

PENERAPAN DATA MINING UNTUK EVALUASI DATA PENJUALAN MENGGUNAKAN METODE *CLUSTERING* DAN *ALGORITMA HIRARKI DIVISIVE*

Yuda Irawan

Sistem Informasi, STMIK Hang Tuah Pekanbaru
Jl. Mustafa Sari No 5 Tangkerang Selatan, Pekanbaru
e-mail: yudairawan89@gmail.com

ABSTRAK

Semakin besar suatu perusahaan, semakin lama perusahaan itu berdiri, semakin banyak perusahaan memiliki cabang, tentunya semakin besar data yang dimiliki. Data-data tersebut bisa berupa data konsumen, data pembelian, data penjualan, data penggajian, dan masih banyak data-data yang lain lagi. Semua data tersebut biasanya akan tersimpan dalam database. Namun banyak perusahaan, bahkan divisi Teknologi Informasi atau Information Technology (IT) yang tidak menyadari betapa berharganya tumpukan data-data lama yang dihasilkan perusahaan dalam bertransaksi dan beraktifitas. Data mining merupakan ilmu yang mempelajari metode untuk menghasilkan pengetahuan atau menemukan pola untuk pengolahan di suatu data. Sehingga tidak hanya menjadi sebuah informasi saja, akan tetapi menjadi sebuah pengetahuan. Data Mining mempunyai beberapa metode diantaranya adalah clustering. Clustering merupakan metode yang sudah cukup dikenal dan banyak dipakai dalam data mining. Tujuan utama dari metode clustering ini adalah pengelompokan sejumlah data/obyek ke dalam cluster (group) sehingga dalam cluster akan berisi data yang sama dengan grupnya masing-masing. Dalam penelitian ini digunakan Algoritma Hirarki Divisive untuk membentuk klaster-klaster. Dari pola yang diperoleh diharapkan dapat memberikan pengetahuan untuk perusahaan Media World Pekanbaru sebagai alat pendukung untuk mengambil kebijakan.

Kata Kunci: Algoritma Hirarki Divisive, Clustering, Data Mining.

ABSTRACT

The larger the firm, the longer the company it stands, more and more companies have branches, of course, the greater the data that owned. These data can include customer data, purchasing data, sales data, payroll data, and many other data again. All data will usually be stored in the database. But many companies, even division of Information Technology or Information Technology (IT) are not aware of how precious pile of old data produced by the company in the transaction and activity. Data mining is the study of methods to produce knowledge or find patterns in the data for processing. Thus, not only be an information only, but will be a knowledge. Data Mining has several methods which are clustering. Clustering is a method that is well known and widely used in data mining. The main objective of this clustering method is the grouping of a number of data / objects into clusters (groups) so that the cluster will contain the same data to each group. This study used Hierarchical divisive algorithm to form clusters. From the pattern obtained is expected to provide knowledge to the company Media World Pekanbaru as a support tool for the take policy.

Keywords: Clustering, Data Mining, Hierarchical divisive algorithms.

I. PENDAHULUAN

MEDIA World Pekanbaru merupakan perusahaan *supplier* yang bergerak dibidang pelayanan jasa dan penjualan mesin dan bahan-bahan kebutuhan *digital printing*. Pelanggan merupakan target utama perusahaan dalam meningkatkan penjualan mesin dan jasa untuk mencari keuntungan perusahaan. Untuk mengolah data penjualan mesin dan bahan-bahan *digital printing* selama ini di perusahaan Media World Pekanbaru menggunakan aplikasi *Accurate 4*. Dengan menggunakan aplikasi ini perusahaan bisa mengolah data penjualan mulai dari penginputan hingga ke laporan data. Data yang diinput tersimpan di dalam *database*. Oleh karena itu, data yang ada di dalam *database* berjumlah ribuan data dan perusahaan memandangnya hanya sebagai arsip atau bahkan yang sudah menjadi berkas-berkas tidak terpakai dan bisa dihancurkan kapan saja. Hal tersebut tentu saja merupakan pandangan yang salah, sebab dengan penanganan data yang tepat, data-data tersebut dapat diberdayakan dan nantinya bisa digunakan untuk memprediksi strategi bisnis yang dijalankan di masa yang akan datang dan berguna untuk perusahaan tersebut.

