yang diakibatkan tidak jalannya sistem fifo (first in first out). Untuk menghadapi masalah ini, perusahaan harus mengambil keputusan dengan tepat dalam menetapkan strategi produk yang dijual. Dan mendapatkan keputusan yang tepat diperlukan data barang yang cukup untuk dianalisa. Tujuan dari analisis data ini adalah sebagai sarana untuk membentuk suatu keputusan agar dapat memprediksi barang rusak yang ada pada perusahaan dengan menggunakan metode algoritma C4.5 yang mana saat ini sedang banyak digunakan untuk analisa. Dalam perhitungannya, data mining dengan menggunakan algoritma C4.5 ini dapat menghasilkan nilai ukur yang dapat memprediksi barang yang riskan cacat atau rusak dan mengimplementasikannya menggunakan aplikasi RapidMiner Karena selama ini, barang yang rusak tidak diketahui sebelumnya dan hasilnya menyebabkan customer return akibat menerima barang yang rusa.

Kata Kunci: *Data Barang, Data Mining, System Fifo (first in first out), Algoritma C4.5, RapidMiner*

Abstract

The use of data in the information system to support decision-making activities is not enough to rely solely on operational data, it requires a data analysis to explore the potential of existing information. Decision makers try to utilize the data warehouse that they already have to dig up useful information to help make decisions, this encourages the emergence of new branches of science to overcome the problem of extracting information or important patterns or interesting from large amounts of data, which is called data mining. The damage to the goods at PT Home Center Indonesia was also caused by not having a fifo system (first in first out). To overcome this problem the company must make the right decision in determining the strategy of the product being sold. And getting the right decision requires sufficient item data to be analyzed. The purpose of this data analysis is as a means to form a decision in order to predict the damaged goods that exist in the company using the C4.5 algorithm method which is currently being used for analysis. In calculations, data mining using the C4.5 algorithm can produce a measure that can predict items that are at risk of defects or damage and implement them using RapidMiner applications. Because all this time, damaged items are not known beforehand and the results cause customer returns due to receiving damaged products.

Keywordad: *Product data, data mining, system fifo (first in first out), Algoritma C4.5, RapidMiner*

1. Pendahuluan

PT Home Center Indonesia merupakan perusahaan yang bergerak dibidang distributor. Selain PT Home Center Indonesia, masih cukup banyak perusahaanperusahaan lain yang bergerak di bidang serupa. Hal tersebut dapat menimbulkan persaingan antara bisnis perusahaan. Dan salah satu faktor yang menunjang akan kemajuan perusahaan salah satunya adalah kualitas barang yang akan dijual.

Terjadinya kerusakan barang di PT Home Center Indonesia ini juga diakibatkan tidak jalanya system fifo (first in first out). Untuk menghadai masalah ini perusahaan harus mengambil keputusan dengan tepat dalam menetapkan strategi produk yang dijual. Dan mendapatkan keputusan yang tepat diperlukan data barang yang cukup untuk dianalisa. Dan pertumbuhan yang sangat pesat dalam dunia teknologi dari akumulasi data telah menciptakan kondisi kaya akan data tetapi minim informasi.

Data Mining merupakan penambangan atau penemuan informasi baru dengan mencari pola atau aturan tertentu dari sejumlah data dalam jumlah yang besar yang diharapkan dapat mengatasi kondisi tersebut. Data mining sendiri memiliki beberapa teknik salah satunya yaitu klasifikasi. Teknik klasifikasi terdiri dari beberapa metode, dan Decision Tree adalah bagian dari metode klasifikasi. Terdapat berbagai macam metode untuk melakukan Data Mining yang digunakan seperti ID3, Neighbor, C4.5. Metode yang terdapat didalam data mining yang akan digunakan dalam penelitian ini adalah algoritma C4.5, dimana agoritma C4.5 merupakan algoritma yang digunakan untuk membentuk pohon keputusan (Decision Tree).

2. Tinjauan Studi

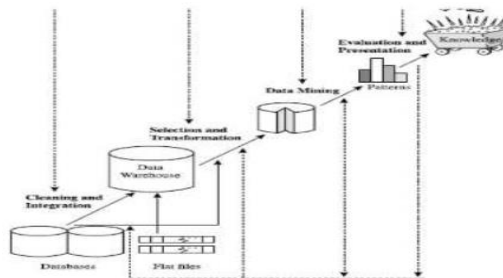
2.1. Data Mining

Data Mining adalah langkah analisis terhadap proses penemuan pengetahuan didalam basis data atau knowledge discovery in databases disingkat KDD. Pengetahuanya bisa berupa pola data atau relasi antar data yang valid yang berarti tidak diketahui sebelumnya. Data mining juga merupakan gabungan sejumlah disiplin ilmu komputer yang didefinisikan sebagai proses penemuan pola-pola baru dari kumpulan-kumpulan data yang sangat besar, meliputi metode-metode yang merupakan irisan dari artificial intelligence, machine learning, statistics dan databases system (Suyanto, 2017).

