



**PENERAPAN *DATA MINING* PADA *TREND* PENJUALAN MENGGUNAKAN METODE *CLUSTERING***

**STUDY KASUS PT.INDOMARCO PALEMBANG**

**PROPOSAL PENELITIAN**

Diajukan guna melakukan penelitian skripsi

**DISUSUN OLEH :**

**SUTRISNO**

**09142305**

**PROGRAM STUDI TEKNIK INFORMATIKA**

**FAKULTAS ILMU KOMPUTER**

**UNIVERSITAS BINA DARMA**

**2013**

**PROPOSAL PENELITIAN**

**ABSTRAK**

Dari tahun ketahun perkembangan sistem informasi semakin meningkat dan data mining merupakan salah satu bagian dari sistem informasi itu sendiri. Data mining adalah proses pencarian secara otomatis informasi yang berguna dalam tempat penyimpanan data berukuran besar. Data mining memiliki beberapa teknik salah satunya teknik *clustering*, dengan melakukan teknik *clustring* data-data akan dikelompokkan kedalam beberapa *cluster* yang sudah didefenisikan sebelumnya. Pada penelitian ini yang akan dibahas adalah *trend* penjualan makanan dan minuman pada PT.Indomarco palembang. PT. Indomarco mendistribusikan makanan dan minuman ke mitra-mitra yang sudah terjalin kerjasama dengan PT.Indomarco. Salah satu mitra yang dibahas pada kasus ini adalah PT. Indomaret cabang Palembang. Perusahaan ingin mengetahui jumlah penjualan makanan dan minuman pada masing-masing toko indomaret dengan melihat trend penjualan pada toko indomaret PT.Indomarco dapat memperoleh analisa data penjualan yang paling banyak diminati masyarakat terhadap penjualan makanan dan minuman.

Kata Kunci : *Data Mining, Clustering*, Persediaan Barang

**KATA PENGANTAR**

 Puji Syukur saya panjatkan kepada Allah SWT yang telah memberikan rahmat dan Hidayah-Nya sehingga penulis dapat meneyelesaikan laporan Proposal dengan judul “Penerapan data *mining* pada *trend* penjualan menggunakan metode *clustering* pada PT.Indomarco Palembang “.

Penulis mengucapkan banyak terimakasih kepada :

1. Prof. Dr. Ir. Buchori Rahman, MSc Selaku Rektor Universitas Bina Darma Palembang.
2. M Izman Herdiansyah ST. MM. Ph. D selaku dekan Ilmu Komputer.
3. Syahril Rizal, ST, MM, M.Kom selaku kepala program Studi Teknik Informatika
4. Afriyudi, M.Kom selaku dosen pembimbing utama pembuatan laporan skripsi.
5. Widyanto, M.M., M.Kom sebagai dosen pembimbing ke dua laporan skripsi ini.
6. Segenap Dosen fakultas ilmu komputer Universitas Bina Darma Palembang
7. Rekan-rekan Program Studi teknik informatika Universitas Bina Darma Palembang
8. Kedua orang tuaku yang selalu mendoakan disetiap langkahku.
9. Dan semua pihak yang tidak dapat disebutkan satu persatu yang telah membantu penulisan laporan Proposal ini

Penulis juga menyadari laporan ini banyak kekurangan baik didalam penyusunan maupun penulisan. Oleh karena itu saran dan kritik sangat penulis harapkan demi kesempurnaan laporan berikutnya.

Akhirnya penulis hanya berharap semoga laporan proposal ini dapat bermanfaat bagi kita semua. Amin

