

PENERAPAN ALGORITMA C4.5 PADA ANALISIS KERUSAKAN BARANG JADI (STUDI KASUS: PT KAYU LAPIS ASLI MURNI)

Ivan Oktana, Seng Hansun

Program Studi Teknik Informatika, Universitas Multimedia Nusantara, Tangerang, Indonesia
ivan.oktana@rocketmail.com, hansun@umn.ac.id

Diterima 15 April 2015

Disetujui 3 Juni 2015

Abstract—A Data mining is the activity that includes the collection, the use of historical data to discover regularity, patterns and relationship in a large data set. The usefulness of data mining is to specify a pattern to be found in the data mining task. The presences of data mining is motivated by the problem of data explosion which had been experienced lately these day where many organization or company collect so many years of data (purchasing data, sales data, damage data item, transaction data, and so on). In this paper data mining methods been used to analyze the damage data of finished products, with the goal of producing a pattern of the damage product. Based on the pattern from product's damage, can be see and the aspects that affect to the damaged product. The purpose of this study is to show information about the relationship of damage data of finished goods using C4.5 algorithms in PT. Kayu Lapis Asli Murni, and display results in the form of a decision tree mining.

Index Terms—data mining, C4.5 algorithms, decision tree, damage data of finished products, PT. Kayu Lapis Asli Murni

I. PENDAHULUAN

A. Latar Belakang

Perusahaan-perusahaan pada saat ini sudah memiliki *database* yang berukuran sangat besar pada bagian kerusakan barang dan dapat terus bertambah ukurannya. Selama ini, masalah akan muncul dalam melakukan analisis data kerusakan barang karena jumlah data yang sangat banyak dan keterbatasan pengolahan data yang tidak berfungsi secara maksimal. Dalam menghadapi persaingan bisnis dan meningkatkan pendapatan perusahaan, pihak terkait dalam perusahaan dituntut untuk dapat mengambil keputusan yang tepat dalam menentukan strategi produknya. Untuk mencapai hal tersebut melalui kebutuhan bisnis, solusinya dengan pengurangan biaya operasi perusahaan. Untuk memenuhi kebutuhan tersebut, diperlukan langkah-langkah cara menganalisis data perusahaan.

PT.Kayu Lapis Asli Murni merupakan perusahaan yang bergerak dalam bidang industri *plywood*. Selain PT.Kayu Lapis Asli Murni, masih cukup banyak perusahaan-perusahaan lain yang bergerak di bidang serupa. Hal tersebut dapat menimbulkan persaingan bisnis antar perusahaan.

Untuk menghadapi persaingan bisnis antar perusahaan dan meningkatkan pendapatan, perusahaan harus dapat mengambil keputusan dengan tepat dalam menetapkan strategi produk yang dijual. Untuk mendapatkan keputusan yang tepat diperlukan data yang cukup untuk dianalisa.

Ketersediaan data yang melimpah akan mendukung pengambilan keputusan untuk membuat solusi bisnis di bidang teknologi informasi yang melatarbelakangi lahirnya teknologi *data mining*. Dengan *data mining* perusahaan dapat menganalisis data perusahaan dan mengambil keputusan dengan tepat yang pada akhirnya bisa meningkatkan keuntungan atau mengurangi kerugian perusahaan.

Terdapat berbagai macam metode untuk melakukan *data mining* yang digunakan seperti ID3, *Neighbor*, C4.5. Metode yang terdapat di dalam *data mining* yang akan digunakan dalam penelitian ini adalah algoritma C4.5, dimana algoritma C4.5 merupakan algoritma yang digunakan untuk membentuk pohon keputusan. Berdasarkan penelitian sebelumnya yang dilakukan oleh Hasan [1] dikatakan metode yang digunakan dalam *data mining* adalah *Nearest Neighbor* dimana metode tersebut memberikan kesimpulan bahwa rata-rata keberhasilan K-NN dalam melakukan klasifikasi data mencapai akurasi di atas 90%. Hal ini menunjukkan bahwa K-NN memiliki performa yang handal dalam melakukan klasifikasi. Berdasarkan penelitian yang dilakukan oleh Kusriani, Sri Hartati, Retno Wardoyo, Agus Harjoko [2] mengatakan bahwa dengan menggunakan metode *nearest neighbor*, proses *testing* memerlukan waktu yang lebih lama

dibanding dengan menggunakan algoritma C4.5. Dari sini dapat dilihat bahwa algoritma C4.5 lebih cepat digunakan dari algoritma *nearest neighbor*.