Pada penelitian ini digunakan teknik *data mining* untuk menemukan pengetahuan pelanggan-pelanggan yang loyal terhadap perusahaan Media World Pekanbaru, kemudian bisa dijadikan informasi untuk pengambilan keputusan dan kebijakan oleh perusahaan dalam mengembangkan penjualan mesin dan bahan-bahan *digital printing* ataupun jasa baik jangka pendek, jangka menengah, dan jangka panjang.

Data mining merupakan bagian dari tahapan proses Knowledge Discovery in Database (KDD). Dengan data mining, kita dapat melakukan pengklasifikasian, memprediksi, memperkirakan dan mendapatkan informasi lain yang bermanfaat dari kumpulan data dalam jumlah yang besar [5]. Salah satu metode yang terdapat dalam *data mining* yang digunakan untuk penelitian ini adalah metode *clustering* karena merupakan metode yang sudah cukup dikenal dan banyak dipakai dalam *data mining*. Tujuan utama dari metode *clustering* ini adalah pengelompokan sejumlah data/obyek ke dalam *cluster (group)* sehingga dalam *cluster* akan berisi data yang sama dengan grupnya masing-masing. Dalam *clustering* metode ini berusaha untuk menempatkan obyek yang mirip (jaraknya dekat) dalam satu *cluster* dan membuat jarak antar *cluster* sejauh mungkin. Digunakannya metode *clustering* pada data penjualan mesin dan bahan-bahan digital printing untuk menjadikan data tersebut *cluster-cluster*, sehingga nantinya akan mendapatkan pengetahuan berupa pelanggan-pelanggan yang loyal pada perusahaan Media World Pekanbaru.

II. TINJAUAN PUSTAKA

A. Data Mining

Data mining adalah proses yang mempekerjakan satu atau lebih teknik pembelajaran komputer (*machine learning*) untuk menganalisis dan menghasilkan pengetahuan (*knowledge*) secara otomatis. Definisi lain diantaranya adalah pembelajaran berbasis induksi (*induction-based learning*) adalah proses pembentukan definisi-definisi konsep umum yang dilakukan dengan cara mengobservasi contoh-contoh spesifik dari konsep-konsep yang akan dipelajari. *Knowledge Discovery in Database (KDD)* adalah penerapan metode saintifik pada *data mining*. Dalam konteks ini *data mining* merupakan satu langkah dari proses *Knowledge Discovery in Database (KDD)* [4].

Data mining dan knowledge discovery in database (KDD) sering kali digunakan secara bergantian untuk menjelaskan proses penggalian informasi tersembunyi dalam suatu basis data yang besar. Sebenarnya kedua istilah tersebut memiliki konsep yang berbeda, tetapi berkaitan satu sama lain. Dan salah satu tahapan dalam keseluruhan proses KDD adalah data mining [1].

Kemampuan Data mining untuk mencari informasi bisnis yang berharga dari basis data yang sangat besar, dapat dianalogikan dengan penambangan logam mulia dari lahan sumbernya, teknologi ini dipakai untuk: 1. Prediksi trend dan sifat-sifat bisnis, dimana data mining mengotomatisasi proses pencarian informasi pemrediksi di dalam basis data yang besar. 2. Penemuan pola-pola yang tidak diketahui sebelumnya, dimana data mining “menyapu” basis data, kemudian mengidentifikasi pola-pola yang sebelumnya tersembunyi dalam satu sapanu [2].

