

MODUL DATA MINING

UNIVERSITAS TEKNOLOGI MATARAM

DAFTAR ISI

COVER

KATA PENGANTAR

BAB I : Pendahuluan

BAB II : Data Dan Karakteristik Data Mining

BAB III : Fungsi Minor Estimasi

BAB IV : Fungsi Mayor Metode Klasifikasi

BAB V : Algoritma Nearest Neighbor

BAB VI : Algoritma Naive Bayesian Clasifier

BAB VII : Fungsi Minor Metode Prediksi

BAB VIII : Fungsi Mayor Clustering

BAB IX : Teknik Single Linkage

BAB X : Metode K-Means Clustering

BAB XI : Fungsi Minor Asosiasi

BAB XII : Metode Multiple Regression

BAB XIII : Pengujian Dengan Software Pendukung

BAB XIV : Pengujian Lanjutan



BAB I : PENDAHULUAN

1.1 Definisi Dan Konsep Data Mining

Kalau kita membahas tentang Data Mining, tentulah kita harus mengetahui terlebih dahulu definisi dari Data Mining. Secara umum Data Mining terbagi atas 2(dua) kata yaitu:

- Data yaitu Kumpulan Fakta yang terekam atau sebuah entitas yang tidak memiliki arti dan selama ini terabaikan.
- Mining yaitu proses Penambangan
- Sehingga Data Mining itu dapat diartikan sebagai proses penambangan data yang menghasilkan sebuah ouput (keluaran) berupa pengetahuan. Selain itu juga Definisi Data Mining dapat dikutip dari beberapa sumber yaitu:

Menurut Pramudiono:

Data Mining adalah analisis otomatis dari data yang berjumlah besar atau kompleks dengan tujuan untuk menemukan pola atau kecenderungan yang penting yang biasanya tidak disadari keberadaanya. (Pramudiono,2006)

Menurut Larose:

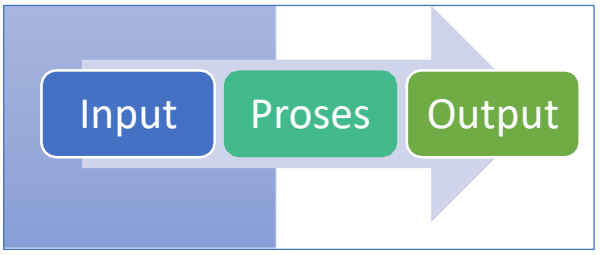
Data Mining merupakan analisis dari peninjauan kumpulan data untuk menemukan hubungan yang tidak diduga dan meringkasdata dengan cara berbeda dengan cara yang berbeda dengan sebelumnya, yang dapat dipahami dan bermanfaat bagi pemilik data. (Larose, 2005)

Data Mining merupakan bidang dari beberapa bidang keilmuan yang menyatukan teknik dari pembelajaran mesin, pengenalan pola, *statistic,database*, dan visualisasi untuk penanganan permasalahan pengambilan informasi dari database yang besar. (Larose, 2005)

Mengapa kita perlu memahami *Data Mining?* Karena manusia menghasilkan banyak sekali Data yang sangat besar baik dalam bidang Bisnis, Kedokteran, Cuaca,

Olahraga, Politik dan sebagainya. Contohnya dalam Dunia Olahraga kita mengetahui Dari FIFA berapa banyak Lionel Messi Mencetak Gol selama semusim, berapa banyak Lionel Messi memberikan Asisst. Pada Bidang Bisnis khususnya Saham, kita memperolehnya dari Bursa Efek Jakarta, kapan Harga Saham Naik maupun Turun. Pada Bidang Cuaca kita mengetahui data tentang Curah Hujan, Suhu, Kelembaban dan lain sebagainya.

Kita mengetahui bahwa setiap proses terdiri dari 3 (tiga) fase yaitu:



Gambar 1.1: Siklus Penyelesaian Dari Input Ke Output

Dari gambar di atas bahwa mengetahui suatu hal itu dapat diselesaikan dimulai dengan sebuah Inputan (data) kemudian di Proses sehingga menghasilkan sebuah keluaran. Tentunya di dalam data mining juga mengalami fase tersebut. Yang membedakannya adalah pada *Data Mining* yang menjadi Input adalah Himpunan Data, Prosesnya adalah Algoritma atau metode dalam *Data Mining* itu sendiri, dan Keluarannya adalah berupa Pengetahuan dalam bentuk Pola, *Decision Tree, Cluster* dan lain-lain.Untuk lebih jelas memahaminya berikut ini dapat dijelaskan pada gambar di bawah ini:



Gambar 1.2 : Perbedaan *Gold Mining* dan *Data Mining*

Keterangan gambar:

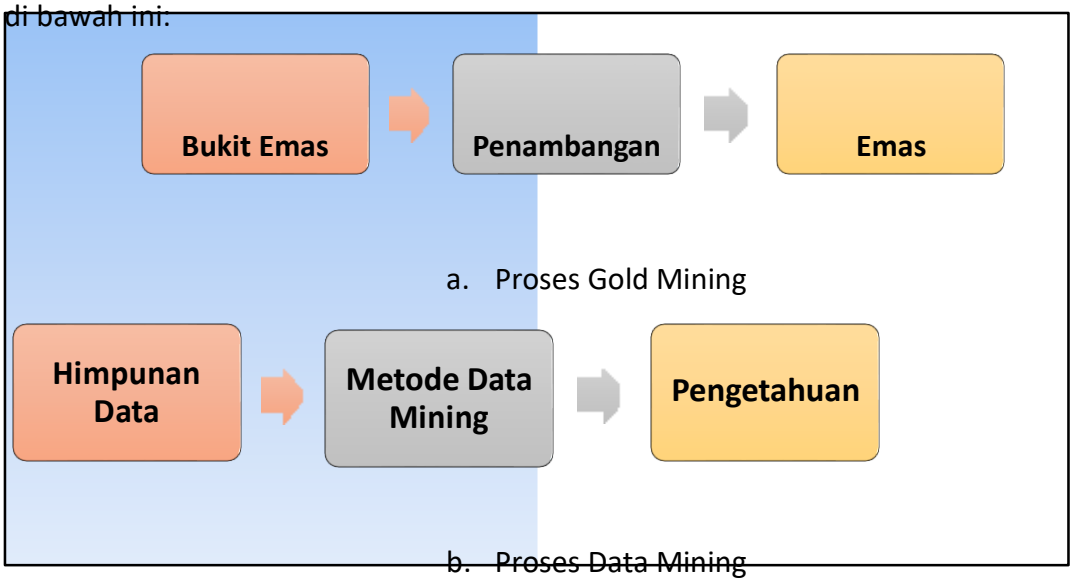
1. *Gold Mining*.

Gambar Tersebut diatas menjelaskan tentang beberapa orang sedang mencari dan melakukan penambangan emas. Jadi dalam hal ini yang menjadi Input adalah Bukit Emas sedangkan proses yang K adalah Penambangan yang dimulai dengan Identifikasi titik pada Bukit Emas tersebut dimana Emas itu berada dan dilanjutkan dengan proses Penggalian sampai mendapatkan Emas sebagai *Output*nya.

2. *Data Mining*:

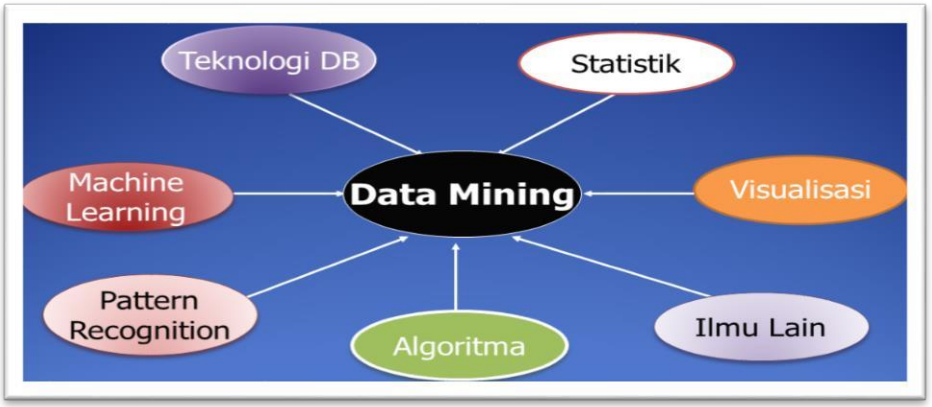
Gambar tersebut di atas menjeleskan sebuah Himpunan Data yang menjadi *Input* kemudian dilakukan pencarian pengetahuan menggunakan Metode *Data Mining* sehingga pada akhirnya didapatkan Pengetahuan sebagai *Output*.

Selain itu proses *Gold Mining* dan *Data Mining* dapat digambarkan pada gambar di bawah ini:



Gambar 1.3 : Proses *Gold Mining* Dan *Data Mining*

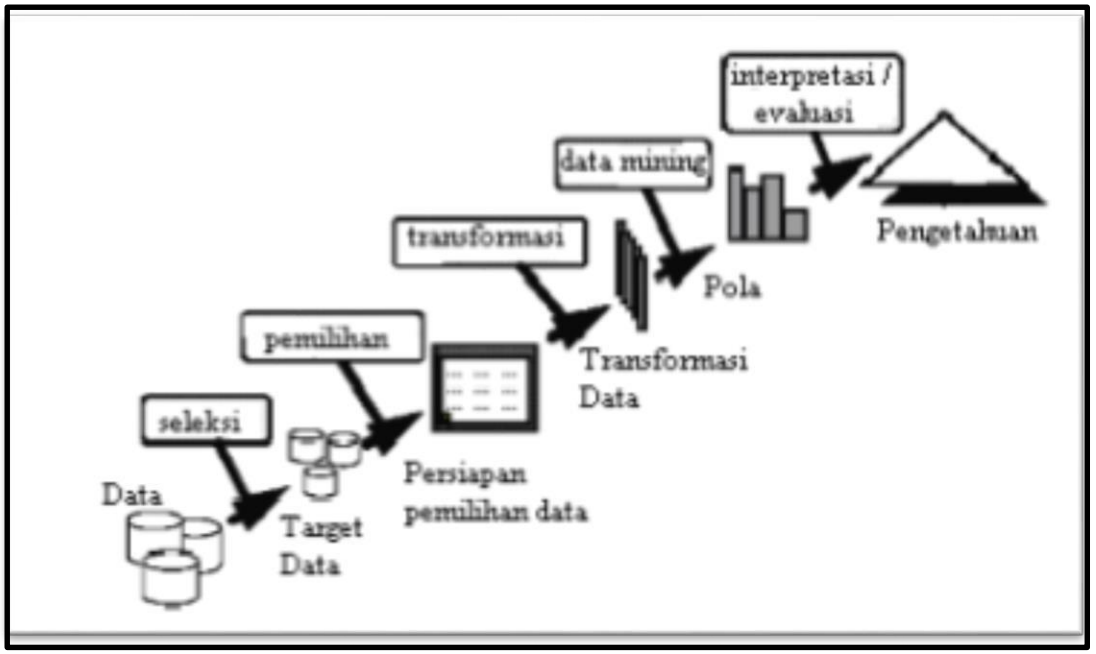
Berdasarkan proses di atas untuk menunjang pemahaman tentang *Data Mining* kita harus memahami beberapa disiplin ilmu lain seperti terlihat pada gambar di bawah ini:



Gambar 1.4 : Disiplin Ilmu Data Mining

1.2 Knowledge Discovery Database(KDD)

Pada proses *Data Mining* yang biasa disebut Knowledge *Discovery Database(KDD)* terdapat beberapa proses seperti terlihat pada gambar di bawah ini:



Gambar 1.5 : Proses *Knowledge Discovery Database* (KDD)

Penjelasan gambar:

Pada proses *Knowledge Discovery Database (KDD)*terdapat beberapa fase yaitu sebagai berikut:

Pada proses *Knowledge Discovery Database (KDD)*terdapat beberapa fase yaitu sebagai berikut:

1. ***Seleksi Data (Selection)***

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

2. ***Pemilihan Data (Preprocessing/Cleaning)***

   Proses *Preprocessing*mencakup antara lain membuang duplikasi data, memeriksa data yang inkonsisten, dan memperbaiki kesalahan pada data, seperti kesalahan cetak (tipografi). Juga dilakukan proses enrichment, yaitu proses "memperkaya" data yang sudah ada dengan data atau informasi lain yang relevan dan diperlukan untuk KDD, seperti data atau informasi eksternal.

3. ***Transformasi (Transformation)***

   Pada fase ini yang dilakukan adalah mentransformasi bentuk data yang belum memiliki entitas yang jelas kedalam bentuk data yang valid atau siap untuk dilakukan prose *Data Mining*

4. ***Data Mining***

   Pada fase ini yang dilakukan adalah menerapkan algoritma atau metode pencarian pengetahuan.

5. ***Interpretasi/Evaluasi (Interpratation/Evaluation)***

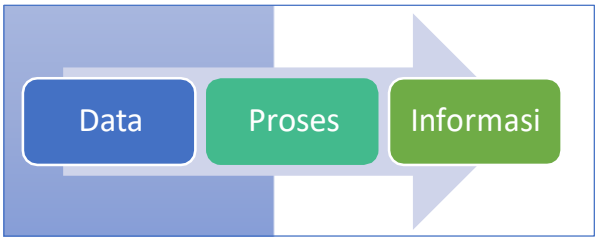
   Pada fase terakhir ini yang dilakukan adalah proses pembentukan keluaran yang mudah dimengerti yang bersumber pada proses *Data Mining* Pola informasi.

BAB 2

BAB II : Data Dan Karakteristik Metode Data Mining

2.1 Data

Data merupakan kumpulan fakta yang direpresentasikan ke dalam bentuk karakter baik huruf, angka dan lainnya yang dapat diproses menjadi sebuah informasi. Sesuai dengan kaidah penelitian untuk *Data Collecting* (pengumpulan data) bisa melalui observasi, angket, wawancara dengan *stakeholder* dan lain-lain. Secara definitif kita mengetahui bahwa Data adalah kumpulan Fakta yang terekam dan tidak memiliki arti. Selain itu data dapat diartikan sebagai kumpulan fakta-fakta yang direpresentasikan kedalam beberapa bentuk baik karakter : Angka, huruf maupun simbol yang diproses sehingga menghasilkan sebuah informasi. Atau data dapat dinterpretasikan sebagai Entitas yang tidak memiliki arti yang selamai ini terabaikan. Data juga dapat di analogi pada dunia pabrikasi yaitu sebagai "Bahan Mentah" sedang hasil pengolahan Produksinya yang disebut "Bahan Jadi" yaitu berupa Informasi. Untuk lebih jelasnya dapat dilihat pada gambar di bawah ini:



Gambar 1.6 : Proses Terbentuknya Informasi

Data data mining tentulah kita semua mengetahui bahwa yang akan ditambang atau digali dalam tanda kutip adalah Himpunan Data / Basis Data (*database*) ,yang kemudian akan diekstraksi menjadi sebuah pengetahuan baik Pola, Klaster, *Decision Tree* dan lain-lain.Sebelum kita melakukan proses data mining tentunya kita terlebih dahulu mengetahui beberapa elemen dalam sebuah himpunan data seperti pada gambar di bawah ini:

Tabel 1.1: Contoh Himpunan Data

Attribut / Feature/Selection

Label/Class/ Target

No	NAMA	NILAI IPK	NILAI ABSENSI	NILAI ETIKA	Ket
1	Dini	0.25	73.6	79.3	Gagal
2	Dino	3.75	98.9	87	Lulus
3	Dina	3.85	99	85	Lulus
4	Dani	0.56	60.3	65	Gagal
5	Dana	3.15	95.7	84.3	Lulus
6	Danu	0.35	52.6	56	Gagal
7	Doni	1.72	68.3	73	Gagal

- **Attribut** adalah deskripsi data yang bisa mengidentifikasikan entitas Field adalah lokasi penyimpanan *Record* adalah kumpulan dari berbagai field yang saling berhubungan.

- **Class / Label / Target** bisa disebut sebagai atribut keputusan.

Pada *Data Mining* secara garis besar terdapat 2(dua) tipe data yang harus dipahami yaitu:

1. ***Numeric*** merupakan tipe data yang bisa di kalkulasi
2. ***Nominal*** merupakan tipe data yang tidak bisa di kalkulasi baik tambah, kurang, kali maupun bagi.

   Untuk contoh pemanfaatan tipe data dapat terlihat pada tabel di bawah ini:

Tabel 1.2 : Tipe Data Dalam *Data Mining*

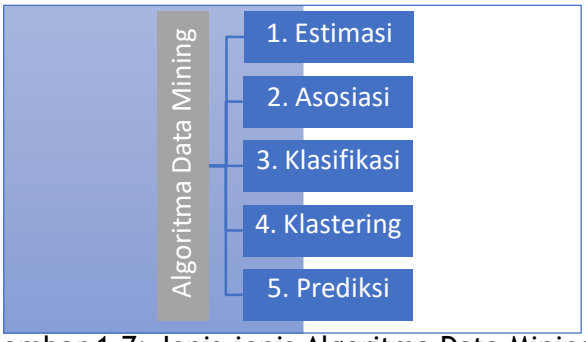
No	NAMA	V1	V2	V3	Ket
1	Dini	0.25	73.6	79.3	Gagal
2	Dino	3.75	98.9	87	Lulus
3	Dina	3.85	99	85	Lulus
4	Dani	0.56	60.3	65	Gagal
5	Dana	3.15	95.7	84.3	Lulus
6	Danu	0.35	52.6	56	Gagal
7	Doni	1.72	68.3	73	Gagal
8	Dono	0.75	79.4	80	Gagal

Nominal

Numeric

2.1 Jenis Algoritma dan Metode *Data Mining*

Pada proses pemecahan masalah dan pencarian pengetahuan baru terdapat beberapa klasifikasi secara umum yaitu:



Gambar 1.7: Jenis-jenis Algoritma Data Mining

1. **Estimasi**

   Digunakan untuk melakukan estimasi terhadap sebuah data baru yang tidak memiliki keputusan berdasarkan histori data yang telah ada. Contohnya ketika melakukan Estimasi Pembiayaan pada saat pembangunan sebuah Hotel baru pada Kota yang berbeda.

2. **Asosiasi**

   Digunakan untuk mengenali kelakuan dari kejadian-kejadian khusus atau proses dimana hubungan asosiasi muncul pada setiap kejadian. Adapun metode pemecahan masalah yang sering digunakan seperti Algoritma Apriori. Contoh pemanfaatan Algoritma Asosiasi yaitu pada Bidang Marketing ketika sebuah Minimarket melakukan Tata letak produk yang dijual berdasarkan Produk-produk mana yang paling sering dibeli konsumen, selain itu seperti tata letak buku yang dilakukan pustakawan di perpustakaan

3. **Klasifikasi**

   Suatu teknik dengan melihat pada kelakuan dan atribut dari kelompok yang telah didefinisikan. Teknik ini dapat memberikan klasifikasi pada data baru dengan memanipulasi data yang ada yang telah diklasifikasi dan dengan menggunakan hasilnya untuk memberikan sejumlah aturan. Salah satu contoh yang mudah dan popular adalah dengan Decision tree yaitu salah satu metode klasifikasi yang paling populer karena mudah untuk interpretasi seperti *Algoritma C4.5*, *ID3* dan

lain-lain. Contoh pemanfaatannya adalah pada bidang Akademik yaitu Klasifikasi siswa yang layak masuk kedalam kelas unggulan atau akselerasi di sekolah tertentu.

4. Klastering

Digunakan untuk menganalisis pengelompokkan berbeda terhadap data, mirip dengan klasifikasi, namun pengelompokkan belum didefinisikan sebelum dijalankannya *tool data mining*. Biasanya menggunkan metode neural network atau statistik, analitikal hierarki *cluster*.*Clustering* membagi item menjadi kelompok-kelompok berdasarkan yang ditemukan *tool data mining*.

5. Prediksi

Algoritma prediksi biasanya digunakan untuk memperkirakan atau *forecasting* suatu kejadian sebelum kejadian atau peristiwa tertentu terjadi. Contohnya pada bidang Klimatologi dan Geofisika, yaitu bagaimana Badan Meterologi Dan Geofisika (BMKG) memperkirakan tanggal tertentu bagaimana Cuacanya, apakah Hujan, Panas dan lain sebagainya. Ada beberapa metode yang sering igunakan salah satunya adalah Metode Rough Set. Di dalam data mining juga sama halnya dengan konsep Neural Network mengandung 2(dua) pengelompokkan yaitu:

1. ***Supervised Learning*** yaitu pembelajaran menggunakan guru dan biasanya ditandai dengan adanya Class/Label/Target pada himpunan data. Adapun metode-metode yang digunakan yang bersifat *supervised learning* seperti Metode Prediksi dan Klasifikasi seperti Algoritma C4.5, Metode *Rough Set* dan Lain-lain.
2. ***Unsupervised Learning*** yaitu pembelajaran tanpa menggunakan guru dan biasanya ditandai pada himpunan datanya tidak memiliki attribut keputusan atau *Class/Label/Target*. Adapun metode-metode yang bersifat Unsupervised Learning yaitu Metode Estimasi, *Clustering*, Dan Asosiasi seperti *Regresi Linier*, *Analytical Hierarchy Clustering* dan lain-lain.

BAB 3

BAB III : FUNGSI MINOR ESTIMASI

3.1 Pendahuluan

Metode Estimasi merupakan salah satu metode yang ada dalam Data Mining. Ada hal yang perlu dipahami bahwasanya metode ini dapat bekerja apabila himpunan data sebagai sampel data yang akan di proses bersifat numerik dan memiliki label. Biasanya metode ini tidak memiliki rumus yang pasti karena bersifat Regresi. Artinya dalam penentuan sebuah keputusan dari sebuah sampel baru berasal dari sebuah rumus yang terbentuk berdasarkan parameter-parameter himpunan data.

Dalam metode estimasi terdapat beberapa algoritma yang dapat dijadikan sebagai Learning Algorithma diantaranya yaitu Regresi Linier.

3.1 Metode Regresi Linear Sederhana

Regresi Linier merupakan suatu alat ukur yang dapat digunakan untuk mengetahui adanya korelasi antara beberapa variabel. Dalam Regresi Linier ada beberapa hal yang harus dipahami diantaranya Variabel Terikat, Variabel Bebas, Konstanta dan Koefisien Regresi. Kalau ditinjau keakurasiannya dalam pemecahan sebuah kasus, regresi memiliki tingkat akurasi yang lebih baik di dalam konsep analisis sebuah hubungan antara 1(satu) variabel dengan variabel lainnya .

Untuk fungsi regresi terdapat beberapa rumus yang dapat dijadikan untuk pembentukan rumus regresi baru yaitu:

Y = a+bX

Selain itu juga persamaan Regresi Linier dapat dituliskan dengan rumus sebagai berikut:

$$Y = \left| \frac{\Sigma xy}{\left(\Sigma x^2\right)} \right| x$$

Keterangan:

Y = variabel terikat

X = variabel bebas

a = konstanta (intersep)

b = koefisien Regresi (slop)

Untuk mencari nilai a (konstanta) dan b(koefisien Regresi) maka ada beberapa rumus yang dapat digunakan yaitu:

$$a = \frac{(\Sigma Y)(\Sigma X^2) - (\Sigma X)(\Sigma XY)}{(n)(\Sigma X^2) - (\Sigma X)^2}$$

$$b = \frac{(n)(\Sigma XY) - (\Sigma X)(\Sigma Y)}{(n)(\Sigma X^2) - (\Sigma X)^2}$$

Atau

$$b = \frac{(n)(\Sigma XY) - (\Sigma X)(\Sigma Y)}{(n)(\Sigma X^2) - (\Sigma X)^2}$$

$$a = \overline{Y - b}.X$$

Untuk lebih jelasnya berikut ini adalah contoh kasus dan bagaimana konsep pemecahan masalahnya dengan Regresi Linier.

1. Contoh Kasus Dan Konsep Pemecahan Masalah

Bagian Sumber Daya Manusia (SDM) pada sebuah perusahaan Ritel di Indonesia ingin membuat sebuah penelitian berkenaan dengan produktivitas bekerja karyawan lama dan baru. Perusahaan ini melihat sejauh mana produktivitas karyawan lama dan baru ini berdasarkan umur. Setelah ditelusuri pada arsip bagian Sumber Daya Manusia (SDM) terdapat track record penjualan khususnya bagian marketing yang saya hubungkan atau korelasikan berdasarkan pengalaman kerja karyawan.

Berikut ini adalah sampel data yang telah ada pada histori atu arsip bagian Sumber Daya Manusia.

Tabel 3.1 : Sampel Data Regresi Linier

Karyawan	Pengalaman Kerja (X)	Omzet Penjualan (Y)
1	10 tahun	10.000
2	8 Tahun	8.000
3	7 Tahun	7.000
4	4 Tahun	4.000
5	3 Tahun	2.000

Soal: Apabila ada karyawan 6 (baru) yang memiliki Pengalaman kerja selama 8.5 tahun, maka Estimasinya Omzet Penjualannya sebesar?

Maka,

Langkah awalnya adalah menentukan nilai a dan b dengan cara:

X	y	X^2	y^2	XY
6	**10**	**36**	**100**	**60**
8	**8**	**64**	**64**	**64**
7	**7**	**49**	**49**	**49**
4	**4**	**16**	**16**	**16**
3	**3**	**9**	**9**	**9**

Rumus 1

- **Menghitung Nilai Rata-rata**

$$\overline{X} = \frac{28}{5} = 5.6 \quad \overline{Y} = \frac{32}{5} = 6.4$$

- **Menghitung Nilai a dan b**

$$a = \frac{(\Sigma Y)(\Sigma X^2) - (\Sigma X)(\Sigma XY)}{(n)(\Sigma X^2) - (\Sigma X)^2}$$

$$b = \frac{(n)(\Sigma XY) - (\Sigma X)(\Sigma Y)}{(n)(\Sigma X^2) - (\Sigma X)^2}$$

Maka,

$$a = \frac{(32)(174)) - ((28)(198))}{((5)(174)) - (28)^2}$$

a = 5568-5544

870-784

a = 24

86

a = 0.279

dan

b = ((5)(198))-((28)(32))

((5)(174))-$(28)^2$

= 990-896

870-784

= 94/86

= 1.093

Maka persamaan regresinya adalah:

Y = a+bx

Y =0.279+ 1.093x

Jadi,

Y = 0.279 + 1.093x

Y = 0.279 + 1.093 (8.5)

Y = 0.279 + 9.2905

Y = 9.569 (ribuan)

Y = 9569

Rumus 2:

$$b = \frac{(n)(\Sigma XY) - (\Sigma X)(\Sigma Y)}{(n)(\Sigma X^2) - (\Sigma X)^2}$$

$$a = \overline{Y - b}.\,X$$

-Menghitung Nilai a dan b

b = (5)(198)-(28)(32))

(10 (55^2)-(55))^2

= 990-896

870-784

= 94/86

= 1.093

a = 6.4-1.093(5.6)

= 6.4-6.1208

= 0.279

Maka persamaan Regresinya adalah:

Y = a+bx

Y =0.279+ 1.093x

Jadi,

Y = 0.279 + 1.093x

Y = 0.279 + 1.093 (8.5)

Y = 0.279 + 9.2905

Y = 9.569 (ribuan)

Y = 9569

Dengan melihat rumus yang ada, memiliki kesamaan yang sama nilai akhirnya yaitu dengan Pengalaman **8.5 tahun** di Estimasi mendapatkan omset sebesar **9569.** Agar dapat melihat prosentasenya lebih baik, Untuk menentukan nilai prosentase Koefisien Determinasinya yaitu:

$$R^2 = \frac{((n)(\Sigma XY) - (\Sigma X)(\Sigma Y))^2}{(n(\Sigma X^2) - (\Sigma X)^2\ (n(\Sigma Y^2) - (\Sigma Y)^2)}$$

R^2 = ((5)(198)-(28)(32))^2

(5(174)-(28)^2(5(238)-(32)^2)

R^2 = (990-896)^2

(870-784)(1190-1024)

R^2 = (94)^2

(86)(166)

R^2 = 8.836/14.276

R^2 = 0.6189

Maka Nilai Determinasi Koefisien (R^2) =**0.6189**, artinya sumbangan untuk pengaruh pengalaman terhadap hasil kinerja pegawai yang berhubungan dengan naik turunnya omset penjualan perusahaan adalah **61.89 %** sisanya sebesar **38.11%** berhubungan dengan faktor yang lain.

Latihan:

Berikut ini adalah sampel data *Delivery Order* pada PT. Jago Ayam Friedi Chicken

Tabel 3.2 : Sampel Data Regresi Linier

Customer	Jumlah Pesanan (JP)	Jumlah Trafic Light (JT)	Jarak (J)	Waktu Tempuh (T)
C1	**3 paket**	**3 unit**	**3 km**	**16 menit**
C2	**1 paket**	**7 unit**	**4 km**	**20 menit**
C3	**2 paket**	**4 unit**	**6 km**	**18 menit**
C4	**4 paket**	**6 unit**	**8 km**	**36 menit**
C5	**2 paket**	**4 unit**	**2 km**	**12 menit**

Soal: Apabila ada Customer C6 dengan JP=4 , JT=2 dan J=5 maka Estimasi Jarak Tempuh yang dibutuhkan (T) untuk pengiriman pesanan selama?



BAB III : FUNGSI MAYOR METODE KLASIFIKASI

3.1 Pendahuluan

Metode Estimasi merupakan salah satu metode yang ada dalam Data Mining. Ada hal yang perlu dipahami bahwasanya metode ini dapat bekerja apabila himpunan data sebagai sampel data yang akan di proses bersifat numerik dan memiliki label. Biasanya metode ini tidak memiliki rumus yang pasti karena bersifat Regresi. Artinya dalam penentuan sebuah keputusan dari sebuah sampel baru berasal dari sebuah rumus yang terbentuk berdasarkan parameter-parameter himpunan data.

Dalam metode estimasi terdapat beberapa algoritma yang dapat dijadikan sebagai Learning Algorithma diantaranya yaitu Regresi Linier.

Klasifikasi merupakan sebuah proses training (pembelajaran) suatu fungsi tujuan (target) yang digunakan untuk memetakan tiap himpunan atribut suatu objek ke satu dari label kelas tertentu yang di definisikan sebelumnya. Teknik Klasifikasi ini cocok digunakan dialam mendeskripsikan data-set dengan tipe data dari suatu himpunan data yaitu biner atau nominal. Adapun kekurangan dari teknik ini yaitu tidak tepat untuk himpunan data ordinal karena pendekatan-pendekatan yang digunakan secara implisit dalam kategori data.

Ada beberapa teknik klasifikasi yang digunakan sebagai solusi pemecahan kasus diantaranya yaitu:

- Algoritma C4.5
- Algoritma K-Nearest Neighbor
- ID3
- Naïve Bayesian Clasification
- CART (*Clasification And Regression Tree*)

Dan lain-lain

Output atau keluaran dari metode klasifikasi ini biasanya dalam bentu "*Decision Tree* (pohon keputusan)". Dalam pembahasan kali ini saya mencoba untuk membahas tentang Algoritma C4.5.

3.2 Algoritma C4.5

Algoritma C4.5 merupakan salah satu solusi pemecahan kasus yang sering digunakan dalam pemecahan masalah pada teknik klasifikasi. Keluaran dari algoritma C4.5 itu berupa sebuah decision tree layaknya teknik klasifikasi lain. Sebuah pohon keputusan adalah sebuah struktur yang dapat digunakan untuk membagi kumpulan data yang besar menjadi himpunan-himpunan *record* yang lebih kecil dengan menerapkan serangkaian aturan keputusan. Dengan masing-masing rangkaian pembagian, anggota himpunan hasil menjadi mirip satu dengan yang lain (Berry & Linoff, 2004).