2.2. Tahapan Data Mining

Knowledge Discovery In Database (KDD), adalah keseluruhan proses nontrivial untuk mencari dan mengidentifikasi pola (pattern) dalam data, dimana pola yang ditemukan bersifat sah, baru, dapat bermanfaat dan dapat dimengerti. KDD berhubungan dengan teknik integrasi dan penemuan ilmiah, interpretasi dan visualisasi dari pola-pola sekumpulan data. Pada proses data mining terdapat beberapa tahapan yang harus dilakukan. Menurut (Retno, 2017) terdapat beberapa tahapan , antara lain sebagai berikut :

1. Data Selection
2. Data Preprocessing
3. Integrasi Data
4. Transformasi Data
5. Aplikasi Teknik Data Mining
6. Interpretasi dan Evaluasi



Gambar 1. Tahapan Data Mining (Retno, 2017)

2.3. Teknik Data Mining

2.3.1. Klasifikasi (Classification)

Classification (zaki, 2013) yaitu bagaimana mempelajari sekumpulan data sehingga dihasilkan aturan yang bisa diklasifikasi atau mengenali data-data baru yang belum pernah dipelajari. Klasifikasi dapat didefinisikan sebagai proses untuk menyatakan suatu objek data sebagai salah satu kategori (kelas) yang telah didefinisikan sebelumnya. Klasifikasi banyak digunakan dalam berbagai aplikasi, diantaranya adalah deteksi kecurangan (fraud detection), pengelolaan pelanggan, diagnosis medis, prediksi penjualan, dan sebagainya.

2.3.2. Regresi (Regression)

Regresi pada dasarnya mirip dengan klasifikasi, yakni memerlukan data pelatihan yang sudah diberi label. Bedanya, output klasifikasi adalah nilai diskrit, sedangkan output dari regresi adalah nilai kontinyu. Regresi ini mencari model hubungan antara atribut predictor dan atribut dependent, dimana atribut

dependennya juga berupa nilai kontinyu. Contoh regresi adalah memprediksi nilai kurs rupiah terhadap dollar (Budi dkk, 2018).

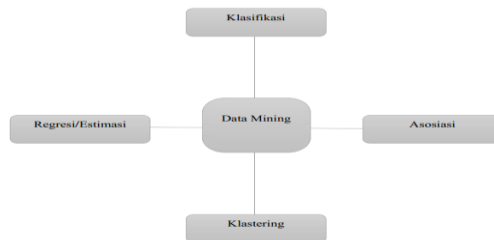
2.3.3. Klastering(Clustering)

Klastering merupakan suatu pengelompokan obyek kedalam beberapa kelompok berdasarkan kemiripan antar obyek, dimana dalam satu klaster harus berisi obyek yang saling mirip dan antara klaster obyek saling tidak mirip.

Teknik klastering bisa digunakan untuk mereduksi data, ide dasarnya sangatlah sederhana. Obyek data dipartisi kedalam sejumlah kelompok atau cluster, dimana obyek-obyek yang sangat mirip dikelompokkan kedalam satu cluster yang sama sedangkan obyek-obyek yang berbeda di cluster dengan obyek yang berbeda. Kemiripan antar obyek data dihitung dengan menggunakan fungsi jarak yang berupa similarity atau dissimilarity (Suyanto, 2017)

2.3.4. Asosiasi (Association Rule Mining)

Asosiasi dilakukan dengan cara menghitung berapa kali dalam suatu set data suatu transaksi yang mengandung dua item atau lebih yang berhubungan satu sama lain. Sering ada yang menyebut Market Basket Analysis. (Budi dkk, 2018)



Gambar 2. Kajian Data Mining (budi dkk, 2018)

2.4. Pohon Keputusan (Decision Tree)

Decision Tree adalah pemetaan mengenai alternatif-alternatif pemecahan masalah yang dapat diambil dari masalah tersebut. Pohon tersebut juga melihatkan factor-faktor kemungkinan/probabilitas yang akan mempengaruhi alternatif keputusan tersebut, disertai dengan estimasi hasil akhir yang akan didapat bila kita mengambil alternatif keputusan tersebut (Retno, 2017)

2.5. Metode Algoritma C4.5

Algoritma data mining C4.5 merupakan salah satu algoritma yang digunakan untuk melakukan klasifikasi atau segmentasi atau pengelompokan yang bersifat prediktif. Klasifikasi merupakan suatu proses data mining yang bertujuan untuk menemukan pola yang berharga dari data yang berukuran relatif21 besar hingga sangat besar. Algoritma C4.5 sendiri merupakan pengembangan dari algoritma ID3 (Retno, 2017).

Secara umum algoritma C4.5 untuk membangun pohon keputusan adalah sebagai berikut :

1. Pilih atribut sebagai node akar.
2. Buat cabang untuk tiap-tiap nilai.
3. Bagi kasus dalam cabang.
4. Ulangi proses untuk setiap cabang sampai semua kasus pada cabang memiliki kelas yang sama (Kusrini & Luthfi, 2011)

Untuk memilih atribut sebagai node akar, didasarkan pada nilai Gain tertinggi dari atribut-atribut yang ada. Untuk menghitung Gain digunakan rumus seperti tertera dalam persamaan berikut :

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

Keterangan :

S : himpunan kasus

A: Atribut

n : jumlah partisi atribut

|Si| : jumlah kasus pada partisi ke -i

|S| : jumlah kasus dalam S

Setelah mendapatkan nilai Gain, ada satu hal lagi yang perlu kita lakukan perhitungan, yaitu mencari nilai Entropy. Entropy digunakan untuk menentukan seberapa informatif sebuah masukan atribut untuk menghasilkan keluaran atribut. Rumus dasar dari Entropy tersebut adalah sebagai berikut :

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

Keterangan :

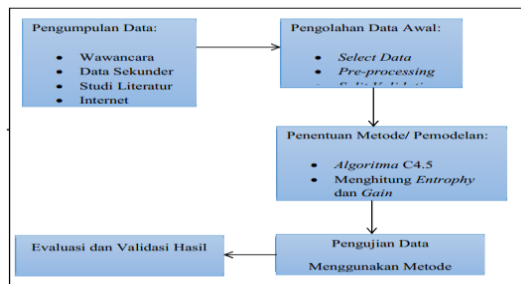
- S : Himpunan kasus
- A : Fitur
- n : Jumlah partisi S
- pi : Proporsi dari Si terhadap S

2.6. Produk Barang Rusak

Menurut Mulyadi (2012) produk rusak adalah produk yang tidak memenuhi standar mutu yang telah ditetapkan yang secara ekonomis tidak dapat diperbaiki menjadi produk yang baik dan produk rusak merupakan produk yang telah menyerap biaya bahan baku, biaya tenaga kerja dan biaya overhead pabrik.