B. Clustering

Pengelompokan (*clustering*) adalah bagian dari data mining yang bersifat tanpa arahan (*unsupervised*). *Clustering* merupakan suatu proses pembagian data kedalam kelas atau disebut *cluster* yang berdasarkan tingkat kemiripan. Dalam *clustering*, data-data yang mempunyai kemiripan digabungkan kedalam *cluster* yang sama, begitu juga sebaliknya data yang tidak mempunyai kemiripan akan dimasukkan dalam *cluster* yang berbeda [3].

C. Algoritma Hirarki Divisive

Langkah awal yang dilakukan dalam algoritma hirarki divisive adalah membentuk satu klaster yang besar yang dapat ditempati oleh semua obyek data. Pada langkah berikutnya, satu klaster besar tersebut dipisah-pisah menjadi beberapa klaster yang lebih kecil dengan karakteristik data yang mempunyai lebih besar kesamaan satu dengan yang lainnya, sehingga data yang tidak memiliki kemiripan yang cukup besar berada pada klaster yang terpisah. Algoritma Hirarki Divisive memiliki 3 tahapan:

1. Tahap Pertama

Pada tahap pertama, data pertama sekali diurutkan dari nilai paling rendah ke nilai paling tinggi.

2. Tahap Kedua

Tahap selanjutnya adalah penarikan anggota dari titik pusat klaster terdekat, anggota yang masuk merupakan nilai yang berada diantara batas bawah dengan batas atas di setiap titik klaster yang telah ditentukan pada langkah pertama.

3. Tahap Ketiga

Tahap ketiga merupakan tahap pembuktian dan perbandingan nilai titik pusat klaster yang terbentuk. Nilai titik pusat klaster yang terbentuk pada langkah kedua akan dibandingkan dengan nilai yang didapat pada langkah pertama.

III. METODOLOGI PENELITIAN

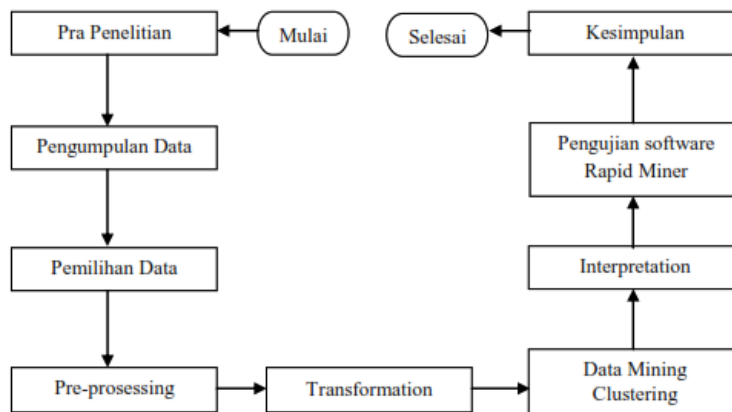
A. Metode Pengumpulan Data

Metode yang digunakan dalam pengambilan dan pengumpulan data adalah sebagai berikut :

- a. Pengamatan Langsung
Melakukan pengamatan langsung ke perusahaan Media World Pekanbaru untuk mendapatkan data yang dibutuhkan.
- b. Wawancara
Mengadakan wawancara dengan pihak-pihak yang ada berhubungan dengan penjualan mesin dan bahan-bahan *digital printing* di perusahaan Media World Pekanbaru yang berkaitan langsung dengan permasalahan yang sedang diteliti pada tugas akhir ini untuk memperoleh hasil yang lebih baik.
- c. Studi Pustaka
Studi pustaka merupakan sumber utama yang dapat dijadikan sebagai rujukan dari sumber data tersebut atau sebagai literatur-literatur.
- d. Browsing.
Sumber data melalui browsing merupakan salah satu pengumpulan data yang bersumber dari internet.

B. Tahap Penelitian

Tahapan yang dilakukan dalam penelitian ini adalah seperti pada Gambar 1.



