

PENERAPAN DATA MINING DALAM MEMREDIKSI PEMBELIAN *T-SHIRT* MENGGUNAKAN ALGORITMA C4.

Muhammad Makmun Effendi¹, Wahyu Hadikristanto², Rosyati Adelia

Program Studi Teknik Informatika Sekolah Tinggi Teknologi Pelita Bangsa
¹effendiy@pelitabangsa.ac.id, ²wahyu.hadikristanto@pelitabangsa.ac.id

Disetujui, 05 Maret 2018

Abstrak

Penjualan *t-shirt* pada distro sangat diminati bagi kalangan muda karena harga yang ditawarkan relatif lebih murah. Banyaknya pengusaha distro disatu daerah maka diperlukan strategi bisnis yang baik dan benar. Dengan adanya prediksi pada stok *t-shirt* dapat digunakan sebagai bahan pertimbangan dalam pengambilan keputusan dan manajemen bisnis. Dari data penjualan *t-shirt* yang banyak, maka dapat dilakukan *data mining* dengan menggunakan Algoritma C4.5. Algoritma C4.5 merupakan algoritma klasifikasi data bertipe pohon keputusan. Dari hasil dari *data mining* ini diharapkan dapat memberikan sebuah keputusan untuk melihat pola prediksi penjualan *t-shirt*. Hasil yang diperoleh dalam prediksi mengacu pada atribut harga, karena dalam pembelian barang (*t-shirt*) pembeli umumnya mempertimbangkan harga.

Kata kunci: *Data Mining, Penjualan T-shirt, Decision Tree, Algoritma C4.5*

Abstract

Sales of t-shirts at the distribution are very attractive to young people because the prices offered are relatively cheaper. The number of distro entrepreneurs in one region is needed a good and correct business strategy. With the prediction of t-shirt stock, it can be used as consideration in business decision making and management. From the many t-shirt sales data, data mining can be performed using the C4.5 Algorithm. Algoritma C4.5 is a classification tree of decision tree type data algorithms. From the results of this data mining is expected to provide a decision to see the prediction pattern of t-shirt sales. The results obtained in the prediction refer to the price attribute, because in the purchase of goods (t-shirts) the buyer generally considers the price.

Keywords: *Data Mining, T-shirt Sales, Decision Tree, C4.5 Algorithm*

1. Pendahuluan

Dalam perkembangannya penjualan merupakan peluang usaha yang cukup tinggi, salah satunya penjualan pakaian. Terbukti dengan semakin banyaknya usaha pakaian yang ada, mulai dari departement store hingga ke distro-distro yang ada di berbagai daerah di Indonesia

tanpa terkecuali di Cikarang. Saat ini di wilayah Cikarang banyak sekali distro-distro anak muda yang kualitasnya tidak kalah jika dibandingkan dengan produk-produk pakaian yang ada di departement store. Dengan harga yang relatif lebih murah dibandingkan produk yang dijual di departement store, kalangan muda di daerah Cikarang umumnya membeli produk yang dijual di distro. Distro (*Distribution Outlet*) adalah jenis toko yang menjual pakaian dan aksesoris yang dititipkan oleh pembuat pakaian atau diproduksi sendiri.

Banyaknya pengusaha distro dalam satu daerah maka diperlukan strategi bisnis yang baik dan benar. Dalam strategi bisnis para pengusaha distro perlu adanya memiliki bayangan terhadap penjualan barang yang hendak dijual yaitu dengan melakukan prediksi penjualan. Hasil dari prediksi penjualan dapat digunakan sebagai bahan pertimbangan dalam mengambil keputusan dalam manajemen bisnis. Keputusan yang tepat akan sangat berpengaruh terhadap kelangsungan suatu usaha tersebut.

Dalam penelitian ini, obyek penelitian adalah Distro Crunchz yang merupakan salah satu distro yang bergerak dibidang pakaian dan aksesoris yang berdiri pada Agustus 2016 dan terletak di Jalan Raya Ciantra, Cikarang selatan. Distro crunchz itu sendiri menawarkan suatu

produk yang bermutu dengan kualitas yang terbaik seperti t-shirt, kemeja, jacket, tas, sepatu dan aksesoris lainnya. Distro ini harus mampu menyediakan barang yang sesuai dengan kebutuhan pasar. Untuk mencapai penjualan yang diharapkan dengan keuntungan yang besar maka distro crunchz co ini perlu menggunakan strategi bisnis yang benar dalam pengambilan keputusan. Pengambilan keputusan yang benar akan sangat berpengaruh terhadap kelangsungan distro tersebut. Dengan beberapa masalah penjualan yang dihadapi seperti halnya barang yang kurang begitu diminati oleh pelanggan sehingga barang tersebut tidak laku dijual dan menyebabkan masalah pada pendapatan distro crunchz co maka penulis melakukan analisis lebih lanjut sehingga dalam penelitian ini peneliti dapat mengambil judul “**Penggunaan Data Mining Dalam Memrediksi Pembelian T-Shirt Menggunakan Algoritma C4.5 Pada Distro Crunchz Co**”.

