PENERAPAN *DATA MINING* PADA PENJUALAN UNIT PESAWAT DI PT. DIRGANTARA INDONESIA BANDUNG DENGAN METODE *CLUSTERING*

Beni Harta Ginting[1)], Selvia Lorena Br.Ginting [2)]
[1), 2)]Jurusan Sistem Komputer Universitas Komputer Indonesia Bandung
[1]binben182@gmail.com

ABSTRAK

Dalam dunia bisnis yang selalu dinamis dan penuh persaingan, para pelakunya harus senantiasa memikirkan cara-cara untuk terus survive dan jika mungkin mengembangkan skala bisnis mereka PT.Dirgantara Indonesia merupakan perusahaan yang bergerak dalam bidang pembuatan unit pessawat . Tidak hanya PT.Dirgantara Indonesia, masih cukup banyak perusahaan-perusahaan lain yang bergerak dibidang serupa. Hal tersebut tentu saja menimbulkan persaingan bisnis antar perusahaan Dalam rangka menghadapi persaingan bisnis dan meningkatkan pendapatan perusahaan, pihak terkait dalam perusahaan tersebut dituntut untuk dapat mengambil keputusan yang tepat dalam menentukan strategi pemasaran unit pesawat yang akan dijualnya. Ketersediaan data yang melimpah, kebutuhan akan informasi (atau pengetahuan) sebagai pendukung pengambilan keputusan untuk membuat solusi bisnis, dan dukungan infrastruktur di bidang teknologi informasi merupakan cikal-bakal dari lahirnya teknologi data mining. Data mining dimaksudkan untuk memberikan solusi nyata bagi para pengambil keputusan di dunia bisnis, untuk mengembangkan bisnis mereka. Salah satu metode yang terdapat dalam data mining yang digunakan dalam penelitian ini adalah pengelompokan (Clustering) dimana metode tersebut mengidentifikasi objek yang memiliki kesamaan karakteristik tertentu, dan kemudian menggunakan karakteristik tersebut sebagai "vektor karakteristik" atau "centroid". Hasil dari penelitian ini adalah aplikasi yang dapat mempermudah menganalisis sejumlah data yang besar guna membantu memberikan informasi berharga sebagai gambaran dasar pengambilan keputusan perusahaan.

Kata kunci : *persaingan bisnis, solusi bisnis, teknologi informasi, data mining, pengelompokan(clustering)*

1. PENDAHULUAN

Dalam dunia bisnis yang selalu dinamis dan penuh persaingan, para pelakunya harus senantiasa memikirkan cara-cara untuk survive dan jika mungkin mengembangkan skala bisnis mereka, untuk mencapai hal itu, ada tiga kebutuhan bisnis yang dapat dilakukan, yaitu penambahan jenis maupun peningkatan kapasitas produk, pengurangan biaya operasi perusahaan, dan peningkatan efektifitas pemasaran dan keuntungan. Agar bisa memenuhi kebutuhan-kebutuhan bisnis di atas banyak cara yang dapat ditempuh, salah satunya adalah dengan melalukan analisis data perusahaan. PT. Dirgantara Indonesia merupakan perusahaan yang bergerak dalam bidang industri pembuatan unit pesawat. Sebagai perusahaan besar, PT Dirgantara Indonesia menghasilkan berbagai macam unit pesawat dengan nama dan jenis yang berbeda. Tidak hanya PT Dirgantara Indonesia masih cukup banyak juga perusahaan lain yang bergerak di bidang yang sama. Hal tersebut tentu saja menimbulkan persaingan bisnis antar perusahaan. Produk yang diproduksi oleh PT Dirgantara Indonesia selain memiliki nama yang berbeda, setiap nama dari unit tersebut pun memiliki jenis yang beragam

sesuai dari permintaan konsumen. Produk yang diproduksi tersebut dipasarkan ke banyak kota-kota di Indonesia bukan hanya di Indonesia tetapi dipasarkan juga ke luar negeri Hal ini tentu akan menghasilkan data penjualan produk yang begitu banyak.