Palembang, Mei 2013

 Penulis

**DAFTAR ISI**

**Halaman**

**HALAMAN JUDUL i**

**LEMBAR PENGESAHAN ii**

**ABSTRAK iii**

**KATA PENGANTAR iv**

**DAFTAR ISI vi**

**DAFTAR GAMBAR vii**

**DAFTAR TABEL viii**

**BAB I PENDAHULUAN 1**

 1.1 Latar Belakang 1

 1.2 Rumusan Masalah 3

 1.3 Batasan Masalah 3

 1.4 Tujuan dan Manfaat penelitian 3

 1.4.1 Tujuan 3

 1.4.2 Manfaat 4

 1.5 Model pembangunan perangkat lunak 4

**BAB II TINJAUAN PUSTAKA 6**

 2.1 Landasan Teori 6

 2.1.1 Pengertian Data Mining 6

 2.1.2 Tahapan Data Mining 7

 2.1.3 Pengelompokkan (Clustering) 9

 2.2 Penelitian Sebelumnya 12

**BAB III METODOLOGI PENELITIAN 13**

 3.1 Waktu dan Tempat Penelitian 13

 3.2 Alat dan Bahan Penelitian 14

 3.3 Metode Pengumpulan Data 14

 3.4 Metode penerapan Data Mining 15

 3.4.1 Teknik Clustering 15

 3.4.2 *Agglomerative Hierarchical Clustering Algorithm*(*AHC*

 *algorithm*) 15

 3.4.3 Contoh Penerapan *Clustering* Menggunakan

 Algoritma AHC 16

**DAFTAR PUSTAKA 17**

**LAMPIRAN-LAMPIRAN**

**DAFTAR GAMBAR**

1 Paradigma Waterfall 5

2 Tahapan pada proses *knowledge discovery* 8

3 Ilustrasi *Clustering* 11

4 Pengelompokkan himpunan data menjadi tiga *cluster* 12

**DAFTAR TABEL**

1 Data Penjualan Makanan dan Minuman Per Wilayah Tanggal 1 Januari 2013 16

2 Pembuatan *cluster* pertama 17

3 Hasil *clustering* sebanyak tiga *cluster* 18

**PENERAPAN *DATA MINING* PADA *TREND* PENJUALAN MENGGUNAKAN METODE *CLUSTERING***

**STUDY KASUS PT.INDOMARCO PALEMBANG**

**1. PENDAHULUAN**

**1.1 Latar Belakang**

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. 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 melakukan analisis data perusahaan.

PT. Indomarco Palembang merupakan perusahaan yang bergerak dalam bidang distribusi makanan dan minuman. Tidak hanya PT. Indomarco Palembang, masih cukup banyak perusahaan-perusahaan lain yang bergerak di bidang serupa. Hal tersebut tentu saja menimbulkan persaingan bisnis antar perusahaan.

PT. Indomarco mendistribusikan makanan dan minuman ke mitra-mitra yang sudah terjalin kerjasama dengan PT.Indomarco. Salah satu mitra yang dibahas pada kasus ini adalah PT. Indomaret cabang Palembang. Perusahaan ingin mengetahui jumlah penjualan makanan dan minuman pada masing-masing toko indomaret dengan melihat trend penjualan pada toko indomaret PT.Indomarco dapat memperoleh analisa data penjualan yang paling banyak diminati masyarakat terhadap penjualan makanan dan minuman.

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.

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 cikalbakal dari lahirnya teknologi *data mining*. *Data mining* dimaksudkan untuk memberikan solusi nyata bagi para pengambil keputusan di dunia bisnis, untuk mengembangkan bisnis mereka.

Untuk itu penulis tertarik mengangkat permasalahan ini kedalam penelian skripsi dengan judul “**PENERAPAN *DATA MINING* PADA *TREND* PENJUALAN MENGGUNAKAN METODE *CLUSTERING* STUDY KASUS PT.INDOMARCO PALEMBANG.**

**1.2 Perumusan Masalah**

Dengan mengacu pada latar belakang masalah diatas, maka permasalahan

yang dibahas dan diteliti adalah :

**“Bagaimana cara menerapkan *Data Mining* pada *trend* penjualan menggunakan metode *clustering* study kasus PT.Indomarco Palembang?”.**

**1.3 Batasan Masalah**

Adapun batasan masalah yang akan dibahas adalah :

1. Data yang akan dianalisa merupakan data penjualan produk makanan dan minuman di PT. Indomarco Palembang.
2. Algoritma yang digunakan dalam metode *clustering* yang dilakukan adalah *Agglomerative Hierarchical Clustering Algorithm (AHC)*.
3. Hasil dari analisa tersebut adalah terbentuknya pola penjualan produk makanan dan minuman di PT. Indomarco Palembang.
4. Metode yang digunakan adalah metode aliran terstruktur dimana *tools* yang digunakan adalah *Data Flows Diagram* (DFD) dan *Entity* *Relationship Diagram* (ERD).