2. Tinjauan Studi

2.1. Teknologi Informasi

Menurut Fandy Ferdian Harryanto dan seng Hansun (2017:8), “Implementasi penggunaan algoritma C4.5 dapat dilakukan prediksi terhadap calon pegawai baru yang ingin masuk ke perusahaan dengan cara pembuatan pohon keputusan berdasarkan data-data yang sudah ada. Berdasarkan metode pengukuran akurasi ten-fold cross validation telah didapatkan hasil pengukuran tingkat keberhasilan prediksi calon pegawai baru sebesar 71 %”.

Menurut Fitriana Harahap (2015:1-6), “Dengan penggunaan Algoritma C4.5 dalam memprediksi penjualan cat pada Home Smart dimana kompetisi supplier dalam memasarkan produknya sehingga didapatkan pengampilan keputusan yang dapat dilihat dari hubungan supplier dan berapa banyaknya dana sponsor yang diberikan kepada perusahaan. Penerapan teknik klasifikasi data mining pada perusahaan digunakan dalam pengambilan keputusan dalam pembelian cat yang efektif”.

Menurut Sularno dan Putri Anggraini (2017:9), “Penggunaan algoritma C4.5 berguna sebagai gambaran jenis hama padi dan tingkat keanasannya. Dengan metode pohon keputusan dapat memberikan informasi sehingga petani dapat mengetahui cara penanganan hama secara tepat agar tidak terjadi kerusakan dan gagal panen dan dalam penggunaannya algoritma C4.5 dapat digunakan dalam pengambilan keputusan untuk mencari alternatif terbaik”.

Menurut Nurul Azwanti (2018:1-6), “Untuk merebut pangsa pasar dibutuhkan keistimewaan produk yang sesuai dengan selera konsumen dengan demikian produk sepeda

motor yang ditawarkan mampu menarik minat konsumen sepeda motor sehingga diperlukan analisa untuk melihat pola dari data penjualan guna menghasilkan prediksi penjualan menggunakan algoritma C4.5 dengan pengujian menggunakan WEKA dan didapat hasil penjualan masih didominasi motor matik”. Menurut Nanang Iriadi dan Nia Nuraeni (2016 : 6) “Untuk menerapkan algoritma C4.5 dengan seleksi atribut sehingga dapat mengurangi dimensi dari data sehingga penelitian kelayakan kredit pada bank mayapada jakarta didapatkan hasil penelitian untuk nilai akurasi algoritma klasifikasi C4.5 adalah 83.67% Hasil evaluasi menggunakan ROC Curve nilai AUC adalah 0.904 dengan tingkat klasifikasi sangat baik”

2.2. Distro

Distro (*Distribution Outlet*) adalah tempat/outlet/toko yang secara khusus mendistribusikan produk dari suatu komunitas. Biasanya berasal dari komunitas musik band-band independen atau istilahnya band indie dan komunitas skateboard. (Ade Rahma, 2017 : 2).

Distro telah menjadi sebuah tren setter yang menghadirkan gaya busana remaja dan anak muda dengan berbagai macam keunikan dan kelebihannya. Distro diartikan sebagai tempat atau toko yang berfungsi untuk menjual dan menerima titipan dari berbagai macam merk clothing company local yang memproduksi produknya sendiri. Distro memiliki sifat eksklusif atau cenderung tidak menjual banyak produk untuk setiap desainnya. (Christopher Sirait, 2017 : 53)

2.3. Definisi Data Mining

Kata *Mining* merupakan kiasan dari bahasa inggris yaitu *mine*. Jika *mine* berarti menambang sumber daya yang tersembunyi didalam tanah, maka *data mining* merupakan penggalian makna yang tersembunyi dari kumpulan data yang sangat besar. (Dennis Aprilla C, Donny Aji Baskoro, Lia Ambarwati, I Wayan Simri Wicaksana, 2013 : 42-43). *Data Mining* adalah studi pengumpulan, pembersihan, pemrosesan, analisis, dan memperoleh manfaat wawasan dari information. Ada banyak variasi dalam hal area masalah, aplikasi, formulasi, dan representasi data yang ditemui dalam aplikasi nyata. Karena itu, "Penambangan information" adalah istilah payung luas digunakan untuk menggambarkan berbagai aspek ini pengolahan information. (Charu.C Aggarwal, 2015 : 1)