Untuk menghadapi persaingan bisnis yang dinamis pihak perusahaan dituntut agar dapat mengambil keputusan yang tepat dalam menentukan strategi pemasaran produk yang akan dijualnya Agar dapat melaksanakan hal tersebut, perusahaan memerlukan informasi yang cukup untuk dapat dianalisa lebih lanjut. Informasi yang dihasilkan tentunya adalah hasil analisis dari pengolahan data penjualan pada perusahaan tersebut. Data penjualan yang sudah ada akan diolah atau dianalisis untuk mengetahui tingkat kecenderungan konsumen di setiap tempat tujuan pemasaran produk pada faktor ketertarikannya. Dari pengolahan data tersebut akan diperoleh suatu pola konsumsi masyarakat terhadap produk dari perusahaan tersebut. Data mining didefinisikan sebagai satu set teknik yang digunakan secara otomatis untuk mengeksplorasi secara menyeluruh dan membawa ke permukaan relasi-relasi yang kompleks pada set data yang sangat besar. Set data yang dimaksud di sini adalah set data yang berbentuk tabulasi, seperti yang banyak diimplementasikan dalam teknologi manajemen basis data relasional. Akan tetapi, teknik-teknik data mining dapat juga diaplikasikan pada representasi data yang lain, seperti domain data spatial, berbasis text, dan multimedia (citra). Data mining dapat juga didefinisikan sebagai "pemodelan dan penemuan pola-pola yang tersembunyi dengan memanfaatkan data dalam volume yang besar" Data mining menggunakan pendekatan discovery-based dimana pencocokan pola (pattern-matching) dan algoritma-algoritma yang lain digunakan untuk menentukan relasi-relasi kunci di dalam data yang diekplorasi Data mining merupakan komponen baru pada arsitektur Sistem Pendukung Keputusan (SPK) di perusahaan-perusahaan.

2. TEORI PENUNJANG

Data Mining (DM) adalah salah satu bidang yang berkembang pesat karena besarnya kebutuhan akan nilai tambah dari database skala besar yang makin banyak terakumulasi sejalan dengan pertumbuhan teknologi informasi. Definisi umum dari DM itu sendiri adalah serangkaian proses untuk menggali nilai tambah berupa pengetahuan yang selama ini tidak diketahui secara manual dari suatu kumpulan data.

Data mining adalah kombinasi secara logis antara pengetahuan data, dan analisa statistik yang dikembangkan dalam pengetahuan bisnis atau suatu proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, tiruan dan machine-learning untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat bagi pengetahuan yang terkait dari berbagai database besar**.**

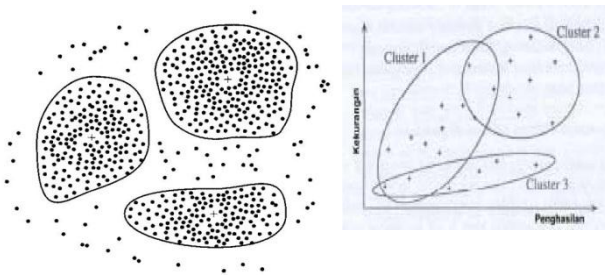
Pengelompokan (Clustering) merupakan tugas deskripsi yang banyak digunakan dalam mengidentifikasi sebuah himpunan terbatas pada kategori atau cluster untuk mendeskripsikan data yang ditelaah. Kategori-kategori ini dapat bersifat eksklusif dan ekshaustif mutual, atau mengandung representasi yang lebih kaya seperti kategori yang hirarkis atau saling menumpuk (overlapping).