**1.4 Tujuan dan Manfaat Penelitian**

**1.4.1 Tujuan Penelitian**

Berdasarkan permasalahan yang diteliti, maka maksud dari penulisan skripsi ini adalah

1. Untuk menerapkan *Data Mining* pada penjualan produk makanan dan minuman di PT. Indomarco Palembang menggunakan metode *clustering*.
2. Untuk melihat trend penjualan yang paling banyak diminati konsumen khususnya penjualan makanan dan minuman.

**1.4.2 Manfaat Penelitian**

Manfaat dari penelitian ini antara lain :

1. Mempermudah menganalisis data yang besar

2. Membantu memberikan informasi dari data penjualan yang diolah

3. Memberikan gambaran dalam pengambilan keputusan

**1.5 Model Pembangunan Perangkat Lunak**

Teknik analisis data dalam pembuatan perangkat lunak menggunakan pemodelan perangkat lunak dengan paradigma *waterfall* seperti terlihat pada Gambar 1, yang meliputi beberapa proses diantaranya:

*a. System / Information Engineering*

Merupakan bagian dari sistem yang terbesar dalam pengerjaan suatu proyek,

dimulai dengan menetapkan berbagai kebutuhan dari semua elemen yang

diperlukan sistem dan mengalokasikannya kedalam pembentukan perangkat

lunak.

*b. Analisis*

Merupakan tahap menganalisis hal-hal yang diperlukan dalam pelaksanaan

proyek pembuatan perangkat lunak.

*c. Design*

Tahap penerjemahan dari data yang dianalisis kedalam bentuk yang mudah

dimengerti oleh *user*.

*d. Coding*

Tahap penerjemahan data atau pemecahan masalah yang telah dirancang

keadalam bahasa pemrograman tertentu.

e. Pengujian

Merupakan tahap pengujian terhadap perangkat lunak yang dibangun.

*f. Maintenance*

Tahap akhir dimana suatu perangkat lunak yang sudah selesai dapat mengalami perubahan–perubahan atau penambahan sesuai dengan permintaan *user.*



 **Sumber:Al-Fatah, (2007:13)**

Gambar 1 Paradigma *Waterfall*

**II. TINJAUAN PUSTAKA**

**2.1 Landasan Teori**

**2.1.1 Pengertian Data Mining**

*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. (Iko : 2009 : 25)

Menurut Abdul Kadir (2007:30),  *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.

 *Data mining* meliputi tugas-tugas yang dikenal sebagai ekstraksi pengetahuan, arkeologi data, eksplorasi dalam pemrosesan pola data dan memanen informasi. Semua aktifitas ini dilakukan secara otomatis dan mengizinkan adanya penemuan cepat bahkan oleh non programmer. *Data mining* cerdas menemukan informasi di dalam data *warehouse* dimana laporan dan query tidak bisa diungkapkan secara efektif. Piranti *data mining* menemukan pola-pola di dalam data dan bahkan menyimpulkan aturan dari data tersebut Definisi umum dari *Data Mining* adalah proses pencarian pola-pola yang menarik (*hidden pattern*) berupa pengetahuan (*knowledge*) yang tidak diketahui sebelumnya dari suatu kumpulan data dimana data tersebut dapat berada dalam *database, data warehouse*, atau media penyimpanan informasi yang lain. *Data Mining* merupakan proses analisis terhadap data dengan penekanan menemukan informasi yang tersembunyi pada sejumlah besar data yang disimpan ketika menjalankan bisnis perusahaan. Sumber : (*Data* *Mining dan Web Mining*, <http://andyku.wordpress.com/2008/04/17/data-mining-dan> web-mining, diakses tanggal 10 April 2013).