Data mining adalah proses yang memperkerjakan satu atau lebih teknik pembelajaran komputer (*machine learning*)

untuk menganalisis dan mengekstraksi pengetahuan (*knowledge*) secara otomatis. Defenisi lain diantaranya adalah pembelajaran berbasis induksi (Inductiaon-based learning) adalah proses pembentukan defenisidefenisi konsep umum yang dilakukan dengan cara mengobservasi contoh-contoh spesifik dari konsepkonsep yang akan dipelajari. Knowledge Discovery in data mining merupakan satu langkah dari proses KDD. (Ira Maya Sari, Nelly Astuti Hasibuan, Natalia Silalahi, 2017 : 2)

Data mining merupakan proses iteratif dan interaktif untuk menemukan pola atau model baru yang sah (sempurna), bermanfaat dan dapat dimengerti dalam suatu database yang sangat besar (massive databases). Data mining digunakan untuk melakukan information discovery yang informasinya lebih ditujukan untuk seorang Data Analyst dan Business Analyst (dengan ditambah visualisasi tentunya). (Fajar Astuti Hermawati, 2013 : 3-4).

Data Mining didefinisikan sebagai proses menemukan pola dalam data. Prosesnya harus otomatis atau (lebih sering) semiotomatis. Pola-pola yang ditemukan harus bermakna bahwa mereka menghasilkan beberapa keuntungan, biasanya yang ekonomis. Data selalu hadir dalam jumlah besar. (Ian H. Witten, Eibe Frank, Mark A. Hall, 2011 : 5).

2.4. Algoritma C4.5

Algoritma C4.5 merupakan salah satu algoritma yang digunakan untuk membuat *decesion tree* berdasarkan *training* data. Algoritma C4.5 ini merupakan pengembangan dari algoritma ID3, algoritma tersebut digunakan untuk membentuk pohon keputusan (*decision tree*). Pohon keputusan dianggap sebagai salah satu pendekatan yang paling populer, dalam klasifikasi pohon keputusan terdiri dari sebuah node yang berbentuk akar, node akar tidak memiliki inputan. (Sularno, Putri Anggraini, 2017 : 2)

Dalam pemilihan atribut akar, yang didasarkan pada nilai *gain* tertinggi dari atribut-atribut yang ada. Adapun beberapa tahapan dalam membuat sebuah *decesion tree* dengan menggunakan algoritma C4.5 yaitu:

1. Mempersiapkan data training. Data training biasanya diambil dari data histori yang pernah terjadi sebelumnya atau disebut data masa lalu dan sudah dikelompokkan dalam kelas kelas tertentu.
2. Menghitung akar dari pohon. Akar akan diambil dari atribut yang dipilih, dengan cara menghitung nilai *gain* dari masing-masing atribut, nilai *gain* yang paling tinggi yang akan menjadi akar pertama. Sebelum

menghitung nilai *gain* dari atribut, hitung terlebih dahulu nilai *entropy*. Untuk menghitung nilai *entropy*

3. Kerangka Konsep

Metode *Data Mining* yang digunakan ialah metode klasifikasi. Pemodelan bertujuan mencari aturan klasifikasi, dimana aturan klasifikasi selanjutnya dijadikan acuan untuk menentukan produk terlaris dari data penjualan selama 8 bulan terakhir. Salah satu contoh yang mudah dan populer ialah dengan menggunakan *Decision Tree* yaitu salah satu metode klasifikasi yang paling populer karena mudah diinterpretasi. *Decision Tree* merupakan model prediksi menggunakan struktur berhirarki. *Decision Tree* merupakan implementasi dari sebuah sistem yang manusia kembangkan dalam mencari dan membuat keputusan untuk masalah-masalah tersebut dengan memperhitungkan berbagai macam faktor yang berkaitan didalam lingkup masalah. *Decision Tree* merupakan metode yang digunakan untuk mengklasifikasi data mining.

Klasifikasi merupakan suatu teknik untuk menentukan pola atau fungsi yang mendeskripsikan serta memisahkan kelas data yang satu dengan yang lainnya. Banyak algoritma yang dapat digunakan untuk membangun suatu *decision tree* salah satunya ialah algoritma C4.5. Algoritma C4.5 merupakan rasio perolehan (*gain ratio*), sebelum mengitung rasio perolehan perlu dilakukannya perhitungan nilai informasi dalam satuan bits dari sekumpulan objek yaitu dengan menggunakan konsep *Entropy*. Dalam tahapan ini akan dilakukan beberapa langkah pengujian data yaitu sebagai berikut :



Gambar 1 Langkah Pengujian Metode

4. Desain Penelitian/Metedologi

Penelitian menggunakan algoritma C4.5 dalam menentukan hasil keputusan dan akurasi. Perhitungan algoritma memiliki beberapa langkah untuk menghasilkan pohon keputusan antara lain menghitung nilai *entropy* dan *gain ratio*. *Gain ratio* paling tinggi akan menjadi akar, kemudian dari akar tersebut akan menghasilkan *internal node* atau *leaf*. Pada *internal node*

terdapat nilai variabel, jika salah satu label menghasilkan 0 (nol) maka terdapat *leaf* tetapi jika nilai pada label variabel tidak ada nilai 0 (nol) maka variabel tersebut menjadi *internal node*. Kemudian *internal node* tersebut dihitung kembali dengan cara yang sama dengan menentukan nilai *entropy* dan *gain ratio* untuk menentukan *leaf*. Langkah tersebut diulang hingga menemukan sebuah keputusan akhir atau jika atribut telah selesai dan belum menentukan keputusan, maka keputusan diambil ada *value* tertinggi.