Berbeda dengan association rule mining dan classification dimana kelas data telah ditentukan sebelumnya, clustering melakukan pengelompokan data tanpa berdasarkan kelas data tertentu. Bahkan clustering dapat dipakai untuk memberikan label pada kelas data yang belum diketahui itu. Karena itu clustering sering digolongkan sebagai metode unsupervised learning. Prinsip dari clustering adalah memaksimalkan kesamaan antar anggota satu kelas dan meminimumkan kesamaan antar kelas/cluster. Clustering dapat dilakukan pada data yang memiliki beberapa atribut yang dipetakan sebagai

ruang multidimensi. Ilustrasi dari clustering dapat dilihat pada Gambar 2.4 dimana lokasi, dinyatakan dengan bidang dua dimensi, dari pelanggan suatu toko dapat dikelompokkan menjadi beberapa cluster dengan pusat cluster ditunjukkan oleh tanda positif (+).

Banyak algoritma clustering memerlukan fungsi jarak untuk mengukur kemiripan antar data, diperlukan juga metode untuk normalisasi bermacam atribut yang dimiliki data.

Beberapa kategori algoritma clustering yang banyak dikenal adalah metode partisi dimana pemakai harus menentukan jumlah k partisi yang diinginkan lalu setiap data di tes untuk dimasukkan pada salah satu partisi, metode lain yang telah lama dikenal adalah metode hierarki yang terbagi dua lagi : bottom-up yang menggabungkan cluster kecil menjadi cluster lebih besar dan top-down yang memecah cluster besar menjadi cluster yang lebih kecil. Kelemahan metode ini adalah bila salah satu penggabungan/pemecahan dilakukan pada tempat yang salah, tidak dapat didapatkan cluster yang optimal. Pendekatan yang banyak diambil adalah menggabungkan metode hierarki dengan metode clustering lainnya seperti yang dilakukan oleh Chameleon.



Gambar 2.4 Ilustrasi *Clustering*

K-Means Clustering merupakan metode untuk mengklasifikasikan atau mengelompokan objek-objek (data) kedalam K-group (cluster) berdasarkan atribut tertentu. Pengelompokan data dilakukan untuk memperhitungkan jarak terdekat antara data-data(objek observasi) dengan pusat cluster (centroid). Prinsip utama dari metode ini adalah menyusun K buah centroid atau rata-rata (mean) dari sekumpulan data berdimensi N, dimana metode ini mensyaratkan nilai K sudah diketahui sebelumnya (apriori). Algoritma K-Means dimulai dengan pembentukan prototype cluster diawal kemudian secara iteratif prototipe cluster tersebut diperbaiki sehingga tercapai kondisi konvergen, yaitu kondisi dimana tidak terjadi perubahan yang signifikan terhadap prototipe cluster. Perubahan ini diukur dengan menggunakan fungsi objektif D yang umumnya didefinisikan sebagai jumblah atau rata-ratajarak tiap item data dengan centroid groupnya.

Algoritma K-Means

K-Means termasuk dalam partitioning clustering yaitu setiap data harus masuk dalam cluster tertentu dan memungkinkan bagi setiap data yang termasuk dalam cluster tertentu pada suatu tahapan proses, pada tahapan berikutnya berpindah ke cluster yang lain. K-Means memisahkan data ke k daerah bagian yang terpisah, dimana k adalah bilangan integer positif. Algoritma K-Means sangat terkenal karena kemudahan dan kemampuannya untuk mengklasifikasi data besar dan outlier dengan sangat cepat. Berikut adalah langkah-langkah algoritma K-Means :

1. Penentuan Pusat Cluster Awal

Dalam menentukan n buah pusat cluster awal dilakukan pembangkitan bilangan random yang merepresentasikan urutan data input. Pusat awal cluster didapatkan dari data sendiri bukan dengan menentukan titik baru, yaitu dengan menrandom pusat awal dari data.

2. Perhitungan Jarak Dengan Pusat Cluster

Untuk mengukur jarak antara data dengan pusat cluster digunakan

Euclidian distance.

1) Algoritma perhitungan jarak data dengan pusat cluster

2) Ambil nilai data dan nilai pusat cluster

Hitung Euclidian distance data dengan tiap pusat cluster

3 Pengelompokkan Data

Jarak hasil perhitungan akan dilakukan perbandingan dan dipilih jarak terdekat antara data dengan pusat cluster, jarak ini menunjukkan bahwa data tersebut berada dalam satu kelompok dengan pusat cluster terdekat.