3. Kerangka Konsep

Pada penelitian ini, tahapan yang akan digunakan untuk melakukan klasifikasi terhadap barang-barang rusak pada PT Home Center Indonesia dapat dilihat pada Gambar 3.1. tahap ini dilakukan untuk mempermudah penelitian sehingga penelitian dapat berjalan dengan lancar, baik dan sistematis. Dan penelitian ini dapat memenuhi tujuan yang diinginkan dan berikut ini beberapa langkah dalam tahapan penelitian.



Gambar 3. Tahapan Penelitian

Pengumpulan data dilakukan dengan mempertimbangkan penggunaan data dari jenis dan sumbernya. Data yang digunakan dalam penelitian ini adalah data sekunder, data yang penulis dapatkan adalah data master barang yang akan dijadikan sebagai bahan penelitian. Data yang akan digunakan dalam penelitian ini adalah data barang yang sudah masuk pada inventory gudang hasil tarikan dalam satu tahun terakhir pada PT Home Center Indonesia. Jumlah data yang digunakan adalah 350 data barang.

Kategori	Uraian	Ud	Ug	Uc	Ua	Ud	Ue	Uf	Ug	Uh	Ui	Uj	Uk	Ul	Um	Un	Uo	Up	Uq	Ur	Us	Ut	Uu	Uv	Uw	Ux	Uy	Uz
11	1000000001	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
12	1000000002	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
13	1000000003	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
14	1000000004	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
15	1000000005	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
16	1000000006	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
17	1000000007	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
18	1000000008	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
19	1000000009	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
20	1000000010	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
21	1000000011	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
22	1000000012	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
23	1000000013	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
24	1000000014	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
25	1000000015	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
26	1000000016	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
27	1000000017	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
28	1000000018	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
29	1000000019	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
30	1000000020	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
31	1000000021	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
32	1000000022	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
33	1000000023	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
34	1000000024	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
35	1000000025	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
36	1000000026	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
37	1000000027	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
38	1000000028	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
39	1000000029	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
40	1000000030	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

Gambar 4. Data Inventory Barang

3.1 Pengolahan Data Awal

Pada tahap ini menjelaskan tentang tahap awal data mining. Data yang telah didapatkan akan diolah ke format yang dibutuhkan, pengelompokan dan penentuan atribut data. Dalam melakukan pengolahan data awal, akan dilakukan beberapa tahapan agar didapatkan data yang bisa digunakan untuk tahap selanjutnya. Beberapa tahapan tersebut yaitu : *select data, pre-processing*, serta akan dilakukan *split validation*.

1. **Select Data**

Pada tahap ini akan dilakukan pemilihan variabel data yang akan dianalisis, karena sering ditemukan bahwa tidak semua data yang dibutuhkan dengan mempertimbangkan tujuan penulisan, sehingga diperoleh beberapa variabel yang akan digunakan untuk menjadi masukan variabel input. Dari data 350 record dengan 18 variabel hanya diambil 8 variabel saja, yaitu kode barang, nama barang, jenis barang, material, manufacturing date, temperature udara, tanggal kedatangan, dan tanggal expired, Data hasil seleksi akan digunakan dalam proses data mining

2. **Pre-processing**

Pada proses pre-processing akan dilakukan pembersihan data untuk membuang data yang missing value yaitu data yang tidak konsisten dan juga memperbaiki data yang rusak. Proses pembersihan data dilakukan secara manual untuk memastikan bahwa data yang telah dipilih layak untuk dilakukan proses permodelan. Pada tahap ini juga dilakukan proses menghilangkan variabel RID data barang dan Stock Kode Unit (SKU) karena tidak akan digunakan pada proses permodelan