Dari latar belakang masalah yang dikemukakan dapat diidentifikasi masalah antara lain sebagai berikut:

1. Bagaimanakah cara menerapkan *data mining* dan menganalisa kerusakan barang jadi (*reject*) pada PT. Kayu Lapis Asli Murni?
2. Bagaimana melihat hubungan antar atribut yang cacat dengan jumlah data yang banyak?

Agar pembahasan penelitian ini tidak menyimpang dari apa yang telah dirumuskan, maka diperlukan batasan-batasan. Batasan-batasan dalam penelitian ini adalah:

1. Dalam pengembangan aplikasi ini, sistem hanya mencakup aspek-aspek yang terkait dengan analisa kerusakan barang jadi yang dilihat dari hubungan produk-produk yang mengalami kerusakan.
2. Pengolahan aplikasi dan data yang akan dikembangkan ditunjukkan khusus untuk PT. Kayu Lapis Asli Murni.
3. Algoritma data mining yang digunakan adalah C4.5.
4. Bahasa *scripting* yang digunakan adalah PHP dan *database* yang digunakan adalah MYSQL.
5. Hasil *grade plywood* yang digunakan untuk hasil *mining* adalah *down grade* dan *reject*.
6. Teknik pengumpulan data yang digunakan adalah *cleaning*, *mining*, *pattern evaluation*, dan *knowledge presentation*.
7. Ukuran *size plywood* menggunakan satu ukuran.

Berdasarkan permasalahan yang diteliti, maka maksud dari penulisan ini adalah untuk membangun aplikasi data *mining* pada kerusakan barang jadi di PT. Kayu Lapis Asli Murni yang memiliki tujuan penelitian untuk:

1. Memberikan hasil *grade plywood* berdasarkan atribut-atribut kerusakan *plywood*.
2. Membantu memberikan informasi mengenai hubungan data kerusakan barang jadi.
3. Melakukan studi terhadap algoritma klasifikasi *data mining*, khususnya C4.5.

Adapun manfaat dari penelitian ini adalah:

1. Perusahaan dapat mengetahui dan mengidentifikasi kerusakan barang jadi.

2. Menyatakan perbedaan antara tipe kerusakan barang jadi bulan ini dan sebelumnya.
3. Memprediksi peluang menjual *plywood* yang cacat ke dalam negeri.

II. TELAHAH LITERATUR

Menurut Tan [3], *data mining* merupakan pengekstrakan informasi baru yang diambil dari bongkahan data besar yang membantu dalam pengambilan keputusan. Pola yang disajikan haruslah mudah dipahami, berlaku untuk data yang akan diprediksi dengan derajat kepastian tertentu, berguna, dan baru. *Data mining* adalah bagian dari *Knowledge Discovery in Database*. *Knowledge Discovery in Database* adalah sebuah proses otomatis atas pencarian data di dalam sebuah memori yang amat besar dari data untuk mengetahui pola.

Tahapan proses KDD menurut Han & Kamber [4] adalah:

1. *Data Cleaning*: Pembersihan data dilakukan untuk menghilangkan data yang tidak konsisten dan mengandung *noise*.
2. *Data Integration*: Proses integrasi data dilakukan untuk menggabungkan data dari berbagai sumber menjadi bentuk sebuah penyimpanan data yang saling berhubungan, seperti dalam *data warehousing*.
3. *Data Selection*: Proses seleksi data mengambil data yang relevan digunakan untuk proses analisis.
4. *Data Transformation*: Proses ini mentransformasikan atau menggabungkan data ke dalam bentuk yang tepat untuk dilakukan proses *mine* dengan cara melakukan peringkasan atau operasi agregasi. Dalam beberapa kasus, proses transformasi dilakukan sebelum proses seleksi, misalnya dalam kasus *data warehouse*.
5. *Data Mining*: *Data mining* merupakan proses yang penting, dimana metode-metode cerdas diaplikasikan untuk mengekstrak pola-pola dalam data.
6. *Pattern Evaluation*: Evaluasi pola diperlukan untuk mengidentifikasi beberapa pola yang menarik dalam merepresentasikan pengetahuan.
7. *Knowledge Presentation*: Penggunaan visualisasi dan teknik representasi untuk menunjukkan pengetahuan hasil penggalian dari tumpukan data kepada pengguna.