Adapun penjelasan tentang Algoritma C4.5 itu sendiri yaitu Salah satu algoritma C4.5 induksi pohon keputusan yaitu ID3 (*Iterative Dichotomiser 3*). input berupa sampel training, label training dan atribut. Algoritma C4.5 merupakan pengembangan dari ID3. Jika suatu set data mempunyai beberapa pengamatan dengan *missing value* yaitu *record* dengan beberapa nilai variable tidak ada, jika jumlah pengamatan terbatas maka atribut dengan *missing value* dapat diganti dengan nilai rata-rata dari variable yang bersangkutan. (Santosa, 2007)

Untuk penyelesaian kasus didalam Algoritma C4.5 ada beberapa elemen yang diketahui yaitu:

1. *Entropy*
2. *Gain*

Entropy(S) merupakan jumlah bit yang diperkirakan dibutuhkan untuk dapat mengekstrak suatu kelas (+ atau -) dari sejumlah data acak pada ruang sampel S. Entropy dapat dikatakan sebagai kebutuhan bit untuk menyatakan suatu kelas. semakin kecil nilai Entropy maka akan semakin Entropy digunakan dalam mengekstrak suatu kelas. Entropi digunakan untuk mengukur ketidakaslian S. Adapun rumus untuk mencari nilai Entropi.

$$Entropy(S) \equiv -p_{\oplus}\, log_2\, p_{\oplus} - p_{\ominus}\, log_2 p_{\ominus}$$

Dimana:

- S : ruang (data) sampel yang digunakan untuk pelatihan

- $p_{\oplus}$:jumlah yang bersolusi positif atau mendukung pada data sampel untuk kriteria tertentu
- $p_{\ominus}$:jumlah yang bersolusi negatif atau tidak mendukung pada data sampel untuk kriteria tertentu.

- Entropi(S) = 0, jika semua contoh pada S berada dalam kelas yang sama.
- Entropi(S) = 1, jika jumlah contoh positif dan negative dalam S adalah sama.
- 0 > Entropi(S) > 1, jika jumlah contoh positif dan negative dalam S tidak sama.

Gain (S,A) merupakan Perolehan informasi dari atribut A relative terhadap output data S. Perolehan informasi didapat dari output data atau variabel dependent S yang dikelompokkan berdasarkan atributA, dinotasikan dengan gain (S,A). Adapun rumus untuk mencari nilai Gain yaitu:

$$Gain(S,A) \equiv Entropy(S) - \sum_{i=1}^{n} \frac{|Si|}{|S|} * Entropy(Si)$$

Dimana:

- A : Atribut
- S : Sampel
- n : Jumlah partisis himpunan atribut A
- |Si| : Jumlah sampel pada pertisi ke –i
- |S| : Jumlah sampel dalam S

Adapun langkah-langkah untuk penyelesaian Algoritma C4.5 terlihat pada siklus di bawah ini:



Gambar 4.1 : Algoritma Penyelesaian Algoritma C4.5

2.Contoh Kasus Dan Teknik Penyelesaian Masalah

Masalah yang akan di analisis adalah untuk mengklasifikasikan calon pendaftar di suatu STMIK xxx dalam hal pemilihan program studi khususnya : Sistem Komputer Atau Sistem Informasi. Adapun data yang digunakan dalam membentuk pohon keputusan untuk menganalisis minat calon mahasiswa baru untuk mendaftar ke STMIK xxx berdasarkan program studi strata 1 adalah nama mahasiswa, minat calon mahasiswa, asal sekolah, jenis kelamin, hobi. Data selajutnya akan dilakukan pra-proses untuk menghasikan data kasus yang siap dibentuk untuk menjadi sebuah pohon keputusan.

Data yang tidak lengkap disebabkan karena ada data yang kosong atau atribut yang salah. Demikian pula dengan data minat calon mahasiswa baru yang mendaftar ke STMIK xxx berdasarkan program studi strata 1, ada sebagian atribut yag tidak perlu sehingga proses *Data Preprocessing* perlu dilakukan sehingga data base sesuai dengan ketentuan yang diperlukan.

Data Preprocessing merupakan hal yang penting dalam proses data mining, hal yang termasuk antara lain:

1. *Data Selection*

Data minat calon mahasiswa/i baru yang mendaftar ke STMIK Triguna Dharma berdasarkan program studi strata 1 tersebut akan menjadi data kasus dalam proses operasional data mining. Dari data yang ada, kolom yang diambil sebagai atribut keputusan adalah ***hasil***, sedangkan kolom yang diambil atribut penentuan dalam pembentukan pohon keputusan adalah:

a. Nama Mahasiswa
b. Minat calon mahasiswa
c. Asal sekolah
d. Jenis kelamin
e. Hobi

2. *Data Preprocessing / Data Cleaning*

   Data *Cleaning* diterapkan untuk menambahkan isi atribut yang hilang atau kosong dan merubah data yang tidak konsisten.

3. *Data Transformation*

   Dalam proses ini, data ditransferkan ke dalam bentuk yang sesuai untuk proses data mining.

4. *Data Reduction*

   Reduksi data dilakukan dengan menghilangkan atribut yang tidak diperlukan sehingga ukuran dari *database* menjadi kecil dan hanya menyertakan atribut yang diperlukan dalam proses *data mining,* karena akan lebih efisien terhadap data yang lebih kecil.

Masalah klasifikasi berakhir dengan dihasilkan sebuah pengetahuan yang dipresentasikan dalam bentuk diagramyang biasa disebut pohon keputusan (*decision tree*). Data berikut ini dipergunakan untuk data latihan. Data selengkapnya tampak pada tabel dibawah ini:

Tabel 4.1 : Sampel Data Yang Digunakan

No.	Nama Mahasiswa	Calon Mahasiswa	Asal Sekolah	Kelamin	Hobi	Hasil
1	Novita Devi Batu Bara	Hardware	SMK Komputer	Laki-Laki	Non	SK
2	Ahmad Riyandi	Hardware	SMK Komputer	Laki-Laki	IT	SK
3	Reza Adriansyah	Umum	SMK Komputer	Laki-Laki	Non	SI
4	Gafar Dwi Satrio	Software	SMA UMUM	Laki-Laki	Non	SI
5	Nur Azizah Dalimunthe	Software	SMK TEKNIK	Perempuan	Non	SI
6	Roy Ishak Permana Barus	Software	SMK TEKNIK	Perempuan	IT	SI
7	Muhammad Rizky Fadly	Umum	SMK TEKNIK	Perempuan	IT	SI
8	Zulfikar Ali	Hardware	SMA UMUM	Laki-Laki	Non	SK
9	Putra Mustaqim	Hardware	SMK TEKNIK	Perempuan	Non	SI
10	Debby Latifah Simatupang	Software	SMA UMUM	Perempuan	Non	SI
11	Daniel Alberto Sihombing	Hardware	SMA UMUM	Perempuan	IT	SI
12	Asri Anzani Br. Tarigan	Umum	SMA UMUM	Laki-Laki	IT	SI
13	Abdul Alim	Umum	SMK Komputer	Perempuan	Non	SI
14	Akbar Widiantara	Software	SMA UMUM	Laki-Laki	IT	SK

Keterangan :

- Untuk Asal Sekolah yang disebut SMK Komputer yaitu yang berasal dari jurusan Teknik Komputer Dan Jaringan, Multimedia, dan Rekayasa perangkat lunak sedangkan yang dikatakan sekolah umum yaitu Sekolah Menengah Atas yang terdiri dari jurusan IPA maupun IPS dan yang dimaksud SMK Teknik adalah yang berasal dari jurusan baik Teknik Elektro, Teknik Mesin, Teknik Listrik dan Lain-lain. SI merupakan Nilai Atribut Hasil Sistem Informasi dan SK merupakan Nilai Atribut Hasil Sistem Komputer.

Setelah kita memperoleh data Minat Calon Mahasiswa/i Baru yang tercantum pada Tabel 3.1. Langkah selanjutnya adalah menentukan nilai *Entropy* dan *Gain*nya:

1. **Nilai *Entropy***

   a. *Entropy* Total=

   $\sum^{n}$

   Entropy Total=((-4/14*log 2 (4/14) + (-10/14*log 2 (10/14))

   = **0.863120569**

b. *Entropy* Minat Calon Mahasiswa

- Nilai atribut "**Hardware**" = ((-3/5)*Log 2(3/5)+(-2/5)*log 2 (2/5)

=**0.970950594**

- Nilai atribut "**Software**" = ((-1/5)*Log 2(1/5)+(-4/5)*log 2 (4/5)
  = **0.721928095**
- Nilai atribut "**Umum**" = ((-0/4)*Log 2(0/4)+(- 4/4)*log 2 (4/4)
  = **0**

c. *Entropy* Histori Pendidikan (Asal Sekolah)

- Nilai atribut **"SMK Komputer"**
  = ((-2/4)*Log 2(2/4)+(-2/4)*log 2 (2/4) = **1**
- Nilai atribut **"SMK Teknik"**
  = ((-0/4)*Log 2(0/4)+(- 4/4)*log 2 (4/4) = **0**
- Nilai atribut **"SMA Umum"**
  =((-2/6)*Log 2(2/6)+(- 4/6)*log 2 (4/6) = **0.918295834**

d. *Entropy* Hobi

- Nilai atribut **"IT"** =
  ((-4/6)*Log 2(4/6)+(- 2/6)*log 2 (2/6) = **0.918295834**
- Nilai atribut **"Non IT"** =
  ((-2/8)*Log 2(2/8)+(-6/8)*log 2 (6/8) = **0.811278124**

e. *Entropy* Jenis Kelamin

- Nilai atribut "1" = ((-4/7)*Log 2(4/7)+(- 3/7)*log 2 (3/7)
  = **0.985228136**
- Nilai atribut "0" = ((-0/7)*Log 2(0/7)+(-7/7)*log 2 (7/7)
  = **0**

2. Nilai *Gain*

Berikut ini adalah nilai *Gain* dari setiap kriteria.

- Nilai *Gain* Minat Calon Mahasiswa

=**0.863120569**-((5/14)* **0.970950594**))+((5/14)*
0.721928095))+((4/14)*0**)))** **= 0.258521037**

- Nilai *Gain* Histori Pendidikan

  =**0.863120569**-((4/14)*__1__))+((4/14)*__0__))+((6/14)* **0.918295834**)))**= 0.183850925**

- Nilai *Gain* Hobi

  =**0.863120569**–((6/14)* **0.918295834**))+((8/14)*0)) **= 0.005977711**

  - Nilai *Gain* Jenis Kelamin

    = **0.863120569**–((7/14)*** 0.985228136**))+((7/14)*****0))

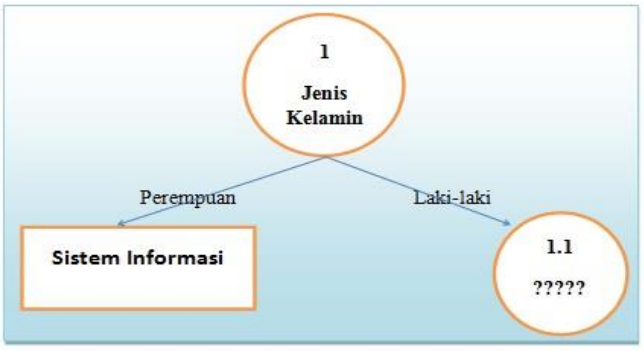
    = **0.005977711**

Setelah di dapatkan nilai *Entropy* dan *Gain* dari sampel data yang dimiliki, berikut ini adalah rekapitulasi perhitungan nilai *Entropy* dan *Gain*nya.

Tabel 4.2 Rekapitulasi Hasil

NODE		Keterangan	Jml Kasus (S)	Sistem Komputer (SK)	Sistem Informasi (SI)	Entropy	Gain
1	**TOTAL**		**14**	**4**	**10**	**0.86312**	
	Minat Calon						**0.25852103**
		Hardware	**5**	**3**	**2**	**0.970950**	
		Software	**5**	**1**	**4**	**0.721928**	
		Umum	**4**	**0**	**4**	**0**	
	Asal Sekolah						**0.18385092**
		SMK Komputer	**4**	**2**	**2**	**1**	
		SMK Teknik	**4**	**0**	**4**	**0**	
		SMA	**6**	**2**	**4**	**0.918295**	
	Jenis Kelamin						**0.37050650**
		Laki-laki	**7**	**4**	**3**	**0.985228**	
		Perempuan	**7**	**0**	**7**	**0**	
	Hobi						**0.00597771**
		IT	**6**	**4**	**2**	**0.918295**	
		Non	**8**	**2**	**6**	**0.811278**	

Tabel di atas menunjukkan bahwasanya kriteria Jenis Kelamin memiliki nilai *Gain* yang paling tinggi. Untuk *fase* selanjutnya adalah pembentukan *Tree* (pohon keputusannya). Berikut ini adalah *Tree* dari rekapitulasi nilai *Entropy* dan *Gain*nya:



Gambar 4.2 : Node

Pohon keputusan di atas belum terlihat keputusan yang dominan dari setiap program studi yang di pilih. Maka kita harus mencari kembali nilai *Entropy* dan *Gain* dari setiap atribut(kritera) Jenis Kelamin = Laki-laki.

1. Nilai *Entropy*

   Berikut ini adalah tabel penyelesainnya.

   Tabel: 4.3 Sampel Data Yang Di Uji Ulang (Kriteria Jenis Kelamin)

Kriteria	Attribut	Jumlah Kasus	SI	SK
Jenis Kelamin	**Laki-laki**	**7**	**4**	**3**

Setelah itu kita hitung nilai *Entropy* dari atribut Jenis Kelamin = Laki-Laki yang memiliki jumlah kasus “7” seperti terlihat pada Tabel di bawah ini.

Tabel 4.4 Atribut Jenis Kelamin

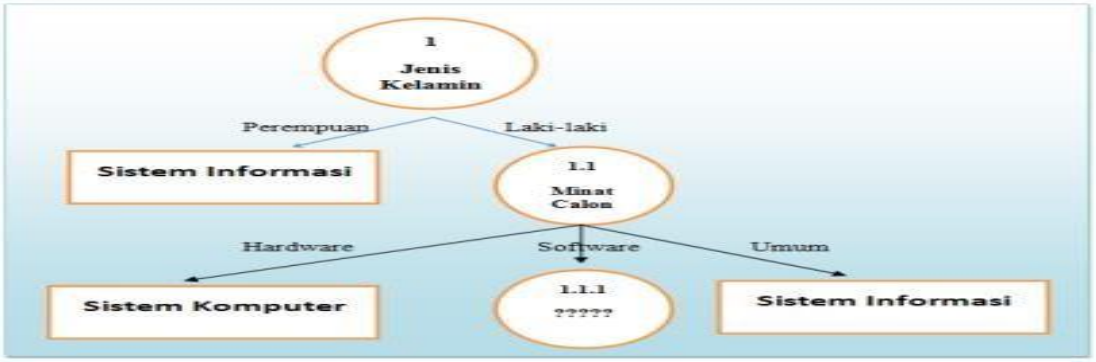
No.	Nama Mahasiswa	Minat Calon Mahasiswa	Asal Sekolah	Jenis Kelamin	Hobi	Hasil
1	Novita Devi Batu Bara	Hardware	SMK Komputer	Laki-Laki	Non	SK
2	Ahmad Riyandi	Hardware	SMK Komputer	Laki-Laki	IT	SK
3	Reza Adriansyah	Umum	SMK Komputer	Laki-Laki	Non	SI
4	Gafar Dwi Satrio	Software	SMA UMUM	Laki-Laki	Non	SI
8	Zulfikar Ali	Hardware	SMA UMUM	Laki-Laki	Non	SK
12	Asri Anzani Br. Tarigan	Umum	SMA UMUM	Laki-Laki	IT	SI
14	Akbar Widiantara	Software	SMA UMUM	Laki-Laki	IT	SK

Langkah selanjutnya menghitung nilainya, berikut ini adalah rekapitulasi nilai *entropy* dan *gain*nya.

Tabel 3. 5 Atribut Jenis Kelamin

NODE		Keterangan	Jml Kasus (S)	Sistem Komputer (SK)	Sistem Informasi (SI)	Entropy	Gain
1.1	**Jenis Kelamin = Laki-laki**		**7**	**4**	**3**	**0.98522**	
	Minat Calon						**0.69951**
		Hardware	**3**	**3**	**0**	**0**	
		Software	**2**	**1**	**1**	**1**	
		Umum	**2**	**0**	**2**	**0**	
	Asal Sekolah						**0.02024**
		SMK Komputer	**3**	**2**	**1**	**0.91829**	
		SMK Teknik	**0**	**0**	**0**	**0**	
		SMA Umum	**4**	**2**	**2**	**1**	
	Hobi						**0.02024**
		IT	**3**	**2**	**1**	**0.91829**	
		Non	**4**	**2**	**2**	**1**	

Berdasarkan tabel di atas terlihat bahwasanya Attribut = Minat Calon memiliki nilai Gain Tertinggi, maka untuk Root selanjutnya pada pohon keputusannya dapat terlihat pada gambar pohon (tree) berikut ini:



Gambar 3.3: Pohon Keputusan

Karena pohon keputusan belum terlihat keseluruhan hasilnya sehingga kita perlu untuk mencari kembali Nilai Gain dan Entropy selanjutnya berikut ini adalah tabelnya.

Tabel 4.6 Data Uji Akhirnya

Kriteria	Attribut	Jumlah Kasus	SK	SI
Minat Calon	**Software**	**2**	**1**	**1**

Setelah itu kita data terlebih dahulu dari atribut Minat Calon = Software yang memiliki jumlah kasus “2” seperti terlihat pada Tabel di bawah ini.

Tabel 4.7 Atribut Minat Calon Siswa

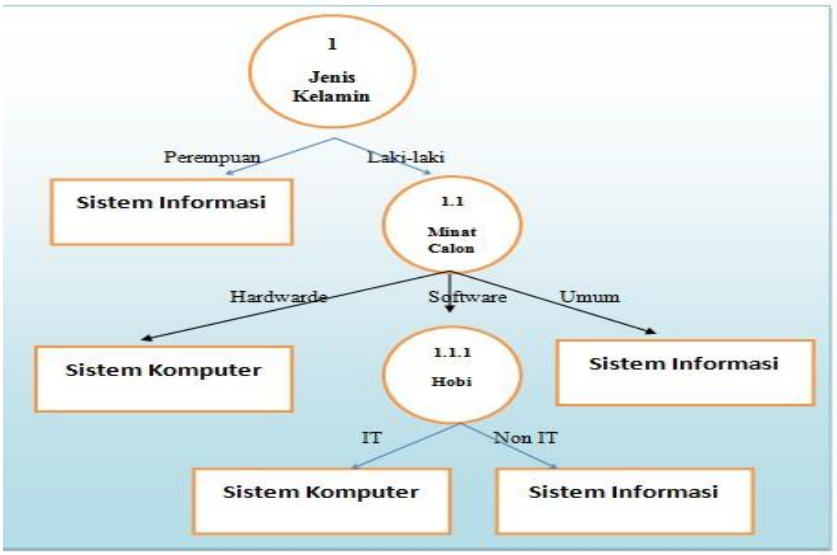
No.	Nama Mahasiswa	Minat Calon	Asal Sekolah	Hobi	Hasil
1	Gafar Dwi Satrio	Software	SMA UMUM	Non	SI
2	Akbar Widiantara	Software	SMA UMUM	IT	SK

Selanjutnya adalah kita menghitung kembali nilai Entropy dan Gainnya seperti terlihat pada tabel di bawah ini:

Tabel 4.8 Hitung Nilai Entropy Minat Calon = Software

NODE		Keterangan	Jml Kasus (S)	Sistem Komputer (SK)	Sistem Informasi (SI)	Entropy	Gain
1.1.1	Jenis Kelamin = Laki-laki Dan Minat Calon= Software		2	1	1	1	
	Asal Sekolah						0
		SMK Komputer	0	0	0	0	
		SMK Teknik	0	0	0	0	
		SMA Umum	2	1	1	1	
	Hobi						1
		IT	1	1	0	0	
		Non IT	1	0	1	0	

Gambar di atas menjelaskan bahwasanya yang memiliki kriteria memiliki nilai Gain tertinggi yaitu : 1 maka node pohon keputusannya adalah sebagai berikut:



Gambar 4.4: Hasil Dari Pohon Keputusan

Maka basis pengetahuan atau rule yang terbentuk yaitu:

1. Jika **Jenis Kelamin** = Perempuan maka **Hasil**= Sistem Informasi
2. Jika **Jenis Kelamin** = Laki-laki dan **Minat Calon**=Hardware maka **Hasil**= Sistem Komputer
3. Jika **Jenis Kelamin** = Laki-laki dan **Minat Calon**=Umum maka **Hasil=** Sistem Informasi
4. Jika **Jenis Kelamin** = Laki-laki dan **Minat Calon=** Software dan **Hobi**=IT maka **Hasil** = Sistem Komputer
5. Jika **Jenis Kelamin** = Laki-laki dan **Minat Calon**= Software dan **Hobi**=Non IT maka **Hasil** = Sistem Informasi



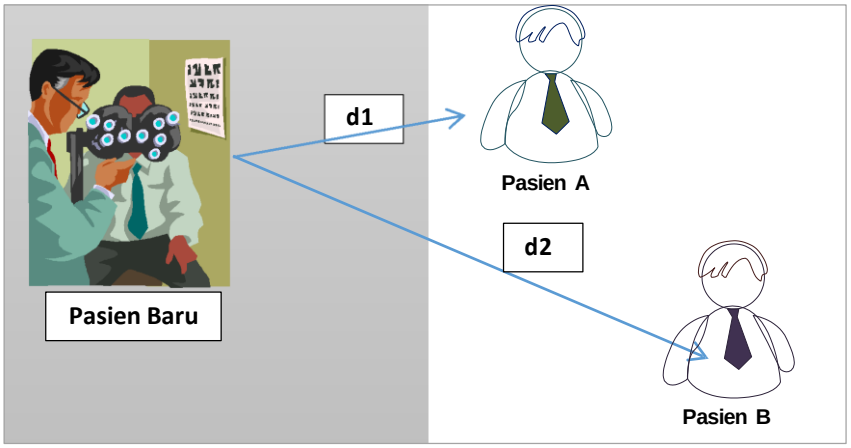
BAB V : ALGORITMA NEAREST NEIGHBOR

5.1 Pendahuluan

Algoritma Nearest Neighbor merupakan salah satu metode klasifikasi yang digunakan untuk pemecahan masalah pada bidang Data Mining. Sama halnya dengan beberapa metode lainnya yang ada pada metode klasifikasi, algoritma ini memiliki ciri yaitu dengan pendekatan untuk mencari kasus dengan menghitung kedekatan kasus yang baru dengan kasus yang lama. Adapun teknik yang digunakan yaitu berdasarkan bobot dari sejumlah objek kasus yang ada.

Contoh di dunia kesehatan yaitu ketika seorang dokter mencari solusi terhadap diagnosis penyakit pasien yang baru berdasarkan hasil diagnosis pasien yang lama.Atau seorang Guru BP yang menangani masalah siswa yang melakukan kesalahan akan ditinjau solusi penyelesaiannya berdasarkan kasus dari histori siswa yang bermasalahan lainnya. Selain itu juga seorang Hakim menetapkan hukuman yang dijatuhkan kepada seorang terpidana kasus kriminalitas selainnya menggunakan Undang-undang sebagai Dasar Hukum tetapi juga menggunakan Histori kriminal yang telah dilakukan seseorang.

Contoh di dunia kesehatan yaitu ketika seorang dokter mencari solusi terhadap diagnosis penyakit pasien yang baru berdasarkan hasil diagnosis pasien yang lama.Atau seorang Guru BP yang menangani masalah siswa yang melakukan kesalahan akan ditinjau solusi penyelesaiannya berdasarkan kasus dari histori siswa yang bermasalahan lainnya. Selain itu juga seorang Hakim menetapkan hukuman yang dijatuhkan kepada seorang terpidana kasus kriminalitas selainnya menggunakan Undang-undang sebagai Dasar Hukum tetapi juga menggunakan Histori kriminal yang telah dilakukan seseorang.



Gambar 5.1: Ilustrasi Kedekatan Kasus

Penjelasan dari gambar di atas dijelaskan bahwasanya ada dua pasien lama yaitu Pasien A dan B. Ketika seorang Dokter memeriksakan Pasien Baru maka solusi yang diambil dari Pasien tersebut berdasarkan hasil diagnosis terdekat dari pasien lama yaitu Pasien A dan Pasien B. Apabila nilai d1 diilustrasikan dengan kedekatan antara Pasien Baru dan Pasien A dan nilai d2 diilustrasikan dengan kedekatan antara Pasien Baru dan Pasien B. Karena nilai d1 memiliki kedekatan dibandingkan dengan nilai d2 terhadap Pasien Baru, maka Diagnosis Pasien Baru digunakan berdasarkan Pasien A. Di dalam Nearest Neighbor terdapat istilah “Similarity” atau kesamaan. Adapun rumus yang digunakan pada nilai Nearest Neighbor yaitu:

$$\text{Similarity } (T,S) = \frac{(\sum^{\square} \ \square(\square\square\square,\square i)*w\square\square)}{wi}$$

Keterangan:

T : Kasus Baru

S : Kasus yang ada dalam penyimpanan

n : jumlah attribut dalam setiap kasus

i : attribut individu antara 1 sampai dengan n

f : fungsi similarity atribut i antara kasus T dan Kasus S

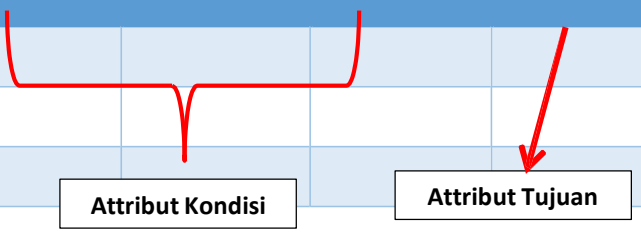
w : bobot yang diberikan pada atribut ke-i

Sebagai penjelasan tambahan bahwasanya untuk nilai Similaritas atau kesamaan berada di antara nilai 1 dan nilai 0. Yang mana untuk nilai 0 memiliki arti: Kasus Mutlak

tidak mirip dan apabila nilai 1 memiliki arti : Kasus Mutlak memiliki kemiripan. Seperti diketahui bahwasanya dalam Data Mining kita bermain dengan istilah Himpunan Data. Dalam Nearest Neighbor kita juga menggunakan himpunan data. Berikut ini adalah contoh Himpunan Data yang ada dalam Nearest Neighbor.

Tabel 5.1 : Himpunan Data Algoritma C4.5

No	Nama Pasien	Kriteria 1	Kriteria 2	Kriteria 3	Keterangan
1	Xxxx				
2	Xxxx				
3	Xxxx				



Untuk lebih jelasnya berikut ini ada sebuah kasus yang dapat dijadikan rujukan terhadap penjelasan rumus di atas.

2. CONTOH KASUS DAN PEMECAHAN

Pada sebuah Bank swasta di Indonesia, terdapat seorang calon nasabah yang ingin mengajukan Kredit Perumahan Rakyat (KPR) . Dalam hal ini terdapat beberapa nasabah yang telah mengajukan KPR ke Bank tersebut. Berikut ini adalah data-data nasabah yang pernah mengajukan KPR di Bank Swasta tersebut.

Tabel 5.2 : Sampel Kasus

No Kasus	Nama Nasabah	Kriteria 1	Kriteria 2	Kriteria 3	Keterangan
1	Dicky	Di bawah 30	Tinggi	Baik	Ya
2	Dicko	Di atas 30	Rendah	Baik	Tidak
3	Dicka	Di atas 30	Rendah	Tidak	Tidak

Keterangan:

Kriteria 1 = Menjelaskan tentang Kriteria “Umur”

Kriteria 2 = Menjelaskan tentang Kriteria “Penghasilan”

Kriteria 3 = Menjelaskan tentang Kriteria “BI Checking”

Dari tabel di atas, untuk mempermudah dalam penghitungan nilai Atributnya berikut ini adalah tabel bobot dari atribut tersebut di atas.

Tabel 5.3 : Sampel Kasus

Attribut	Bobot
Umur	0.5
Penghasilan	0.75
BI Checking	1

Untuk penyelesaian kasus pada fase awalnya kita perlu untuk menentukan Nilai kedekatan antara setiap nilai-nilai atribut. Berikut ini adalah kedekatan Nilai-nilai dari setiap atribut kondisinya.

1. Atribut Kriteria 1 (Umur)

Tabel 5.4: Kedekatan Nilai Atribut Kriteria 1 (Umur)

Nilai 1	Nilai 2	Kedekatan
Di bawah 30	Di bawah 30	1
Di atas 30	Di atas 30	1
Di bawah 30	Di atas 30	0.4
Di atas 30	Di bawah 30	0.4

2. Atribut Kriteria 2 (Penghasilan)

Tabel 5.5: Kedekatan Nilai Atribut Kriteria 2 (Penghasilan)

Nilai 1	Nilai 2	Kedekatan
Tinggi	Tinggi	1
Rendah	Rendah	1
Tinggi	Rendah	0.5
Rendah	Tinggi	0.5

3. Atribut Kriteria 3 (BI Checking)

Tabel 5.6 : Kedekatan Nilai Atribut Kriteria 3 (BI Checking)

Nilai 1	Nilai 2	Kedekatan
Baik	Baik	1
Tidak	Tidak	1
Baik	Tidak	0.75
Tidak	Baik	0.75

Soal: Misalkan terdapat seorang Nasabah baru yang ingin mengajuk Kredit Perumahan Rakyat (KPR) dengan keterangan di bawah ini:

Nama Nasabah	Kriteria 1	Kriteria 2	Kriteria 3
Dian	Di atas 30	Tinggi	Baik

Maka untuk menyelesaikan masalah di atas berikut ini adalah Algoritma penyelesainnya

1. Menghitung Nilai Kedekatan Atribut Kondisi Kasus baru dengan Kasus No 1

Tabel 5.7 : Kedekatan Kasus Baru Dengan Kasus 1

Nama Nasabah	Kriteria 1	Kriteria 2	Kriteria 3
Dian	Di atas 30	Tinggi	Baik
Dicky	Di bawah 30	Tinggi	Baik
Nilai Kedekatan	0.4	1	1
Nilai Atribut	a	c	e

Tabel 5.8: Bobot Kasus 1

Attribut	Bobot	Nilai Atribut
Umur	0.5	b
Penghasilan	0.75	d
BI Checking	1	f

Hitung:

Jarak = $\frac{(a*b) + (c*d) + (e*f)}{b + d + f}$

Jarak = $\frac{(0.4*0.5) + (1*0.75) + (1*1)}{0.5 + 0.75 + 1}$

Jarak = $\frac{2 + 0.75 + 1}{2.25}$

Jarak = 3.75 / 2.25

Jarak = 1.667

2. Menghitung Nilai Kedekatan Atribut Kondisi Kasus baru dengan Kasus No 2

Tabel 5.9 : Kedekatan Kasus Baru Dengan Kasus 2

Nama Pasien	Kriteria 1	Kriteria 2	Kriteria 3
Dian	Di atas 30	Tinggi	Baik
Dicko	Di atas 30	Rendah	Baik
Nilai Kedekatan	1	0.5	1
Nilai Atribut	a	c	e

Tabel 5.10 : Bobot Kasus 2

Attribut	Bobot	Nilai Atribut
Umur	0.5	b
Penghasilan	0.75	d
BI Checking	1	f

Hitung:

Jarak = $\frac{(a*b) + (c*d) + (e*f)}{b + d + f}$

Jarak = $\frac{(1*0.5) + (0.5*0.75) + (1*1)}{}$

0.5 + 0.75 + 1

Jarak = 0.5 + 0.28125+ 1

2.25

Jarak = 1.78125/ 2.25

Jarak = 0.791667

3. Menghitung Nilai Kedekatan Atribut Kondisi Kasus baru dengan Kasus No 3

Tabel 5.11: Kedekatan Kasus Baru Dengan Kasus 2

Nama Nasabah	Kriteria 1	Kriteria 2	Kriteria 3
Dian	Di atas 30	Tinggi	Baik
Dicka	Di atas 30	Rendah	Tidak
Nilai Kedekatan	1	0.5	0.75
Nilai Atribut	a	c	e

Tabel 5.12 : Bobot Kasus 2

Attribut	Bobot	Nilai Atribut
Umur	0.5	b
Penghasilan	0.75	d
BI Checking	1	f

Hitung:

Jarak = (a*b) + (c*d) + (e*f)

b + d + f

Jarak = (1*0.5) + (0.5*0.75) + (0.75*1)

0.5 + 0.75 + 1

Jarak = 0.5 + 0.28125+ 0.75

2.25

Jarak = 1.53125/ 2.25

Jarak = 0.68055

Dari langkah 1, 2 dan 3 dapat diketahui nilai kedekatannya, berikut ini adalah rekapitulasi.