1) Ambil nilai jarak tiap pusat cluster dengan data
2) Cari nilai jarak terkecil
3) Kelompokkan data dengan pusat cluster yang memiliki jarak terkecil.
4. Penentuan pusat cluster baru

Untuk mendapatkan pusat cluster baru bisa dihitung dari rata-rata nilai anggota cluster dan pusat cluster. Pusat cluster yang baru digunakan untuk melakukan iterasi selanjutnya, jika hasil yang didapatkan belum konvergen. Proses iterasi akan berhenti jika telah memenuhi maksimum iterasi yang dimasukkan oleh User atau hasil yang dicapai sudah konvergen (pusat cluster baru sama dengan pusat cluster lama).

Proses Clustering Algoritma K-Means

Pada tahap ini akan dilakukan proses utama yaitu segmentasi data nilai
yang diakses dari database yaitu sebuah metode clustering algoritma K-Means. Berikut ini merupakan diagram flowchart dari algoritma K-Means dengan asumsi bahwa parameter input adalah jumlah data set sebanyak n data dan jumlah inisialisasi centroid K=4 sesuai dengan jumlah jurusan yang ada di PT Dirgantara Indonesia yaitu Boeing, Hovercraft, Helikopter, dan Aircraft
beberapa langkah yang dilalui oleh clustering algoritma K-Means memuat bagian-bagian sebagai berikut ini:

1. N data : data set yang akan diolah sebanyak N data dimana N data tersebut terdiri dari atribut-atributnya N(Boeing,Hovercraft,Helikopter, dan Aircraft) yang berarti data N memiliki atribut sebanyak 4

2. K Centroid : Inisialisasi dari pusat cluster data adalah sebanyak K dimana pusat-pusat awal tersebut digunakan sebagai banyaknya kelas yang akan tercipta. Centroid didapatkan secara random dari N data set yang ada.

3. Euclidian Distance: merupakan jarak yang didapat dari perhitungan antara semua N data dengan K centroid dimana akan memperoleh tingkat kedekatan dengan kelas yang terdekat dengan populasi data tersebut. Jarak euclidian untuk menandai adanya persamaan antar tiap cluster dengan jarak minimum dan mempunyai persamaan yang lebih tinggi. Euclidian matrik.

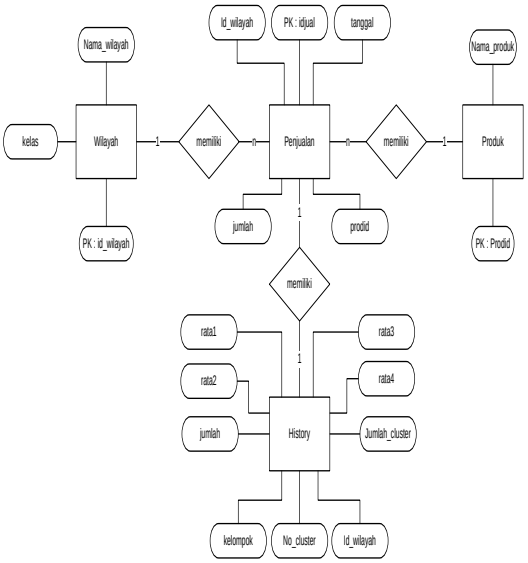
$$d(x,y) = |x - y| = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2} .$$

Dimana :
x : Titik data pertama
y : Titik data kedua,
n : Jumlah karakteristik (attribut) dalam terminologi data mining,
d(x,y) : Euclidian distance yaitu jarak antara data pada titik x dan titik y menggunakan kalkulasi matematika