Menurut Kusnawi [5], algoritma C4.5 merupakan algoritma yang digunakan untuk membentuk pohon keputusan. Metode pohon keputusan mengubah fakta yang sangat besar menjadi pohon keputusan yang

merepresentasikan aturan.

Secara umum algoritma C4.5 untuk membangun pohon keputusan dengan [2]:

- Pilih atribut sebagai akar.
- Buat cabang untuk masing-masing nilai.
- Bagi kasus dalam cabang.
- Ulangi proses untuk masing-masing cabang yang memiliki kelas yang sama.

Untuk memilih atribut sebagai akar, didasarkan pada nilai *gain* tertinggi dari atribut-atribut yang lain. Untuk menghitung *gain* digunakan rumus:

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

dimana:

- S : Himpunan kasus
- A : Atribut
- n : Jumlah partisi atribut A
- |S_i| : Jumlah kasus pada partisi ke-i
- |S| : Jumlah kasus dalam S

Untuk menghitung nilai dari *entropy* dengan rumus:

$$\text{Entropy}(S) = \sum_{i=1}^n - p_i * \log_2 p_i \quad \dots(2)$$

dimana:

- S : Himpunan kasus
- A : Fitur
- n : Jumlah partisi S
- P_i : Proporsi dari S_i terhadap S

III. METODOLOGI PENELITIAN

Metode penelitian yang digunakan adalah:

1. Studi Literatur

Tahap awal dilakukan studi mengenai referensi-referensi yang berhubungan dengan *data mining*. Referensi tersebut dapat berupa jurnal, artikel, buku, dan lain-lain.

2. Observasi Data Kerusakan Barang PT. KAYU LAPIS ASLI MURNI

Observasi dilakukan untuk mendapatkan data-data kerusakan barang jadi. Hasil observasi ditulis dalam *file Microsoft Excel* agar membantu organisasi data dalam membangun *ontology*.

3. Perancangan dan Pembangunan Aplikasi

Aplikasi analisa data kerusakan barang jadi berbasis *data mining* menggunakan bahasa pemrograman *web*.

4. Uji Coba

5. *Knowledge base* dan aplikasi analisa data kerusakan barang jadi berbasis *data mining* yang telah dibuat akan diuji coba. Hal ini bertujuan untuk mencari *bug* yang masih ada pada aplikasi tersebut dan memperbaiki aplikasi tersebut.

Objek penelitian yang digunakan dalam penelitian ini adalah data analisa kerusakan barang jadi. Data analisa kerusakan barang jadi yang diambil sebagai sampel adalah data di bulan Juli 2013.

Variabel adalah objek penelitian, atau apa yang menjadi fokus di dalam suatu penelitian, oleh karena itu *variabel* sangat dibutuhkan dalam penelitian. *Variabel* yang dibahas adalah *variabel* jenis cacat pada kerusakan *plywood*. Jenis cacat adalah atribut-atribut dalam *plywood* yang bisa mendapatkan cacat.

IV. HASIL PENELITIAN

Berdasarkan pengujian yang dilakukan, disimpulkan bahwa aplikasi *data mining* dapat digunakan untuk mengetahui hubungan data kerusakan barang, dengan membandingkan *grade plywood data testing* dengan *grade plywood* dari pohon keputusan yang dibentuk dari *data training*. Pada pengujian ini membandingkan *grade data testing* dengan *grade* dari pohon keputusan diambil dari data pada tanggal 3 September 2013. Diambil empat ratus data analisa kerusakan barang, dimana tiga ratus data untuk data *training* dan seratus data untuk data *testing*. Hasil uji coba dari proses *mining* dan *mining preprunning* memiliki hasil yang sama dikarenakan pohon keputusan yang dibentuk memiliki bentuk yang sama.