Kelasi	Description	Material	Available	Ekspan Date	Ta Ekspan date	Temperature	Status
01.002.01	CARPET GRIZZLY AWM-91 DARK GREY 160X230	Kain wol	35	4/30/2019	30	32°C	Bad
01.002.03	CURTAIN ROD SINGLE 602 BLACK	Besi	120	6/30/2020	730	32°C	Good
01.002.04	CURTAIN ROD SINGLE 1.602 BR.ATK	Besi	480	6/30/2020	730	32°C	Good
05.040.03	CARLO MINI SERVING SQUARE LARGE	Kayu	240	7/29/2023	1835	35°C	Bad
05.040.01	THEO MINI SERVING DOUBLE HANDLE	Kayu	210	7/29/2023	1825	32°C	Good
06.010.06	BONTE BANQUET CHAIR GREEN	Dacron	20	9/14/2021	1095	31°C	Good
06.010.07	LEFT LEG HERNOX DT 200X100X5CM (P2.1)	Kayu	24	9/30/2022	1643	32°C	Good
06.011.01	BONTE BANQUET CHAIR RED	Dacron	30	7/31/2021	1095	32°C	Bad
14.037.04	JUAN STAIR CHAIR B ACK	Dacron	5	10/31/2019	548	33°C	Good
14.037.04	KONTE MANAGERIAL CHAIR HIGH BACK BLACK	Kulit sintetis	6	2/2/2020	548	33°C	Good
14.037.01	ARMREST DAYTON SOFA 3S 580100-1 BLK (P2)	Kain wol	6	1/12/2018	518	33°C	Good
14.037.04	SEAT DAY TON SOFA 2S 580100-2 BK (P12)	Kain wol	2	1/14/2018	548	33°C	Good
14.037.04	ARMREST DAYTON SOFA 3S 580100-3 BLK (P2)	Kain wol	2	1/14/2018	548	33°C	Good
14.037.01	SEAT DAY TON SOFA 2S 580100-2 BLK RED (P)	Kain wol	1	1/14/2018	518	33°C	Good
14.037.05	DRAWER HARVARD 2 TABLE OAK GREY-L (P2.2)	Kulit sintetis	9	7/1/2019	565	34°C	Bad
14.037.06	DRAWER HARVARD 2 TABLE OAK GREY-L (P2.2)	Kayu	68	12/1/2018	565	35°C	Bad
18.039.03	MANITOVA SOFA 3S 4C BROWN#571C	Kulit sintetis	20	12/1/2019	720	33°C	Good
18.039.07	MANITOVA SOFA 3S 4C BROWN#574C	Kulit sintetis	6	11/6/2019	720	35°C	Good
18.039.04	CAMDEN SOFA RC 2.5S CHARCOAL-BV/PVC-068E	Kulit sintetis	1	12/1/2019	565	35°C	Bad
18.039.01	ST JOHN'S SOFA 3S RC BROWN-PVC-N287C-CS	Kulit sintetis	6	6/30/2020	730	35°C	Bad
18.039.04	MANSTON SOFA 3S 13ARK BROWN 20443	Busa	20	12/1/2019	720	35°C	Good
18.039.04	KRINGLE SOFA 3S LIGHT BROWN-CU203-5	Busa	18	6/21/2020	720	35°C	Good
18.039.05	AUSTIN SOFA 3S COFFEE-C5/PVC-N66GB	Kulit sintetis	1	5/2/2020	720	35°C	Bad

Gambar 5. Process Cleaning Data

3. **Split Validation**

Split Validation merupakan teknik validasi yang membagi data menjadi dua bagi secara acak, sebagian data training dan sebagian data testing. Data yang sudah disiapkan untuk klasifikasi dibagi menjadi dua untuk data training dan data testing menggunakan sampling random sistematis (Systematic Random Sampling). Cara penggunaan teknik sampling random sistematis ini perandoman atau pengundian hanya dilakukan satu kali, yakni ketika menentukan unsur pertama dari sampling yang akan diambil. Penentuan unsur sampling selanjutnya ditempuh dengan cara memanfaatkan interval sampel.

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

$Jumlah\ Data\ (N) = 350$

$Jumlah\ Data\ Testing = 20\% \times 350 = 70$

$Jumlah\ Sampel\ (n) = 70$

$Interval\ sampling = N/n = 350/70 = 5$

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

$Unsur\ Kedua = s + k$

$Unsur\ Ketiga = s + 2k$

$Unsur\ Keempat = s + 3k, \text{ dan seterusnya hingga unsur ke-}n4$

Pembagian data menjadi data training dan data testing pada penelitian ini menggunakan split ratio 80% untuk data training dan 20% untuk data testing. Dari hasil diatas diperoleh data testing sebanyak 70 data barang, maka sisanya dijadikan data training 280 data barang.

NO	Lokasi	Description	Material	Available	Expair Date	Type Daring	Temperature	Status
1	01.002.01	CARPET GRIZZLY AWM-91 DARK GREY 160X190	Kain wol	85	4/30/2019	Aknesotia	32°C	Bad
2	01.002.04	CURTAIN ROD SINGLE 602 BK/BLACK	Besi	420	6/30/2020	Aknesotia	32°C	Good
3	01.002.04	CURTAIN ROD SINGLE 602 BLACK	Besi	480	6/30/2020	Aknesotia	32°C	Good
4	05.000.03	CADLE MINI SERVING SQUARE LARGE	Kayu	240	7/20/2023	Aknesotia	35°C	Bad
5	05.000.04	THRO MINI SERVING DOUBLE HANDLE	Kayu	240	7/20/2023	Aknesotia	32°C	Good
6	06.000.06	FRONTE BANQUET CHAIR GREEN	Dacron	20	5/14/2021	Aknesotia	31°C	Good
7	06.000.07	LEFT LEG HERMON DT 200X100X75CM (P2.3)	Kayu	24	9/30/2022	Dining	32°C	Good
8	06.001.01	FRONTE BANQUET CHAIR RED	Dacron	30	7/31/2021	Dining	32°C	Bad
9	14.037.04	ZUAN STAFF CHAIR BLACK	Dacron	5	10/31/2019	Office	33°C	Good
10	14.037.04	COUNCIL MANAGERIAL CHAIR HIGH BACK BK/BLACK	Kulit sintetis	6	3/2/2020	Office	33°C	Good
11	14.037.04	ARMREST DAYTLON SOFA JS 880100-1 BLK (P2)	Kain wol	6	1/14/2018	Living	33°C	Good
12	14.037.04	SEAT DAYTLON SOFA JS 880100-2 BLK (P1.2)	Kain wol	2	1/14/2018	Living	33°C	Good
13	14.037.04	ARMREST DAYTLON SOFA JS 880100-3 BLK (P2)	Kain wol	2	1/14/2018	Living	33°C	Good
14	14.037.04	SPAT DAYTLON SOFA JS 880100-2 BLK RFD (P)	Kain wol	1	1/14/2018	Office	33°C	Good
15	14.037.06	HUBERT OPTIC TBL 64X90X100 BK/BLACK	Kulit sintetis	3	7/1/2019	Meets	34°C	Bad
16	14.037.06	DRAWER HARVARD 2 TABLE OAK GREY-L (P2.2)	Kayu	68	12/1/2018	Living	35°C	Bad
17	18.039.03	MANTOVA SOFA JS RC BROWN/4574PC	Kulit sintetis	20	12/1/2019	Living	33°C	Good
18	18.039.03	MANTOVA SOFA JS RC BROWN/4574PC	Kulit sintetis	6	11/30/2019	Living	35°C	Good
19	18.039.04	CAMDEN SOFA RC 1.5S CIARCOAL/40V-PVC-081E	Kulit sintetis	1	12/1/2019	Living	35°C	Bad
20	18.039.04	ST JOHNS SOFA JS RC BROWN-PVC-N237C-CS	Kulit sintetis	6	6/30/2020	Living	35°C	Bad
21	18.039.04	MANSION SOFA JS DAK BROWN-204H	Busa	20	12/1/2019	Living	35°C	Good
22	18.039.04	KRINER F SOFA JS 1 LEFT BROWN/ACT034-4	Busa	18	6/20/2020	Living	35°C	Good
23	18.039.05	AUSTIN SOFA JS COFFEE/CS-PVC-N60B	Kulit sintetis	1	5/2/2020	Living	35°C	Bad
24	18.039.05	JACOBSEN SOFA RC 2S WALNUT/EN-037	Kulit sintetis	1	10/8/2018	Living	34°C	Good
25	18.039.05	IS RHT MACLAINE SOFA JS BROWN/80(P2.2)	Kulit sintetis	6	9/20/2018	Living	35°C	Good
26	18.039.04	IS RHT MACLAINE SOFA JS BROWN/80(P2.2)	Kulit sintetis	20	9/20/2018	Living	35°C	Good
27	18.039.04	IS RHT MACLAINE SOFA JS BROWN/80(P2.2)	Kulit sintetis	18	9/20/2018	Living	35°C	Good
28	18.039.04	MADISON JS RHT RC SOFA MOD BLACK-2525	Kulit sintetis	1	9/21/2018	Living	35°C	Good
29	18.039.05	1.5S RHT RC MINNESOTA RC CAPPUCINO(P6.0)	Busa	1	9/21/2018	Living	36°C	Bad