Tabel 5.13: Rekapitulasi Nilai Kedekatan

No	Kasus	Nilai Kedekatan
1	Kasus 1	1.667
2	Kasus 2	0.791667
3	Kasus 3	0.68055

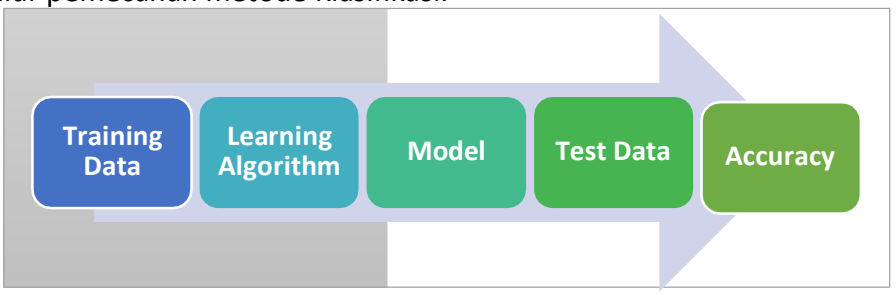
Berdasarkan tabel diatas untuk nilai Kedekatan tertinggi adalah pada nilai Kedekatan dengan Kasus 1. Jadi, untuk soal di atas maka nasabah atas nama "**Dian**" pada nilai atribut keterangannya bernilai **"Ya".**

BAB 6 BAB VI : ALGORITMA NAÏVE BAYESIAN CLASIFIER

6.1 Pendahuluan

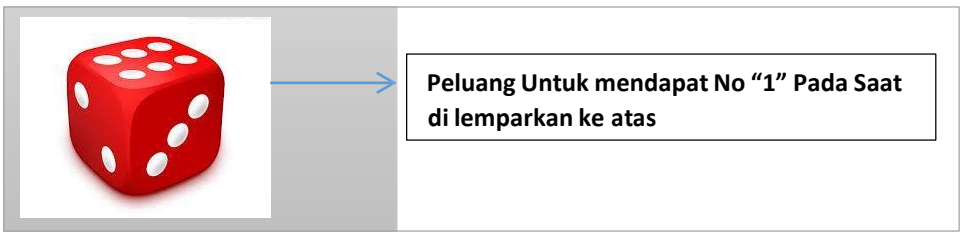
Naïve Bayesian Clasifier merupakan salah satu algoritma pemecahan masalah yang termasuk kedalam Metode Klasifikasi pada Data Mining. Naïve Bayesian Clasifier mengadopsi ilmu statistika yaitu dengan menggunakan teori kemungkinan (Probabilitas) untuk menyelesaikan sebuah kasus *Supervised Learning*, artinya dalam himpunan data terdapat Label, Class atau Target sebagai acuan atau gurunya.

Naïve Bayesian Clasifier dalam konsep penyelesaiannya tidak jauh beda dengan konsep Nearest Neighbor. Seperti kita ketahui bahwasanya dalam metode klasifikasi terdapat beberapa fase penyelesaian yaitu dimulai dari Training dan diakhiri dengan proses Testing sehingga dihasilkan sebuah keputusan yang akurat. Berikut ini adalah gambar alur pemecahan metode Klasifikasi.



Gambar 6.1: Fase Penyelesaian Metode Klasifikasi

Pada *Naïve Bayesian Clasifier* yang dimaksud Learning yaitu proses pembelajaran dengan cara menghitung nilai probabilistik dari suatu kasus. Sedang testing yaitu proses pengujian menggunakan model yang mengadopsi data testing. Adapun contoh teori peluang sehingga kita mudah memahami *Naïve Bayesian Clasifier*dapat terlihat pada gambar dan penjelasan di bawah ini.



Gmbar 6.2 : Ilustrasi Peluang

Dari gambar di atas dapat kita mengetahui secara sederhana bahwasanya peluang untuk mendapatkan no.1 yaitu : 1/6. Dengan asumsi jumlah yang bernilai dadu no.1 ada 1 sedangkah total keseluruhan datu ada 6.

Berikut ini adalah rumus untuk mencari nilai peluang dari Hipotesa benar (valid) untuk data sampel X yaitu:

$$P(H|X)=\frac{P(X|H)P(H)}{P(X)}$$

Dari rumus di atas, sebagai dasar teori bayesian sebagai pemecahan masalah, kita harus mengetahui terlebih dahulu beberapa hal diantaranya yaitu:

X: sampel data yang memiliki kelas (label) yang tidak diketahui

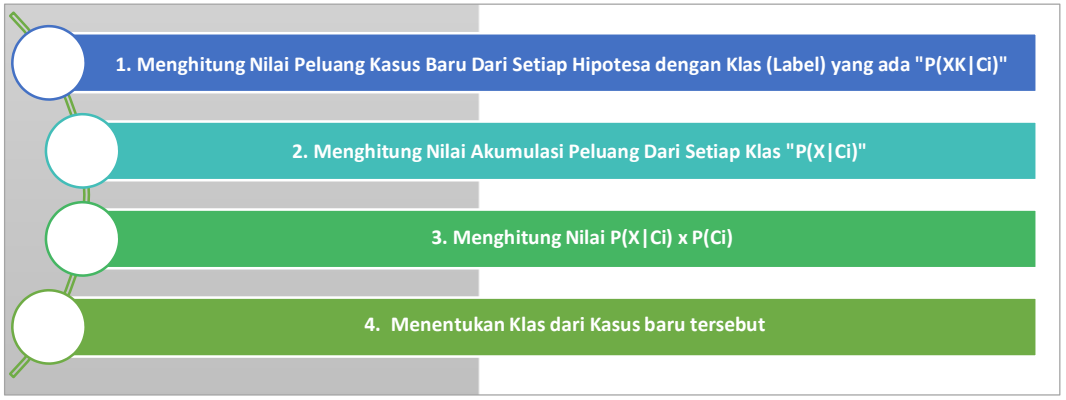
H : hipotesa bahwa x adalah data kelas (label)

P(H) : peluang dari hipotesa H

P(X) : peluang dari data sampel yang di amati

P(X|H) : peluang dari data sampel X bila diasumsikan bahwa hipotesa benar

Sehingga *Naïve Bayesian Clasifier*dapat didefinisikan juga sebagai metode klasifikasi yang berdasarkan teori probabilitas dan teorema bayesian dengan asumsi bahwa setiap variabel atau parameter penentu keputusan bersifat bebas (independence) sehingga keberadaan setiap variabel tidak ada kaitannya dengan keberadaan atribut yang lain. Adapun algoritma penyelesaian dari *Naïve Bayesian Clasifier*dapat di lihat pada gambar di bawah ini:



Gambar 6.3 : Algoritma Naïve Bayesian Clasifier

1. **Contoh Kasus Dan Pemecahan Masalah**

Pada sebuah Bank swasta di Indonesia, terdapat seorang calon nasabah yang ingin mengajukan Kredit Perumahan Rakyat (KPR) . Dalam hal ini terdapat beberapa nasabah yang telah mengajukan KPR ke Bank tersebut. Berikut ini adalah data-data nasabah yang pernah mengajukan KPR di Bank Swasta tersebut.

Tabel 6.1 : Sampel Kasus*Naïve Bayesian Clasifier*

No Kasus	Nama Nasabah	Kriteria 1	Kriteria 2	Kriteria 3	Ket
1	Dicky	Di bawah 30	Tinggi	Baik	Ya
2	Dicko	Di atas 30	Rendah	Baik	Tidak
3	Dicka	Di atas 30	Tinggi	Baik	Ya
4	Dina	Di bawah 30	Tinggi	Tidak	Tidak
5	Dini	Di bawah 30	Sedang	Baik	Ya
6	Dino	Di atas 30	Sedang	Baik	Ya

Keterangan:

Kriteria 1 = Menjelaskan tentang Kriteria “Umur”

Kriteria 2 = Menjelaskan tentang Kriteria “Penghasilan”

Kriteria 3 = Menjelaskan tentang Kriteria “BI Checking”

Soal: Misalkan terdapat seorang Nasabah baru yang ingin mengajuk Kredit Perumahan Rakyat (KPR) dengan keterangan di bawah ini:

Nama Nasabah	Kriteria 1	Kriteria 2	Kriteria 3
Dian	Di atas 30	Sedang	Baik

<u>Penyelesaian:</u>

1. Hitung nilai P(XK|Ci) untuk setiap class i
   - P(Kriteria 1= “Di atas 30” | Keterangan = “Ya”)

     P(Kriteria 1 = 2/4 = 0.5
   - P(Kriteria 1= “Di atas 30” | Keterangan = “Tidak”)

P(Kriteria 1 = 1/2 = 0.5

- P(Kriteria 2= “Sedang” | Keterangan = “Ya”)

  P(Kriteria 2 = 2/4 = 0.5

- P(Kriteria 2= “Sedang” | Keterangan = “Tidak”)

  P(Kriteria 2 = 0/2 = 0

- P(Kriteria 3= “Baik” | Keterangan = “Ya”)

  P(Kriteria 2 = 4/4 = 1

- P(Kriteria 3= “Baik” | Keterangan = “Tidak”)

  P(Kriteria 2 = 1/2 = 0.5

2. Hitung nilai P(X|Ci) untuk setiap Kelas (label)
   - P(X|Keterangan = “Ya”)

     = 0.5 x 0.5 x 1 = 1.25

   - P(X|Keterangan = “Tidak”)

     = 0.5 x 0 x 0.5 = 0
3. Hitung nilai P(X|Ci) * P (Ci)

   -(P(X|Keterangan = “Ya”) x P(Keterangan=Ya”)

   = 1.25 x 4/6 = 0.8333

   -(P(X|Keterangan = “Ya”) x P(Keterangan=Ya”)

   = 0 x 2/6 = 0
4. Menentukan klas dari kasus tersebut

   Berdasarkan perhitungan akhir dengan mengalikan nilai peluang dari kasus yang di angkat, kita melihat bahwa nilai P(X|Keterangan=”Ya”) lebih tinggi dari P(X|Keterangan=”Tidak”) = 0.833 banding 0, maka

Nama Nasabah	Kriteria 1	Kriteria 2	Kriteria 3	Keterangan
Dian	Di atas 30	Sedang	Baik	**Ya**

LATIHAN :

Pada sebuah SMA Swasta XXX ingin menganalisa Kelulusan Siswa di sekolahnya. Adapun sampel Data Siswa SMA Swasta XXX yaitu:

Tabel: Sampel Kasus *Naïve Bayesian Clasifier*

No Kasus	Nama Siswa	Kriteria 1	Kriteria 2	Kriteria 3	Ket
1	Doni	Sangat Baik	Sangat Baik	Sangat Baik	Ya
2	Dono	Baik	Kurang	Kurang	Tidak
3	Dona	Sangat Baik	Sangat Baik	Baik	Ya
4	Dika	Baik	Kurang	Baik	Tidak
5	Diki	Sangat Baik	Baik	Sangat Baik	Ya
6	Diko	Baik	Sangat Baik	Sangat Baik	Ya

Apabila terdapat Nama Siswa = Dodi dengan Kriteria 1 = Baik, Kriteria 2= Kurang, Dan Kriteria 3= Sangat Baik maka mahasiswa tersebut dapat digolongkan kedalam? (Gunakan Naïve Bayesian Clasifier)??

BAB 7

BAB VII : FUNGSI MINOR METODE PREDIKSI

7.1 Pendahuluan

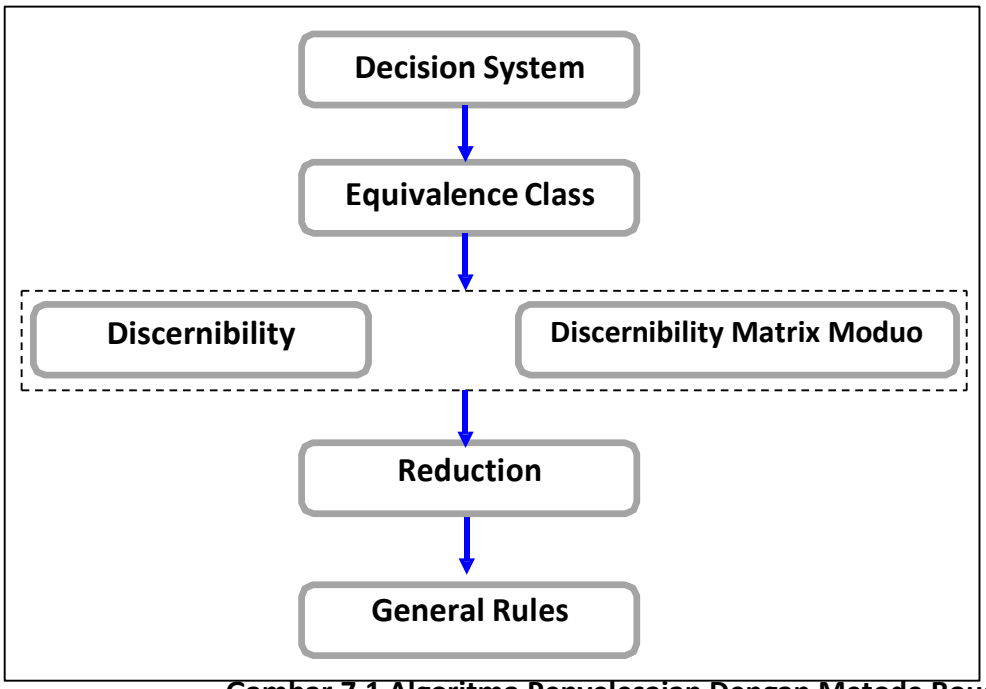
Di beberapa jurnal internasional, definisi *Rough Set* adalah sebagai berikut : *Rough set theory is an elegant and powerful methodology inextracting and minimizing rules from decision tables and Pawlakinformation systems. Its central notions are core, reduct, andknowledge dependency.* (T.Y Lin, 1996).

Melalui jurnal tersebut Metode *Rough Set*, metodologi yang elegan dan kuat dalam penggalian dan meminimalkan aturan dari tabel keputusan dan sistem informasi Pawlak. Gagasan utamanya adalah inti, mengecil, dan ketergantungan pengetahuan.

Pada jurnal internasional lainnya bahwasanya *Rough Set* di definisikan *An approach first forwarded by mathematician Zdzislaw Pawlak at the beginning of theeighties; it is used as a mathematical tool to treat the vague and the imprecise. Rough Set Theory is similar to Fuzzy Set Theory, however the uncertain and imprecision in thisapproach is expressed by a boundary region of a set, and not by a partial membership as inFuzzy Set Theory. Rough Set concept can be defined quite generally by means of interiorand closure topological operations know approximations (*Pawlak,1982*).*

Berdasarkan kutipan jurnal internasional di atas di jelaskan bahwasanya Metode *Rough Set* merupakan sebuah pendekatan pertama diteruskan oleh matematikawan Zdzislaw Pawlak pada awal tahun delapan puluhan, melainkan digunakan sebagai alat matematika untuk mengobati kabur dan tidak tepat. Teori Set kasar mirip dengan Teori Set Fuzzy, namun pasti dan ketidaktepatan dalam pendekatan ini diungkapkan oleh daerah batas set, dan bukan oleh keanggotaan parsial seperti pada Teori Set Fuzzy. Konsep Set kasar dapat didefinisikan secara umum cukup dengan cara operasi interior dan penutupan topologi perkiraan.

Berikut ini adalah Skema penyelesaian menggunakan metode Rough Set yaitu:



Gambar 7.1 Algoritma Penyelesaian Dengan Metode Rough Set

Keterangan:

1. *Decision System*

Decision system adalah *information system* dengan atribut tambahan yang dinamakan dengan *decision atribute*, dalam *data mining* dikenal dengan nama kelas atau target. Atribut ini merepresentasikan hasil dari klasifikasi yang diketahui. *Decision system* merupakan fungsi yang mendeskripsikan *information system,* maka *InformationSystems* (*IS*) menjadi

DS=(U,{A,C}).

Keterangan

U= {x1, x2,…, xm} yang merupakan sekumpulan *example*.

A= {a1, a2, …, an} yang merupakan sekumpulan *attribute* kondisi secara berurutan.

C=*decision attributes* (keputusan).

2. *Equivalence Class*

Equivalence class adalah mengelompokan objek-objek yang sama untuk attribute A (U, A).

3. *Discernibility Matrix*

Definisi *Discerniblity Matrix*: Diberikan sebuah IS A=(U,A) and B Gabung A, *discernibility matrix* dari A adalah MB, dimana tiap-tiap entry MB(I,j) tediri dari sekumpulan attribute yang berbeda antara objek Xi dan Xj.

4. *Discernibility matrix modulo D*

Discernibility matrix modulo D didefinisikan seperti berikut dimana *Modulo* (i,j) adalah sekumpulan *attribute* yang berbeda antara objek x_i dan x_j dan juga berbeda *attribute* keputusan. Diberikan sebuah DS A=(U,A{d{) dan subset dari attribut B Gabung A, *discernibility matrix modulo* D dari A, MBd, didefinisikan seperti berikut dimana MB(I,j) adalah sekumpulan attribute yan berbeda antara objek Xi dan Xj dan juga berbeda attribute keputusan.

5. Reduction

Untuk data yang jumlah variabel yang sangat besar sangat tidak mungkin mencari seluruh kombinasi variabel yang ada, karena jumlah *indiscernibility* yang dicari = (2 n-1-1). Oleh karena itu dibuat satu teknik pencarian kombinasi atribut yang mungkin yang dikenal dengan *QuickReduct*, yaitu dengan cara :

1. Nilai *indiscernibility* yang pertama dicari adalah *indiscernibility* untuk kombinasi atribut yang terkecil yaitu 1.
2. Kemudian lakukan proses pencarian *dependency attributes*. Jika nilai *dependency attributes* yang didapat =1 maka *indiscernibility* untuk himpunan miminal variabel adalah variabel tersebut.
3. Jika pada proses pencarian kombinasi atribut tidak ditemukan *dependency attributes* =1, maka lakukan pencarian kombinasi yang lebih besar, di mana kombinasi *variable* yang dicari adalah kombinasi dari variabel di tahap sebelumnya yang nilai *dependency attributes* paling besar.Lakukan proses (3), sampai didapat nilai *dependency attributes* =1.

6. *General Rules*

Proses utama menemukan pengetahuan dalam database adalah ekstraksi aturan dari sistem pengambilan keputusan. Metode set kasar dalam menghasilkan aturan-aturan keputusan dari tabel keputusan didasarkan pada perhitungan set mengecil.

7.2 Metode Rough Set

Pada sub ini merupakan gambaran proses analisa suatu masalah dan gambaran dari penerapan metode untuk memecahkan masalah yang dihadapi. Untuk menunjang analisis data dalam pencarian *knowledge*, penulis mencoba untuk mentransformasi data manual yang terdapat pada Program Studi Diploma 3(D3) di STMIK XXXXXX

Di dalam *data mining* terdapat beberapa tekhnik pengolahan data agar data yang diolah lebih bermanfaat atau bernilai ilmu. Salah satunya adalah Metode *Rough Set*.

a. Sampel Data Yang Digunakan

Tahapan ini menjelaskan mengenai sampel data yang digunakan untuk proses analisa metode *Rough Set*. Sampel yang digunakan diambil berdasarkan data mahasiswa pada *point* sebelumnya.

Tabel 7.1 Kriteria Penilaian

No	Nilai	Range
1	A	80 – 100
2	B+	75 – 79
3	B	70 – 74
4	C+	65 – 69
5	C	60 – 64
6	D	50 – 59
7	E	0 – 49

Adapun kriteria penilaian yang dinyatakan ”Lulus” yaitu dengan Nilai : A, B+, B, C+, sedangkan mahasiswa yang dikatakan ”Gagal” yaitu dengan Nilai : C, D, dan E.

Berikut ini adalah tabel rekapitulasi sampel data yang digunakan. Adapun data yang digunakan sebagai sampel dilakukan secara acak berdasarkan kriteria yang sama.

Tabel 7.2 Sampel Data Yang Akan Di Proses

NO	Objek	Nilai				
		Kehadiran	Tugas	UTS	UAS	Hasil
1	Dina	100	65	60	75	Lulus
2	Dini	100	70	80	95	Lulus
3	Dino	57	60	0	0	Gagal
4	Doni	0	0	50	0	Gagal
5	Dona	0	40	0	40	Gagal
6	Done	79	60	20	60	Gagal
7	Dian	93	60	60	65	Lulus
8	Dipa	71	58	0	0	Gagal
9	Dipi	0	0	0	0	Gagal
10	Dapi	21	0	0	0	Gagal

b. Analisa Kriteria Penilaian

Studi kasus yang dilakukan pada Program Studi Diploma 3 (D3) di STMIK XXXXXX. Mahasiswa pada Program Studi Diploma 3 (D3) di STMIK XXXXXXyang akan dijadikan sampel untuk menganalisis data Kelulusan Mahasiswa pada suatu mata kuliah di Program Studi Diploma 3 (D3) di STMIK XXXXXX. Berikut ini adalah tabel untuk menganalisis data Mahasiswa pada Program Studi Diploma 3 (D3) di STMIK XXXXXX

Ada beberapa kriteria yang dapat menyimpulkan bahwa Mahasiswa itu adalah Mahasiswa potensial atau tidak, diantaranya:

1. **Nilai Kehadiran**

   Untuk nilai kehadiran sangat menentukan keikutsertaan mahasiswa di dalam mengikuti Ujian Akhir Semester(UAS). Jika kehadiran kurang dari 75% dari total pertemuan di setiap semesternya mahasiswa yang bersangkutan tidak bisa untuk

mengikuti Ujian Akhir Semester(UAS). Adapun klasifikasi dari penilaian kehadiran ada 2(dua) yaitu “Memenuhi Syarat” dan “Tidak Mencukupi”.

Adapun yang memenuhi syarat apabila nilai kehadiran melebihi dari 75% sedangkan kurang dari 75% di klasifikasikan kedalam tidak mencukupi.

2. Nilai Tugas/Quiz

Untuk nilai Tugas/Quiz saya klasifikan ke dalam 2(dua) bagian yaitu: “Di atas Rata-Rata” dan “Di bawah Standar”. Adapun nilai diatas rata-rata memiliki *range* antara 65 smpai dengan 100. Di bawah nilai tersebut dikatakan sebagai di bawah standar.

3. Nilai Ujian Tengah Semester(UTS)

Untuk nilai Ujian Tengah Semester (UTS) saya klasifikan ke dalam 2(dua) bagian yaitu : “Di atas Rata-Rata” dan “Di bawah Standar”. Adapun nilai di atas rata-rata memiliki *range* antara 65 smpai dengan 100. Di bawah nilai tersebut dikatakan sebagai di bawah standar.

4. Nilai Ujian Akhir Semester(UAS)

Untuk Nilai Ujian Akhir Semester (UAS) saya klasifikan ke dalam 2(dua) bagian yaitu : “Di atas Rata-Rata” dan “Di bawah Standar”. Adapun nilai di atas rata-rata memiliki *range* antara 65 smpai dengan 100. Di bawah nilai tersebut dikatakan sebagai di bawah standar.

Rough set menawarkan dua bentuk representasi data yaitu *Information Systems* (IS) dan *Decision Systems* (DS). Definisi *Decision Systems* yaitu Sebuah pasangan *Information System*, di mana”U” adalah Anggota bilangan “n” dan yang merupakan sekumpulan *example* dan *atribute* kondisi secara berurutan. Definisi diatas memperlihatkan bahwa sebuah *Information Systems* yang terdiri dari sekumpulan *example*, seperti {Anggota01, Anggota 02, dan Anggota bilangan-m} dan *atribute* kondisi, seperti Kondisi Awal. Berikut ini adalah contoh analisa penerapan Metode *Rough Set*. Adapun studi kasus yang diangkat yaitu “Analisa Data Mahasiswa Lulus Matakuliah Pada Program Studi Diploma 3 (D3) di STMIK XXXXXX”.

Algoritma penyelesaian masalah dari metode *Rough Set* yaitu sebagai berikut:

1. *Decision System*

Decision System merupakan *Information System* yang telah memiliki keputusan atau hasil berdasarkan asumsi yang telah memenuhi syarat dan ketentuan berdasarkan atributnya. Berikut ini adalah tabel *Decision System* dari data mahasiswa Program Studi Diploma 3 (D3) di STMIK XXXXXX yang menjadi objek penelitian.

Tabel 7.3 *Decision System*

NO	Objek	Nilai				
		Kehadiran	Tugas/Quiz	UTS	UAS	Hasil
1	E1	Memenuhi Syarat	Di atas Rata-Rata	Di bawah Standar	Di atas Rata-Rata	Lulus
2	E2	Memenuhi Syarat	Di atas Rata-Rata	Di atas Rata-Rata	Di atas Rata-Rata	Lulus
3	E3	Tidak Mencukupi	Di bawah Standar	Di bawah Standar	Di bawah Standar	Gagal
4	E4	Tidak Mencukupi	Di bawah Standar	Di bawah Standar	Di bawah Standar	Gagal
5	E5	Tidak Mencukupi	Di bawah Standar	Di bawah Standar	Di bawah Standar	Gagal
6	E6	Memenuhi Syarat	Di bawah Standar	Di bawah Standar	Di bawah Standar	Gagal
7	E7	Memenuhi Syarat	Di bawah Standar	Di bawah Standar	Di atas Rata-Rata	Lulus
8	E8	Memenuhi Syarat	Di bawah Standar	Di bawah Standar	Di bawah Standar	Gagal
9	E9	Tidak Mencukupi	Di bawah Standar	Di bawah Standar	Di bawah Standar	Gagal
10	E10	Tidak Mencukupi	Di bawah Standar	Di bawah Standar	Di bawah Standar	Gagal

Tabel 7.3 memperlihatkan sebuah *DecisionSystems* yang sederhana. Ia hanya terdiri dari *n* objek, E1, E2, E3, E4, E5, E6, E7, E8, E9 dan E10 seperti Nilai Kehadiran, Nilai Tugas/Quiz, Nilai UTS, Nilai UAS Serta Hasil. Dalam tabel ini, *n*-1 Kehadiran, Tugas/Quiz, UTS dan UAS adalah *atribute* kondisi, sedangkan Hasil adalah *DecisionAtribute*.

2. *Equivalen Class*

Equivalence Class adalah mengelompokan objek-objek yang sama untuk *atribute* tertentu. Diberikan *DecisionSystems* pada tabel 7.4, maka dapat diperoleh *equivalence class* (*Equivalence Class* 1 sampai dengan *Equivalence Class 7*) seperti digambarkan pada Tabel 7.4

Tabel 7.4 *Equivalence Class*

Objek	Nilai Kehadiran	Nilai Tugas/Quiz	Nilai UTS	Nilai UAS	Hasil
EC1	Memenuhi Syarat	Diatas Rata-rata	Dibawah Standar	Diatas Rata-rata	Lulus
EC2	Memenuhi Syarat	Diatas Rata-rata	Diatas Rata-rata	Diatas Rata-rata	Lulus
EC3	Tidak Mencukupi	Dibawah Standar	Dibawah Standar	Dibawah Standar	Gagal
EC4	Tidak Mencukupi	Diatas Rata-rata	Dibawah Standar	Dibawah Standar	Gagal
EC5	Memenuhi Syarat	Dibawah Standar	Dibawah Standar	Diatas Rata-rata	Gagal
EC6	Memenuhi Syarat	Dibawah Standar	Diatas Rata-rata	Diatas Rata-rata	Lulus
EC7	Memenuhi Syarat	Dibawah Standar	Dibawah Standar	Dibawah Standar	Gagal

3. *Discernibility Matrix atau Discernibility Matrix Modulo D*

Setelah dilakukan klasifikasi menggunakan *Equivalence Class*. Langkah selanjutnya untuk menganalisa data tersebut adalah dengan salah satu proses antara *Discernibility Matrix* atau *Discernibility Matrix Modulo D*.

Untuk menghitung *Discernibility Matrix* atau *Discernibility Matrix Modulo D*kita menggunakan tabel acuan *Discernibility Matrix* atau *Discernibility Matrix Modulo D*seperti terlihat pada tabel 4.5 di bawah ini.

Tabel 7.5 Tabel Acuan *Discernibility Matrix* atau *Discernibility Matrix Modulo*

Objek	Nilai Kehadiran	Nilai Tugas/Quiz	Nilai UTS	Nilai UAS	Hasil
EC1	Memenuhi Syarat	Diatas Rata-rata	Dibawah Standar	Diatas Rata-rata	Lulus
EC2	Memenuhi Syarat	Diatas Rata-rata	Diatas Rata-rata	Diatas Rata-rata	Lulus
EC3	Tidak Mencukupi	Dibawah Standar	Dibawah Standar	Dibawah Standar	Gagal
EC4	Tidak Mencukupi	Diatas Rata-rata	Dibawah Standar	Dibawah Standar	Gagal
EC5	Memenuhi Syarat	Dibawah Standar	Dibawah Standar	Diatas Rata-rata	Gagal
EC6	Memenuhi Syarat	Dibawah Standar	Diatas Rata-rata	Diatas Rata-rata	Lulus
EC7	Memenuhi Syarat	Dibawah Standar	Dibawah Standar	Dibawah Standar	Gagal

Untuk mendapatkan nilai *discernibilitymatrix*-nya yaitu dengan mengklasifikasikan atribut yang berbeda antara objek *ke-i* dan objek *ke-j* (yang dilihat hanya atribut kondisi saja). Berdasarkan data di atas maka berikut ini adalah *Discernibility Matrix* nya :

Tabel 7.6 Hasil *Dicernibility Matrix*

Objek	EC1	EC2	EC3	EC4	EC5	EC6	EC7
EC1	X	C	ABD	AD	B	BC	BD
EC2	C	X	ABCD	ACD	BC	B	BCD
EC3	ABD	ABCD	X	B	AD	ACD	A
EC4	AD	ACD	D	X	ABD	ABCD	AB
EC5	B	BC	AD	ABD	X	C	D
EC6	BC	B	ACD	ABCD	D	X	CD
EC7	BD	BCD	A	ABD	C	CD	X

Selain itu juga kita dapat menggunakan *Dicernibility Matrix Modulo D. Dicernibility Matrix Modulo D* ini merupakan sekumpulan atribut yang berbeda antara objek *ke-i* dan *ke-j* beserta dengan atribut hasilnya seperti terlihat pada tabel di bawah ini.