A. Flow Chart Perhitungan C4.5

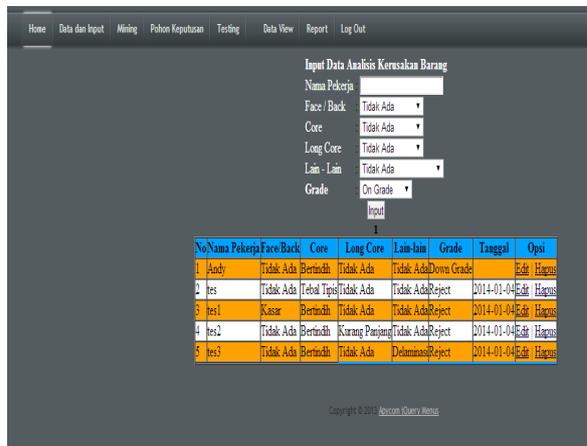


Gambar 1. Flow Chart Perhitungan C4.5

Flow chart aplikasi data mining, proses mining subproses perhitungan C4.5 menggambarkan flow chart untuk proses perhitungan C4.5 yang terdiri dari enam proses, dan sembilan proses untuk data.

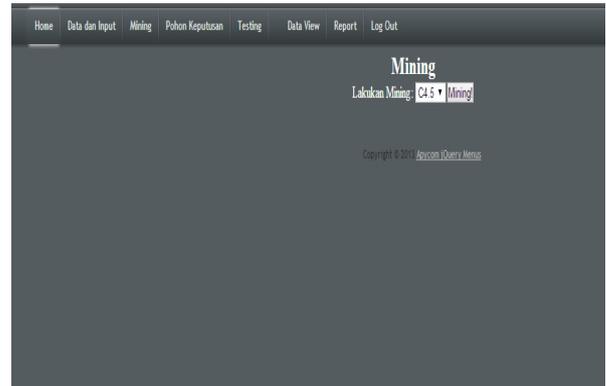
B. Tampilan Halaman Input

Halaman data input (Gambar 2) berisikan masukan untuk data plywood yang baru dan daftar plywood yang sudah ada. Dari halaman utama, user dapat mengakses halaman input data dengan melakukan click pada menu "Data dan Input" yang ada pada menubar di bagian atas aplikasi.



Gambar 2. Tampilan Halaman Input

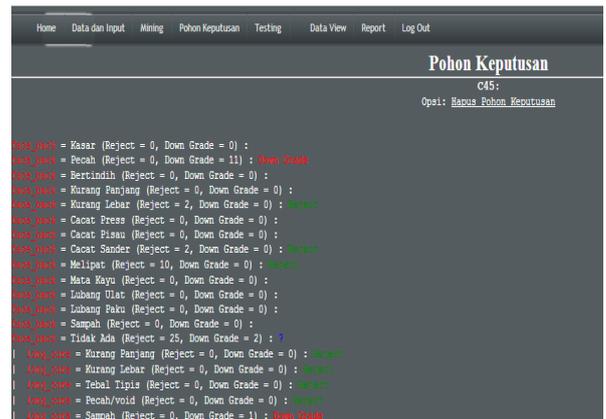
C. Tampilan Halaman Mining



Gambar 3. Tampilan Halaman Mining

Halaman mining (Gambar 3) berisikan tombol untuk melakukan proses mining yang mengolah data yang terdapat pada data training dan akan menghasilkan pohon keputusan di halaman pohon keputusan.

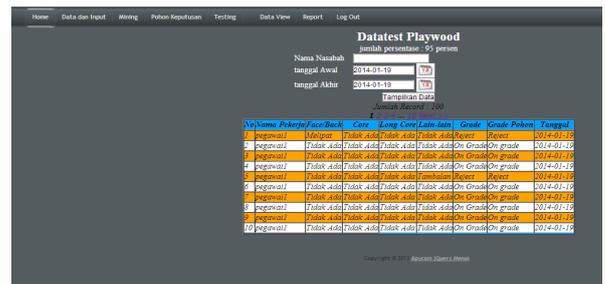
D. Tampilan Halaman Pohon Keputusan



Gambar 4. Tampilan Halaman Pohon Keputusan

Halaman pohon keputusan (Gambar 4) berisikan hasil proses mining dari halaman mining. Halaman pohon keputusan akan menghasilkan tree yang digambarkan seperti gambar di bawah ini.