Gambar. 1. Tahapan Penelitian

Adapun kegiatan-kegiatan penelitian adalah sebagai berikut:

a. Pra Penelitian

Pra penelitian merupakan tahap pertama yang dilakukan sebelum penelitian dimulai yaitu mencari literatur yang berhubungan dengan penelitian, mencari lokasi studi kasus dan pembuatan proposal ke Pusat Penelitian dan Pengabdian Masyarakat Perguruan Tinggi.

b. Pengumpulan Data

Data yang dikumpulkan dan digunakan dalam proses data mining *clustering* ini adalah data penjualan mesin dan bahan-bahan digital printing perusahaan Media World Pekanbaru tahun 2018. Data ini memiliki atribut No. Item, No. Faktur, Tgl Faktur, Kuantitas, Unit Barang, dan Nama Pelanggan.

c. Pemilihan Data

Pemilihan (seleksi) data dari sekumpulan data operasional perlu dilakukan sebelum tahap penggalian informasi dalam *Knowledge Discovery in Database* (KDD) dimulai. Data hasil seleksi yang akan digunakan untuk proses data mining, disimpan dalam suatu berkas, terpisah dari basis data operasional.

Maka hasil titik pusat dari langkah pertama adalah :

- a. Titik Pusat 1 = 17,83823529
- b. Titik Pusat 2 = 34,67647059
- c. Titik Pusat 3 = 196,8382353

Perhitungan rentang tertinggi
 = (Titik Pusat 3 – Titik Pusat 2) / 2
 = (196,8382353 - 34,67647059) / 2
 = 81,08088235

Batas rentang tertinggi dari titik pusat 2
 = (Titik Pusat 3 – jml rentang tertinggi)
 = 196,8382353 – 81,08088235
 = 115,7573529

Sehingga anggota klaster dari titik pusat 2 diambil dari rentang titik frekuensi data $26,25735294 \leq$ Anggota klaster titik pusat 2 < 115,7573529

Langkah 2

Perhitungan langkah kedua berdasarkan data dari langkah pertama yaitu : Titik Pusat 1

Anggota klaster titik pusat satu pada langkah pertama adalah :

[1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,2,2,2,2,2,2,2,2,2,2,2,2,2,3,3,3,3,3,3,3,3,3,3,4,4,4,4,4,4,4,5,5,5,5,6,6,7,7,7,7,7,7,7,8,8,8,9,9,9,10,10,10,10,10,10,11,11,11,11,13,13,13,14,14,14,14,15,15,15,16,18,18,20,20,21,21,22,22,23,24,24,24,25]

Jumlah semua item data pada klaster 1 = 135 item

Kemudian menentukan nilai median pada rentang data di atas untuk menentukan titik pusatnya.
 (N + 1) / 2 = 135 + 1 / 2 = 68

Anggota klaster titik pusat kedua pada langkah pertama adalah :

[27,27,28,28,29,29,29,29,29,29,30,30,30,31,32,32,33,34,34,36,39,40,42,42,42,44,45,46,47,48,48,49,50,51,52,52,54,58,63,63,72,74,78,86,96,96,98,103,103,111,113]

Jumlah semua item data pada klaster 1 = 51 item

Kemudian menentukan nilai median pada rentang data di atas untuk menentukan titik pusatnya.
 (N + 1) / 2 = 51 + 1 / 2 = 26

Nilai yang terletak pada item 26 adalah nilai tengah dari nilai 44
 Jadi nilai titik pusat 2 pada langkah kedua adalah 44

Hasil perhitungan dari titik pusat 2 pada langkah kedua kemudian dibandingkan dengan nilai titik pusat 2 pada langkah pertama, yaitu $44 \neq 34,67647059$, dari keterangan ini dapat disimpulkan bahwa titik pusat masih berubah.

Titik Pusat 3

Anggota klaster titik pusat ketiga pada langkah pertama adalah :

[117,121,123,141,143,150,183,194,200,201,205,206,213,242,252,287,317,358].