Gambar 6. Data Training

NO	Lokasi	Description	Material	Available	Expair Date	To Ekspair date	Temperature	Status
1	01.004.05	CURTAIN ROD SINGLE 619-BLACK	Kain wol	480	4/30/2019	730	32°C	Bad
2	01.004.06	CURTAIN ROD SINGLE 619-BLACK	Besi	480	6/30/2020	730	32°C	Good
3	01.004.07	CURTAIN CINDER DAMASK GREY 140X290CM	Besi	320	6/30/2020	730	32°C	Good
4	01.004.07	BRENDEN SET COTTON 60 DOUBLE	Kayu	21	7/20/2023	730	32°C	Good
5	01.004.03	CURTAIN ROD SINGLE 18010 IVORY	Kayu	1176	7/20/2023	730	32°C	Good
6	06.005.04	HARPER BANQUET CHAIR BLACK	Kayu	10	5/6/2020	730	31°C	Good
7	06.005.04	FRONTE BANQUET CHAIR RED	Kayu/Besi	40	7/31/2021	1095	32°C	Good
8	06.005.05	TIFANY BANQUET CHAIR BROWN WOOD	Kayu	30	4/31/2021	1095	32°C	Bad
9	06.005.06	FRONTE BANQUET CHAIR GREEN	Kayu	40	7/31/2021	1095	33°C	Good
10	06.005.07	CHAIR OLAF DINING SET GREY (P2.2)	Kayu/Kulit	28	10/30/2022	1643	33°C	Good
11	06.005.07	CHAIR OLAF DINING SET GREY (P2.2)	Kayu/Kulit	31	11/30/2022	1643	33°C	Good
12	06.005.07	HEIDIE DINING CHAIR-BLACK (F)	Kayu/Kulit	20	12/30/2020	913	33°C	Good
13	18.042.05	ARMLESS MINNESOTA RC CAPPUCINO (P3.6)	Busa	2	7/29/2023	548	33°C	Good
14	18.042.05	ARMLESS MINNESOTA RC CAPPUCINO (P3.6)	Busa	2	7/29/2023	548	33°C	Good
15	18.042.05	1.5S RHT RC MINNESOTA RC CAPPUCINO(P6.0)	Busa	3	6/6/2020	548	34°C	Bad
16	18.042.05	ARMLESS RC MC LEOD SECT DRK BRN#902 (P4.6)	Busa	1	7/31/2021	548	35°C	Bad
17	18.042.05	MC LEOD SOFA RC JS ROCKER DARK BROWN#902	Busa	1	4/31/2021	365	33°C	Good
18	18.043.02	MADISON JS RHT RC SOFA MOD BLACK 2525	Busa	3	7/31/2021	0	35°C	Good
19	18.043.02	ARMLESS MINNESOTA RC CAPPUCINO (P3.6)	Busa	1	10/30/2022	1100	35°C	Bad
20	18.043.03	MADISON SOFA JS 2S ARMLESS MOD BLACK 2525	Busa	9	11/30/2022	730	35°C	Good
21	14.071.04	LOWER SCHOLASTICA ZENI BLU 120XCM (P1.6)	Dacron	3	5/6/2020	365	35°C	Good
22	14.071.04	HEADBOARD SERENITY W/ RIFLOW TOP 120X120	Dacron/Kayu	108	7/31/2021	365	35°C	Good
23	14.071.04	CUBBON FANCY PRINTED 40X60CM	Busa	5	7/29/2023	365	35°C	Bad
24	14.072.01	COUNCIL MANAGERIAL CHAIR HIGH BACK BLACK	Besi	6	5/6/2020	0	34°C	Good
25	14.037.04	ARMREST DAYTLON SOFA JS 880100-1 BLK (P2)	Kayu/Dacron	6	7/31/2021	0	35°C	Good
26	14.037.04	SEAT DAYTLON SOFA JS 880100-2 BLK (P1.2)	Kayu/Dacron	2	5/31/2021	0	35°C	Good
27	14.037.04	ARMREST DAYTLON SOFA JS 880100-3 BLK (P2)	Kayu/Dacron	2	7/31/2021	0	35°C	Good
28	14.037.04	SEAT DAYTLON SOFA JS 880100-2 BLK RFD (P)	Kayu/Dacron	1	10/30/2022	548	35°C	Good
29	18.047.01	ULVEN SOFA RC JS TAUP#40V-PVC-722A	Busa	55	5/6/2020	365	36°C	Bad