E. Tampilan Halaman Testing Data



Gambar 5. Tampilan Halaman Testing Data

Halaman *testing* data menampilkan persentase dari *grade plywood* dengan *grade* dari pohon keputusan, seperti terlihat pada Gambar 5.

F. Tampilan Hasil Uji

No	Nama Pemesan	Face Back	Case	Lang Core	Lain-lain	Grade	Grade Pohon	Tanggal
1	pegawaian	Tidak Ada	Tidak Ada	Tidak Ada	Tidak Ada	On	grade	2014-01-19
2	pegawaian	Tidak Ada	Tidak Ada	Tidak Ada	Tidak Ada	On	grade	2014-01-19
3	pegawaian	Tidak Ada	Tidak Ada	Tidak Ada	Tidak Ada	On	grade	2014-01-19
4	pegawaian	Tidak Ada	Tidak Ada	Tidak Ada	Tidak Ada	On	grade	2014-01-19
5	pegawaian	Tidak Ada	Tidak Ada	Tidak Ada	Tidak Ada	On	grade	2014-01-19
6	pegawaian	Tidak Ada	Tidak Ada	Tidak Ada	Tidak Ada	On	grade	2014-01-19
7	pegawaian	Tidak Ada	Tidak Ada	Tidak Ada	Tidak Ada	On	grade	2014-01-19
8	pegawaian	Tidak Ada	Tidak Ada	Tidak Ada	Tidak Ada	On	grade	2014-01-19
9	pegawaian	Tidak Ada	Tidak Ada	Tidak Ada	Tidak Ada	On	grade	2014-01-19
10	pegawaian	Tidak Ada	Tidak Ada	Tidak Ada	Tidak Ada	On	grade	2014-01-19

Gambar 6. Tampilan Hasil Uji

Dari hasil uji yang dilakukan dari seratus *data testing* didapat 85 persen hasil *grade* dari data *testing* sama dengan hasil *grade* dari pohon keputusan. Karena data yang diujicoba adalah seratus data, maka data dengan hasil yang sama adalah delapan puluh lima data dan data dengan hasil yang berbeda adalah lima belas data (dapat dilihat pada Gambar 6).

V. KESIMPULAN

Kesimpulan dari pembuatan aplikasi ini adalah aplikasi *data mining* dapat digunakan untuk menampilkan informasi mengenai hubungan data kerusakan barang jadi dengan menggunakan pohon keputusan. Aplikasi *data mining* ini dapat memberikan hasil *grade plywood* berdasarkan atribut-atribut kerusakan *plywood* dengan algoritma C4.5. Untuk

melihat keakuratan aplikasi *data mining* digunakan perbandingan hasil *grade* dari data *plywood* dengan hasil *grade* dari pohon keputusan yang hasilnya ditampilkan dalam persentase *grade*. Semakin tinggi nilai persentase dari data *grade plywood* yang ada dengan *grade* dari pohon keputusan, maka menandakan bahwa pola kerusakan barang jadi dari hari ke hari tidak terlalu banyak berubah, akan tetapi bila persentasenya kecil, pola kerusakan barang jadi berubah sangat besar dari hari sebelumnya.

Hasil dari proses *data mining* ini dapat berguna untuk mempertimbangkan pengambilan keputusan lebih lanjut untuk meminimalisir kerusakan *plywood* dengan melihat tipe kerusakan *plywood* pada PT. Kayu Lapis Asli Murni.

DAFTAR PUSTAKA

- [1] Hasan Indra. 2013. Penerapan Algoritma K-Nearest Neighbor untuk prediksi potensi calon kreditur di xyz finance. Gorontalo : Universitas negeri Gorontalo.
- [2] Kusriani dan Luthfi Emha Taufiq. 2009. *Algoritma Data Mining*. Penerbit Andi Offset, Yogyakarta.
- [3] Tan Pang-Ning. 2006. Introduction to Data Mining Pearson International Edition. Pearson Addison Wesley.
- [4] Han, J & Kamber, M. 2006. *Data Mining Concepts and Techniques*. San Fransisco : Morgan Kaufmann Publishers.
- [5] Kusnawi. 2007. *Pengantar Solusi Data Mining*. Seminar Nasional Teknologi 2007. Yogyakarta: STIMIK AMIKOM.