4. Pengelompokkan Data: setelah sejumlah populasi data tersebut menemukan kedekatan dengan salah satu centroid yang ada maka secara otomatis populasi data tersebut masuk kedalam kelas yang memiliki centroid yang bersangkutan.
5. Update Centroid Baru: tiap kelas yang telah tercipta tadi melakukan update
centroid baru. Hal ini dilakukan dengan menghitung nilai rata-rata dari kelas masing-masing. Apabila belum memenuhi optimal hasil proses pengukuran ecluidian distance dilakukan kembali.
6. Batas Iterasi : apabila dalam proses clustering belum optimal namun sudah memenuhi batas iterasi maka proses dihentikan

3. PERANCANGAN SISTEM

ERD



Gambar 3.1 ERD yang digunakan

4. IMPLEMENTASI DAN PENGUJIAN

Implementasi Aplikasi

a. Implementasi Antar Muka Pengguna (*User Interface*) awal

Berikut adalah tampilan antar muka pengguna (*user interface*) aplikasi.



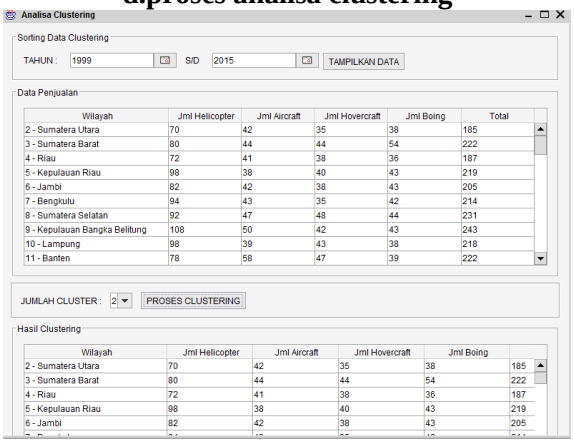
Gambar 4.1 Tampilan awal

b. Implementasi tampilan utama

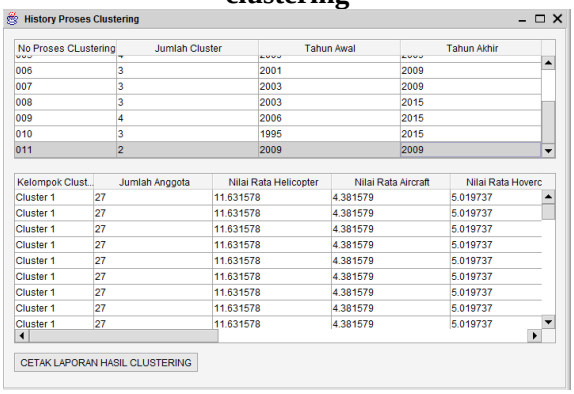


Gambar 4.2 Tampilan utama

d.proses analisa clustering



Gambar 4.3 Tampilan form proses clustering



Gambar 4.4 tampilan form history proses clustering dan laporan akhir



DAFTAR NAMA WILAYAH PENJUALAN PER KELOMPOK
APLIKASI DATA MINING PADA PENJUALAN HELICOPTER, AIRCRAFT, HOVERCRAFT & BOING
PADA PT. DIRGANTARA INDONESIA (INDONESIAN AEROSPACE)
PERIODE TAHUN 2009 S/D 2009

Kelompok : Cluster 1
Jumlah : 27
Rata-Rata : 11.631578, 4.381579, 5.019737, 6.342105