Gambar 7. Data Testing

4. Desain Penelitian/Metodologi

Pada penelitian ini akan dilakukan pemodelan menggunakan metode Algoritma C4.5 untuk pengolahan data barang, kemudian untuk pengolahan *dataset* dan untuk mengolah hasil akurasi dari algoritma yang digunakan. Metode Algoritma C4.5 dipilih karena salah satu kelebihanannya adalah dapat menangani data numerik dan diskret. Algoritma C4.5 menggunakan rasio perolehan (*gain rasio*). Sebelum menghitung rasio perolehan, perlu dilakukan perhitungan nilai informasi dalam satuan bits dari suatu kumpulan objek, yaitu dengan menggunakan konsep entropy untuk membentuk pohon keputusan. Data kemudian dihitung menggunakan algoritma sesuai dengan metodenya kemudian dicari hasil akurasi. Berikut ini tahapan proses pemodelan dalam penelitian ini.

1. **Choosing the appropriate Data Mining task**

Pada tahap ini, penulis memilih jenis data mining yang digunakan untuk menentukan barang yang risan rusak (*expired*) maka dipilih jenis data mining yang akan digunakan adalah klasifikasi.

2. **Choosing the Data Mining Algorithm**

Setelah pemilihan jenis data mining yang akan digunakan yaitu klasifikasi, maka selanjutnya menentukan algoritma klasifikasi yang akan digunakan. Pada penelitian ini algoritma yang dipilih untuk klasifikasi adalah Algoritma C4.5

3. **Employing the Data Mining Algorithm**

Tahap ini dilakukan untuk pengolahan data dengan algoritma yang telah dipilih yaitu dengan menggunakan Algoritma C4.5.

4. **Evaluation**

Dalam tahap ini dilakukan evaluasi dan menafsirkan pola yang didapatkan dari hasil algoritma yang dipakai untuk mengetahui aturan, kehandalan, dan lainlain. Evaluasi dilakukan dengan menerapkan pola yang didapat dari proses sebelumnya terhadap data testing yang disediakan. Evaluasi dilakukan dengan confusion matrix dan kurva ROC.

5. **Using the Discovered Knowledge**

Pada tahap ini menggunakan pengetahuan yang diperoleh dari proses data mining untuk penerapan pada aplikasi atau lainnya. Pengetahuan klasifikasi memprediksi barang rusak (*expired*) diterapkan pada data baru untuk membuat klasifikasi barang yang good dan Bad/Not Good.

5. Hasil Penilitia Dan Pembahasan

5.1. Hasil Penelitian

5.1.1 Perhitungan Jumlah Kasus Data Keseluruhan, Data Training dan Data Testing

Data keseluruhan yang digunakan pada penelitian ini sejumlah 350 kasus, kemudian dilakukan pembagian data menjadi data training sebanyak 280 kasus dan data testing sebanyak 70 kasus. Berikut hasil perhitungan kasus pada data keseluruhan.

NO	ATRIBUT	VALUE	JUMLAH KASUS	GOOD (S1)	NOT GOOD (S2)		
1	TOTAL		350	282	68		
2	DEPARTMEN	Accessories	31	22	9		
		Dining	63	53	10		
		Living	188	162	26		
		Office	20	5	15		
		Matress	46	38	8		
3	MATERIAL	Besi	13	10	3		
		Besi Busa	124	109	15		
		Dacron	8	4	4		
		Kaca	7	6	1		
		Kain Printing	4	4	0		
		Kain Wall	4	3	1		
		Kayu	30	27	3		
		Kayu Busa	131	95	36		
		Kulit Kayu Busa	6	6	0		
		Plastik	1	1	0		
		Plastik Busa	22	17	5		
		4	LOKASI	01	100	53	47
				02	41	41	0
03	67			66	1		
04	45			45	0		
05	62			34	28		
06	15			6	9		
07	20			7	13		
31°C	12			7	5		
5	TEMPERATURE	32°C	74	58	16		
		33°C	92	75	17		
		34°C	26	21	5		
		35°C	146	121	25		
		OK	283	234	49		
6	STATUS	DMG	67	48	19		
		0	19	3	16		
		365	10	8	2		
7	TO EXPIRED DATE	548	35	29	6		
		720	31	31	0		
		730	142	109	33		
		913	7	5	2		
		1095	21	18	3		
		1100	4	4	0		
		1643	11	11	0		
		1825	10	10	0		
		3650	60	54	6		

Gambar 8. Jumlah Nilai Atribut Keseluruhan

Gambar 8 memperlihatkan semua kasus pada data keseluruhan yang berjumlah 350 dengan keputusan **GOOD** sebanyak 282 dan **NOT GOOD** 68. Data keseluruhan yang dibagi menjadi data training dan data testing dihitung kembali jumlah kasusnya, tabel 4.3 dibawah ini adalah jumlah kasus pada data training dan data ini akan dijadikan sebagai data sekunder dalam pembuatan pohon keputusan prediksi barang rusak pada PT Home Center Indonesia. Proses perhitungan klasifikasi prediksi barang rusak menggunakan Algoritma C4.5 adalah sebagai berikut :