**2.1.2 Tahapan Data Mining**

Menururut Syaifullah (2010:15), Dalam aplikasinya, *data mining* sebenarnya merupakan bagian dari proses *Knowledge Discovery in Database* (KDD), bukan sebagai teknologi yang utuh dan berdiri sendiri. Data mining merupakan suatu bagian langkah yang penting dalam proses KDD terutama berkaitan dengan ekstraksi dan penghitungan pola-pola dari data yang ditelaah, seperti ditunjukan oleh gambar 1 dibawah ini :



**Sumber:Al-Fatah, (2007:13)**

**Gambar 2. Tahapan pada proses *knowledge discovery***

1. *Data cleaning*

Untuk menghilangkan data *noise* (data yang tidak relevan/berhubungan langsung dengan tujuan akhir proses *data* *mining*, misal: *data mining* yang bertujuan untuk menganalisa hasil penjualan, maka data-data dalam kumpulan seperti ”nama pegawai”, ”umur”, dan sebagainya dapat di-*ignore*) dan tidak konsisten.

1. *Data integration*

Untuk menggabungkan *multiple data source.*

1. *Data selection*

Untuk mengambil data yang sesuai untuk keperluan analisa.

1. *Data transformation*

Untuk mentransformasikan data ke dalam bentuk yang lebih sesuai untuk di *mining.* *Data mining* Proses terpenting dimana metode tertentu diterapkan untuk menghasilkan *data pattern*.

1. *Pattern evaluation*

Untuk mengidentifikasi apakah *interenting patterns* yang didapatkan sudah cukup mewakili *knowledge* berdasarkan perhitungan tertentu.

1. *Knowledge presentation*

Untuk mempresentasikan *knowledge* yang sudah didapatkan dari *user.*

**2.1.3 Pengelompokkan (*Clustering*)**

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 atau mengandung representasi yang lebih kaya seperti kategori yang hirarkis atau saling menumpu (*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*. (Abdul Kadir : 2009 : 35)

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 2.2 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. (Abdul Kadir :2009:45).

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. (Abdul Kadir:2009:48)

****

 **Sumber: Abdul Kadir(2009:45)**

**Gambar 3. Ilustrasi *Clustering***

****

**Sumber: Abdul Kadir(2009:45)**

**Gambar 4. Pengelompokkan himpunan data menjadi tiga *cluster***

Ada beberapa teknik dalam *clustering*, yaitu :

1. Metode partisi : membangun berbagai partisi dan kemudian mengevaluasi partisi tersebut dengan beberapa kriteria.

2. Metode hierarki : membuat suatu penguraian secara hierarkikal dari himpunan data menggunakan beberapa kriteria.

3. Metode Berbasis densitas : berdasarkan konektivitas dan fungsi densitas.

4. Metode Berbasis grid : berdasarkan suatu struktur granularitas multilevel.

5. Metode Berbasis model : suatu model dihipotesakan untuk masing – masing *cluster* dan ide untuk mencari *best fit* dari model tersebut untuk masing – masing yang lain.

**2.2 Penelitian Sebelumnya**

Tahta Alfina (2012 : 2), **“Analisa Perbandingan Metode *Hierarchical Clustering*, K-means dan Gabungan Keduanyadalam *Cluster* Data(Studi kasus : Problem Kerja Praktek JurusanTeknik Industri ITS)**”. Konsep data mining semakin dikenal sebagai *tools* penting dalam manajemen informasi karena jumlah informasi yang semakin besar jumlahnya. Salah satu teknik yang dikenal dalam data mining adalah *clustering*, berupa proses pengelompokan sejumlah data

atau objek ke dalam *cluster* (*group*) sehingga setiap dalam *cluster* tersebut akan berisi data yang semirip mungkin dan berbeda dengan objek dalam *cluster* yang lainnya. *Clustering* memiliki dua metode, yaitu partisi dan hierarki. Dua metode ini memiliki kelebihan dan kekurangan masing-masing, dan dengan menggabungkan keduanya dapat diperoleh hasil *cluster* yang lebih baik.

Dari hasil *cluster* dengan menggunakan data problem Kerja Praktek Jurusan Teknik Industri ITS, maka diperoleh hasil bahwa gabungan metode *Single Linkage Clustering* dan K-means memberikan hasil *cluster* yang lebih baik dengan parameter uji *cluster variance* dan metode *silhouette* coefisien. Kombinasi algoritma *hierarchical clustering* dan kmeans menghasilkan pengelompokan data yang lebih baik jika dibandingkan dengan k-means dalam semua pengujian.