Wilayah	Jml Helicopter	Jml Aircraft	Jml Hovercraft	Jml Boing	Rata Helicopter	Rata Aircraft	Rata Hovercraft	Rata Boing
Samarinda	6	1	1	6	4.382078573671	4.877559008342	5.541431155406	5.171362262486
Banjar	2	2	4	2	8.827312807248	3.285343259771	3.442036391949	3.786936577351
Kalimantan Barat	4	2	3	4	6.087284242188	2.39411809707	3.451978284429	3.227139233048
Kupang	2	2	3	2	8.578331751877	2.408664963294	2.486527811541	3.05761243479
Nusa Tenggara Barat	4	2	5	4	6.908071481082	4.552671054603	5.024886986485	4.734519728629
Mataram	4	1	5	4	7.284886349007	4.566973667862	5.257451257016	5.067201380922
Papua	4	1	4	3	7.211283588147	3.659314367217	4.136116336862	3.891426041263
Jawa Tengah	10	5	2	3	3.456102009582	8.767487779263	8.01304600728	6.488429596022
Jawa Barat	6	6	1	4	4.122719421748	8.017065203606	5.197846015197	4.169198501212
Magelang	8	2	1	3	3.956612927601	8.021639886658	5.631709729474	4.345379975514
Sulawesi Tenggara	2	1	2	2	8.752291787246	1.709945930734	2.215278434598	3.08997795687
Cimahi	2	3	3	5	7.378396590096	3.01048927558	3.783615221142	4.322976316344
Balikpapan	10	3	1	2	4.424209112054	8.225755015886	7.513896054762	6.007339205258
Jambi	4	1	1	3	6.708398606644	2.095375151467	2.471327349355	2.478547283112
Tanjung Pinang	2	2	3	2	8.578331751877	2.408664963294	2.486527811541	3.05761243479
Pontianak	2	1	2	2	8.752291787246	1.709945930734	2.215278434598	3.08997795687

5. KESIMPULAN DAN SARAN

Setelah melalui tahap implementasi dan pengujian sistem dilakukan maka dapat diambil beberapa kesimbulan sebagai berikut:

Setelah melakukan analisis, perancangan, dan pengujian, maka dapat diperoleh kesimpulan sebagai berikut :

1. Aplikasi yang dibangun dapat membantu PT. Dirgantara Indonesia sebagai gambaran bagi pengambilan keputusan perusahaan dalam rangka mendapatkan pola penjualan produk yang dihasilkan.
2. Pengolahan data yang dilakukan dapat menghasilkan informasi yang cukup untuk dapat dianalisa lebih lanjut.
3. Aplikasi yang dibangun dapat mengurangi penumpukan data yang kurang dimanfaatkan sebelumnya.
4. Aplikasi yang dibangun dapat memberi gambaran dari penjualan produk ,wilayah dan tahun penjualanya dengan grafik dan diagram yang ada.
5. Aplikasi yang dibangun dapat Memudahkan perusahaan dalam mengehtahui informasi dari penjualan unit di PT Dirgantara Indonesia

Adapun saran untuk mengoptimalkan aplikasi di masa mendatang adalah sebagai berikut:

1. Berdasarkan kesimpulan di atas, hal yang diharapkan kedepan adalah agar aplikasi ini dapat dikembangkan lebih jauh dengan pengolahan data yang lebih besar dan luas sehingga aplikasi ini benar-benar dapat digunakan sebagai salah satu gambaran dalam pengambilan keputusan perusahaan yang lebih akurat dan berguna.

6. DAFTAR PUSTAKA

[1] Andi (Rabu, 4 Maret 2009), Data Mining dan Web Mining, http://andyku.wordpress.com/2008/04/17/data-mining-dan-web-mining/

[2] Andri Heryandi, S.T. SQL (Structured Query Language) dengan delphi,

Diktat Delphi, 1-55

[3] Iko Pramudiono. (Rabu, 4 Maret 2009), Pengantar Data Mining: Menambang Permata Pengetahuan di Gunung Data, http://ilmukomputer.org/2008/11/25/pengantardata-mining/

[4] Kadir, Abdul,(2001), Konsep Dan Tuntunan Basis Data , Yogyakarta : Andi.

[5] Sandy Kurniawan, Taufiq Hidayat (Rabu, 5 Agustus 2009), PENERAPAN DATA MINING

DENGAN METODE INTERPOLASI UNTUK MEMPREDIKSI MINAT KONSUMEN ASURANSI (Studi

[7] Susanti, Azhar, (2003), Sistem Informasi Management, Yogyakarta: Andi.

Yusta Noverison. (Selasa, 3 Maret 2009), Data Mining, http://yustanoverison.blogspot.com/2008/05/data-mining.html