NO	ATRIBUT	VALUE	JUMLAH KASUS	GOOD (S1)	NOT GOOD (S2)	Entropy	Gain
1	Total		280	236	44	0.65966744	
2	Department	Accessories	19	18	1	0.305559617	
		Dining	56	49	7	0.569613146	
		Living	152	134	18	0.44934818	0.242677629
		Office	37	30	7	0.735212822	
		Matress	16	5	11	1.978116282	
3	Material	Besi	6	6	0	0	
		Besi Busa	106	95	11	0.501816247	
		Dacron	4	3	1	0.839598958	
		Kaca	7	6	1	0.621449159	
		Kain Printing	4	4	0	0	0.64406196
		Kain Wall	2	2	0	0	
		Kayu	27	25	2	0.394315266	
		Kayu Busa	108	83	25	0.813368012	
		Kulit Kayu Busa	0	0	0	0	
		Plastik	1	1	0	0	
		Plastik Busa	15	11	4	0.8583399	
4	Location	01	81	70	11	0.601519546	0.269688327
		02	35	35	0	0	
		03	55	54	1	0.132562889	
		04	36	36	0	0	
		05	49	28	21	0.772623784	
		06	10	6	4	0.832154357	
		07	14	7	7	0.5	
5	Temperature	31°C	8	6	2	0.839598958	
		32°C	58	48	10	0.697411841	
		33°C	70	58	12	0.695072792	0.656030132
		34°C	20	17	3	0.660913173	
		35°C	124	107	17	0.66522903	
6	Status	OK	221	200	21	0.471770276	0.108655381
		DMG	59	36	23	0.847832977	
7	To Expired Date	0	14	3	11	-	
		365	10	8	2	0.757542476	
		548	26	21	5	0.741817848	
		720	29	29	0	0	
		730	100	81	19	0.736942879	0.507670036
		913	7	5	2	0.875504686	
		1095	21	18	3	0.621449159	
		1100	2	2	0	0	
		1643	11	11	0	0	
		1825	10	10	0	0	
		3650	60	48	12	0.247579046	

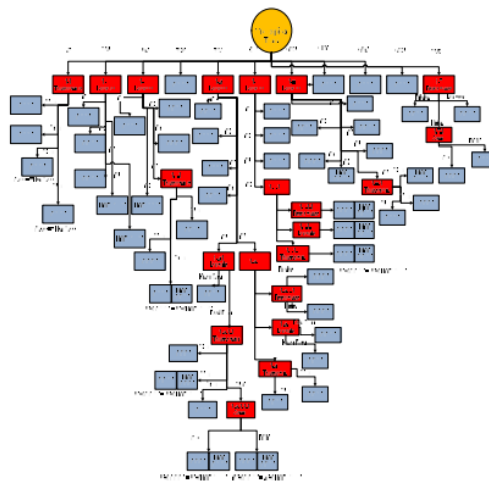
Gambar 9. Perhitungan Data Training

Dari gambar 9 memperlihatkan bahwa pada baris total diketahui jumlah kasus (S) adalah 280, jumlah kasus keputusan GOOD (S1) ada 236, dan jumlah kasus keputusan NOT GOOD (S2) adalah 44. Perhitungan entropy total pada tabel 9 dapat dihitung dengan menggunakan persamaan sebagai berikut :

$$\begin{aligned}
 Entropy(Total) &= (236/280 * \text{Log}_2(236/280)) + (44/280 * \text{Log}_2(44/280)) \\
 Entropy(Total) &= 0.65966744
 \end{aligned}$$

Dan untuk mencari nilai entropy pada masing-masing atribut dilakukan sama dengan mencari nilai entropy pada total atribut. Sementara itu, nilai gain pada baris Departmen dihitung dengan menggunakan persamaan sebagai berikut:

$$\begin{aligned}
 Gain(\text{Departmen}) &= Entropy(Total) - \sum \frac{|\text{Departmen}|}{|\text{Total}|} * Entropy(\text{Departmen}) \\
 Gain(\text{Departmen}) &= (0.65966744) - ((19/280 * 0.305559617) - \\
 &((56/280 * 0.569615146) - ((152/280 * 0.549341818) - ((37/280 * 0.735212822) - \\
 &((16/280 * 1.978110282) \\
 Gain(\text{Departmen}) &= 0.242677629
 \end{aligned}$$

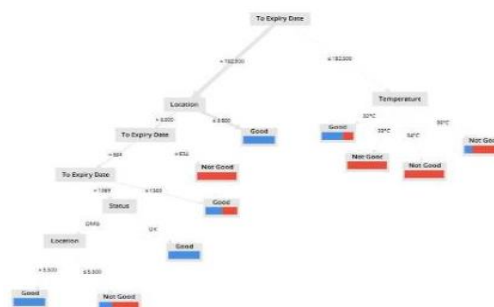


Gambar 10. Pohon Keputusan

5.2. Hasil Pengujian

Setelah analisis dan mengklasifikasikan barang rusak di PT Home Center Indonesia menggunakan algoritma C4.5 dihitung dan pembuatan pohon keputusan menggunakan perhitungan manual, maka untuk tahap berikutnya adalah pembuktian dari analisis dan perhitungan manual tersebut. Adapun aplikasi yang digunakan dalam klasifikasi barang rusak menggunakan aplikasi Rapid Miner 7.6.003

Adapun hasil dari pembuatan pohon keputusan menggunakan aplikasi Rapid Miner 7.6.003 dapat dilihat pada Gambar 11.