Dengan evalusi koefisien *cophenetic*, metode *clustering* terbaik dihasilkan oleh *average linkage* *clustering*

Dalam studi kasus Problem Kerja Praktek Jurusan Teknik Industri ITS, dari kombinasi *hierarchical* *clustering* dan K-means yang ada, kombinasi *single*

*linkage clustering* dan K-means menghasilkan pengelompokan data yang terbaik dibandingkan dengan metode hierarki yang lainnya.

Enur Irdiansyah (2010 : 2), **“Penerapan Data Mining Pada Penjualan Produk Minuman Di PT.Pepsi Cola Indobeverages menggunakan metode *clustering*”.** Aplikasi yang dibangun dapat membantu PT. Pepsi Cola Indobeverages sebagai gambaran bagi pengambilan keputusan perusahaan dalam rangka mendapatkan pola penjualan produk yang dihasilkan, pengolahan data yang dilakukan dapat menghasilkan informasi yang cukup untuk dapat di analisa lebih lanjut, aplikasi yang dibangun dapat mengurangi penumpukkan dana yang kurang dimanfaatkan sebelumnya.

**III. METODOLOGI PENELITIAN**

**3.1 Waktu dan Tempat Penelitian**

Penelitian ini dilakukan mulai bulan maret 2013 sampai dengan Agustus 2013.

**3.2 Alat dan Bahan penelitian**

Adapun alat-alat yang digunakan dalam penelitian ini adalah :

1. Alat yaitu :

a. *Processor Intel core i3*

b. *RAM* 2 GB

c. *Hardisk* 320 GB

d. *Printer*

2. Bahan Yaitu :

a. *Windows 7*

b. *Microsoft Word 2007*

**3.3 Metode Pengumpulan Data**

Metode pengumpulan data yang digunakan untuk mendapatkan data dan informasi, maka metode yang digunakan dalam proses pengumpulan data sebagai berikut :

1. Metode Observasi

Dalam hal ini yang akan dilakukan adalah melihat serta mempelajari permasalahan yang ada dilapangan yang erat kaitannya dengan objek yang diteliti.

2. Metode Studi Pustaka

Metode yang dilakukan adalah dengan cara mencari bahan yang mendukung dalam pendefenisian masalah melalui buku-buku, internet.

**3.4. Metode Penerapan Data Mining**

Metode yang digunakan untuk penerapan data mining adalah metode *clustering* (pengelompokkan).

**3.4.1 Teknik *Clustering***

Secara garis besar metode *clustering* dibagi dalam 2 tipe yaitu : *hierarchical* dan *non hierarchical*. *Hierarchical* menggunakan *N x N* *similarity matrix*, sedangkan *non hierarchical* membagi dataset menjadi sebuah level *single* partisi, dengan atau tanpa pencocokan antara *clusters*. Selain itu hal mendasar yang membedakan kedua metode ini adalah : metode pengelompokan hirarki digunakan apabila belum ada informasi jumlah kelompok, sedangkan metode pengelompokan non hirarki bertujuan mengelompokan *n* objek ke dalam *k* kelompok (k<n). (Andi : 2009 : 27)

**3.4.2 *Agglomerative Hierarchical Clustering Algorithm*(*AHC algorithm*)**

Proses *Clustering* yang akan dilakukan menggunakan *Agglomerative Hierarchical Clustering Algorithm* (*AHC algorithm*). Walaupun lebih lambat, hasil akhir pengelompokan *agglomerative hierarchical clustering* lebih baik dibandingkan *kmeans* yang sering digunakan, karena lebih efisien . Proses dasar dari *hierarchical* *clustering* adalah:

1. Mulailah dengan mendefinisikan setiap item sebagai *cluster*, jadi data yang ada sejumlah N *items,* maka sekarang *cluster* yang ada sejumlah N *items* tersebut. Jarak antara *clusters* sama dengan jarak antara *items* yang ada.
2. Selanjutnya, mencari pasangan *cluster* yang paling dekat dan menggabungkan keduanya dalam satu *cluster*.
3. Hitung jarak antara *cluster* yang baru dibentuk dengan *cluster* yang sudah ada sebelumnya. Ulangi tahap b dan c sampai *cluster* membentuk N *cluster* yang diinginkan.