Tabel 7.7 Hasil *Dicernibility Matrix Modulo D*

Objek	EC1	EC2	EC3	EC4	EC5	EC6	EC7
EC1	X	X	ABD	AD	B	X	BD
EC2	X	X	ABCD	ACD	BC	X	BCD
EC3	ABD	ABCD	X	X	X	ACD	X
EC4	AD	ACD	X	X	X	ABCD	X
EC5	B	BC	X	X	X	C	X
EC6	X	X	ACD	ABCD	C	X	CD
EC7	BD	BCD	X	X	X	CD	X

Adapun penulis menggunakan *Dicernibility Matrix* sebagai acuan untuk melakukan proses *Reduction.*

4. Reduction

Untuk data yang jumlah variabel yang sangat besar sangat tidak mungkin mencari seluruh kombinasi variabel yang ada, karena jumlah *indiscernibility* yang dicari = (2 n-1-1). Oleh karena itu dibuat satu teknik pencarian kombinasi atribut yang mungkin yang dikenal dengan *QuickReduct*, yaitu dengan cara :

1. Nilai *indiscernibility* yang pertama dicari adalah *indiscernibility* untuk kombinasi atribut yang terkecil yaitu 1.
2. Kemudian lakukan proses pencarian *dependency atributes*. Jika nilai *dependency atributes* yang didapat =1 maka *indiscernibility* untuk himpunan miminal variabel adalah variabel tersebut.
3. Jika pada proses pencarian kombinasi atribut tidak ditemukan *dependency atributes* =1, maka lakukan pencarian kombinasi yang lebih besar, di mana kombinasi *variable* yang dicari adalah kombinasi dari variabel di tahap sebelumnya yang nilai *dependency atributes* paling besar. Lakukan proses (3), sampai didapat nilai *dependency atributes* =1.

Berdasarkan proses *Reduction* di atas berikut ini adalah hasil tabel penyeleksian.

Tabel 7.8 Proses Penyeleksian

Class	CNF of Boolean Function	Prime Implicant	Reduct
EC1	C ∧ (A∨B∨D)∧ (A∨D)∧ B∧ (B∨C) ∧ (B∨D)	C∧ (A∨D) ∧ B	{C}, {A}, {B}
EC2	C ∧ (A∨B∨C∨D) ∧ (A∨C∨D) ∧ (B∨C)∧ B ∧ (B∨C∨D)	C∧ B	{C}, {B}
EC3	(A∨B∨D) ^ (A∨B∨C∨D) ^ B ^ (A∨D) ^ (A∨C∨D) ^ A	B∧ A	{B}, {A}
EC4	(A∨D) ^ (A∨C∨D) ^ D ^ (A∨B∨D) ^ (A∨B∨C∨D) ^ (A∨B)	(A∨D) ^ D	{D}, {A}, {B}
EC5	(B∨C) ^ B ^ (A∨C∨D) ^ (ABCD) ^ D ^ (C∨D)	B ^ (C∨D)	{B}, {C}, {D}
EC6	(B∨D) ^ (B∨C∨D) ^ A ^ (A∨B∨D) ^ C ^ (C∨D)	(B∨D)	{B}, {D}
EC7	(B∨D) ^ (B∨C∨D) ^ A ^ (A∨B∨D) ^ C ^ (C∨D)	(B∨D)	{B}, {D}

5. *General Rules*

Setelah didapatkan hasil dari *Reduction*, maka langkah terakhir untuk menentukan *GeneralRule* nya. Adapun *GeneralRule* dari hasil *Reduction* yang dideskripsikan pada tabel penyeleksian adalah sebagai berikut:

Tabel 7.9 Keputusan Atau *Knowledge* Baru

No	Nilai Atribut	Nama Atribut Dan Kondisi	Keputusan
1	A	Kehadiran = Memenuhi Syarat	Hasil= Lulus atau Hasil= Gagal
2	A	Kehadiran = Tidak Mencukupi	Hasil= Gagal
3	B	Tugas = Di Atas Rata-Rata	Hasil=Lulus
4	B	Tugas = Di Bawah Standar	Hasil = Lulus or Hasil = Gagal
5	C	UTS = Di Bawah Standar	Hasil = Lulus or Hasil = Gagal
6	C	UTS = Di Atas Rata-Rata	Hasil = Lulus
7	D	UAS = Di Atas Rata-Rata	Hasil = Lululs
8	D	UAS = Di Bawah Standar	Hasil = Gagal

Berdasarkan tabel di atas dapat dideskripsikan sebagai berikut:

1. ***If*** Kehadiran = Memenuhi Syarat ***then*** Hasil = Lulus ***or*** Hasil = Gagal
2. ***If*** Kehadiran = Tidak Mencukupi ***then*** Hasil = Gagal
3. ***If*** Tugas = Di Atas Rata-rata ***then*** Hasil = Lulus
4. ***If*** Tugas = Di Bawah Standar ***then*** Hasil = Lulus ***or*** Hasil = Gagal
5. ***If*** UTS = Di Bawah Standar ***then*** Hasil = Lulus ***or*** Hasil = Gagal
6. ***If*** UTS = Di Atas Rata-rata ***then*** Hasil = Lulus
7. ***If*** UAS = Di Atas Rata-rata ***then*** Hasil = Lulus
8. ***If*** UAS = Di Bawah Standar ***then*** Hasil = Gagal

Latihan:

Berikut ini adalah Data Kegempaan yang di himpun dari Badan Meterologi Dan Geofisika.

Tabel : Faktor Kegempaan

No	Sampel	K1	K2	K3	K4	K5	Prediksi
1	E1	MS	MS	MS	MS	MS	Gempa
2	E2	TMS	TMS	TMS	TMS	TMS	Tidak
3	E3	MS	MS	TMS	TMS	TMS	Tidak
4	E4	MS	TMS	TMS	TMS	TMS	Tidak
5	E5	MS	TMS	TMS	TMS	TMS	Tidak
6	E6	TMS	MS	MS	MS	MS	Gempa
7	E7	MS	MS	MS	MS	MS	Gempa
8	E8	MS	TMS	MS	MS	MS	Gempa
9	E9	TMS	MS	MS	MS	MS	Gempa
10	E10	TMS	TMS	TMS	TMS	MS	Tidak

Keterangan:

- K1 = Geodesi
- K2 = Geomagnet
- K3 = Geologi
- K4 = Seismologi
- K5 = Seismic Velocity
- MS = Memenuhi Syarat
- TMS = Tidak Memenuhi Syarat

Dengan menggunakan Metode Rough Set Maka Pengetahuan Baru Yang Di Dapat (General Rules-nya) adalah?

BAB 8

BAB VIII : FUNGSI MAYOR CLUSTERING

8.1 Pendahuluan

Analisis *cluster* merupakan salah satu teknik *data mining* yang bertujuan untuk mengidentifikasi sekelompok obyek yang mempunyai kemiripan karakteristik tertentu yang dapat dipisahkan dengan kelompok obyek lainnya, sehingga obyek yang berada dalam kelompok yang sama relatif lebih homogeni daripada obyek yang berada pada kelompok yang berbeda. Jumlah kelompok yang dapat diidentifikasi tergantung pada banyak dan variasi data obyek. Tujuan dari pengelompokan sekumpulan data obyek kedalam beberapa kelompok yang mempunyai karakteristik tertentu dan dapat dibedakan satu sama lainnya adalah untuk analis dan interprestasi lebih lanjut sesuai dengan tujuan penelitian yang dilakukan.

Model yang diambil diasumsikan bahwa data yang dapat digunakan adalah data yang berupa interval, frekuensi dan biner. Set data obyek harus mempunyai peubah dengan tipe yang sejenis tidak campur antara tipe yang satu dengan lainnya.Analisis *cluster* dapat diterapkan pada bidang apa saja.

Namun pemakaian teknik ini lebih familiar pada bidang pemasaran karena memang salah satu kegiatan yang dilakukan dalam pemasaran adalah pengelompokan, yang disebut segementasi pasar. Tujuan analisis *cluster* di dalam pemasaran adalah sebagai berikut :

1. Membuat segmen pasar (*segmenting the market*)

   Pelanggan atau pembeli sering diklasterkan berdasarkan manfaat atau keuntungan yang diperoleh dari pembelian barang. Setiap *cluster* akan terdiri dari pelanggan/pembeli yang relatif homogen, dinyatakan dalam manfaat yang dicari.
2. Memahami perilaku pembeli

   Analisis *cluster* digunakan untuk mengenali/mengidentifikasi kelompok pembeli yang homogen/relatif homogen. Kemudian perilaku untuk dalam setiap kelompok perlu dikaji secara terpisah. Responden (pembeli) dikelompokkan didasarkan pada *self-*

reported importance yang terkait pada setiap faktor pilihan yang digunakan untuk memilih toko atau *mall* dimana para pembeli membeli barang yang dibutuhkan.

3. Mengenali peluang produk baru

   Dengan mengklasterkan merk dan produk, *competitive set* di dalam pasar bisa ditentukan. Merek di dalam klaster yang sama bersaing sengit satu sama lain, daripada merek dari klaster lain.

4. Mereduksi data

   Analisis*cluster* digunakan sebagai suatu alat mereduksi data secara umum,untuk mengembangkan klaster atau sub-group dari data yang mudah dikelola dari kumpulan data asli, secara individual.Berikut ini adalah prosedur analisis *cluster*, yaitu:



Gambar 8.1: Prosedur Analis *Cluster*

Secara umum proses dimulai dengan merumuskan masalah pengklasteran dengan mendefinisikan variabel-variabel yang dipergunakan untuk dasar pengklasteran/pembentukan klaster. Kemudian pengambilan p pengukuran peubah pada n obyek pengamatan. Data tersebut dijadikan matriks data mentah m x p. Matrik tersebut ditransformasikan ke dalam bentuk matriks similaritas (kemiripan) berupa n x

n yang dihitung berdasarkan pasangan-pasangan obyek p peubah. Konsep dasar pengukuran analis *cluster* adalah konsep pengukuran jarak (*distance*) dan kesamaan (*similarity*). *Distance* adalah ukuran tentang jarak pisah antar obyek sedangkan *similarity* adalah ukuran kedekatan. Konsep ini penting karena pengelompokan pada analis *cluster*didasarkan pada kedekatan. Pengukuan jarak (*distance type measure*) digunakan untuk data-data yang bersifat matriks, sedangkan pengukuran kesesuaian (*matching type measure*) digunakan untuk data-data yang bersifat kualitatif.

Adapun beberapa teknik pengukuran jarak sebagai berikut :

1. *Euclidean Distance*

   Merupakan ukuran jarak antara dua item X dan Y.

   $D(X, Y) = \sqrt{\sum (Xi - Yi)\ 2}$

2. *Squared Euclidean Distance*

   Merupakan ukuran jarak antara dua item X dan Y.

   $D(X, Y) = \sum (Xi - Yi)\ 2$

3. *Person Correlation*

   Korelasi antara vector nilai :

   $$S(X,Y) = \frac{\sum Z_{xi}Z_{yi}}{(N - 1)}$$

   Dimana Z_{xi} adalah nilai x yang telah distandarkan untuk item ke-I dan N adalah jumlah itemnya.

4. *Chebychev*

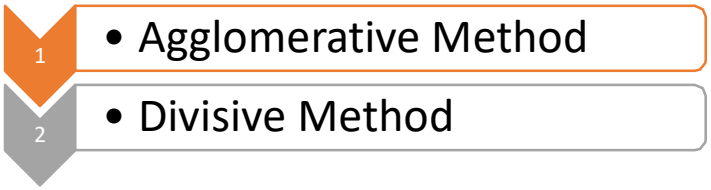
   $D(X,Y) = \max_i \mid X_i - Y_i \mid$

5. *Block*

   $D(X,Y) = \sum \mid X_i - Y_i \mid$

6. *Minkowski*
7. *Chi-Square*
8. *Phi-Square*
9. *Hamming*

Teknik hirarki (*hierarchical methods*) adalah teknik *clustering* membentuk kontruksi hirarki atau berdasarkan tingkatan tertentu seperti struktur pohon (struktur pertandingan). Dengan demikian proses pengelompokkannya dilakukan secara bertingkat atau bertahap. Hasil dari pengelompokan ini dapat disajikan dalam bentuk dendogram. Metode-metode yang digunakan dalam teknik hirarki :



Analisis *cluster* merupakan salah satu teknik *data mining* yang bertujuan untuk mengidentifikasi sekelompok obyek yang mempunyai kemiripan karakteristik tertentu yang dapat dipisahkan dengan kelompok obyek lainnya, sehingga obyek yang berada dalam kelompok yang sama relatif lebih homogeni daripada obyek yang berada pada kelompok yang berbeda. Jumlah kelompok yang dapat diidentifikasi tergantung pada banyak dan variasi data obyek. Tujuan dari pengelompokan sekumpulan data obyek kedalam beberapa kelompok yang mempunyai karakteristik tertentu dan dapat dibedakan satu sama lainnya adalah untuk analis dan interprestasi lebih lanjut sesuai dengan tujuan penelitian yang dilakukan.

Model yang diambil diasumsikan bahwa data yang dapat digunakan adalah data yang berupa interval, frekuensi dan biner. Set data obyek harus mempunyai peubah dengan tipe yang sejenis tidak campur antara tipe yang satu dengan lainnya.Analisis *cluster* dapat diterapkan pada bidang apa saja.

Namun pemakaian teknik ini lebih familiar pada bidang pemasaran karena memang salah satu kegiatan yang dilakukan dalam pemasaran adalah pengelompokan, yang disebut segementasi pasar. Tujuan analisis *cluster* di dalam pemasaran adalah sebagai berikut :

5. Membuat segmen pasar (*segmenting the market*)

   Pelanggan atau pembeli sering diklasterkan berdasarkan manfaat atau keuntungan yang diperoleh dari pembelian barang. Setiap *cluster* akan terdiri dari pelanggan/pembeli yang relatif homogen, dinyatakan dalam manfaat yang dicari.

6. Memahami perilaku pembeli

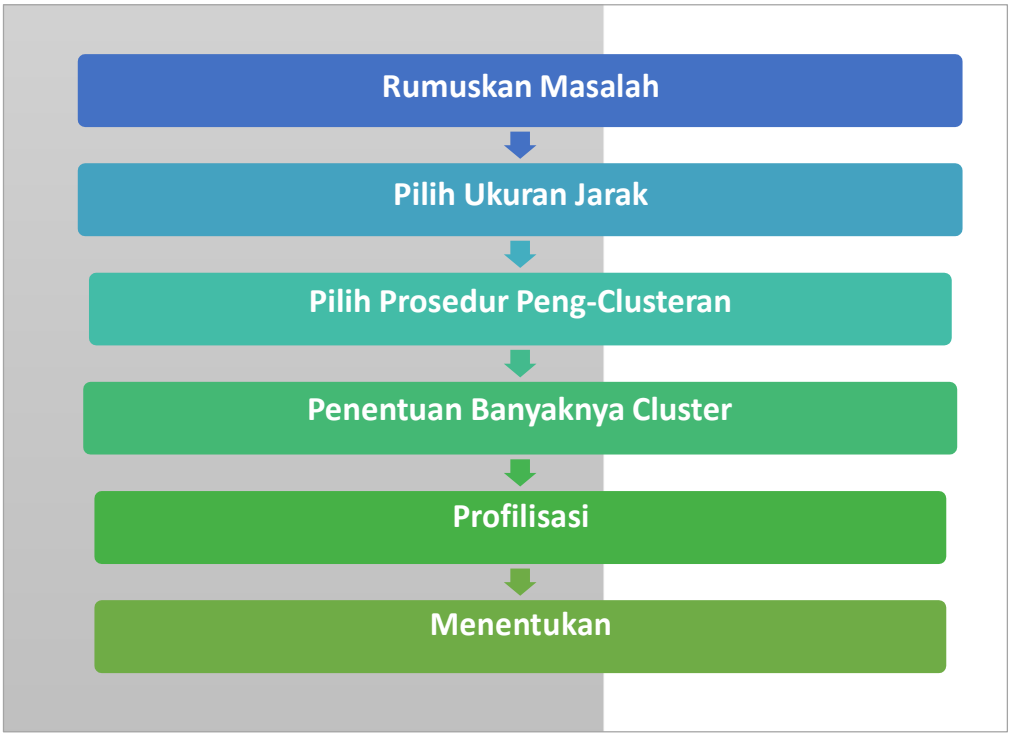
Analisis *cluster* digunakan untuk mengenali/mengidentifikasi kelompok pembeli yang homogen/relatif homogen. Kemudian perilaku untuk dalam setiap kelompok perlu dikaji secara terpisah. Responden (pembeli) dikelompokkan didasarkan pada *self-reported importance* yang terkait pada setiap faktor pilihan yang digunakan untuk memilih toko atau *mall* dimana para pembeli membeli barang yang dibutuhkan.

7. Mengenali peluang produk baru

   Dengan mengklasterkan merk dan produk, *competitive set* di dalam pasar bisa ditentukan. Merek di dalam klaster yang sama bersaing sengit satu sama lain, daripada merek dari klaster lain.

8. Mereduksi data

   Analisis*cluster* digunakan sebagai suatu alat mereduksi data secara umum,untuk mengembangkan klaster atau sub-group dari data yang mudah dikelola dari kumpulan data asli, secara individual.Berikut ini adalah prosedur analisis *cluster*, yaitu :



Gambar 5.1: Prosedur Analis *Cluster*

Secara umum proses dimulai dengan merumuskan masalah pengklasteran dengan mendefinisikan variabel-variabel yang dipergunakan untuk dasar pengklasteran/pembentukan klaster. Kemudian pengambilan p pengukuran peubah pada n obyek pengamatan. Data tersebut dijadikan matriks data mentah m x p. Matrik tersebut ditransformasikan ke dalam bentuk matriks similaritas (kemiripan) berupa n x n yang dihitung berdasarkan pasangan-pasangan obyek p peubah. Konsep dasar pengukuran analis *cluster* adalah konsep pengukuran jarak (*distance*) dan kesamaan (*similarity*). *Distance* adalah ukuran tentang jarak pisah antar obyek sedangkan *similarity* adalah ukuran kedekatan. Konsep ini penting karena pengelompokan pada analis *cluster*didasarkan pada kedekatan. Pengukuan jarak (*distance type measure*) digunakan untuk data-data yang bersifat matriks, sedangkan pengukuran kesesuaian (*matching type measure*) digunakan untuk data-data yang bersifat kualitatif.

Adapun beberapa teknik pengukuran jarak sebagai berikut :

10. *Euclidean Distance*

    Merupakan ukuran jarak antara dua item X dan Y.

    $D(X, Y) = \sqrt{\sum} (Xi - Yi)\ 2$

11. *Squared Euclidean Distance*

    Merupakan ukuran jarak antara dua item X dan Y.

    $D(X, Y) = \sum (Xi - Yi)\ 2$

12. *Person Correlation*

    Korelasi antara vector nilai :

    $$S(X,Y) = \frac{\sum Z_{xi}Z_{yi}}{(N-1)}$$

    Dimana Z_{xi} adalah nilai x yang telah distandarkan untuk item ke-I dan N adalah jumlah itemnya.

13. *Chebychev*

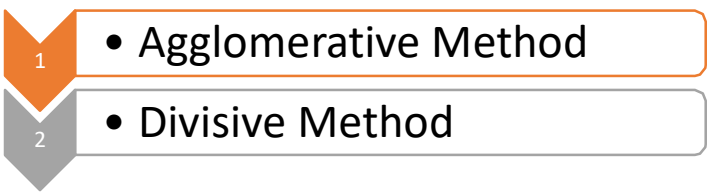
    $D(X,Y) = \max_i \mid X_i - Y_i \mid$

14. *Block*

    $D(X,Y) = \sum \mid X_i - Y_i \mid$

15. *Minkowski*
16. *Chi-Square*
17. *Phi-Square*
18. *Hamming*

Teknik hirarki (*hierarchical methods*) adalah teknik *clustering* membentuk kontruksi hirarki atau berdasarkan tingkatan tertentu seperti struktur pohon (struktur pertandingan). Dengan demikian proses pengelompokkannya dilakukan secara bertingkat atau bertahap. Hasil dari pengelompokan ini dapat disajikan dalam bentuk dendogram. Metode-metode yang digunakan dalam teknik hirarki :



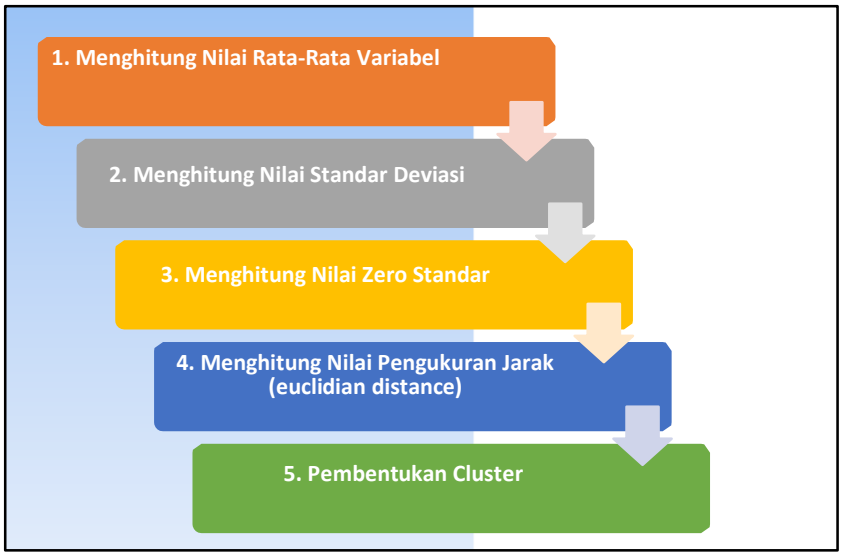
8.2 Agglomerative Method (Analisa dan Pemecahan Masalah)

Metode ini dimulai dengan kenyataan bahwa setiap obyek membentuk clusternya masing-masing. Kemudian dua obyek dengan jarak terdekat bergabung. Selanjutnya obyek ketiga akan bergabung dengan *cluster* yang ada atau bersama obyek lain dan membentuk *cluster* baru. Hal ini tetap memperhitungkan jarak kedekatan antar obyek. Proses akan berlanjut hingga akhirnya terbentuk satu *cluster* yang terdiri dari keseluruhan obyek.



Gambar 8.2 : Jenis-jenis Hierarchy Agglomerathip Cluster

Pada pembahasan kali ini penulis mencoba untuk menjelaskan tentang penyelesaian Teknik Single Linkage di dalam proses Clustering dengan menggunakan pendekata Euclidian Distance.Berikut ini adalah proses penyelesaian Clustering dengan Teknik Single Linkage dengan pendekatan Euclidian Distance:



Gambar 8.3 : Algoritma Penyelesaian Clustering Dengan Teknik Single Linkage Dengan Euclidian Distance.

Berdasarkan gambar dari algoritma penyelesaian clustering di atas maka kita harus mengetahui rumus (formula) penyelesaiannya. Berikut ini adalah rumus untuk menyelesaiakan metode clustering dengan Pendekatan Teknik Single Linkage.

1. **Menghitung Nilai Rata-rata Variabel (**V1)

Rumus : $\bar{X} = \frac{\sum_{i=1}^{n} Xi}{n}$

Keterangan:

$\bar{X}$ = Nilai Rata-rata Variabel

Xi = Nilai Variabel

N = Jumlah atau kuantitas dari objek

2. **Menghitung Nilai Standar Deviasi (Std(x))**

Rumus : $Std(X) = \sqrt{\frac{\sum_{i=1}^{n} (Xi - \bar{X})^2}{n-1}}$

Keterangan:

Std (X) = Nilai Rata-rata Variabel

Xi = Nilai Variabel

N = Jumlah atau kuantitas dari objek

3. **Menghitung Nilai Zero Standar (Zi)**

Rumus: $Zi = \frac{Xi - \bar{X}}{Std(X)}$

Keterangan:

Zi= Nilai Zero Standar

$\bar{X}$= Nilai Rata-rata Variabel

Std (X) = Nilai Rata-rata Variabel

Xi = Nilai Variabel

BAB 9 BAB IX : TEKNIK SINGLE LINKAGE

9.1 Contoh Soal Dan Penyelesaiannya

Bab ini merupakan kelanjutan dari bab sebelumnya. Pada bab ini akan diberikan contoh penyelesaian sebuah kasus dengan menggunakan teknik single linkage. Berikut ini adalah contoh soalnya:

Algoritma Nearest Neighbor merupakan salah satu metode klasifikasi yang digunakan untuk pemecahan masalah pada bidang Data Mining. Sama halnya dengan beberapa metode lainnya yang ada pada metode klasifikasi, algoritma ini memiliki ciri yaitu dengan pendekatan untuk mencari kasus dengan menghitung kedekatan kasus yang baru dengan kasus yang lama. Adapun teknik yang digunakan yaitu berdasarkan bobot dari sejumlah objek kasus yang ada.

Pada sebuah Sekolah Tinggi Manajemen dan Informatika Komputer di dalam peningkatan kualitas Akademik dan memaksimalkan kualitas Lulusan maka STMIK tersebut ingin membuat sebuah langkah dengan cara pembentukan kelas unggulan. Dalam pembentukan kelas unggulan tersebut mengalami kendala selama ini, sehingga di butuhkan sebuah cara untuk mewujudkan hal tersebut maka digunakan Metode Clustering dengan Pendekatan Euclidian Distance menggunakn Teknik Single Linkage. Adapun sampel data yang akan di proses terlihat pada tabel di bawah ini:

Tabel 9.1 : Sampel Data Untuk Pengujian

No	NAMA	V1	V2	V3
1	Afdilla Ramadhani Arhas	0.25	73.6	79.3
2	Atikah Khairani Rangkuti	3.75	98.9	87
3	Ayumi Sari Putri	3.85	99	85
4	Daniel Sastra Novindo Valente	0.56	60.3	65
5	Dewi Purnama Nasution	3.15	95.7	84.3
6	Dian Srihartati Simanjorang	0.35	52.6	56
7	Fauzi Pratama	1.72	68.3	73
8	Hendi Thien Sulkanando	0.75	79.4	80
9	Irvan Yoanda Ginting	3.18	84	83
10	Juli Nanda Sari	3.86	100	86.1
11	Misna Wati Br Meliala	1.9	78.6	74.3
12	Muhammad Iksan Rasta T.	0.94	80	78.3
13	Nika Depri Sitepu	3.16	83	86
14	Qaswin Ilham	1.95	87.5	70
15	Rinaldy Fadly	3.08	83.8	82.5
16	Rita Yanti Purba	3.76	97.8	85.6
17	Rizky Ramadhan	3.81	97.5	81.9
18	Sabariah Solin	3.21	85	84.6
19	Susi Sitangang	0.1	60	60
20	Yolanda Sherpita Br. Ginting	0.3	63.9	76

Keterangan:

V1= Indeks Prestasi Kumulatif (IPK)

V2 = Nilai Absensi

V3 = Nilai Etika

Untuk mempermudah kita dalam memproses data yang akan di adopsi dalam pengujian menggunakan metode maka langkah awal (preprosessing) yang harus

dilakukan adalah membuat tabel asumsi terhadap objek. Berikut ini adalah tabel asumsi objek dari sampel data yang digunakan.