Gambar 11. Pohon Keputusan Hasil Rapid Miner

Berdasarkan pohon keputusan (decision tree) yang terbentuk pada Gambar 4.20 di atas, didapat aturan-aturan/rule model dalam penentuan barang rusak. Ada 5 aturan yang terbentuk, maka dapat dibuat rule sebagai berikut :

Rule Number	Rule Details
1	<i>If to Expired Date = 0 and Than = Good</i>
2	<i>If To Expired Date = 365 and Location = >4.500 Then = Not Good</i>
3	<i>If to Expired Date = 548 and Location = 06 and Temperature = 35°C and Than = Good</i>
4	<i>If to Expired Date = 548 and Location = 06 and Status = OK and Than = Good</i>
5	<i>If to Expired Date = 548 and Location = 06 and Status = DMG and Than = Not Good</i>

Gambar 12. Rule Hasil Pohon Keputusan

Pengujian berdasarkan pengukuran Confusion Matrix menghasilkan nilai accuracy, precision dan recall yang tinggi dengan nilai accuracy sebesar 98.57%, precision sebesar 100.00% dan recall sebesar 85.71% . Dimana sebelumnya data barang yang ada belum diketahui tingkat akurasi, dan sekarang sudah dapat dilihat seberapa tinggi nilai akurasi yang diperoleh. Dan pengujian berdasarkan pengukuran kurva ROC/ AUC sebesar 1.000 dengan tingkat akurasi Excellent Classification

6. Kesimpulan

Berdasarkan hasil penelitian ini maka didapat beberapa kesimpulan sebagai berikut:

1. Dengan menggunakan metode Algoritma C4.5 dalam klasifikasi barang rusak, metode Algoritma C4.5 mampu menentukan penerima bantuan dengan baik dan cepat.
2. Evaluasi hasil klasifikasi data barang dengan metode Algoritma C4.5 yang dievaluasi dengan confusion matrix menghasilkan tingkat akurasi sebesar 98.44%, sedangkan evaluasi dengan kurva ROC dengan akurasi Excellent Classification sebesar 1.000
3. Terbentuk 5 rule model yang dapat digunakan sebagai data testing
4. Penerapan data mining menggunakan metode Algoritma C4.5 mampu mempercepat waktu pengambilan keputusan dalam menentukan barang yang rusak

Daftar Pustaka

- Aydin, R., Brown, A., Badurdeen, F., & Rouch, K. E. (2018). Jurnal Sistem Manufaktur dampak mengukur ketidakpastian produk pengembalian pertunjukan ekonomi dan lingkungan dari produk con fi desain gurasi, (November 2017). <https://doi.org/10.1016/j.jmsy.2018.04.009>
- Budi, S. Ardian, U. (2018). Data Mining dan Big Data Analytics. Penebar Media Pustaka, Yogyakarta.
- Harani, W. M., Mulyanto, E., Nuswantoro, U. D., Nakula, J., Semarang, I. N., & Kunci, K. (n.d.). PREDIKSI PENJUALAN BARANG PADA ALFAMART REMBANG MENGGUNAKAN EXPONENTIAL SMOOTHING, 1–31.
- Hastuti, K. (2012). Analisis komparasi algoritma klasifikasi data mining untuk prediksi mahasiswa non aktif, 2012(Semantik), 241–249.
- Kori, A. (2017). Comparative Study of Data Classifiers Using Rapidminer, 5(2), 1041–1043.
- Oktana, I., & Hansun, S. (2015). Penerapan Algoritma C4.5 pada Analisis Kerusakan Barang J adi, VII(1), 24–28. Teknologi, M. K. (2018). Jurnal iptek, 79–86.
- Praja, E., Mandala, W., Kom, S., & Kom, M. (2015). DATA MINING MENGGUNAKAN BAYESIAN CLASSIFIER UNTUK MENENTUKAN KELAYAKAN KENDARAAN YANG AKAN DIJUAL PADA SHOWROOM MOTOR BEKAS, 1(Senatkom). Sebagai, D., Satu, S., Memperoleh, U., & Sarjana, G. (2010). APLIKASI DATA MINING UNTUK MENAMPILKAN INFORMASI TINGKAT KELULUSAN MAHASISWA (Studi Kasus di Fakultas MIPA Universitas Diponegoro). Universitas Stuttgart
- Retno, T., Vlandari 2017. Data Mining, Teori dan Aplikasi RapidMiner. Penerbit Gava Media, Yogyakarta.
- Salam, A., & Nuswantoro, U. D. (2018). Implementasi Algoritma Apriori untuk Mencari Asosiasi Barang yang Dijual di E-commerce OrderMas, 17(2), 158– 170.
- Suyanto. (2017). Data Mining Untuk Klasifikasi dan Klasterisasi Data. Informatika Bandung, Bandung.
- Tampubolon, K., Saragih, H., Reza, B., Epicentrum, K., Asosiasi, A., & Apriori, A. (2013). IMPLEMENTASI DATA MINING ALGORITMA APRIORI PADA SISTEM PERSEDIAAN ALAT-ALAT KESEHATAN, 93–106.
- Universitas, M., Yogyakarta, N., & Uny, K. (2017). Jurnal SCRIPT Vol . 5 No . 1 Desember 2017 ISSN : 2338-6304 Jurnal SCRIPT Vol . 5 No . 1 Desember 2017, 5(1), 39–45.
- Santos, Singgih, *Structural Equation Modelling (SEM) Konsep dan Aplikasi dengan AMOS18*, ElexMedia Komputindo, Jakarta, 2011.