Sumber : ( Iko Pramudiono. (Rabu, 4 Maret 2009), Pengantar Data Mining:

Menambang Permata Pengetahuan di Gunung Data, <http://ilmukomputer.org/2008/11/25/p> engantardata-mining).

Ada tiga metode yang ada dalam *AHC algorithm* dan yang akan digunakan untuk melakukan *clustering* adalah metode *single linkage*. Metode *single lingkage* menggunakan prinsip jarak minimum. Dibawah ini adalah *psedeocode* dari algoritma *Agglomerative Hierarchical* *Clustering*.

**Agglomerative Hierarchical Algorithm:**

**Given:**

a set *X* of objects {*x*1,...,*xn*}

a distance function *dis*(*c*1,*c*2)

1. **for** *i* = 1 to *n*

*ci* = {*xi*}

**end for**

2. *C* = {*c*1,...,*cb*}

3. *l* = *n*+1

4. **while** *C*.size > 1 **do**

a) (*cmin*1,*cmin*2) = minimum *dis*(*ci*,*cj*) for all *ci*,*cj* in *C*

b) remove *cmin*1 and *cmin*2 from *C*

c) add {*cmin*1,*cmin*2} to *C*

d) *l* = *l* + 1

**end while**

Contoh : diberikan kumpulan X(dapat dilihat pada gambar ), yang sudah direpresentasikan dalam bentuk matrix. Dengan menggunakan fungsi *Euclidean distance* dan *single link method* untuk menghitung jarak antara *clusters*.



a. pertama, masing – masing anggota (Xi) dari X, ditempatkan dalam cluster

 Ci,dimana Ci adalah anggota dari cluster C.

 C = {{x1},{x2},{x3}, {x4},{x5},{x6},{x7}, {x8},{x9},{x10}}

b. set l = 11.

c. (iterasi pertama dari penggulangan menggunakan while) C.size = 10. Setelah

 dilakukan perhitungan jarak maka didapatkan jarak minimum adalah hasil

 perhitungan jarak antara C2 dan C10. (cmin1,cmin2) = (c2,c10)

d. Dan perhitungan terus dilakukan sampai membentuk N *cluster* yang diinginkan.

e. Kesimpulan dari hasil *clustering* yang didapatkan.

Tahapan akhir dari proses *clustering* adalah kesimpulan dari *clustering* yang didapatkan. Enur Irdiansyah.(2010:6).

**3.4.3 Contoh Penerapan *Clustering* Menggunakan Algoritma AHC**

Penulis mencoba menganalisa data penjualan dan minuman yang didapatkan dari PT.Indomarco Palembang. Dari data yang didaptkan penulis mencoba melakukan simulasi perhitungan jumlah mnuman dan makanan yang dijual ke masing-masing indomaret yang tersebar di wilayah Sumatera Selatam, Penulis hanya mengambil 11 wilayah untuk dijadikan contoh. Didapatkan sample data penjualan produk minuman dari perusahaan PT. Indomarco Palembang. Dari data penjualan tersebut akan dicoba dilakukan pengelompokan (*clustering*) menggunakan algoritma AHC. Adapun data penjualan tersebut dapat dilihat pada Tabel 1 dibawah.

**Tabel 1. Data Penjualan Makanan dan Minuman Per Wilayah Tanggal 1 Januari 2013**

|  |  |  |
| --- | --- | --- |
| **No** | **Wilayah Sumbagsel** | **Jumlah Penjualan** |
| 1 | Palembang | 30787 |
| 2 | Muara enim | 20567 |
| 3 | Prabumulih | 9655 |
| 4 | Indralaya | 8555 |
| 5 | Sungai lilin | 7659 |
| 6 | Jambi | 22643 |
| 7 | Pagaralam | 7643 |
| 8 | Kayu Agung | 11823 |
| 9 | Sekayu | 17273 |
| 10 | Banyuasin | 5890 |
| 11 | Tugu Mulyo | 14759 |