Tabel 9.2: Asumsi Objek

No. Urut Mahasiswa	Kode Huruf	Nama Mahasiswa
1	A	Afdilla Ramadhani Arhas
2	B	Atikah Khairani Rangkuti
3	C	Ayumi Sari Putri
4	D	Daniel Sastra Novindo Valente
5	E	Dewi Purnama Nasution
6	F	Dian Srihartati Simanjorang
7	G	Fauzi Pratama
8	H	Hendi Thien Sulkanando
9	I	Irvan Yoanda Ginting
10	J	Juli Nanda Sari
11	K	Misna Wati Br Meliala
12	L	Muhammad Iksan Rasta T.
13	M	Nika Depri Sitepu
14	N	Qaswin Ilham
15	O	Rinaldy Fadly
16	P	Rita Yanti Purba
17	Q	Rizky Ramadhan
18	R	Sabariah Solin
19	S	Susi Sitangang
20	T	Yolanda Sherpita Br. Ginting

Berikut ini adalah langkah-langkah penyelesaiannya menggunakan Metode Clustering dengan Pendekatan Euclidian Distance menggunakn Teknik Single Linkage yaitu sebagai berikut:

1. **Menghitung Nilai Rata-rata setiap variabel**

Melakukan standarisasi data untuk setiap nilai variabel pada masing-masing mahasiswa dengan N adalah jumlah mahasiswa yaitu 20. Standarisasi data dilakukan dengan menggunakan persamaan (3). Untuk mahasiswa A dengan Variabel 1 yang nilainya adalah 20 (selanjutnya disebut data V11), maka standarisasi dengan menghitung nilai Rata-rata adalah sebaga berikut:

Rumus : $\bar{X} = \frac{\sum_{i=1}^{n} Xi}{n}$

Maka,

$$\bar{V1} = \frac{V1_1 + V1_2 + V1_3 + V1_4 + V1_5 + V1_6 + V1_7 + V1_8 + V1_9 + V1_{10} + V1_{11} + V1_{12} + V1_{13} + V1_{14} + V1_{15} + V1_{16} + V1_{17} + V1_{18} + V1_{19} + V1_{20}}{N}$$

$$= \frac{0.25 + 3.75 + 3.85 + 0.56 + 3.15 + 0.35 + 1.72 + 0.75 + 3.18 + 3.86 + 1.9 + 0.94 + 3.16 + 1.95 + 3.08 + 3.76 + 3.81 + 3.21 + 0.1 + 0.3}{20}$$

= **2.182**

Untuk Nilai$\bar{V2}$ dan $\bar{V3}$ Silahkan dihitung dengan konsep perhitungan **V1**

2. **Menghitung Nilai Standar Deviasi Variabel**

Rumus : $Std(X) = \sqrt{\frac{\sum_{i=1}^{n} (Xi - \bar{X})^2}{n-1}}$

Maka,

$$Std(V1) = \sqrt{(V1_1 - \bar{V1})^2 + (V1_2 - \bar{V1})^2 + (V1_3 - \bar{V1})^2 + (V1_4 - \bar{V1})^2 + (V1_5 - V1)^2 + (V1_6 - \bar{V1})^2 + (V1_7 - \bar{V1})^2 + (V1_8 - \bar{V1})^2 + (V1_9 - V1)^2 + (V1_{10} - V1)^2 + (V1_{11} - \bar{V1})^2 + (V1_{12} - \bar{V1})^2 + (V1_{13} - \bar{V1})^2 + (V1_{14} - V1)^2 + (V1_{15} - V1)^2 + (V1_{16} - V1)^2 + (V1_{17} - V1)^2 + (V1_{18} - V1)^2 + (V1_{19} - V1)^2 + (V1_{20} - V1)^2}$$

$$\text{N-1}$$

$$\text{Std(V1)} = \sqrt{\frac{\begin{aligned}&(0.25 - 2.182)^2 + (3.75 - 2.182)^2 + (3.85 - 2.182)^2 + (0.56 - 2.182)^2 + \\ &(3.15 - 2.182)^2 + (0.35 - 2.182)^2 + (1.72 - 2.182)^2 + (0.75 - 2.182)^2 + \\ &(3.18 - 2.182)^2 + (3.86 - 2.182)^2 + (1.9 - 2.182)^2 + (0.94 - 2.182)^2 + \\ &(3.16 - 2.182)^2 + (1.95 - 2.182)^2 + (3.08 - 2.182)^2 + (3.76 - 2.182)^2 + \\ &(3.81 - 2.182)^2 + (3.21 - 2.182)^2 + (0.1 - 2.182)^2 + (0.3 - 2.182)^2\end{aligned}}{20 - 1}}$$

$$= \sqrt{\frac{\begin{aligned}&(-1.932)^2 + (1.569)^2 + (1.669)^2 + (-1.622)^2 + (0.969)^2 + (-1.832)^2 + (-0.462)^2 \\ &+ (-1.432)^2 + (0.999)^2 + (1.679)^2 + (-0.282)^2 + (-1.242)^2 + (0.979)^2 + (-0.232)^2 \\ &+ (0.899)^2 + (1.579)^2 + (1.629)^2 + (1.029)^2 + (-2.082)^2 + (-1.882)^2\end{aligned}}{20 - 1}}$$

$$= \sqrt{\frac{\begin{aligned}&3.731 + 2.460 + 2.784 + 2.629 + 0.938 + 3.354 + 0.213 + 2.049 + 0.997 + \\ &2.817 + 0.079 + 1.541 + 0.957 + 0.054 + 0.807 + 2.492 + 2.652 + 1.058 + \\ &4.333 + 3.540\end{aligned}}{19}}$$

=**1.442**

Untuk Nilai **Std(V2)** dan **Std(V3)** Silahkan dihitung dengan konsep perhitungan **Std(V3).**

3. **Menghitung Nilai Zero Standard (Zi) dari setiap Objek**

Rumus $Z_i = \dfrac{X_i - \bar{X}}{Std(X)}$

Maka,

$Z(V1_1) = \dfrac{V1_1 - \bar{V}1}{Std(V1)}$ = (**0.25 – 2.182) / 1.442 = -1.340**

Untuk Nilai

- $Z(V1_2)$ sampai dengan $Z(V1_{20})$
- $Z(V2_1)$ sampai dengan $Z(V2_{20})$
- $Z(V3_1)$ sampai dengan $Z(V3_{20})$

Silahkan dihitung dengan konsep perhitungan $Z(V1_1)$. Setelah semua data distandarkan dengan cara yang sama, maka didapatkan tabel baru yang berisi data yang telah standar, yaitu:

Tabel 9.3: Nilai Zero Standar Keseluruhan

OBJ/VAR	V1	V2	V3
A	-1.340	-0.538	0.156
B	1.088	1.196	1.008
C	1.157	1.203	0.787
D	-1.125	-1.449	-1.428
E	0.672	0.977	0.709
F	-1.270	-1.977	-2.424
G	-0.320	-0.901	-0.542
H	-0.993	-0.140	0.233
I	0.693	0.175	0.565
J	1.164	1.272	0.908
K	-0.195	-0.195	-0.398
L	-0.861	-0.099	0.045
M	0.679	0.107	0.897
N	-0.161	0.415	-0.874
O	0.623	0.161	0.510
P	1.095	1.121	0.853
Q	1.130	1.100	0.443
R	0.713	0.244	0.742
S	-1.444	-1.470	-1.981
T	-1.305	-1.202	-0.210

4. Menghitung Nilai Pengukuran Jaraknya

Membuat matriks jarak untuk objek mahasiswa dengan menggunakan metode Euclidean Single Linkage. Caranya adalah dengan menggunakan persamaan (4). Misal, untuk mencari jarak kedekatan antar mahasiswa A dan B.

$$dAB = \sqrt{(V1_1 - V1_2)^2 + (V2_1 - V2_2)^2 + (V3_1 - V3_2)^2}$$

$$dAB = \sqrt{((-1.340) - 1.088)^2 + ((-0.538) - 1.196)^2 + (0.156 - 1.008)^2}$$

$$dAB = \sqrt{(5.894)^2 + (3.007)^2 + (0.727)^2}$$

dAB = **3.103**

Langkah yang sama dilakukan pada seluruh data mahasiswa, sehingga pada akhir perhitungan didapatkan matriks seperti berikut:

Tabel 9.4 : Matriks Jarak

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
A	0	3.103	3.109	1.839	2.578	2.955	1.288	0.533	2.192	3.180	1.317	0.659	2.245	1.833	2.114	3.027	2.977	2.274	2.333	0.759
B	3.103	0	0.232	4.222	0.557	5.235	2.964	2.592	1.181	0.146	2.358	2.531	1.169	2.390	1.239	0.172	0.574	1.058	4.738	3.601
C	3.109	0.232	0	4.141	0.541	5.130	2.894	2.595	1.150	0.140	2.278	2.514	1.202	2.262	1.203	0.123	0.359	1.058	4.644	3.584
D	1.839	4.222	4.141	0	3.699	1.137	1.316	2.119	3.148	4.254	1.870	2.015	3.328	2.171	3.066	4.091	3.884	3.018	0.639	1.256
E	2.578	0.557	0.541	3.699	0	4.724	2.465	2.061	0.815	0.608	1.831	1.987	0.891	1.875	0.841	0.470	0.544	0.735	4.207	3.083
F	2.955	5.235	5.130	1.137	4.724	0	2.367	3.242	4.174	5.252	2.904	3.129	4.378	3.059	4.095	5.092	4.843	4.346	0.695	2.346
G	1.288	2.964	2.894	1.316	2.465	2.367	0	1.277	1.846	3.004	0.731	1.131	2.021	1.367	1.768	2.835	2.660	2.007	1.912	1.082
H	0.533	2.592	2.595	2.119	2.061	3.242	1.277	0	1.747	2.665	1.019	0.233	1.816	1.492	1.667	2.517	2.468	1.822	2.622	1.192
I	2.192	1.181	1.150	3.148	0.815	4.174	1.846	1.747	0	1.242	1.361	1.661	0.339	1.690	0.090	1.067	1.030	0.191	3.709	2.547
J	3.180	0.146	0.140	4.254	0.608	5.252	3.004	2.665	1.242	0	2.389	2.594	1.262	2.380	1.298	0.175	0.497	1.135	4.761	3.670
K	1.317	2.358	2.278	1.870	1.831	2.904	0.731	1.019	1.361	2.389	0	0.805	1.591	0.775	1.273	2.227	2.035	1.523	2.385	1.511
L	0.659	2.531	2.514	2.015	1.987	3.129	1.131	0.233	1.661	2.594	0.805	0	1.772	1.265	1.577	2.443	2.358	1.756	2.514	1.216
M	2.245	1.169	1.202	3.328	0.891	4.378	2.021	1.816	0.339	1.262	1.591	1.772	0	1.984	0.395	1.097	1.182	0.210	3.908	2.622
N	1.833	2.390	2.262	2.171	1.875	3.059	1.367	1.492	1.690	2.380	0.775	1.265	1.984	0	1.610	2.249	1.967	1.845	2.535	2.090
O	2.114	1.239	1.203	3.066	0.841	4.095	1.768	1.667	0.090	1.298	1.273	1.577	0.395	1.610	0	1.123	1.069	0.263	3.625	2.469
P	3.027	0.172	0.123	4.091	0.470	5.092	2.835	2.517	1.067	0.175	2.227	2.443	1.097	2.249	1.123	0	0.412	0.963	4.603	3.505
Q	2.977	0.574	0.359	3.884	0.544	4.843	2.660	2.468	1.030	0.497	2.035	2.358	1.182	1.967	1.069	0.412	0	0.998	4.371	3.414
R	2.274	1.058	1.058	3.018	0.735	4.346	2.007	1.822	0.191	1.135	1.523	1.756	0.210	1.845	0.263	0.963	0.998	0	3.874	2.659
S	2.333	4.738	4.644	0.639	4.207	0.695	1.912	2.622	3.709	4.761	2.385	2.514	3.908	2.535	3.625	4.603	4.371	3.874	0	1.797
T	0.759	3.601	3.584	1.256	3.083	2.346	1.082	1.192	2.547	3.670	1.511	1.216	2.622	2.090	2.469	3.505	3.414	2.659	1.797	0

5. **Melakukan pengelompokan menggunakan *Euclidean Single Linkage***

a. Mencari nilai terkecil dari matriks jarak

I dan O mempunyai nilai terkecil, yaitu **0.090** maka obyek I dan O bergabung menjadi satu *cluster*

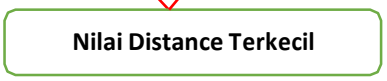
b. Menghitung jarak antar *cluster* IO dengan obyek lainnya.

1) d(IO)A = min {dIA,dOA} = dOA = 2.114
2) d(IO)B = min {dIB,dOB} = dIA = 1.181
3) d(IO)C = min {dIC,dOC} = dIC = 1.150
4) d(IO)D = min {dID,dOD} = dOD = 3.066
5) d(IO)E = min {dIE,dOE} = dIE = 0.815
6) d(IO)F = min {dIF,dOF} = dOF = 4.095
7) d(IO)G = min {dIG,dOG} = dOG = 1.768
8) d(IO)H = min {dIH,dOH} = dOH = 1.667
9) d(IO)J = min {dIJ,dOJ} = dIJ =1.242
10) d(IO)K = min {dIK,dOK} = dOK = 1.273
11) d(IO)L = min {dIL,dOL} = dOL = 1.577
12) d(IO)M = min {dIM,dOM} = dIM = 0.339
13) d(IO)N = min {dIN,dON} = dON = 1.610
14) d(IO)P = min {dIP,dOP} = dIP = 1.067
15) d(IO)Q = min {dIQ,dOQ} = dOQ = 1.030
16) d(IO)R = min {dIR,dOR} = dIR = 0.191
17) d(IO)S = min {dIS,dOS} = dOS = 3.625
18) d(IO)T = min {dIT,dOT} = dOT = 2.469

Dengan demikian terbentuk matriks jarak yang baru yang menjadi sebuah *cluster* antara I dan O pada tabel *cluster* 1 di bawah ini:

Tabel 9.5 : Matriks Jarak Turunan Pertama 1 (*Cluster 1*)

	IO	A	B	C	D	E	F	G	H	J	K	L	M	N	P	Q	R	S	T
IO	0	2.114	1.181	1.150	3.066	0.815	4.095	1.768	1.667	1.242	1.273	1.577	0.339	1.610	1.067	1.030	0.191	3.625	2.469
A	2.114	0	3.103	3.109	1.839	2.578	2.955	1.288	0.533	3.180	1.317	0.659	2.245	1.833	3.027	2.977	2.274	2.333	0.759
B	1.181	3.103	0	0.232	4.222	0.557	5.235	2.964	2.592	0.146	2.358	2.531	1.169	2.390	0.172	0.574	1.058	4.738	3.601
C	1.150	3.109	0.232	0	4.141	0.541	5.130	2.894	2.595	0.140	2.278	2.514	1.202	2.262	0.123	0.359	1.058	4.644	3.584
D	3.066	1.839	4.222	4.141	0	3.699	1.137	1.316	2.119	4.254	1.870	2.015	3.328	2.171	4.091	3.884	3.018	0.639	1.256
E	0.815	2.578	0.557	0.541	3.699	0	4.724	2.465	2.061	0.608	1.831	1.987	0.891	1.875	0.470	0.544	0.735	4.207	3.083
F	4.095	2.955	5.235	5.130	1.137	4.724	0	2.367	3.242	5.252	2.904	3.129	4.378	3.059	5.092	4.843	4.346	0.695	2.346
G	1.768	1.288	2.964	2.894	1.316	2.465	2.367	0	1.277	3.004	0.731	1.131	2.021	1.367	2.835	2.660	2.007	1.912	1.082
H	1.667	0.533	2.592	2.595	2.119	2.061	3.242	1.277	0	2.665	1.019	0.233	1.816	1.492	2.517	2.468	1.822	2.622	1.192
J	1.242	3.180	0.146	0.140	4.254	0.608	5.252	3.004	2.665	0	2.389	2.594	1.262	2.380	0.175	0.497	1.135	4.761	3.670
K	1.273	1.317	2.358	2.278	1.870	1.831	2.904	0.731	1.019	2.389	0	0.805	1.591	0.775	2.227	2.035	1.523	2.385	1.511
L	1.577	0.659	2.531	2.514	2.015	1.987	3.129	1.131	0.233	2.594	0.805	0	1.772	1.265	2.443	2.358	1.756	2.514	1.216
M	0.339	2.245	1.169	1.202	3.328	0.891	4.378	2.021	1.816	1.262	1.591	1.772	0	1.984	1.097	1.182	0.210	3.908	2.622
N	1.610	1.833	2.390	2.262	2.171	1.875	3.059	1.367	1.492	2.380	0.775	1.265	1.984	0	2.249	1.967	1.845	2.535	2.090
P	1.067	3.027	0.172	0.123	4.091	0.470	5.092	2.835	2.517	0.175	2.227	2.443	1.097	2.249	0	0.412	0.963	4.603	3.505
Q	1.030	2.977	0.574	0.359	3.884	0.544	4.843	2.660	2.468	0.497	2.035	2.358	1.182	1.967	0.412	0	0.998	4.371	3.414
R	0.191	2.274	1.058	1.058	3.018	0.735	4.346	2.007	1.822	1.135	1.523	1.756	0.210	1.845	0.963	0.998	0	3.874	2.659
S	3.625	2.333	4.738	4.644	0.639	4.207	0.695	1.912	2.622	4.761	2.385	2.514	3.908	2.535	4.603	4.371	3.874	0	1.797
T	2.469	0.759	3.601	3.584	1.256	3.083	2.346	1.082	1.192	3.670	1.511	1.216	2.622	2.090	3.505	3.414	2.659	1.797	0



c. Mencari nilai terkecil dari matriks jarak, obyek C dan P mempunyai nilai terkecil yaitu 0.123 maka obyek C dan P bergabung menjadi satu *cluster*.

1) d(CP)IO = min {dCIO,dPIO} = dPIO = 1.067
2) d(CP)A = min {dCA,dPA} = dPA = 3.027
3) d(CP)B = min {dCB,dPB} = dPB = 0.172
4) d(CP)D = min {dCD,dPD} = dPD = 4.091
5) d(CP)E = min {dCE,dPE} = dPE = 0.470
6) d(CP)F = min {dCF,dPF} = dPF = 5.092
7) d(CP)G = min {dCG,dPG} = dPG = 2.835
8) d(CP)H = min {dCH,dPH} = dPH = 2.517
9) d(CP)J = min {dCJ,dPJ} = dCJ = 0.140
10) d(CP)K = min {dCK,dPK} = dPK = 2.227
11) d(CP)L = min {dCL,dPL} = dPL = 2.443
12) d(CP)M = min {dCM,dPM} = dPM = 1.097
13) d(CP)N = min {dCN,dPN} = dPN = 2.249
14) d(CP)Q = min {dCQ,dPQ} = dCQ = 0.359
15) d(CP)R = min {dCR,dPR} = dPR = 0.963

16) d(CP)S = min {dCS,dPS} = dPS = 4.603

17) d(CP)T = min {dCT,dPT} = dPT = 3.505

Dengan demikian terbentuk matriks jarak yang baru yang menjadi sebuah *cluster* antara C dan P pada tabel *cluster* 2 di bawah ini:

Tabel 9.6 : Cluster 2

	IO	CP	A	B	D	E	F	G	H	J	K	L	M	N	Q	R	S	T
IO	0	1.067	2.114	1.181	3.066	0.815	4.095	1.768	1.667	1.242	1.273	1.577	0.339	1.610	1.030	0.191	3.625	2.469
CP	1.067	0	3.027	0.172	4.091	0.470	5.092	2.835	2.517	0.140	2.227	2.443	1.097	2.249	0.359	0.963	4.603	3.505
A	2.114	3.027	0	3.103	1.839	2.578	2.955	1.288	0.533	3.180	1.317	0.659	2.245	1.833	2.977	2.274	2.333	0.759
B	1.181	0.172	3.103	0	4.222	0.557	5.235	2.964	2.592	0.146	2.358	2.531	1.169	2.390	0.574	1.058	4.738	3.601
D	3.066	4.091	1.839	4.222	0	3.699	1.137	1.316	2.119	4.254	1.870	2.015	3.328	2.171	3.884	3.018	0.639	1.256
E	0.815	0.470	2.578	0.557	3.699	0	4.724	2.465	2.061	0.608	1.831	1.987	0.891	1.875	0.544	0.735	4.207	3.083
F	4.095	5.092	2.955	5.235	1.137	4.724	0	2.367	3.242	5.252	2.904	3.129	4.378	3.059	4.843	4.346	0.695	2.346
G	1.768	2.835	1.288	2.964	1.316	2.465	2.367	0	1.277	3.004	0.731	1.131	2.021	1.367	2.660	2.007	1.912	1.082
H	1.667	2.517	0.533	2.592	2.119	2.061	3.242	1.277	0	2.665	1.019	0.233	1.816	1.492	2.468	1.822	2.622	1.192
J	1.242	0.140	3.180	0.146	4.254	0.608	5.252	3.004	2.665	0	2.389	2.594	1.262	2.380	0.497	1.135	4.761	3.670
K	1.273	2.227	1.317	2.358	1.870	1.831	2.904	0.731	1.019	2.389	0	0.805	1.591	0.775	2.035	1.523	2.385	1.511
L	1.577	2.443	0.659	2.531	2.015	1.987	3.129	1.131	0.233	2.594	0.805	0	1.772	1.265	2.358	1.756	2.514	1.216
M	0.339	1.097	2.245	1.169	3.328	0.891	4.378	2.021	1.816	1.262	1.591	1.772	0	1.984	1.182	0.210	3.908	2.622
N	1.610	2.249	1.833	2.390	2.171	1.875	3.059	1.367	1.492	2.380	0.775	1.265	1.984	0	1.967	1.845	2.535	2.090
Q	1.030	0.359	2.977	0.574	3.884	0.544	4.843	2.660	2.468	0.497	2.035	2.358	1.182	1.967	0	0.998	4.371	3.414
R	0.191	0.963	2.274	1.058	3.018	0.735	4.346	2.007	1.822	1.135	1.523	1.756	0.210	1.845	0.998	0	3.874	2.659
S	3.625	4.603	2.333	4.738	0.639	4.207	0.695	1.912	2.622	4.761	2.385	2.514	3.908	2.535	4.371	3.874	0	1.797
T	2.469	3.505	0.759	3.601	1.256	3.083	2.346	1.082	1.192	3.670	1.511	1.216	2.622	2.090	3.414	2.659	1.797	0

d. Kemudian mencari nilai terkecil dari matriks jarak, obyek CP dan J mempunyai nilai terkecil yaitu **0.140** maka obyek CP dan J bergabung menjadi satu *cluster*.

1) d(CPJ)IO = min {dCPIO,dJIO} = dCPIO = 1.067
2) d(CPJ)A = min {dCPA,dJA} = dCPA = 3.027
3) d(CPJ)B = min {dCPB,dJB} = dJB = 0.146
4) d(CPJ)D = min {dCPD,dJD} = dCPD = 4.091
5) d(CPJ)E = min {dCPE,dJE} = dCPE = 0.470
6) d(CPJ)F = min {dCPF,dJF} = dCPF = 5.092
7) d(CPJ)G = min {dCPG,dJG} = dCPG = 2.835
8) d(CPJ)H = min {dCPH,dJH} = dCPH = 2.517
9) d(CPJ)K = min {dCPK,dJK} = dCPK = 2.227

10) d(CPJ)L = min {dCPL,dJL} = dCPL = 2.443

11) d(CPJ)M = min {dCPM,dJM} = dCPM = 1.097

12) d(CPJ)N = min {dCPN,dJLN = dCPN = 2.249

13) d(CPJ)Q = min {dCPQ,dJQ} = dCPQ = 0.359

14) d(CPJ)R = min {dCPR,dJR} = dCPR = 0.963

15) d(CPJ)S = min {dCPS,dJS} = dCPS = 4.603

16) d(CPJ)T = min {dCPT,dJT} = dCPT = 3.505

Dengan demikian terbentuk matriks jarak yang baru yang menjadi sebuah *cluster* antara CP dan J pada tabel *cluster* 3 di bawah ini.

Tabel 9.7 : Cluster 3

	IO	CPJ	A	B	D	E	F	G	H	K	L	M	N	Q	R	S	T
IO	0	1.067	2.114	1.181	3.066	0.815	4.095	1.768	1.667	1.273	1.577	0.339	1.610	1.030	0.191	3.625	2.469
CPJ	1.067	0	3.027	0.146	4.091	0.470	5.092	2.835	2.517	2.227	2.443	1.097	2.249	0.359	0.963	4.603	3.505
A	2.114	3.027	0	3.103	1.839	2.578	2.955	1.288	0.533	1.317	0.659	2.245	1.833	2.977	2.274	2.333	0.759
B	1.181	0.146	3.103	0	4.222	0.557	5.235	2.964	2.592	2.358	2.531	1.169	2.390	0.574	1.058	4.738	3.601
D	3.066	4.091	1.839	4.222	0	3.699	1.137	1.316	2.119	1.870	2.015	3.328	2.171	3.884	3.018	0.639	1.256
E	0.815	0.470	2.578	0.557	3.699	0	4.724	2.465	2.061	1.831	1.987	0.891	1.875	0.544	0.735	4.207	3.083
F	4.095	5.092	2.955	5.235	1.137	4.724	0	2.367	3.242	2.904	3.129	4.378	3.059	4.843	4.346	0.695	2.346
G	1.768	2.835	1.288	2.964	1.316	2.465	2.367	0	1.277	0.731	1.131	2.021	1.367	2.660	2.007	1.912	1.082
H	1.667	2.517	0.533	2.592	2.119	2.061	3.242	1.277	0	1.019	0.233	1.816	1.492	2.468	1.822	2.622	1.192
K	1.273	2.227	1.317	2.358	1.870	1.831	2.904	0.731	1.019	0	0.805	1.591	0.775	2.035	1.523	2.385	1.511
L	1.577	2.443	0.659	2.531	2.015	1.987	3.129	1.131	0.233	0.805	0	1.772	1.265	2.358	1.756	2.514	1.216
M	0.339	1.097	2.245	1.169	3.328	0.891	4.378	2.021	1.816	1.591	1.772	0	1.984	1.182	0.210	3.908	2.622
N	1.610	2.249	1.833	2.390	2.171	1.875	3.059	1.367	1.492	0.775	1.265	1.984	0	1.967	1.845	2.535	2.090
Q	1.030	0.359	2.977	0.574	3.884	0.544	4.843	2.660	2.468	2.035	2.358	1.182	1.967	0	0.998	4.371	3.414
R	0.191	0.963	2.274	1.058	3.018	0.735	4.346	2.007	1.822	1.523	1.756	0.210	1.845	0.998	0	3.874	2.659
S	3.625	4.603	2.333	4.738	0.639	4.207	0.695	1.912	2.622	2.385	2.514	3.908	2.535	4.371	3.874	0	1.797
T	2.469	3.505	0.759	3.601	1.256	3.083	2.346	1.082	1.192	1.511	1.216	2.622	2.090	3.414	2.659	1.797	0

e. Kemudian mencari nilai terkecil dari matriks jarak, obyek CPJ dan B mempunyai nilai terkecil yaitu **0.146** maka obyek CPJ dan B bergabung menjadi satu *cluster*.

1) d(CPJB)IO = min {dCPJIO,dBIO} = dCPJIO = 1.067

2) d(CPJB)A = min {dCPJA,dBA} = a = 3.027

3) d(CPJB)D = min {dCPJD,dBD} = dCPJD = 4.091

4) d(CPJB)E = min {dCPJE,dBE} = dCPJE = 0.470

5) d(CPJB)F = min {dCPJF,dBF} = dCPJF = 5.092

6) d(CPJB)G = min {dCPJG,dBG} = dCPJG = 2.835

7) d(CPJB)H = min {dCPJH,dBH} = dCPJH = 2.517

8) d(CPJB)K = min {dCPJK,dBK} = dCPJK = 2.227

9) d(CPJB)L = min {dCPJL,dBL} = dCPJL = 2.443

10) d(CPJB)M = min {dCPJM,dBM} = dCPJM= 1.097

11) d(CPJB)N = min {dCPJN,dBN} = dCPJN = 2.249

12) d(CPJB)Q = min {dCPJQ,dBQ} = dCPJQ = 0.359

13) d(CPJB)R = min {dCPJR,dBR} = dCPJR = 0.963

14) d(CPJB)S = min {dCPJS,dBS} = dCPJS = 4.603

15) d(CPJB)T = min {dCPJT,dBT} = dCPJT = 3.505

Dengan demikian terbentuk matriks jarak yang baru yang menjadi sebuah *cluster* antara CPJ dan B pada tabel *cluster* 4.

Tabel 9.8 : Cluster 4

	CPJB	IO	A	D	E	F	G	H	K	L	M	N	Q	R	S	T
CPJB	0	1.067	3.027	4.091	0.470	5.092	2.835	2.517	2.227	2.443	1.097	2.249	0.359	0.963	4.603	3.505
IO	1.067	0	2.114	3.066	0.815	4.095	1.768	1.667	1.273	1.577	0.339	1.610	1.030	0.191	3.625	2.469
A	3.027	2.114	0	1.839	2.578	2.955	1.288	0.533	1.317	0.659	2.245	1.833	2.977	2.274	2.333	0.759
D	4.091	3.066	1.839	0	3.699	1.137	1.316	2.119	1.870	2.015	3.328	2.171	3.884	3.018	0.639	1.256
E	0.470	0.815	2.578	3.699	0	4.724	2.465	2.061	1.831	1.987	0.891	1.875	0.544	0.735	4.207	3.083
F	5.092	4.095	2.955	1.137	4.724	0	2.367	3.242	2.904	3.129	4.378	3.059	4.843	4.346	0.695	2.346
G	2.835	1.768	1.288	1.316	2.465	2.367	0	1.277	0.731	1.131	2.021	1.367	2.660	2.007	1.912	1.082
H	2.517	1.667	0.533	2.119	2.061	3.242	1.277	0	1.019	0.233	1.816	1.492	2.468	1.822	2.622	1.192
K	2.227	1.273	1.317	1.870	1.831	2.904	0.731	1.019	0	0.805	1.591	0.775	2.035	1.523	2.385	1.511
L	2.443	1.577	0.659	2.015	1.987	3.129	1.131	0.233	0.805	0	1.772	1.265	2.358	1.756	2.514	1.216
M	1.097	0.339	2.245	3.328	0.891	4.378	2.021	1.816	1.591	1.772	0	1.984	1.182	0.210	3.908	2.622
N	2.249	1.610	1.833	2.171	1.875	3.059	1.367	1.492	0.775	1.265	1.984	0	1.967	1.845	2.535	2.090
Q	0.359	1.030	2.977	3.884	0.544	4.843	2.660	2.468	2.035	2.358	1.182	1.967	0	0.998	4.371	3.414
R	0.963	0.191	2.274	3.018	0.735	4.346	2.007	1.822	1.523	1.756	0.210	1.845	0.998	0	3.874	2.659
S	4.603	3.625	2.333	0.639	4.207	0.695	1.912	2.622	2.385	2.514	3.908	2.535	4.371	3.874	0	1.797
T	3.505	2.469	0.759	1.256	3.083	2.346	1.082	1.192	1.511	1.216	2.622	2.090	3.414	2.659	1.797	0

f. Kemudian mencari nilai terkecil dari matriks jarak, obyek IO dan R mempunyai nilai terkecil yaitu **0.191** maka obyek IO dan R bergabung menjadi satu *cluster*.

1) d(IOR)CPBJ = min {dIOCPBJ,dRCPBJ} = dRCPBJ = 0.963

2) d(IOR)A = min {dIOA,dRA} = dIOA = 2.114

3) d(IOR)D = min {dIOD,dRD} = dRD = 3.018

4) d(IOR)E = min {dIOE,dRE} = dRE = 0.735

5) d(IOR)F = min {dIOF,dRF} = dIOF = 4.095

6) d(IOR)G = min {dIOG,dRG} = dIOG = 1.768

7) d(IOR)H = min {dIOH,dRH} = dIOH = 1.667

8) d(IOR)K = min {dIOK,dRK} = dIOK = 1.273

9) d(IOR)L = min {dIOL,dRL} = dIOL = 1.577

10) d(IOR)M = min {dIOM,dRM} = dRM = 0.210

11) d(IOR)N = min {dION,dRN} = dION = 1.610

12) d(IOR)Q = min {dIOQ,dRQ} = dRQ = 0.998

13) d(IOR)S = min {dIOS,dRS} = dIOS = 3.625

14) d(IOR)T = min {dIOT,dRT} = dIOT = 2.469

Dengan demikian terbentuk matriks jarak yang baru yang menjadi sebuah *cluster* antara IO dan R pada tabel *cluster* 5:

Tabel 9.9: Matriks Jarak Turunan 5

	CPJB	IOR	A	D	E	F	G	H	K	L	M	N	Q	S	T
CPBJ	0	0.963	3.027	4.091	0.470	5.092	2.835	2.517	2.227	2.443	1.097	2.249	0.359	4.603	3.505
IOR	0.963	0	2.114	3.018	0.735	4.095	1.768	1.667	1.273	1.577	0.210	1.610	0.998	3.625	2.469
A	3.027	2.114	0	1.839	2.578	2.955	1.288	0.533	1.317	0.659	2.245	1.833	2.977	2.333	0.759
D	4.091	3.018	1.839	0	3.699	1.137	1.316	2.119	1.870	2.015	3.328	2.171	3.884	0.639	1.256
E	0.470	0.735	2.578	3.699	0	4.724	2.465	2.061	1.831	1.987	0.891	1.875	0.544	4.207	3.083
F	5.092	4.095	2.955	1.137	4.724	0	2.367	3.242	2.904	3.129	4.378	3.059	4.843	0.695	2.346
G	2.835	1.768	1.288	1.316	2.465	2.367	0	1.277	0.731	1.131	2.021	1.367	2.660	1.912	1.082
H	2.517	1.667	0.533	2.119	2.061	3.242	1.277	0	1.019	0.233	1.816	1.492	2.468	2.622	1.192
K	2.227	1.273	1.317	1.870	1.831	2.904	0.731	1.019	0	0.805	1.591	0.775	2.035	2.385	1.511
L	2.443	1.577	0.659	2.015	1.987	3.129	1.131	0.233	0.805	0	1.772	1.265	2.358	2.514	1.216
M	1.097	0.210	2.245	3.328	0.891	4.378	2.021	1.816	1.591	1.772	0	1.984	1.182	3.908	2.622
N	2.249	1.610	1.833	2.171	1.875	3.059	1.367	1.492	0.775	1.265	1.984	0	1.967	2.535	2.090
Q	0.359	0.998	2.977	3.884	0.544	4.843	2.660	2.468	2.035	2.358	1.182	1.967	0	4.371	3.414
S	4.603	3.625	2.333	0.639	4.207	0.695	1.912	2.622	2.385	2.514	3.908	2.535	4.371	0	1.797
T	3.505	2.469	0.759	1.256	3.083	2.346	1.082	1.192	1.511	1.216	2.622	2.090	3.414	1.797	0

g. Kemudian mencari nilai terkecil dari matriks jarak, obyek IOR dan M mempunyai nilai terkecil yaitu 0.210 maka obyek IOR dan M bergabung menjadi satu *cluster*.

1) d(IORM)CPBJ = min {dIORCPBJ,dMCPBJ} = dIORCPBJ = 0.963

2) d(IORM)A = min {dIORA,dMA} = dIORA = 2.114

3) d(IORM)D = min {dIORD,dMD} = dIORD = 3.018

4) d(IORM)E = min {dIORE,dME} = dIORE = 0.735

5) d(IORM)F = min {dIORF,dMF} = dIORF = 4.095

6) d(IORM)G = min {dIORG,dMG} = dIORG = 1.768

7) d(IORM)H = min {dIORH,dMH} = dIORH = 1.667

8) d(IORM)K = min {dIORK,dMK} = dIORK = 1.273

9) d(IORM)L = min {dIORL,dML} = dIORL = 1.577

10) d(IORM)N = min {dIORN,dMN}= dIORN = 1.610

11) d(IORM)Q = min {dIORQ,dMQ} = dIORQ = 0.998

12) d(IORM)S = min {dIORS,dMS} = dIORS = 3.625

13) d(IORM)T = min {dIORT,dMT} = dIORT = 2.469

Dengan demikian terbentuk matriks jarak yang baru yang menjadi sebuah *cluster* antara IOR dan M pada tabel *cluster* 6.