Jumlah semua item pada klaster 3 = 18 item

Median pada rentang data klaster 2 (N + 1) / 2 = 18 + 1 / 2
 = 9,5

Nilai yang terletak pada item 9,5 adalah nilai tengah dari nilai 200 dan 201
 = (200 + 201) / 2 = 200,5

Perhitungan jarak rentang tersebut adalah :

Titik Pusat 1

Rentang titik pusat 1

$$\begin{aligned} &= (\text{titik pusat 2} - \text{titik pusat 1}) / 2 \\ &= (44 - 3) / 2 \\ &= 20,5 \end{aligned}$$

Maka batas rentang anggota dari titik pusat 1 pada langkah 2 adalah:

$$\begin{aligned} &= (\text{titik pusat 1} + \text{jml rentang titik pusat 1}) \\ &= 3 + 20,5 \\ &= 23,5 \end{aligned}$$

Anggota kluster 1 pada langkah 2 < data frekuensi bernilai 23,5

Titik pusat 2

Rentang titik pusat 2

$$\begin{aligned} &= (\text{titik pusat 2} - \text{titik pusat 1}) / 2 \\ &= (44 - 3) / 2 \\ &= 20,5 \end{aligned}$$

Perhitungan rentang terendah

$$\begin{aligned} &= (\text{titik pusat 2} - \text{titik pusat 1}) / 2 \\ &= (44 - 3) / 2 \\ &= 20,5 \end{aligned}$$

Maka batas rentang terendah titik pusat 2 adalah

$$\begin{aligned} &= (\text{titik pusat 1} + \text{jml rentang titik pusat 2}) \\ &= 3 + 20,5 \\ &= 23,5 \end{aligned}$$

Batas rentang tertinggi dari titik pusat 2 adalah

$$\begin{aligned} &= (\text{titik pusat 3} - \text{jml rentang tertinggi}) \\ &= 200,5 - 78,25 \\ &= 122,25 \end{aligned}$$

Titik Pusat 3

Rentang titik pusat 3

$$\begin{aligned} &= (\text{titik pusat 3} - \text{titik pusat 2}) / 2 \\ &= (200,5 - 44) / 2 \\ &= 78,25 \end{aligned}$$

Maka batas rentang anggota dari titik pusat 3 pada langkah 2 adalah :

$$\begin{aligned} &= (\text{titik pusat 2} + \text{jml rentang titik pusat 3}) \\ &= 44 + 78,25 \\ &= 122,25 \end{aligned}$$

Anggota kluster 3 pada langkah 2 < data frekuensi bernilai 122,25.

Langkah 3

Langkah ketiga merupakan langkah yang digunakan jika hasil titik pusat pada perhitungan kedua masih berubah dari perhitungan pertama. Fungsi perhitungan yang digunakan sama dengan fungsi pada perhitungan kedua, yaitu dengan cara mencari nilai tengah (median).

Titik Pusat 1

Anggota kluster titik pusat satu pada langkah pertama adalah:

DAFTAR PUSTAKA

- [1] Nasari, F., Darma, S., & Informasi, S. (2015). Penerapan K-Means Clustering Pada Data Penerimaan Mahasiswa Baru, 6–8.
- [2] Fitri, Y. (2018). Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Pada Penerimaan Mahasiswa Baru (Studi Kasus : Universitas Islam Indragiri), 238-249
- [3] Khotimah, T., Teknik, D. F., Studi, P., Informatika, T., & Kudus, U. M. (2014). Pengelompokan Surat Dalam AL QUR ' AN Menggunakan Algoritma K-Means, 5(1), 83–88
- [4] Maburur A.G, R. L. (2012). Penerapan Data Mining untuk Memprediksi Program Studi Teknik Informatika Jurnal Komputer dan Informatika (Komputa), 1.
- [5] Yuli, M. (2017). Data Mining : Klasifikasi Menggunakan Algoritma C4.5, V2.i2 (213-219)