Sesuai dengan algoritma yang diterapkan, maka setiap item wilayah pada table diatas akan dijadikan sebagai *cluster*, jadi data yang ada sejumlah 11 *items,* maka sekarang *cluster* yang ada sejumlah 11 *items* tersebut. Selanjutnya, dari data penjualan (jumlah) dicari pasangan *cluster* yang paling dekat (selisih terkecil) dan menggabungkan keduanya dalam satu *cluster*. Pembuatan *cluster* pertama ini dapat dilihat pada Tabel 2 berikut :

**Tabel 2. Pembuatan *cluster* pertama**

|  |  |  |
| --- | --- | --- |
| **No** | **Wilayah Sumbagsel** | **Jumlah Penjualan** |
| 1 | Kelompok A1 Jumlah Anggota : 1  |
|  | Palembang | 30787 |
| 2 | Kelompok A2 Jumlah Anggota : 1 |
|  | Jambi | 22643 |
| 3 | Kelompok A3 Jumlah Anggota : 1 |
|  | Muara Enim | 20567 |
| 4 | Kelompok A4 Jumlah Anggota : 1 |
|  | Sekayu | 17273 |
| 5 | Kelompok A5 Jumlah Anggota : 1 |
|  | Tugu Mulyo | 14759 |
| 6 | Kelompok A6 Jumlah Anggota : 1 |
|  | Kayu Agung | 11823 |
| 7 | Kelompok A7 Jumlah Anggota : 1 |
|  | Prabumulih | 9655 |
| 8 | Kelompok A8 Jumlah Anggota : 1 |
|  | Indralaya | 8555 |
| 9 | Kelompok A9 Jumlah Anggota : 2 |
|  | Sungai Lilin | 7659 |
|  | Pagaralam | 7643 |
| 10 | Kelompok A10 Jumlah Anggota : 1 |
|  | Banyuasin | 5890 |

Dari tabel diatas dapat terlihat bahwa pasangan *cluster* yang paling dekat yang pertama kali membentuk *cluster* adalah wilayah Sungai Lilin dan Pagaralam yang merupakan anggota kelompok A9. Sedangkan wilayah lainnya masih merupakan cluster dari dirinya sendiri. Setelah jarak terdekat membentuk satu *cluster* pertama, kemudian kita harus menghitung jarak antara *cluster* yang baru dibentuk dengan *cluster* yang sudah ada sebelumnya. Tahap ini akan diulangi sampai *cluster* membentuk N *cluster* yang diinginkan. Dalam kasus ini, misalnya kita akan membentuk tiga *cluster* dari data yang ada, maka hasil dari algoritma AHC yang digunakan adalah seperti terlihat pada Tabel 3 berikut :

**Tabel 3. Hasil *clustering* sebanyak tiga *cluster***

|  |  |  |
| --- | --- | --- |
| **No** | **Wilayah Sumbagsel** | **Jumlah Penjualan** |
| 1 | Kelompok A1 Jumlah Anggota : 1  |
|  | Palembang | 30787 |
| 2 | Kelompok A2 Jumlah Anggota : 5 |
|  | Kayu Agung | 11823 |
|  | Tugu Mulyo | 14759 |
|  | Sekayu | 17273 |
|  | Muara Enim | 20567 |
|  | Jambi | 22643 |
| 3 | Kelompok A3 Jumlah Anggota : 5 |
|  | Banyuasin | 5890 |
|  | Sungai Lilin | 7659 |
|  | Pagaralam | 7643 |
|  | Indralaya | 8555 |
|  | Prabumulih | 9655 |

Berdasarkan tabel diatas dapat kita lihat pada wilayah palembang jumlah makanan dan minuman yang terjual per 1 januari 2013 berjumlah 30787. Berdasarkan data tersebut kita tidak tahu jenis makanan dan minuman apa saja yang menjadi *trend* atau diminati masyarakat palembang. Untuk itulah penulis bermaksut membuat aplikasi sebagai simulasi datanya, sehingga hasil akhir nantinya akan berbentuk grafik atau angka yang menunjukkan jumlah makanan dan minuman yang paling banyak diminati masyarakat.