Tabel 9.10 : Matrik Perbandingan Turunan 6 (Cluster 6)

	IORM	CPBJ	A	D	E	F	G	H	K	L	N	Q	S	T
IORM	0	0.963	2.114	3.018	0.735	4.095	1.768	1.667	1.273	1.577	1.610	0.998	3.625	2.469
CPBJ	0.963	0	3.027	4.091	0.470	5.092	2.835	2.517	2.227	2.443	2.249	0.359	4.603	3.505
A	2.114	3.027	0	1.839	2.578	2.955	1.288	0.533	1.317	0.659	1.833	2.977	2.333	0.759
D	3.018	4.091	1.839	0	3.699	1.137	1.316	2.119	1.870	2.015	2.171	3.884	0.639	1.256
E	0.735	0.470	2.578	3.699	0	4.724	2.465	2.061	1.831	1.987	1.875	0.544	4.207	3.083
F	4.095	5.092	2.955	1.137	4.724	0	2.367	3.242	2.904	3.129	3.059	4.843	0.695	2.346
G	1.768	2.835	1.288	1.316	2.465	2.367	0	1.277	0.731	1.131	1.367	2.660	1.912	1.082
H	1.667	2.517	0.533	2.119	2.061	3.242	1.277	0	1.019	0.233	1.492	2.468	2.622	1.192
K	1.273	2.227	1.317	1.870	1.831	2.904	0.731	1.019	0	0.805	0.775	2.035	2.385	1.511
L	1.577	2.443	0.659	2.015	1.987	3.129	1.131	0.233	0.805	0	1.265	2.358	2.514	1.216
N	1.610	2.249	1.833	2.171	1.875	3.059	1.367	1.492	0.775	1.265	0	1.967	2.535	2.090
Q	0.998	0.359	2.977	3.884	0.544	4.843	2.660	2.468	2.035	2.358	1.967	0	4.371	3.414
S	3.625	4.603	2.333	0.639	4.207	0.695	1.912	2.622	2.385	2.514	2.535	4.371	0	1.797
T	2.469	3.505	0.759	1.256	3.083	2.346	1.082	1.192	1.511	1.216	2.090	3.414	1.797	0

Dengan cara yang sama seperti mencari nilai Cluster 0 s/d 6 maka untuk mencari nilai Cluster 7 s/d Cluster Akhir di dapatkan seperti pada tabel di bawah ini:

Tabel 9.11 : Matrik Perbandingan Turunan 7 (Cluster 7)

	HL	IORM	CPBJ	A	D	E	F	G	K	N	Q	S	T
HL	0	1.577	2.443	0.533	2.015	1.987	3.129	1.131	0.805	1.265	2.358	2.514	1.192
IORM	1.577	0	0.963	2.114	3.018	0.735	4.095	1.768	1.273	1.610	0.998	3.625	2.469
CPBJ	2.443	0.963	0	3.027	4.091	0.470	5.092	2.835	2.227	2.249	0.359	4.603	3.505
A	0.533	2.114	3.027	0	1.839	2.578	2.955	1.288	1.317	1.833	2.977	2.333	0.759
D	2.015	3.018	4.091	1.839	0	3.699	1.137	1.316	1.870	2.171	3.884	0.639	1.256
E	1.987	0.735	0.470	2.578	3.699	0	4.724	2.465	1.831	1.875	0.544	4.207	3.083
F	3.129	4.095	5.092	2.955	1.137	4.724	0	2.367	2.904	3.059	4.843	0.695	2.346
G	1.131	1.768	2.835	1.288	1.316	2.465	2.367	0	0.731	1.367	2.660	1.912	1.082
K	0.805	1.273	2.227	1.317	1.870	1.831	2.904	0.731	0	0.775	2.035	2.385	1.511
N	1.265	1.610	2.249	1.833	2.171	1.875	3.059	1.367	0.775	0	1.967	2.535	2.090
Q	2.358	0.998	0.359	2.977	3.884	0.544	4.843	2.660	2.035	1.967	0	4.371	3.414
S	2.514	3.625	4.603	2.333	0.639	4.207	0.695	1.912	2.385	2.535	4.371	0	1.797
T	1.192	2.469	3.505	0.759	1.256	3.083	2.346	1.082	1.511	2.090	3.414	1.797	0

Tabel 9.12 : Matrik Perbandingan Turunan 8 (Cluster 8)

	CPBJQ	HL	IORM	A	D	E	F	G	K	N	S	T
CPBJQ	0	2.358	0.963	2.977	3.884	0.470	4.843	2.660	2.035	1.967	4.371	3.414
HL	2.358	0	1.577	0.533	2.015	1.987	3.129	1.131	0.805	1.265	2.514	1.192
IORM	0.963	1.577	0	2.114	3.018	0.735	4.095	1.768	1.273	1.610	3.625	2.469
A	2.977	0.533	2.114	0	1.839	2.578	2.955	1.288	1.317	1.833	2.333	0.759
D	3.884	2.015	3.018	1.839	0	3.699	1.137	1.316	1.870	2.171	0.639	1.256
E	0.470	1.987	0.735	2.578	3.699	0	4.724	2.465	1.831	1.875	4.207	3.083
F	4.843	3.129	4.095	2.955	1.137	4.724	0	2.367	2.904	3.059	0.695	2.346
G	2.660	1.131	1.768	1.288	1.316	2.465	2.367	0	0.731	1.367	1.912	1.082
K	2.035	0.805	1.273	1.317	1.870	1.831	2.904	0.731	0	0.775	2.385	1.511
N	1.967	1.265	1.610	1.833	2.171	1.875	3.059	1.367	0.775	0	2.535	2.090
S	4.371	2.514	3.625	2.333	0.639	4.207	0.695	1.912	2.385	2.535	0	1.797
T	3.414	1.192	2.469	0.759	1.256	3.083	2.346	1.082	1.511	2.090	1.797	0

Tabel 9.13 : Matrik Perbandingan Turunan 9 (Cluster 9)

	CPBJQE	HL	IORM	A	D	F	G	K	N	S	T
CPBJQE	0	1.987	0.735	2.578	3.699	4.724	2.465	1.831	1.875	4.207	3.083
HL	1.987	0	1.577	0.533	2.015	3.129	1.131	0.805	1.265	2.514	1.192
IORM	0.735	1.577	0	2.114	3.018	4.095	1.768	1.273	1.610	3.625	2.469
A	2.578	0.533	2.114	0	1.839	2.955	1.288	1.317	1.833	2.333	0.759
D	3.699	2.015	3.018	1.839	0	1.137	1.316	1.870	2.171	0.639	1.256
F	4.724	3.129	4.095	2.955	1.137	0	2.367	2.904	3.059	0.695	2.346
G	2.465	1.131	1.768	1.288	1.316	2.367	0	0.731	1.367	1.912	1.082
K	1.831	0.805	1.273	1.317	1.870	2.904	0.731	0	0.775	2.385	1.511
N	1.875	1.265	1.610	1.833	2.171	3.059	1.367	0.775	0	2.535	2.090
S	4.207	2.514	3.625	2.333	0.639	0.695	1.912	2.385	2.535	0	1.797
T	3.083	1.192	2.469	0.759	1.256	2.346	1.082	1.511	2.090	1.797	0

Tabel 9.14 : Matrik Perbandingan Turunan 10 (Cluster 10)

	HLA	EQCPBJ	IORM	D	F	G	K	N	S	T
HLA	0	1.987	1.577	1.839	2.955	1.131	0.805	1.265	2.333	0.759
EQCPBJ	1.987	0	0.735	3.699	4.724	2.465	1.831	1.875	4.207	3.083
IORM	1.577	0.735	0	3.018	4.095	1.768	1.273	1.610	3.625	2.469
D	1.839	3.699	3.018	0	1.137	1.316	1.870	2.171	0.639	1.256
F	2.955	4.724	4.095	1.137	0	2.367	2.904	3.059	0.695	2.346
G	1.131	2.465	1.768	1.316	2.367	0	0.731	1.367	1.912	1.082
K	0.805	1.831	1.273	1.870	2.904	0.731	0	0.775	2.385	1.511
N	1.265	1.875	1.610	2.171	3.059	1.367	0.775	0	2.535	2.090
S	2.333	4.207	3.625	0.639	0.695	1.912	2.385	2.535	0	1.797
T	0.759	3.083	2.469	1.256	2.346	1.082	1.511	2.090	1.797	0

Tabel 9.15 : Matrik Perbandingan Turunan 11 (Cluster 11)

	DS	HLA	EQCPBJ	IORM	F	G	K	N	T
DS	0	1.839	3.699	3.018	0.695	1.316	1.870	2.171	1.256
HLA	1.839	0	1.987	1.577	2.955	1.131	0.805	1.265	0.759
EQCPBJ	3.699	1.987	0	0.735	4.724	2.465	1.831	1.875	3.083
IORM	3.018	1.577	0.735	0	4.095	1.768	1.273	1.610	2.469
F	0.695	2.955	4.724	4.095	0	2.367	2.904	3.059	2.346
G	1.316	1.131	2.465	1.768	2.367	0	0.731	1.367	1.082
K	1.870	0.805	1.831	1.273	2.904	0.731	0	0.775	1.511
N	2.171	1.265	1.875	1.610	3.059	1.367	0.775	0	2.090
T	1.256	0.759	3.083	2.469	2.346	1.082	1.511	2.090	0

Tabel 9.16 : Matrik Perbandingan Turunan 12 (Cluster 12)

	DSF	HLA	EQCPBJ	IORM	G	K	N	T
DSF	0	1.839	3.699	3.018	1.316	1.870	2.171	1.256
HLA	1.839	0	1.987	1.577	1.131	0.805	1.265	0.759
EQCPBJ	3.699	1.987	0	0.735	2.465	1.831	1.875	3.083
IORM	3.018	1.577	0.735	0	1.768	1.273	1.610	2.469
G	1.316	1.131	2.465	1.768	0	0.731	1.367	1.082
K	1.870	0.805	1.831	1.273	0.731	0	0.775	1.511
N	2.171	1.265	1.875	1.610	1.367	0.775	0	2.090
T	1.256	0.759	3.083	2.469	1.082	1.511	2.090	0

Tabel 9.17 : Matrik Perbandingan Turunan 13 (Cluster 13)

	GK	DSF	HLA	EQCPBJ	IORM	N	T
GK	0	1.316	0.805	1.831	1.273	0.775	1.082
DSF	1.316	0	1.839	3.699	3.018	2.171	1.256
HLA	0.805	1.839	0	1.987	1.577	1.265	0.759
EQCPBJ	1.831	3.699	1.987	0	0.735	1.875	3.083
IORM	1.273	3.018	1.577	0.735	0	1.610	2.469
N	0.775	2.171	1.265	1.875	1.610	0	2.090
T	1.082	1.256	0.759	3.083	2.469	2.090	0

Tabel 9.18 : Matrik Perbandingan Turunan 14 (Cluster 14)

	EQCPBJIORM	GK	DSF	HLA	N	T
EQCPBJIORM	0	1.273	3.018	1.577	1.610	2.469
GK	1.273	0	1.316	0.805	0.775	1.082
DSF	3.018	1.316	0	1.839	2.171	1.256
HLA	1.577	0.805	1.839	0	1.265	0.759
N	1.610	0.775	2.171	1.265	0	2.090
T	2.469	1.082	1.256	0.759	2.090	0

Tabel 9.19 : Matrik Perbandingan Turunan 15 (Cluster 15)

	HLAT	EQCPBJIORM	GK	DSF	N
HLAT	0	1.577	0.805	1.256	1.265
EQCPBJIORM	1.577	0	1.273	3.018	1.610
GK	0.805	1.273	0	1.316	0.775
DSF	1.256	3.018	1.316	0	2.171
N	1.265	1.610	0.775	2.171	0

Tabel 9.20 : Matrik Perbandingan Turunan 16 (Cluster 16)

	GKN	HLAT	EQCPBJIORM	DSF
GKN	0	0.805	1.273	1.316
HLAT	0.805	0	1.577	1.256
EQCPBJIORM	1.273	1.577	0	3.018
DSF	1.316	1.256	3.018	0

Tabel 9.21 : Matrik Perbandingan Turunan 17 (Cluster 17)

	GKNHALT	EQCPBJIORM	DSF
GKNHALT	0	1.273	1.256
EQCPBJIORM	1.273	0	3.018
DSF	1.256	3.018	0

Tabel 9.22: Matriks Perbandingan Turunan 18 (Cluster 18)

	DSFGKNHALT	EQCPBJIORM
DSFGKNHALT	0	1.273
EQCPBJIORM	1.273	0

Tabel 9.23 : Matriks Perbandingan Turunan 18 (Cluster 18)

0	DSFGKHALT-EQCPBJIORM
DSFGKHALT-EQCPBJIORM	0

Tabel 9.24 Data Keseluruhan *Cluster*

Cluster	Kode Cluster	Keterangan	Distance
1	I-O	Irvan Yoanda Ginting (2013010127), Rinaldy Fadly (2013010051)	0.090
2	C-P	Ayumi Sari Putri (2013010119), Rita Yanti Purba (2013010058)	0.123
3	CP-J	Ayumi Saru Putri (2013010119), Rita Yanti Purba (2013010058), Juli Nanda Sari (2013010104)	0.140
4	CPJ-B	Ayumi Saru Putri (2013010119), Rita Yanti Purba (2013010058), Juli Nanda Sari (2013010104), Atikah Khairani Rangkuti (2013010118)	0.146
5	IO-R	Irvan Yoanda Ginting (2013010127), Rinaldy Fadly (2013010051), Sabariah Solin (2013010047)	1.191
6	IOR-M	Irvan Yoanda Ginting (2013010127), Rinaldy Fadly (2013010051), Sabariah Solin (2013010047), Nika Depri Sitepu (2013010032)	0.210
7	H-L	Hendi Thien Sulkanando (2013010112), Muhammad Iksan Rasta Tarigan (2013010226)	0.233
8	CPBJ-Q	Ayumi Saru Putri (2013010119), Rita Yanti Purba (2013010058), Juli Nanda Sari (2013010104), Atikah Khairani Rangkuti (2013010118), Rizky Ramadhan (2013010274)	0.359
9	CPBJQ-E	Ayumi Saru Putri (2013010119), Rita Yanti Purba (2013010058), Juli Nanda Sari (2013010104), Atikah Khairani Rangkuti (2013010118), Rizky Ramadhan (2013010274), Dewi Purnama Nasution (2013010062)	0.470
10	HL-A	Hendi Thien Sulkanando (2013010112), Muhammad Iksan Rasta Tarigan (2013010226), Afdillah Ramadhani Arhas (2013010077)	0.533
11	D-S	Daniel Sastra Novindo Valente (2013010247), Susi Sitanggang (2013010228)	0.639
12	DS-F	Daniel Sastra Novindo Valente (2013010247), Susi Sitanggang (2013010228), Dian Srihartati Simanjorang (2013010055)	0.695
13	G-K	Fauzi Pratama (2013010264), Misna Wati Br Meliala (2013010200)	0.731

Tabel 9.24 Data Keseluruhan *Cluster* (Lanjutan)

14	EQCPBJ-IORM	Ayumi Saru Putri (2013010119), Rita Yanti Purba (2013010058), Juli Nanda Sari (2013010104), Atikah Khairani Rangkuti (2013010118), Rizky Ramadhan (2013010274), Dewi Purnama Nasution (2013010062), Irvan Yoanda Ginting (2013010127), Rinaldy Fadly (2013010051), Sabariah Solin (2013010047), Nika Depri Sitepu (2013010032)	0.735
15	HLA-T	Hendi Thien Sulkanando (2013010112), Muhammad Iksan Rasta Tarigan (2013010226), Afdillah Ramadhani Arhas (2013010077), Yolanda Sherpita Br. Ginting (2013010102)	0.759
16	GK-N	Fauzi Pratama (2013010264), Misna Wati Br Meliala (2013010200), Qaswin Ilham (2013010212)	0.775
17	GKN-HLAT	Fauzi Pratama (2013010264), Misna Wati Br Meliala (2013010200), Qaswin Ilham (2013010212), Hendi Thien Sulkanando (2013010112), Muhammad Iksan Rasta Tarigan (2013010226), Afdillah Ramadhani Arhas (2013010077), Yolanda Sherpita Br, Ginting (2013010102)	0.805
18	DSF-GKNHLAT	Daniel Sastra Novindo Valente (2013010247), Susi Sitanggang (2013010228), Dian Srihartati Simanjorang (2013010055), Fauzi Pratama (2013010264), Misna Wati Br Meliala (2013010200), Qaswin Ilhan (2013010212), Hendi Thien Sulkanando (2013010112), Muhammad Iksan Rasta Tarigan (2013010226), Afdillah RamadhaniArhas (2013010077), Yolanda Sherpita Br Ginting (2013010102)	1.256

Adapun mahasiswa yang dibagi berdasarkan hasil *cluster* di atas adalah sebagai berikut:

Tabel 9.25 Hasil *Cluster*

Nomor *Cluster*	Nirm	Nama
***Cluster* 0**	2013010247	Daniel Sastra Novindo Valente
	2013010228	Susi Sitanggang
	2013010055	Dian Srihartati Simanjorang
	2013010264	Fauzi Pratama
	2013010200	Misna Wati Br Meliala
	2013010212	Qaswin Ilhan
	2013010112	Hendi Thien Sulkanando
	2013010226	Muhammad Iksan Rasta Tarigan
	2013010077	Afdillah Ramadhani Arhas
	2013010102	Yolanda Sherpita Br Ginting
***Cluster* 1**	2013010119	Ayumi Saru Putri
	2013010058	Rita Yanti Purba
	2013010104	Juli Nanda Sari
	2013010118	Atikah Khairani Rangkuti
	2013010274	Rizky Ramadhan
	2013010062	Dewi Purnama Nasution
	2013010127	Irvan Yoanda Ginting
	2013010051	Rinaldy Fadly
	2013010047	Sabariah Solin
	2013010032	Nika Depri Sitepu

Berdasarkan data IPK pada tabel 5.25 dapat disimpulkan dari data *cluster* 0 adalah kelompok mahasiswa yang tidak unggul dan *cluster* 1 kelompok mahasiswa yang unggul, karena data mahasiswa yang berada di *cluster* 0 adalah mahasiswa yang memiliki nilai rendah **(IPK < 2.0),** sedangkan data mahasiswa yang berada di *cluster* 1 adalah mahasiswa yang memiliki nilai tinggi **(IPK > 3.0)**.

LATIHAN: Berikut ini adalah Sampel Data Untuk Pengujian:

No	NAMA	V1	V2	V3
1	Dicky	3.4	80	70
2	Dini	4.00	70	90
3	Dina	3.4	80	80
4	Dino	1.8	70	80
5	Dian	3.4	80	80

Keterangan:

V1= Indeks Prestasi Kumulatif (IPK)

V2 = Nilai Absensi

V3 = Nilai Etika

Dan Berikut ini adalah Tabel Asumsi Objek yaitu:

Tabel: Asumsi Objek

No. Urut Mahasiswa	Kode Huruf	Nama Mahasiswa
1	A	Dicky
2	B	Dini
3	C	Dina
4	D	Dino
5	E	Dian

Dan berdasarkan Data Di atas Carilah Klaster dari Data di atas dengan menggunakan Teknik Single Linkage?

BAB 10

BAB X : METODE K-MEANS CLUSTERING

10.1 Pendahuluan

Algoritma *K-Means* merupakan algoritma pengelompokan iteratif yang melakukan partisi set data kedalam sejumlah K *cluster* yang sudah ditetapkan di awal. Algoritma sederhana untuk diimplementasikan dan dijalankan, relatif cepat, mudah beradaptasi, umum penggunaanya dalam praktik. Secara historis, *K-Means* menjadi salah satu algoritma yang paling penting dalam bidang data mining (Wu dan Kumar, dalam Eko Prasetyo 2014:189).

Algoritma *K-Means* merupakan salah satu metode data *clustering non-hierarki* yang mengelompokkan data dalam bentuk satu atau lebih *cluster* atau kelompok. Langkah-langkah melakukan *clustering* dengan metode *K-Means* adalah :

1. Pilih jumlah *cluster* K.
2. Inisialisasi K pusat *cluster* ini bisa dilakukan dengan berbagai cara.
3. Alokasi semua data/objek ke *cluster* terdekat.
4. Hitung kembali pusat *cluster* dengan keanggotaan *cluster* yang sekarang.
5. Tugaskan lagi setiap objek memakai pusat *cluster* yang baru.

Langkah-langkah *K-Means* adalah sebagai berikut :

1. Langkah pertama: Tanyakan kepada pemakai Algoritma *K-Means*, sebutlah sebanyak K kelompok.
2. Langkah kedua: Secara sembarang, pilihlah K buah catatan (dari sekian catatan yang ada) sebagai pusat-pusat kelompok awal.
3. Langkah ketiga: Untuk setiap catatan, tentukan pusat kelompok terdekatnya dan tetapkan catatan tersebut sebagai anggota dari kelompok yang terdekat pusat kelompoknya. Hitung rasio antara besaran *Between Cluster Variation* dengan *Within Cluster Variation*, lalu bandingkan rasio tersebut dengan rasio sebelumnya (bila sudah ada). Jika rasio tersebut membesar, lanjutkan ke langkah keempat. Jika tidak, hentikan prosesnya.

4. Langkah keempat: Perbaharui pusat-pusat kelompok (berdasarkan kelompok yang didapat dari langkah ketiga) dan kembalilah ke langkah ketiga.

10.2 Contoh Kasus Dan Penyelesainnya

Salah toko penjualan buku sedang melakukan suatu analisi penjualan. Dalam hal ini toko buku ingin melihat item-item apa saja yang menjadi tolak ukur penjualan. Data ini bersumber dari Toko xxx yang pemiliknya langsung memberikan data yang berkenaan dengan penjualan buku di tokonya.Berikut ini adalah daftar buku yang menjadi sampel penelitian di Toko xxx. dimana data tersebut yang akan diolah menggunakan Algoritma *K-Means*. Berikut ini adalah daftar buku yang menjadi sampel penelitian di Toko xxx, dimana data tersebut akan diolah menggunakan algoritma *K-Means*.

Tabel 10.1 Sampel Data

No.	Judul Buku	Penulis	Penerbit
1	Tuntunan Sholat Wanita Lengkap	Fatihuddin Abul Yasin	Terbit Terang
2	Khutbah Jum'at	Syaikh Hamid	Karya Agung
3	Kamus Bahasa Inggris-Indonesia	Adi Gunawan	Kartika
4	Kamus Bahasa Arab-Indonesia Al Manar	Idrus Alkaf	Karya Utama
5	Psikologi Umum	Alex Sabar	Pustaka Setia
6	2700 Peribahasa Indonesia	Nur Arifin Chaniago	Pustaka Setia
7	Ilmu Tajwid	Abu Rifqi al Hanif	Terbit Terang
8	Sirah Nabawiyah	Syaikh Syaifurahman	Pustaka Al Kausar
9	Abu Bakar As Siddiq	M. Ihsan	Literia
10	Bulughul Maram	Imam AL Hafiz	Mutiara Ilmu
11	Ilmu Al Qur'an dan Tafsir	Teungku M. Hasbi	Pustaka Rizki Putra
12	Fiqih Praktis Akhwat	Ahmar Sarwat	Tauhid Media Center
13	Fiqih Munakahat 2	Beni Ahmad Saebani	Pustaka Setia
14	Filsafat Pendidikan	Anas Salahuddin	Pustaka Setia
15	Pasar Modal Syariah	Khaerul Umam	Pustaka Setia
16	Ekonomi Makro	Asfia Murni	Aditama
17	Nama-nama Islami	M. Fairuz	Pustaka Media
18	Tuntunan Mandi Wajib	M. Anwar	Sumber Ilmu Jaya
19	Surat Yaasiin	M. Anwar	Sumber Ilmu Jaya
20	Juz 'Amma dan Terjemah	Abu Ali Zainal	Raja Publishing

Tabel 10.2 Dafta Harga Dan Jumlah Penjualan

Potongan Harga				
No.	**Harga**	**10% (Umum)**	**25% (Pengajar)**	**Jumlah Penjualan**
1	Rp15,000	Rp13,500	Rp11,250	6 Buah
2	Rp17,000	Rp15,300	Rp12,750	5 Buah
3	Rp92,000	Rp82,800	Rp69,000	7 Buah
4	Rp38,500	Rp34,650	Rp28,875	7 Buah
5	Rp70,000	Rp63,000	Rp52,500	5 Buah
6	Rp20,000	Rp18,000	Rp15,000	10 Buah
7	Rp5,000	Rp4,500	Rp3,750	15 Buah
8	Rp110,000	Rp99,000	Rp82,500	4 Buah
9	Rp100,000	Rp90,000	Rp75,000	3 Buah
10	Rp58,000	Rp52,200	Rp43,500	8 Buah
11	Rp60,000	Rp54,000	Rp48,000	6 Buah
12	Rp35,000	Rp31,500	Rp26,250	10 Buah
13	Rp35,000	Rp31,500	Rp26,250	10 Buah
14	Rp32,000	Rp28,800	Rp24,000	3 Buah
15	Rp65,000	Rp58,500	Rp48,750	3 Buah
16	Rp75,000	Rp67,500	Rp56,250	5 Buah
17	Rp10,000	Rp9,000	Rp7,500	10 Buah
18	Rp5,000	Rp4,500	Rp3,750	12 Buah
19	Rp5,000	Rp4,500	Rp3,750	30 Buah
20	Rp10,000	Rp9,000	Rp7,500	35 Buah

2. **Penyelesaian Kasus**

   Langkah pertama untuk menyelesaikan perhitungan ini adalah memilih jumlah *cluster* K. *Cluster* K dari sumber data diatas adalah Penulis, Harga, dan Jumlah Penjualan. Kemudian masuk ke langkah kedua yaitu inisialisasi atau asumsi K pusat *cluster*. Asumsi ini bertujuan untuk mempermudah perhitungan dan mengingat nama objek. Berikut ini adalah tabel asumsi untuk judul buku :

Tabel 10.3 Asumsi Judul Buku

Judul Buku	Asumsi Judul Buku
Tuntunan Sholat Wanita Lengkap	A
Khutbah Jum'at	B
Kamus Bahasa Inggris-Indonesia	C
Kamus Bahasa Arab-Indonesia Al Manar	D
Psikologi Umum	E
2700 Peribahasa Indonesia	F
Ilmu Tajwid	G
Sirah Nabawiyah	H
Abu Bakar As Siddiq	I
Bulughul Maram	J
Ilmu Al Qur'an dan Tafsir	K
Fiqih Praktis Akhwat	L
Fiqih Munakahat 2	M
Filsafat Pendidikan	N
Pasar Modal Syariah	O
Ekonomi Makro	P
Nama-nama Islami	Q
Tuntunan Mandi Wajib	R
Surat Yaasiin	S
Juz 'Amma dan Terjemah	T

Selanjutnya asumsi untuk penulis yang merupakan *cluster* pertama. Berikut tabel asumsi untuk penulis :

Tabel 10.4 Asumsi Penulis

Penulis	Asumsi
Fatihuddin Abul Yasin	1
Syaikh Hamid	2
Adi Gunawan	3
Idrus Alkaf	4
Alex Sabar	5
Nur Arifin Chaniago	6
Abu Rifqi al Hanif	7
Syaikh Syaifurahman	8
M. Ihsan	9
Imam AL Hafiz	10

Teungku M. Hasbi	11
Ahmar Sarwat	12
Beni Ahmad Saebani	13
Anas Salahuddin	14
Khaerul Umam	15
Asfia Murni	16
M. Fairuz	17
M. Anwar	18
M. Anwar	19
Abu Ali Zainal	20

Cluster kedua adalah harga. Di asumsi ini menggunakan interval agar harga mudah diingat dengan cara dikelompokkan. Berikut tabel asumsi harga :

Tabel 10.5 Asumsi Harga

Interval Harga	Asumsi
Rp5.000-Rp20.000	1
Rp21.000-Rp35.000	2
Rp36.000-Rp50.000	3
Rp51.000-Rp65.000	4
Rp66.000-Rp80.000	5
Rp81.000-Rp95.000	6
Rp96.000-Rp110.000	7

Dan ini adalah *cluster* terakhir yaitu jumlah penjualan. Di tabel ini juga interval jumlah penjualannya. Berikut tabel asumsi jumlah penjualan :

Tabel 10.6 Asumsi Jumlah Penjualan

Interval Jumlah Pemjumlahan	**Asumsi**
1-5	1
6-10	2
11-15	3
16-20	4
21-25	5
26-30	6
31-35	7

Dengan melihat tabel ini terlihat jelas semua *cluster* sudah diasumsikan sehingga mempermudah dalam perhitungan nantinya. Berikut ini adalah tabel dari asumsi buku :

Tabel 10.7 Asumsi Buku

No.	Judul Buku	Asumsi Judul Buku
1	Tuntunan Sholat Wanita Lengkap	A
2	Khutbah Jum'at	B
3	Kamus Bahasa Inggris-Indonesia	C
4	Kamus Bahasa Arab-Indonesia Al Manar	D
5	Psikologi Umum	E
6	2700 Peribahasa Indonesia	F
7	Ilmu Tajwid	G
8	Sirah Nabawiyah	H
9	Abu Bakar As Siddiq	I
10	Bulughul Maram	J
11	Ilmu Al Qur'an dan Tafsir	K
12	Fiqih Praktis Akhwat	L
13	Fiqih Munakahat 2	M
14	Filsafat Pendidikan	N
15	Pasar Modal Syariah	O
16	Ekonomi Makro	P
17	Nama-nama Islami	Q
18	Tuntunan Mandi Wajib	R
19	Surat Yaasiin	S
20	Juz 'Amma dan Terjemah	T

Tabel 10.8 Asumsi Data Analisis

Penulis	V1	Harga	V2	Jumlah Penjualan	V3
Fatihuddin Abul Yasin	1	Rp15,000	1	6 Buah	2
Syaikh Hamid	2	Rp17,000	1	5 Buah	1
Adi Gunawan	3	Rp92,000	6	7 Buah	2
Idrus Alkaf	4	Rp38,500	3	7 Buah	2
Alex Sabar	5	Rp70,000	5	5 Buah	1
Nur Arifin Chaniago	6	Rp20,000	1	10 Buah	3
Abu Rifqi al Hanif	7	Rp5,000	1	15 Buah	3
Syaikh Syaifurahman	8	Rp110,000	7	4 Buah	1
M. Ihsan	9	Rp100,000	7	3 Buah	1
Imam AL Hafiz	10	Rp58,000	4	8 Buah	2
Teungku M. Hasbi	11	Rp60,000	4	6 Buah	2
Ahmar Sarwat	12	Rp35,000	2	10 Buah	2
Beni Ahmad Saebani	13	Rp35,000	2	10 Buah	2
Anas Salahuddin	14	Rp32,000	2	3 Buah	1
Khaerul Umam	15	Rp65,000	4	3 Buah	1
Asfia Murni	16	Rp75,000	5	5 Buah	1
M. Fairuz	17	Rp10,000	1	10 Buah	2
M. Anwar	18	Rp5,000	1	12 Buah	3
M. Anwar	18	Rp5,000	1	30 Buah	6
Abu Ali Zainal	19	Rp10,000	1	35 Buah	7

Keterangan:

V1: Asumsi Penulis

V2: Asumsi Harga

V3: Asumsi Jumlah Penjualan

Selanjutnya masuk pada tahap normalisasi semua objek dan *cluster.* Berikut tabel normalisasi dari buku :

Tabel 10.9 Normalisasi

No.	Asumsi Judul	V1	V2	V3
1	A	1	1	2
2	B	2	1	1
3	C	3	3	2
4	D	4	5	2
5	E	5	1	1
6	F	6	1	3
7	G	7	7	3
8	H	8	7	1
9	I	9	4	1
10	J	10	4	2
11	K	11	2	2
12	L	12	2	2
13	M	13	2	2
14	N	14	4	1
15	O	15	5	1
16	P	16	1	1
17	Q	17	1	2
18	R	18	1	3
19	S	18	1	6
20	T	19	1	7

Setelah proses inisialisasi selesai masuk ke langkah ketiga yaitu penentuan titik pusat dari setiap *cluster* seperti dibawah ini.