5. Hasil Penelitian

Dalam pengujian akurasi data menggunakan *rapidminer*, akurasi yang dihasilkan dari data mining dalam prediksi stok *t-shirt* pada distro *crunchz co* ini adalah 55 %. Berikut adalah perhitungan akurasi tersebut :

The screenshot shows the Rapid Miner interface with the following content:

Akurasi = $\frac{t.laris + t.k.laris}{t.laris + f.k.laris + f.laris + t.k.laris} \times 100\%$

$= \frac{54+34}{54+26+46+34} \times 100\%$

$= 55\%$

	true Laris	true Kurang Laris	class precision
pred. Laris	54	46	54.00%
pred. Kurang Laris	26	34	55.67%
class recall	67.50%	42.50%	

Gambar 2. Gambar akurasi data

Hasil dari pohon keputusan menggunakan *rapidminer* dengan perhitungan manual yaitu hampir sama yaitu hasil *root node* pada pohon keputusan “Nama Brand”, *internal node* pertama yaitu “harga”, *internal node* kedua yaitu “Type T- shirt” dan *internal node* selanjutnya yaitu “ukuran”. Perhitungan menggunakan *rapidminer* hasil perhitungan mengacu pada attribute harga, karena dalam pembelian barang (*T-Shirt*) pembeli masih mempertimbangkan harga baru kemudian pemilihan jenis dan juga ukuran yang sesuai.

6. Kesimpulan

Berdasarkan hasil penelitian prediksi stok *t-shirt* pada distro *crunchz co* menggunakan *Algoritma C4.5* dapat diambil beberapa kesimpulan diantaranya ialah sebagai berikut :

1. Penggunaan *Algoritma C4.5* dapat digunakan sebagai prediksi pada pembelian *t-shirt* dengan *root node* yaitu nama brand kemudian diikuti dengan *internal node* yaitu harga. Beberapa brand berbeda dalam tiap pembelannya sehingga pemilih distro dapat mengetahui bayangan pembelian para customer sebelumnya

sebagai keputusan dalam penjualan *t- shirt* dimasa yang akan datang.

2. Hasil perhitungan menggunakan *rapidminer* mengacu pada attribute harga, karena dalam pembelian barang (*t-shirt*) pembeli masih mempertimbangkan harga baru kemudian pemilihan jenis dan juga ukuran yang sesuai dengan apa yang akan dibeli dengan nilai *accuracy* 55%.

Daftar Pustaka

- Aprilla, Dennis C, dkk. (2013). *Belajar Data Mining Dengan Rapid Miner*.
- Aggarwal, Charu C. (2015). *Data Mining The Text Book*. New York.
- Azwanti, Nurul. (2018). *Analisa Algoritma C4.5 Untuk Memprediksi PenjualanMotor Pada PT. Capella Dinamik Nusantara Cabang Muka Kuning*.
- Hermawati, FA. 2013. *Data Mining*. Yogyakarta: Penerbit Andi
- Harahap, Fitriana. (2015). *Penerapan Data Mining Dalam Memprediksi Pembelian Cat*.
- Harryanto, Fandy Ferdian dan Hansun,Seng. (2017). *Penerapan Algoritma C4.5 Untuk Memprediksi Penerimaan Calon Pegawai Baru Di PT. WISE*.
- Iriadi, Nandang dan Nia Nuraeni. (2016). *Kajian penerapan Metode Klasifikasi Data Mining Algoritma C4.5 Untuk Memprediksi Kelayakan Kredit Pada Bank Mayapada Jakarta*.
- Rismayanti. (2016). *Implementasi Algoritma C4.5 Untuk Menentukan PenerimaBeasiswa Di STT Harapan Medan*.
- Sularno dan Anggraini, Putri. (2017). *Penerapan Algoritma C4.5 Untuk Klasifikasi Tingkat Keganasan Hama Pada Tanaman Padi*.
- Sari, Ira Maya, dkk. (2017). *Implementasi Data Mining Penjualan Underware Menggunakan Metode Market Basket Analysis Pada PT. Suriatama Mahkota Kencana*.
- Witten, Ian H, dkk. (2011). *Data Mining Pratical Machine Learning Tools And Techniques Third Edition*.