Pusat 1: T {19, 1, 7} sebagai m1

Pusat 2: B {2,1,1} sebagai m2

Pusat 3: I {9, 7, 1} sebagai m3

Selanjutnya menghitung jarak terdekat dari setiap objek atau judul buku. Berikut perhitungannya :

Rumus menghitung jarak yaitu Euclidian=$\sqrt{\sum_{i=1}^{n} (Xi - Yi)^2}$

1. Objek A
   a. Jarak objek A ke pusat 1

      d(A,T)

      = √(1-19)² + (1-1)² + (2-7)²

      =√(-18)² + (0)² + (-5)²

      = √324 + 0 + 25

      = √349 = 18,681

   b. Jarak objek A ke pusat 2

      d(A,B)

      = √(1-2)²+ (1-1)² + (2-1)²

      = √(- 1)² + (0)² + (1)²

      = √1 + 0 + 1

      = √2 = 1,414

Perhitungan selanjutnya dapat dilihat di lampiran. Adapun hasil dari perhitungan iterasi 1 dapat dilihat pada tabel dibawah ini:

Tabel 10.10 : Iterasi 1

Asumsi Judul	m1	m2	m3	Jarak Terdekat	WCV
A	18.681	1.414	10.049	B	2
B	18.027	0	9.219	B	0
C	17.492	5.196	6.164	B	27
D	15.937	3	6.480	B	9
E	15.748	5	4.472	I	20
F	13.601	4.472	7	B	20
G	12.649	5.385	6.633	B	29
H	13.892	8.485	1	I	1
I	13.114	9.219	0	I	0
J	10.723	8.602	3.316	I	11
K	9.899	9.539	3.741	I	14
L	8.660	10.099	5.916	I	35
M	7.874	11.090	6.480	I	42
N	7.874	12.041	7.071	I	50
O	7.810	13.341	6.708	I	45
P	7.810	14.560	7.280	I	53
Q	5.385	15.033	10.049	T	29
R	4.123	16.124	11	T	17
S	1.414	16.763	11.916	T	2
T	0	18.027	13.114	T	0

Total: m1 = 210.720 Rata-rata: m1= 10.536

m2 = 187.396 m2= 9.369

m3 = 137.615 m3= 6.880

BCV: 40.362

WCV: 406

Rasio: BCV/WCV

40.362/406 = 0.099

Keterangan:

BCV: *Between Cluster Variation*

WCV: *Within Cluster Variation*

Kesimpulan pada iterasi 1 dapat dilihat seperti dibawah ini :

1. Iterasi 1 mempunyai tiga (3) pusat yang menghasilkan tiga (3) *cluster* yaitu :

Tabel 10.11 Kesimpulan Iterasi 1

Nama *Cluster*	Jumlah Anggota *Cluster*	Anggota *Cluster*	Inisial
m1	4	Q = Nama-nama Islami	ISBN-17
		R = Tuntutan Mandi Wajib	ISBN-18
		S = Surat Yaasiin	ISBN-19
		T = Juz ‘Amma dan Terjemah	ISBN-20
m2	6	A = Tuntunan Sholat Wanita Lengkap	ISBN-01
		B = Khutbah Jumat	ISBN-02
		C = Kamus Bahasa Inggris Indonesia	ISBN-03
		D = Kamus Bahasa Arab-Indonesia Al-Manar	ISBN-04
		F = 2700 Peribahasa Indonesia	ISBN-06
		G = Ilmu Tajwid	ISBN-07
m3	10	E = Psikologi Umum	ISBN-05
		H = Sirah Nabawiyah	ISBN-08
		I = Abu Bakar As Siddiq	ISBN-09
		J = Bulughul Maram	ISBN-10
		K = Ilmu Al Qur'an dan Tafsir	ISBN-11
		L = Fiqih Praktis Akhwat	ISBN-12
		M = Fiqih Munakahat 2	ISBN-13
		N = Filsafat Pendidikan	ISBN-14
		O = Pasar Modal Syariah	ISBN-15
		P = Ekonomi Makro	ISBN-16

2. Dari ketiga *cluster* diatas yang paling potensial adalah m3 mempunyai 10 (sepuluh) anggota yang didominasi buku mengenai agama Islam. Jadi Toko Buku Sinar Ilmu sebaiknya lebih banyak menjual buku tentang tokoh Islam dan Fiqih.

Mengingat langkah sebelumnya belum mendapatkan rasio ini, maka perbandingan rasio sebelumnya dilanjutkan dengan pembaruan pusat-pusat kelompok seperti berikut ini :

1. Rata-rata m1 (Q, R, S, T) = 18; 1; 4,5
2. Rata-rata m2 (E, H, I, J, K, L, M, N, O, P) = 11,3 ; 4,2 ; 1,4
3. Rata-rata m3 (A, B, C, D, F, G) = 2,83 ; 2, 17 ; 2,17

Selanjutnya hitung iterasi 2 seperti halnya iterasi 1 hingga mendapatkan nilai rasio yang sama dengan nilai rasio sebelumnya. Dibawah ini adalah perhitungannya iterasi 2:

Rumus menghitung jarak terdekat setiap objek adalah

Euclidian=$\sqrt{\sum_{i=1}^{n} (Xi - Yi)^2}$

1. Objek A
   a. Jarak pusat m1 ke objek A

      d(m1,A)

      = $\sqrt{(18-1)^2 + (1-1)^2 + (4,5-2)^2}$

      =$\sqrt{(17)^2 + (0)^2 + (2,5)^2}$

      = $\sqrt{289 + 0 + 6,25}$= $\sqrt{295,25}$ = 17,182

   b. Jarak pusat m2 ke objek A

      d(m2,A)

      = $\sqrt{(11,3-1)^2 + (4,2-1)^2 + (1,4-2)^2}$

      =$\sqrt{(10,3)^2 + (3,2)^2 + (-0,6)^2}$

      = $\sqrt{106,09 + 10,24 + 0,36}$= $\sqrt{213,58}$ = 14,614

Perhitungan selanjutnya dapat dilihat di lampiran. Adapun hasil dari perhitungan iterasi 2 dapat dilihat pada tabel dibawah ini:

Tabel 10.12 Iterasi 2

Asumsi Judul	m1	m2	m3	Jarak Terdekat	WCV
A	17.182	14.614	3.068	m3	9.416
B	16.408	13.014	2.466	m3	6.083
C	15.337	11.295	3.926	m3	15.416
D	14.534	9.776	0.866	m3	0.7499
E	14.044	9.021	3.278	m3	10.749
F	13.086	8.495	2.598	m3	6.750
G	12.619	7.083	3.476	m3	12.083
H	12.698	5.672	6.487	m2	32.179
I	12.539	3.921	7.170	m2	15.38
J	12.298	1.606	6.435	m2	2.579
K	12.459	2.485	7.399	m2	6.18
L	12.776	3.712	8.170	m2	13.78
M	13.238	4.514	9.169	m2	20.38
N	14.044	5.116	10.234	m2	26.18
O	14.739	6.064	11.376	m2	36.78
P	15.532	8.122	12.546	m2	65.979
Q	16.224	10.038	13.219	m2	100.780
R	17.066	11.574	14.239	m2	133.98
S	17.066	13.310	14.722	m2	177.18
T	18.200	10.044	15.960	m2	100.890

Total: m1 = 292.099 Rata-rata: m1 = 14.604

m2 = 159.487 m2 = 7.974

m3 = 156.811 m3 = 7.840

BCV: 26.536

WCV: 793.5

Rasio: BCV/WCV

26.536/793.5 = 0.033

Keterangan:

BCV: *Between Cluster Variation*

WCV: *Within Cluster Variation*

Kesimpulan pada iterasi 2 dapat dilihat seperti dibawah ini :

1. Iterasi 2 mempunyai tiga (3) pusat yang menghasilkan dua (2) *cluster* yaitu :

Tabel 10.13 Kesimpulan Iterasi 2

Nama *Cluster*	Jumlah Anggota *Cluster*	Anggota *Cluster*	Inisial
m2	7	A = Tuntunan Sholat Wanita Lengkap	ISBN-01
		B = Khutbah Jumat	ISBN-02
		C = Kamus Bahasa Inggris Indonesia	ISBN-03
		D = Kamus Bahasa Arab-Indonesia Al-Manar	ISBN-04
		E = Psikologi Umum	ISBN-05
		F = 2700 Peribahasa Indonesia	ISBN-06
		G = Ilmu Tajwid	ISBN-07
m3	13	H = Sirah Nabawiyah	ISBN-08
		I = Abu Bakar As Siddiq	ISBN-09
		J = Bulughul Maram	ISBN-10
		K = Ilmu Al Qur'an dan Tafsir	ISBN-11
		L = Fiqih Praktis Akhwat	ISBN-12
		M = Fiqih Munakahat 2	ISBN-13
		N = Filsafat Pendidikan	ISBN-14
		O = Pasar Modal Syariah	ISBN-15
		P = Ekonomi Makro	ISBN-16
		Q = Nama-nama Islami	ISBN-17
		R = Tuntutan Mandi Wajib	ISBN-18
		S = Surat Yaasiin	ISBN-19
		T = Juz 'Amma dan Terjemah	ISBN-20

2. Dari kedua *cluster* diatas yang paling potensial adalah m2 mempunyai 13 (tiga belas) anggota yang didominasi buku mengenai agama Islam. Jadi Toko Buku Sinar Ilmu sebaiknya lebih banyak menjual buku tentang tokoh Islam dan Fiqih.

Tampaknya rasio ini (0,033) tidak sama dengan rasio sebelumnya (0,099) maka algoritma dilanjutkan dengan pembaruan pusat-pusat kelompok seperti berikut ini :

1. Rata-rata m1 (A, B, C, D, E, F, G, H) = 4; 2,57;

2. Rata-rata m2 (G, H, I, J, K, L, M, N, O, P, Q, R, S, T) = 13,84 ; 3,15

Selanjutnya hitung iterasi 3 dan tabel dibawah ini menunjukkan nilai rasio ini (0,033) sudah sama dengan rasio sebelumnya (0,033), oleh karena itu algortima pun dihentikan.

Euclidian=$\sqrt{\sum_{=}}$ $(\mathrm{Xi} - \mathrm{Yi})^{2}$

1. Objek A

a. Jarak pusat m1 ke objek A

d(m1,A)

$= \sqrt{(4-1)^2 + (2{,}25-1)^2 + (2-2)^2}$

$= \sqrt{(3)^2 + (1{,}25)^2 + (0)^2}$

$= \sqrt{9 + 1{,}58 + 0}$

$= \sqrt{10{,}58} = 3{,}25$

b. Jarak pusat m2 ke objek A

d(m2,A)

$= \sqrt{(13{,}84-1)^2 + (3{,}15-1)^2 + (2{,}38-2)^2}$

$= \sqrt{(12{,}84)^2 + (2{,}15)^2 + (0{,}38)^2}$

$= \sqrt{169{,}8} = 13.03$

Perhitungan selanjutnya dapat dilihat di lampiran. Adapun hasil dari perhitungan iterasi 3 dapat dilihat pada tabel dibawah ini:

Tabel 10.14 Iterasi 3

Asumsi Judul	m1	m2	Jarak Terdekat	WCV
A	3.252	13.031	m1	10.580
B	2.565	12.119	m1	6.580
C	3.874	11.219	m1	15.008
D	0.742	9.854	m1	0.551
E	3.085	9.142	m1	9.523
F	2.565	8.159	m1	6.580
G	3.402	7.203	m1	11.580
H	8.284	7.133	m2	39.494
I	6.963	6.339	m2	40.195
J	6.248	3.956	m2	15.656
K	7.213	2.994	m2	8.964
L	8.004	2.210	m2	4.887
M	9.003	1.481	m2	81.066
N	10.053	1.808	m2	3.272
O	11.182	1.991	m2	3.964
P	12.350	3.156	m2	9.964
Q	13.060	3.838	m2	14.733
R	14.091	4.719	m2	22.272
S	14.614	5.913	m2	34.964
T	15.861	7.245	m2	52.503

Total: m1 = 156.420 Rata-rata: m1 = 7.821

m2 = 123.521 m2 = 6.176

BCV: 13.031

WCV: 392.345

Rasio: BCV/WCV

13.031/392.345 = 0.033

Keterangan:

BCV: *Between Cluster Variation*

WCV: *Within Cluster Variation*

Kesimpulan pada iterasi 3 dapat dilihat seperti dibawah ini :

1. Iterasi 3 mempunyai tiga (3) pusat yang menghasilkan dua (2) *cluster* yaitu :

Tabel 10.15 Kesimpulan Iterasi 3

Nama *Cluster*	**Jumlah Anggota *Cluster***	**Anggota *Cluster***	**Inisial**
m1	7	A = Tuntunan Sholat Wanita Lengkap	ISBN-01
		B = Khutbah Jumat	ISBN-02
		C = Kamus Bahasa Inggris Indonesia	ISBN-03
		D = Kamus Bahasa Arab-Indonesia Al-Manar	ISBN-04
		E = Psikologi Umum	ISBN-05
		F = 2700 Peribahasa Indonesia	ISBN-06
		G = Ilmu Tajwid	ISBN-07
m2	13	H = Sirah Nabawiyah	ISBN-08
		I = Abu Bakar As Siddiq	ISBN-09
		J = Bulughul Maram	ISBN-10
		K = Ilmu Al Qur'an dan Tafsir	ISBN-11
		L = Fiqih Praktis Akhwat	ISBN-12
		M = Fiqih Munakahat 2	ISBN-13
		N = Filsafat Pendidikan	ISBN-14
		O = Pasar Modal Syariah	ISBN-15
		P = Ekonomi Makro	ISBN-16
		Q = Nama-nama Islami	ISBN-17
		R = Tuntutan Mandi Wajib	ISBN-18
		S = Surat Yaasiin	ISBN-19
		T = Juz ‘Amma dan Terjemah	ISBN-20

2. Dari kedua *cluster* diatas yang paling potensial adalah m1 mempunyai 13 (tiga belas) anggota yang didominasi buku mengenai agama Islam. Jadi Toko xxx sebaiknya lebih banyak menjual buku tentang tokoh Islam dan Fiqih.



BAB XI : FUNGSI MINOR ASOSIASI

5.1 Pendahuluan

Metode asosiasi termasuk salah satu metode dalam data mining yang anyak dipergunakan saat ini. Sebagai contoh web e-commerce yang menggunakan metode ini untuk membuat produk terkait. Selain itu di tempat perbelanjaan bagaimana manajer penjualan dapat mengestimasi produk apa yang persediaannya perlu ditingkatkan dan sebaliknya. Dalam metode ini terdapat beberapa teknik yang disebut Association Rule dengan Algoritma Apriori.

Association Rule Mining adalah teknik *data mining* untuk menemukan aturan asosiatif antara suatu kombinasi item. Contoh aturan asosiatif dari analisis pembelian di suatu pasar swalayan adalah dapat diketahuinya berapa besar kemungkinan seorang pelanggan membeli roti bersamaan dengan susu. Penting tidaknya suatu aturan asosiasi dapat diketahui dengan dua parameter yaitu:

a. ***Support*** **:**

Suatu ukuran yang menunjukkan seberapa besar tingkat dominasi suatu *item/itemset* dari keseluruhan transaksi. Ukuran ini menentukan apakah suatu item/itemset layak untuk dicari *confidence* tersebut (misal, dari keseluruhan transaksi yang ada, seberapa besar tingkat dominasi yang menunjukkan *item* A dan B dibeli bersamaan).

b. ***Confidence*** **:**

Suatu ukuran yang menunjukkan hubungan antar 2 *item* secara *conditional* (misal, seberapa sering item B dibeli jika orang membeli item A) ‘(Kusrini dan Emha, 2009:150).

Algoritma *Apriori* termasuk jenis aturan asosiasi pada data mining. Selain *apriori*, yang termaksuk pada golongan ini adalah metode *Generalized Rule Induction* dan Algoritma *Hash Based*.

Aturan yang menyatakan asosiasi antara beberapa atribut, sering disebut *affanity analysis* atau *market basket analysis*.

Analisis asosiasi atau *association rule mining* adalah teknik *data mining* untuk menemukan aturan asosiatif antara suatu kombinasi *item*. Contoh aturan asosiatif dari analisis pembelian di suatu pasar swalayan adalah dapat diketahuinya berapa besar kemungkinan seorang pelanggan membeli roti bersamaan dengan susu. Dengan pengetahuan tersebut, pemilik pasar swalayan dapat mengatur penempatan barangnya atau merancang kampanye pemasaran dengan memakai kupon diskon untuk kombinasi barang tertentu. Karena analisis asosiasi menjadi terkenal karena aplikasinya untuk menganalisis isi keranjang belanja di pasar swalayan, analisis asosiasi juga sering disebut *market basket analysis.*

Analisis asosiasi dikenal juga sebagai salah satu teknik *data mining* yang menjadi dasar dari berbagai teknik data mining lainnya. Secara khusus, salah satu tahap analisis asosiasi yang menarik perhatian banyak peneliti untuk menghasilkan algoritma yang efisien adalah analisis pola frekuensi tinggi *(frequent pattern mining).*

Penting tidaknya suatu aturan asosiatif dapat diketahui dengan dua parameter, yaitu *support* dan *confidence. Support* (nilai penunjang) adalah persentase kombinasi item tersebut dalam *database*, sedangkan *confidence* (nilai kepastian) adalah kuatnya hubungan antar *item* dalam aturan asosiasi.

Aturan asosiasi biasanya dinyatakan dalam bentuk:

{roti, mentega} → {susu} (support = 40%, confidence = 50%)

Aturan tersebut berarti “50% dari transaksi di database yang memuat item roti dan mentega juga memuat item susu. Sedangkan 40% dari seluruh transaksi yang ada di database memuat ketiga item tersebut.”

Dapat juga diartikan: “Seorang konsumen yang membeli roti dan mentega punya kemungkinan 50% untuk juga membeli susu. Aturan ini cukup signifikan karena mewakili 40% dari catatan transaksi selama ini.”

Analisis asosiasi didefinisikan suatu proses untuk menemukan semua aturan asosiasi yang memenuhi syarat minimum untuk support *(minimum support)* dan syarat minimum untuk *confidence (minimum confidence).* (Kusrini dan Emha, 2009:149)

Adapun langkah-langkah penyelesaiannya adalah sama dengan Assosiat

a. Analisis pola frekuensi tinggi

Tahap ini mencari kombinasi item yang memenuhi syarat minimum dari nilai support dalam database, yang dapat dirumuskan sebagai berikut:

$$\text{Support}(A \cap B) = \frac{\sum \text{Transaksi yang mengandung A dan B}}{\sum \text{Transaksi}} * 100\%$$

b. Pembentukan aturan asosiatif

Setelah semua pola frekuensi tinggi ditemukan, barulah dicari aturan asosiatif yang memenuhi syarat minimum untuk *confidence* dengan menghitung *confidence* aturan *asosiatif* A → B dengan menggunakan rumus berikut:

$$\text{Confidence} = \frac{\sum \text{Transaksi yang mengandung A dan B}}{\sum \text{Transaksi mengandung A}} * 100\%$$

Contoh Kasus Dan Penyelesaiannya

Sebuah Apotik XXX memiliki banyak sekali Stok obatnya sehingga terkadang pemiliknya kesulitan di dalam pencarian obat yang diminta konsumen. Salah satunya adalah pasangan obat yang cocok antara satu dengan yang lainnya. Sebagai contoh kalau obat Batuk maka obatnya adalah “AAA” dan pasangannya sebagai obat tambahan yaitu obat “BBB”.

Analisis kebutuhan adalah suatu proses yang sistematik dari pengembangan kebutuhan. Analisis kebutuhan merupakan cara yang efektif untuk mengidentifikasi masalah-masalah yang muncul. Dalam analisis kebutuhan perlu dilakukan suatu perencanaan, pengumpulan data dan analisis data.

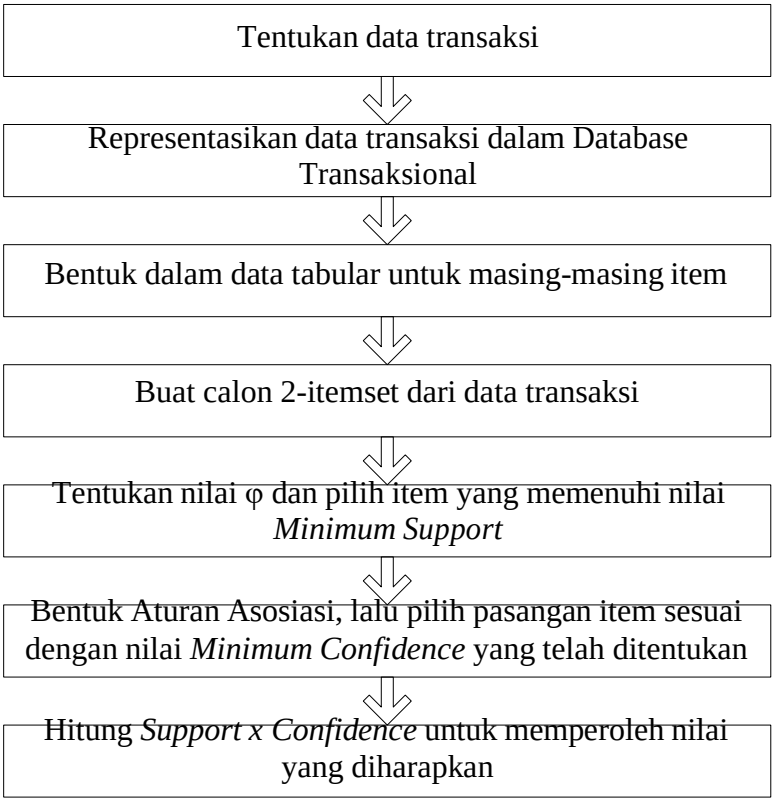
A. Analisis Data

Berikut ini daftar nama – nama obat generik yang ada di Apotek Puskesmas Kota Matsum Medan.

Tabel 11.1 Daftar Stok Obat Generik

No	Nama Obat
1	Acyclovir Cream
2	Acyclovir Tablet 200 mg
3	Acyclovir Tablet 400 mg
4	Albendazol Tablet 400 mg
5	Allupurinol Tablet 100 mg
6	Ambroxol Tablet 30 mg
7	Ambroxol Syrup
8	Aminophillin Injeksi 24 mg/ml – 10 ml
9	Aminophillin Tablet 30 mg
10	Amitriptillin Tablet Salut 25 mg
11	Amlodipin 5 mg Tablet
12	Amlodipin 10 mg Tablet
13	Amoksisilin 125 mg per Sach
14	Amoksisilin Kaplet 500 mg
15	Amoksisilin Kapsul 250 mg
16	Amoksisilin Sirup Kering 125 mg/5 ml
17	Ampicillin Serbuk Injeksi
18	Analgetika Tab
19	Antalgin Tablet 500 mg
...	
256	Zinc Tablet 20 mg

Sebelum membahas mengenai Analisis Data, terlebih dahulu dipaparkan sebuah diagram penyelesaian masalah menggunakan Algoritma *Apriori* pada gambar 11.1 untuk memudahkan pemahaman dan alur proses yang akan dibahas.



Gambar 11.1 Diagram Penyelesaian Masalah Apriori

Analisa data dilakukan setelah data terkumpul dan sesuai dengan kebutuhan sistem ini. Oleh sebab itu, untuk menghasilkan kesimpulan berdasarkan aturan (*rule*) pada analisis data, diperlukan data resep yang telah dibeli konsumen. Analisis data tersebut dilakukan berdasarkan teknik aturan asosiasi menggunakan Algoritma *Apriori* dengan beberapa iterasi atau langkah-langkah. Data yang diambil merupakan data – data resep obat pada tahun 2014 pada Apotek di Puskesmas Kota Matsum Medan. Data tersebut adalah data sampel dari data resep sebanyak 20 transaksi resep dan dapat dilihat pada Tabel 6.1.

Tabel 11.2 Data – Data Resep Obat

Resep	Obat yang dibeli
1	Kotrimoksazol suspensi, Parasetamol syrup 120 mg, Obat Batuk Hitam (OBH), Vitamin B Compleks tab
2	Tetrasiklin HCL kapsul 500 mg, Antalgin tablet 500 mg, Vitamin B Compleks tab
3	Parasetamol syrup 120 mg, Benedict, Tetrasiklin HCL kapsul 500 mg
4	Parasetamol syrup 120 mg, Cetirizin tablet 10 mg, Gliseril guaikolat tablet 100 mg, Vitamin B Compleks tab
5	Antalgin tablet 500 mg, Vitamin B Compleks tab, Tetrasiklin HCL kapsul 500 mg
6	Amoksisilin kaplet 500 mg, Parasetamol tablet 100 mg, Dexametason tablet 0,5 mg, Asam askorbat (Vit C) tablet 50 mg
7	Amoksisilin kaplet 500 mg, Vitamin B Compleks tab, Antalgin tablet 500 mg
8	Cetirizin tablet 10 mg, Dexametason tablet 0,5 mg, Asam askorbat (Vit C) tablet 50 mg
9	Parasetamol syrup 120 mg, Obat Batuk Hitam (OBH), Cetirizin tablet 10 mg, Amoksisilin kaplet 500 mg
10	Amoksisilin kaplet 500 mg, Asam askorbat (Vit C) tablet 50 mg , Parasetamol syrup 120 mg
11	Amoksisilin kaplet 500 mg, Parasetamol syrup 120 mg
12	Thiamin HCL (Vit B1) tablet 50 mg, Antalgin tablet 500 mg, Cetirizin tablet 10 mg
13	Parasetamol syrup 120 mg, Domperidon sirup, Thiamin HCL (Vit B1) tablet 50 mg, Cetirizin tablet 10 mg
14	Parasetamol syrup 120 mg, Gliseril guaikolat tablet 100 mg, Vitamin B Compleks tab
15	Antalgin tablet 500 mg, Vitamin B Compleks tab
16	Ambroxol tablet 30 mg, Cetirizin tablet 10 mg, Asam askorbat (Vit. C) tablet 50 mg, Vitamin B Compleks tab
17	Aminophillin tablet 200 mg, Antalgin tablet 500 mg, Cetirizin tablet 10 mg, Vitamin B Compleks tab
18	Aminophillin tablet 200 mg, Antalgin tablet 500 mg, Thiamin HCL (Vit B1) tablet 50 mg, Metformin tablet 500 mg
19	Aminophillin tablet 200 mg, Metformin tablet 500 mg, Cetirizin tablet 10 mg, Vitamin B Compleks tab
20	Obat Batuk Hitam (OBH), Cetirizin tablet 10 mg, Asam askorbat (Vit C) tablet 50 mg

B. Pengelompokkan Daftar Produk Berdasarkan Data Transaksi

Dari Tabel 11.1 terdapat 20 resep yang masing - masing resep terdapat beberapa *item* yang dibeli oleh konsumen pada Apotek di Puskesmas Kota Matsum Medan. Untuk itu, *item* yang dibeli akan di hitung jumlah item yang keluar dari 20 resep dan akan di hitung nilai support nya untuk 1 *item set*.

Tabel 11.3 Calon 1 *itemset*

No	Nama Obat	Jumlah	Support
1	Kotrimoksazol suspense	1	0,05
2	Parasetamol syrup 120 mg	9	0,45
3	Obat Batuk Hitam (OBH)	3	0,15
4	Vitamin B Compleks tab	10	0,5
5	Tetrasiklin HCL kapsul 500 mg	3	0,15
6	Antalgin tablet 500 mg	7	0,35
7	Benedict	1	0,05
8	Cetirizin tablet 10 mg	9	0,45
9	Gliseril guaikolat tablet 100 mg	2	0,1
10	Amoksisilin kaplet 500 mg	5	0,25
11	Dexametason tablet 0,5 mg	2	0,1
12	Asam askorbat (Vit C) tablet 50 mg	5	0,25
13	Thiamin HCL (Vit B1) tablet 50 mg	3	0,15
14	Domperidon sirup	1	0,05
15	Ambroxol tablet 30 mg	1	0,05
16	Aminophillin tablet 200 mg	3	0,15
17	Metformin tablet 500 mg	2	0,1

C. Analisis Pola Frekuensi Tinggi

Dari Tabel 11.1 Mencari *Quantity* (Jumlah) diberikan nilai minimum Φ >= 2 dari 20 resep.

Langkah 1: Representasi Data Transaksi dalam Database Transaksional seperti berikut:

Dari tabel 11.3, selanjutnya menentukan Calon 2-*itemset* dari data transaksi dengan menggabungkan item-item secara erpasangan.

Tabel 11.4 Calon *2-itemset*

No	2 Kombinasi Item		Qty	No	2 Kombinasi Item		Qty
1	Kotrimoksazol	Parasetamol	1	**24**	**Parasetamol**	**Amoksisilin**	**4**
2	Kotrimoksazol	OBH	1	25	Parasetamol	Dexametason	1
3	Kotrimoksazol	Vit B Komp	1	**26**	**Parasetamol**	**Asam ask**	**2**
4	Kotrimoksazol	Tetrasiklin	1	27	Parasetamol	Thiamin	1
5	Kotrimoksazol	Antalgin	0	28	Parasetamol	Domperidon	1
6	Kotrimoksazol	Benedict	0	29	Parasetamol	Ambroxol	0
7	Kotrimoksazol	Cetirizin	0	30	Parasetamol	Aminophilin	0
8	Kotrimoksazol	Gliseril	0	31	Parasetamol	Metformin	0
9	Kotrimoksazol	Amoksisilin	0	32	OBH	Vit B Komp	1
10	Kotrimoksazol	Dexametason	0	33	OBH	Tetrasiklin	1
11	Kotrimoksazol	Asam ask	0	34	OBH	Antalgin	0
12	Kotrimoksazol	Thiamin	0	35	OBH	Benedict	0
13	Kotrimoksazol	Domperidon	0	**36**	**OBH**	**Cetirizin**	**2**
14	Kotrimoksazol	Ambroxol	0	37	OBH	Gliseril	0
15	Kotrimoksazol	Aminophilin	0	38	OBH	Amoksisilin	1
16	Kotrimoksazol	Metformin	0	39	OBH	Dexametason	0
17	**Parasetamol**	**OBH**	2	40	OBH	Asam ask	0
18	**Parasetamol**	**Vit B Komp**	3	41	OBH	Thiamin	0
19	Parasetamol	Tetrasiklin	1	42	OBH	Domperidon	0
20	Parasetamol	Antalgin	0	43	OBH	Ambroxol	0
21	Parasetamol	Benedict	1	44	OBH	Aminophilin	0
22	**Parasetamol**	**Cetirizin**	**3**	45	OBH	Metformin	0
23	Parasetamol	Gliseril	1	**46**	**Vit B Komp**	**Tetrasiklin**	**2**
47	**Vit B Komp**	**Antalgin**	**5**	70	Tetrasiklin	Metformin	0
48	Vit B Komp	Benedict	0	71	Antalgin	Benedict	0
49	**Vit B Komp**	**Cetirizin**	**4**	**72**	**Antalgin**	**Cetirizin**	**2**
50	Vit B Komp	Gliseril	1	73	Antalgin	Gliseril	0
51	Vit B Komp	Amoksisilin	1	74	Antalgin	Amoksisilin	1
52	Vit B Komp	Dexametason	0	75	sntalgin	Dexametason	0

53	Vit B Komp	Asam ask	0	76	Antalgin	Asam ask	0
54	Vit B Komp	Thiamin	0	**77**	**Antalgin**	**Thiamin**	**2**
55	Vit B Komp	Domperidon	0	78	Antalgin	Domperidon	0
56	Vit B Komp	Ambroxol	1	79	Antalgin	Ambroxol	0
57	**Vit B Komp**	**Aminophilin**	**2**	**80**	**Antalgin**	**Aminophilin**	**2**
58	Vit B Komp	Metformin	1	81	Antalgin	Metformin	1
59	**Tetrasiklin**	**Antalgin**	**2**	82	Benedict	Cetirizin	0
60	Tetrasiklin	Benedict	1	……	…….	……	…..
61	Tetrasiklin	Cetirizin	0	**136**	**Aminophilin**	**Metformin**	**2**

Langkah 2 :

Berdasarkan Tabel 6.4 berisi *item-item* dengan *Quantity* yang dimilikinya, selanjutnya cari L1={*large* 1-*itemset*} dengan memilih *item* yang memenuhi nilai *minimum support* ≥ 10% seperti pada tabel 3.5 kemudian dilakukan pencarian nilai *support* pada masing-masing item dengan rumus :

$$\textit{Support} = \frac{\text{Jumlah Transaksi mengandung A dan B}}{\text{Total Transaksi}} \text{ x } 100\%$$

Dari rumus diatas, nilai *Support* diperoleh dengan *sample* perhitungan sebagai berikut.

$$\textit{Support} = \frac{\text{Jumlah Transaksi mengandung Kotrimoksazol dan Parasetamol}}{\text{Total Transaksi}} \text{ x } 100\%$$

Sehingga,

Support = $\frac{1}{20}$ x 100 % = 5 %

Tabel 11.5: L1 *Large* 1-*itemset*

No.	Pola kombinasi 2 item	Qty	Support AnB (100%)
1	Parasetamol, Obat Batuk Hitam	2	10
2	Parasetamol, Vitamin B Kompleks	3	15
3	Parasetamol, Cetirizin	3	15
4	Parasetamol, Amoksisilin	4	20
5	Parasetamol, Asam askorbat	2	10
6	Obat Batuk Hitam, Cetirizin	2	10
7	Vitamin B Kompleks, Tetrasiklin	2	10
8	Vitamin B Kompleks, Antalgin	5	25
9	Vitamin B Kompleks, Cetirizin	4	20
10	Vitamin B Kompleks, Aminophilin	2	10
11	Tetrasiklin, Antalgin	2	10
12	Antalgin, Cetirizin	2	10
13	Antalgin, Thiamin	2	10
14	Antalgin, Aminophilin	2	10
15	Cetirizin, Asam askorbat	3	15
16	Cetirizin, Thiamin	2	10
17	Cetirizin, Aminophilin	2	10
18	Amoksisilin, Asam askorbat	2	10
19	Dexametason, Asam askorbat	2	10
20	Aminophilin, Metformin	2	10

D. Pembentukan Aturan Asosiasi (Association Rule)

Untuk mencari aturan asosiasi dari iterasi terhadap langkah-langkah yang dilakukan sebelumnya, kemudian akan dihitung nilai *confidence* dari setiap item yang terdapat pada L1 berdasarkan rumus berikut:

$$\text{Confidence} = P(A \rightarrow B) = \frac{\text{Jumlah Transaksi mengandung A dan B}}{\text{Jumlah Transaksi mengandung A}} \times 100\%$$

Dari rumus diatas, nilai *Confidence* diperoleh dengan *sample* perhitungan sebagai berikut.

$$\text{Confidence} = \frac{\text{Jumlah Transaksi mengandung Parasetamol dan OBH}}{\text{Jumlah Transaksi mengandung Parasetamol}}$$

Sehingga,

100%

Confidence = $\frac{2}{9}$ x 100 % = 22 %

Tabel 6.6 L1 dengan nilai C*onfidence*

No	Pola Kombinasi *2-itemsets*	*Confidence*	
1	Parasetamol ☆ Obat Batuk Hitam	2/9	22%
2	Parasetamol ☆Vitamin B Kompleks	3/9	33%
3	Parasetamol ☆ Cetirizin	3/9	33%
4	Parasetamol ☆ Amoksisilin	4/9	44%
5	Parasetamol ☆Asam askorbat	2/9	22%
6	**Obat Batuk Hitam ☆ Cetirizin**	**2/3**	**67%**
7	Vitamin B Kompleks ☆ Tetrasiklin	2/10	20%
8	**Vitamin B Kompleks ☆ Antalgin**	**5/10**	**50%**
9	Vitamin B Kompleks ☆ Cetirizin	4/10	40%
10	Vitamin B Kompleks ☆ Aminophilin	2/10	20%
11	**Tetrasiklin ☆ Antalgin**	**2/3**	**67%**
12	Antalgin ☆ Cetirizin	2/7	29%
13	Antalgin ☆ Thiamin	2/7	29%
14	Antalgin ☆ Aminophilin	2/7	29%
15	Cetirizin ☆ Asam askorbat	3/9	33%
16	Cetirizin ☆ Thiamin	2/9	22%
17	Cetirizin ☆ Aminophilin	2/9	22%
18	Amoksisilin ☆ Asam askorbat	2/5	40%
19	**Dexametason ☆ Asam askorbat**	**2/2**	**100%**
20	**Aminophilin ☆ Metformin**	**2/3**	**67%**

Selanjutnya dimisalkan nilai minimum *confidence* = 50%, maka aturan asosiasi yang mungkin terbentuk adalah sebagai berikut:

Tabel 6.7 Aturan Asosiasi (*Association Rule*)

No	Pola Kombinasi *2-itemsets*	*Support*	*Confidence*
1	**Obat Batuk Hitam ✯ Cetirizin**	**10%**	**67%**
2	**Vitamin B Kompleks ✯ Antalgin**	**25%**	**50%**
3	**Tetrasiklin ✯ Antalgin**	**10%**	**67%**
4	**Dexametason ✯ Asam askorbat**	**10%**	**100%**
5	**Aminophilin ✯ Metformin**	**10%**	**67%**

Dari tahap-tahap yang telah dilakukan di atas, maka item yang memenuhi *Support x Confidence* terbesar dan *minimum confidence* = 50% pada Tabel 6.7 pada tulisan yang bercetak tebal, sehingga berdasarkan aturan asosiasi yang terbentuk maka dapat diambil kesimpulan sebagai berikut:

1. **Jika** membutuhkan Obat Batuk Hitam **maka** akan dipasangkan bersama Citirizen dengan *support* 10% dan *confidence* 67%
2. **Jika** membutuhkan Antalgin **maka** akan dipasangkan bersama Vitamin B Kompleks dengan *support* 25% dan *confidence* 50%
3. **Jika** membutuhkan Dexametason **maka** dipasangkan bersama Asam Askorbat dengan *support* 10% dan *confidence* 100%
4. **Jika** membutuhkan Aminophilin **maka** dipasangkan bersama Metformin dengan *support* 10% dan *confidence* 67%

BAB 12

BAB V : METODE MULTIPLE REGRESSION

12.1 Pendahuluan

Metode Multiple Regression merupakan salah satu metode statistika yang dalam bidang data mining digunakan untuk metode estimasi. Adapun Kegunaan Analisis Regresi Linear Berganda (*multiple regression*) adalah untuk mengukur pengaruh antara lebih dari satu variabel prediktor (variabel bebas) terhadap variabel terikat. Adapun rumus yang digunakan dalam Multiple Regression atau Regresi Linier Berganda adalah:

Rumus:

$$Y = a + b_1X_1+b_2X_2+\ldots+b_nX_n$$

Y = variabel terikat

a = konstanta

b_1,b_2 = koefisien regresi

X_1, X_2 = variabel bebas

$$\sum Y = an + b_1\sum X_1 + b_2\sum X_2$$
$$\sum X_1Y = a\sum X_1 + b_1\sum X^2{}_1 + b_2\sum X_1X_2$$
$$\sum X_2Y = a\sum X_2 + b_1\sum X_1X_2 + b_2\sum X_2{}^2$$

Contoh Soal Dan Penyelesaiannya

Seorang Manajer Pemasaran deterjen merek “XXX” ingin mengetahui apakah Promosi dan Harga berpengaruh terhadap keputusan konsumen membeli produk tersebut?

Hipotesis:

Ho : $\beta_1 = \beta_2 = 0$, Promosi dan Harga tidak berpengaruh signifikan terhadap keputusan konsumen membeli deterjen merek "XXX".

Ha : $\beta_1 \neq \beta_2 \neq 0$, Promosi dan Harga berpengaruh signifikan terhadap keputusan konsumen membeli deterjen merek "XXX".

Tabel 12.1 Data Kasus

No. Responden	Promosi (X1)	Harga (X2)	Keputusan Konsumen (Y)
1	10	7	23
2	2	3	7
3	4	2	15
4	6	4	17
5	8	6	23
6	7	5	22
7	4	3	10
8	6	3	14
9	7	4	20
10	6	3	19
Jumlah	60	40	170

Tabel 12.2 : Tabel Bantu

No. Resp.	X_1	X_2	Y	X_1Y	X_2Y	X1X2	X_12	X_22
1	10	7	23	230	161	70	100	49
2	2	3	7	14	21	6	4	9
3	4	2	15	60	30	8	16	4
4	6	4	17	102	68	24	36	16
5	8	6	23	184	138	48	64	36
6	7	5	22	154	110	35	49	25
7	4	3	10	40	30	12	16	9
8	6	3	14	84	42	18	36	9
9	7	4	20	140	80	28	49	16
10	6	3	19	114	57	18	36	9
Jumlah	60	40	170	1122	737	267	406	182

Berdasarkan Rumus

$$\sum Y = an + b_1 \sum X_1 + b_2 \sum X_2$$
$$\sum X_1 Y = a \sum X_1 + b_1 \sum X^2_1 + b_2 \sum X_1 X_2$$
$$\sum X_2 Y = a \sum X_2 + b_1 \sum X_1 X_2 + b_2 \sum X_2^2$$

Maka,

170 = 10 a + 60 b_1 + 40 b_2…… (1)

1122 = 60 a + 406 b_1 + 267 b_2…….. (2)

737 = 40 a +267 b_1 + 182 b_2…….. (3)

Persamaan (1) dikalikan 6, persamaan (2) dikalikan 1:

1020 = 60 a + 360 b_1 + 240 b_2

35163 = 60 a + 406 b_1 + 267 b_2 _
-102 = 0 a + -46 b_1 + -27 b_2
-102 = -46 b_1-27 b_2…… (4)

Persamaan (1) dikalikan 4, persamaan (3) dikalikan 1:

680 = 40 a + 240 b_1 + 160 b2

737 = 40 a + 267 b_1 + 182 b2 _
-57 = 0 a + -27 b1 + -22 b2
-57 = -27 b1 – 22 b2…… (5)

Persamaan (4) dikalikan 27, persamaan (5) dikalikan 46:

-2754 = -1242 b1 - 729 b2

-2622 = -1242 b1 - 1012 b2 _

-132 = 0 b1 + 283 b2

b_2 = -132:283 = -0,466

Harga b_2 dimasukkan ke dalam salah satu persamaan (4) atau (5):

-102 = -46 b1- 27 (-0,466)

-102 = -46 b1+ 12,582

46 b1 = 114,582

b1 = 2,4909

Harga b_1 dan b_2 dimasukkan ke dalam persamaan 1:

170 = 10 a + 60 (2,4909) + 40 (-0,466)

170 = 10 a + 149,454 –

18,640 10 a = 170 –

149,454 + 18,640

a = 39,186 : 10 = 3,9186

Jadi:

a = 3,9186

b1 =

2,4909

b2 = -0,466

Keterangan:

a = konstanta
b1 = koefisien regresi X1
b2 = koefisien regresi X2
Persamaan regresi:

$Y = 3,9186 + 2,4909\ X_1 – 0,466\ X_2$

PENGUJIAN HIPOTESIS

Koefisien Korelasi Berganda (R)

$$R = \frac{b_1 \sum XY \; b_1 + {}_2\sum X \; Y_2}{\sum Y^2}$$

$$R = \frac{\frac{2,4909.1122 + -0,466.737}{3162}}{\frac{2794,7898 + -343,442}{3162}} =$$

$$= = 0,775252308$$

Koefisien Determinasi (R^2)

R^2 = $(0,775252308)^2$
= 0,60

F Hitung

$$F\ Hitung = \frac{R\ N\ k^2(\ -\ -1)}{k(1-R\ )}{}_2$$

$$= \frac{0,60(10-2-1)}{2(1-0,60)}$$

= 5,25

Ket:

K = jumlah variable bebas

F Tabel

Dk Pembilang = k
= 2
Dk Penyebut = n-k-1
= 10-2-1
= 7
F tabel = 4,74

Hipotesis

Ho : β1 = β2 = 0, Variabel Promosi Dan Harga Tidak Berpengaruh Signifikan Terhadap Keputusan Konsumen Membeli Deterjen Merek ”Xxx”

Ha : β1 ≠ β2 ≠ 0, Variabel Promosi Dan Harga Berpengaruh Signifikan Terhadap Keputusan
Konsumen Membeli Deterjen Merek ”Xxx”

Kriteria:

F hitung ≤ F tabel = Ho diterima

F hitung > F tabel = Ho ditolak, Ha diterima

F hitung (5,25) > F tabel (4,74) = Ho ditolak, Ha Diterima

Jadi, dapat disimpulkan bahwa Promosi dan Harga berpengaruh signifikan terhadap keputusan konsumen membeli deterjen merek “XXX”.

BAB 13

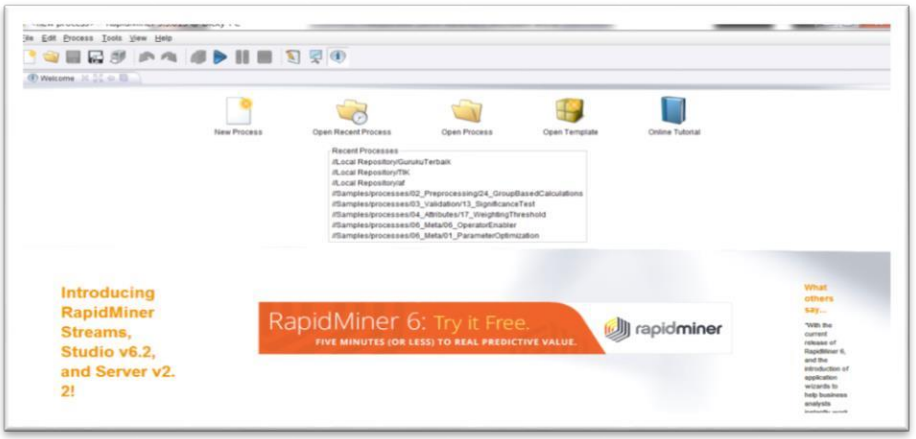
BAB XIII : PENGUJIAN DENGAN APLIKASI PENUNJANG

13.1 Pendahuluan

Pada Bab ini penulis mencoba untuk melakukan pengujian kasus menggunakan aplikasi perancangan dan perangkat lunak *enterprise*. Adapun pengujian Algoritma Diantaranya yaitu:

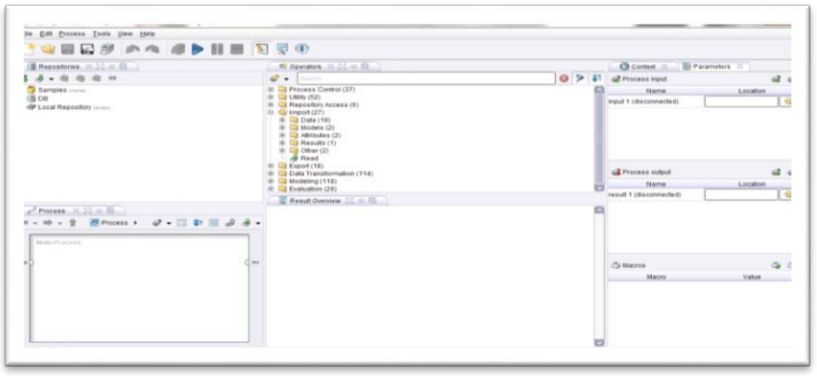
1. **Pengujian Algoritma C4.5**

Untuk pengujian algoritma C4.5 ini, dalam hal ini kita menggunakan *Rapid Miner 6.3*. Berikut ini adalah langkah-langkah pengujian menggunakan Rapid Miner. Jalankan perangkat lunak Rapid Minernya sehingga akan tampak seperti gambar di bawah ini:



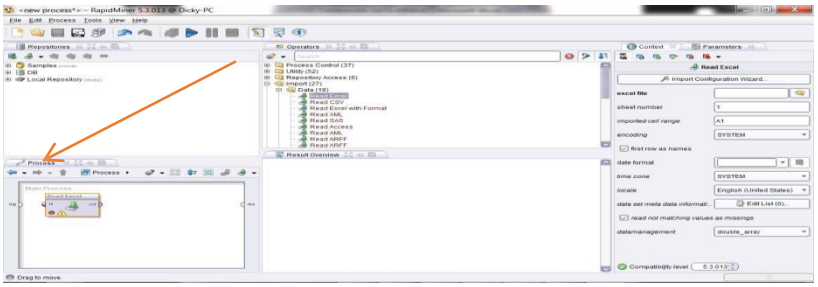
Gambar 13.1 : Tampilan Awal Rapid Miner

Setelah tampak awal *Rapid Miner*-nya maka untuk melakukan pengujian akan dilanjutkan dengan memilih opsi New Process karena kita melakukan proses baru seperti terlihat pada gambar di bawah ini.



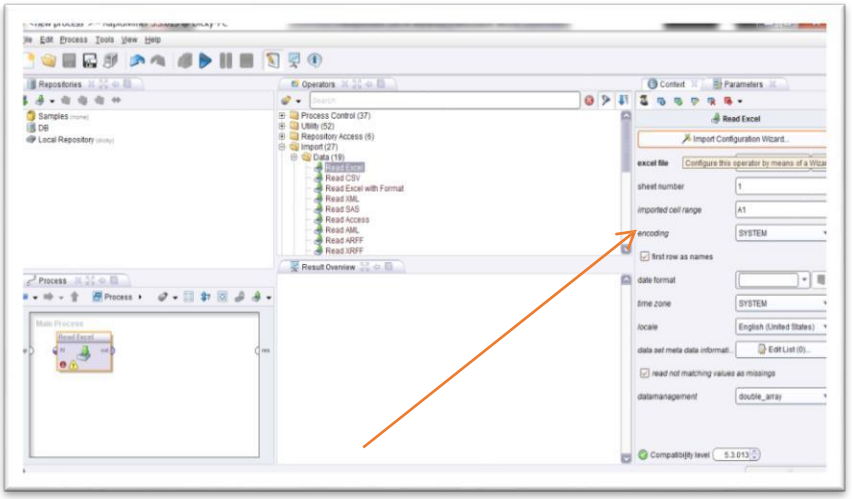
Gambar 13.2 : Tampilan Menu Utama Rapid Miner

1. Mengimport Data yang akan kita gunakan sebagai bahan untuk proses pengujian. (Pada Operator Klik Import-> Data->Read Excel) Kemudian *Drag Read Excel* ke *Main Process* seperti terlihat pada gambar di bawah ini.



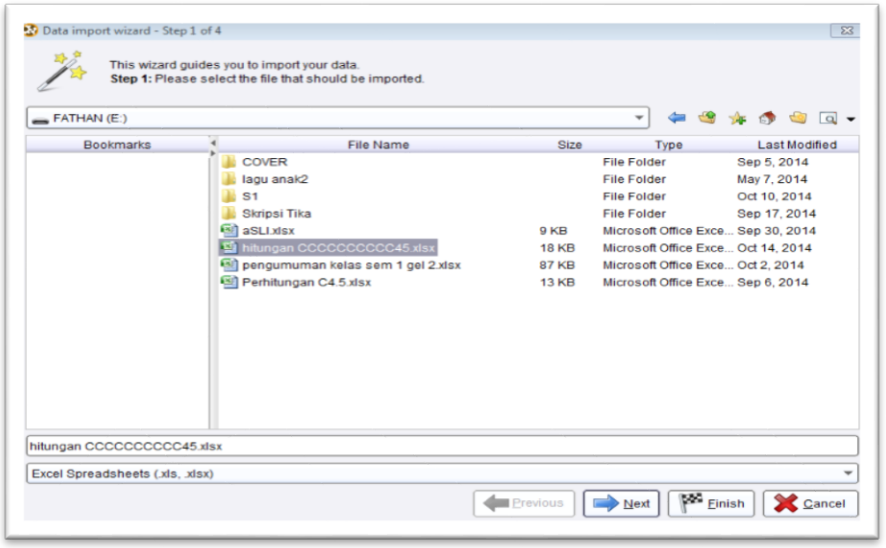
Gambar 13.3 : Proses Awal Untuk Mengimpor Data

2. Pada Menu Parameter ->Klik Menu *Import Configuration Wizard*



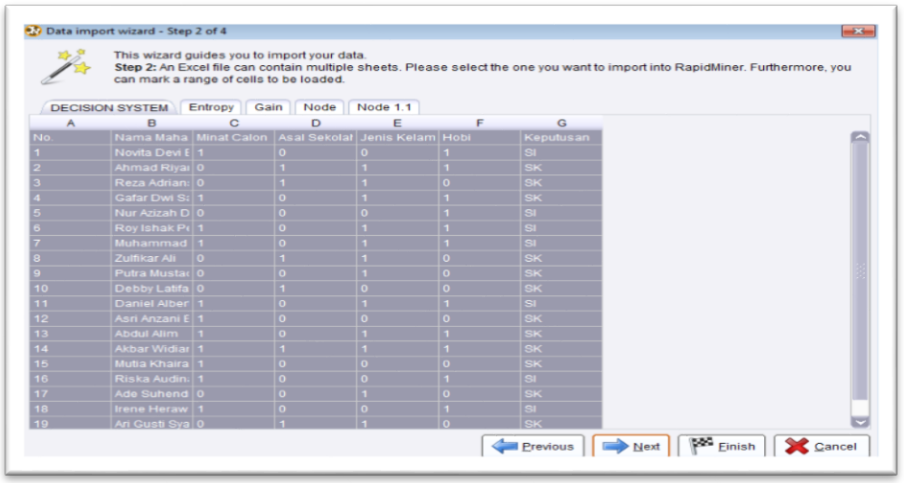
Gambar 13.4 : Import Configuration Wizard

Setelah itu langkah selanjutnya ada memilih data yang akan diproses. Dengan cara memilih *direktori* dimana letak data yang akan di gunakan sebagai sumber pengolahan yaitu:



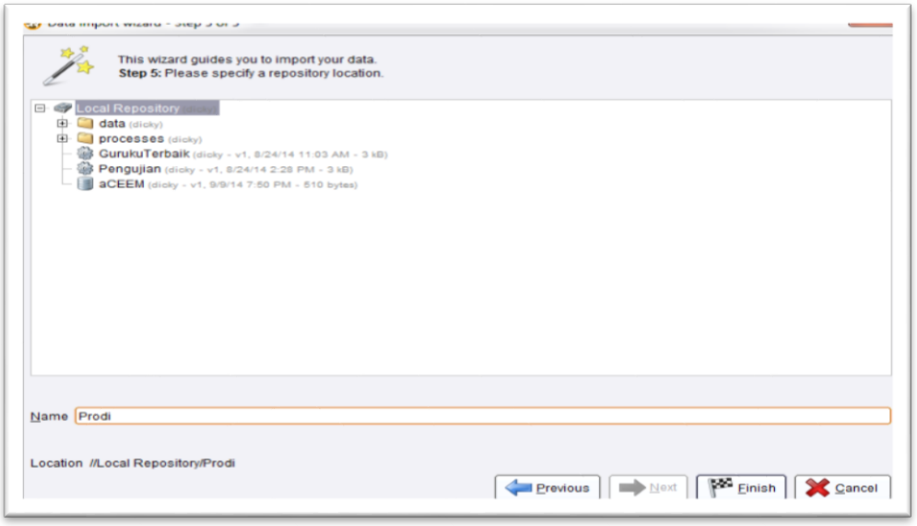
Gambar 13.5 : Direktori Sumber Data

Hasil dari pemilihan *direktori* dapat terlihat pada gambar di bawah ini:



Gambar 13.6: Decision System

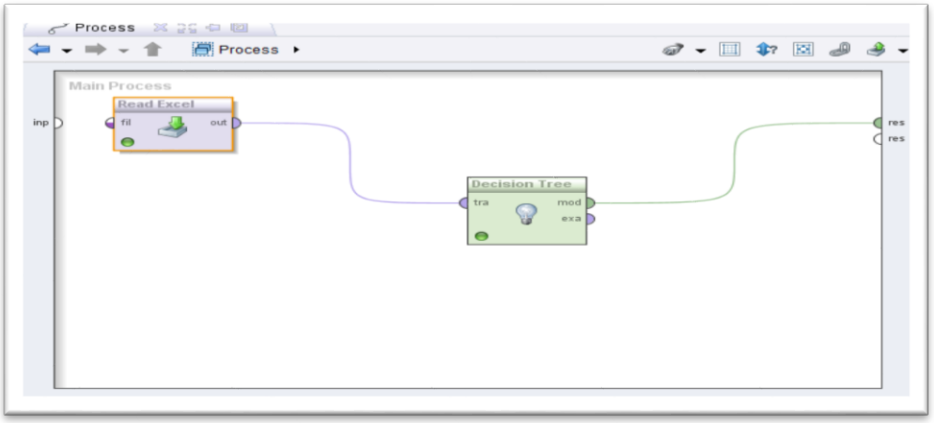
Gambar di atas menjelaskan tentang *Decision Sit*em yang akan diberikan operator untuk pengolahan selanjutnya. Setelah itu *tools* akan memberikan informasi untuk memasukkan *file name repository* dari datanya seperti terlihat pada gambar di bawah ini.



Gambar 13.7: Repositori Data

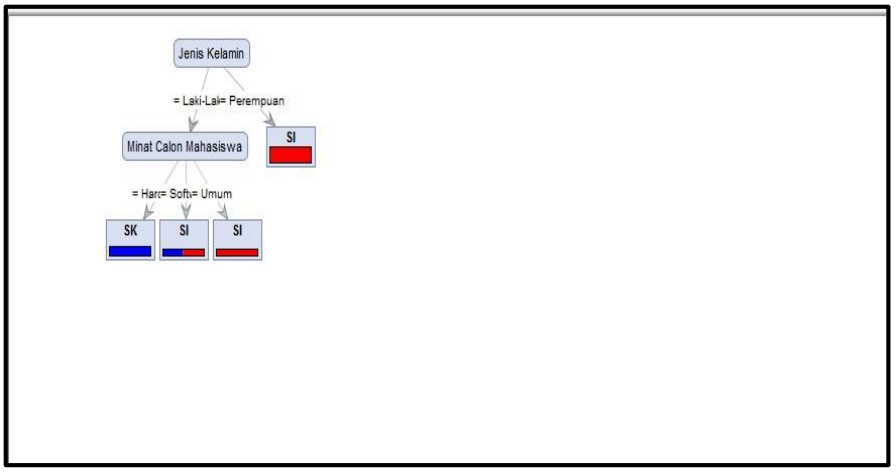
3. Menambahkan Operator

Berikut ini adalah hasil penambahan dari operator dan penyelesaian masalah menggunakan metode *Decision Tree* yaitu terlihat pada gambar di bawah ini.



Gambar 13.8: Pemrosesan Dan Penambahan Tree

Setelah di bentuk diagram dari proses *decision tree* yang terlihat pada gambar di atas langkah selanjutnya adalah melihat *tree view* atau pohon keputusan dari hasil penelitian seprti terlihat pada gambar di bawah ini:



Gambar 13.9 :Pohon Keputusan Berdasarkan Pengujian

Keterangan: warna merah adalah untuk menjelaskan nilai keputusan pemilihan jurusan sistem informasi sedangkan warna biru menjelaskan nilai keputusan pemilihan jurusan sistem komputer.

Berdasarkan gambar di atas pengetahuan baru (*knowledge)* yang ter bentuk dalam aturan baru (*rule*) yaitu sebagai berikut:

Tree

```
Jenis Kelamin = Laki-Laki
|   Minat Calon Mahasiswa = Hardware: SK {SK=3, SI=0}
|   Minat Calon Mahasiswa = Software: SI {SK=1, SI=1}
|   Minat Calon Mahasiswa = Umum: SI {SK=0, SI=2}
Jenis Kelamin = Perempuan: SI {SK=0, SI=7}
```

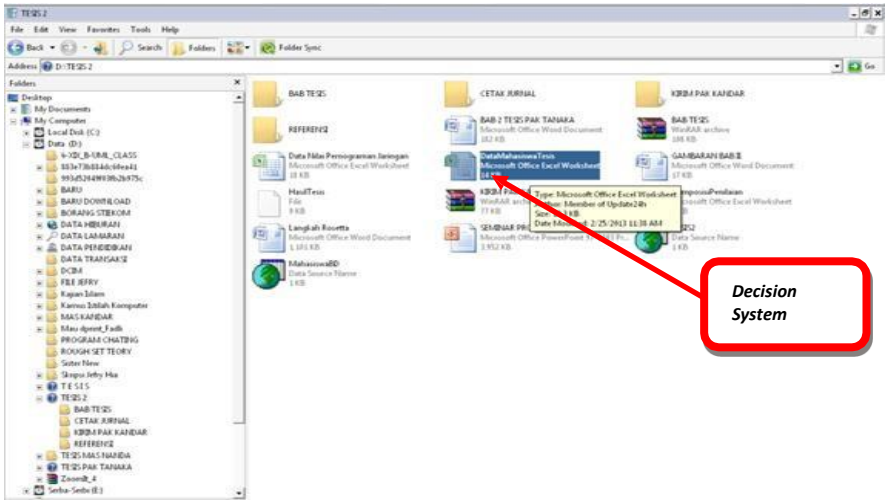
Gambar 13.10 :Hasil Pengujian

2. Pengujian Metode *Rough Set*

Proses pengujian dilakukan untuk mendapatkan data mengenai mahasiswa yang "Lulus" atau "Gagal" pada suatu mata kuliah dalam Program Studi Diploma 3(D3) STMIK xxx ini menggunakan alat bantu perangkat lunak *Rosetta Versi 1.4.41.*

Proses analisa terhadap mahasiswa yang dikategorikan "Lulus" atau "Gagal" pada suatu mata kuliah menggunakan metode *Rough Set,* dimana metode tersebut digunakan untuk menentukan mahasiswa yang lulus atau gagal pada suatu matakuliah pada Program Studi Diploma 3(D3) STMIK xxxx yang diambil sebagai sampel berdasarkan attribut dan karakteristik mahasiswa pada STMIK xxxxx khususnya pada mata kuliah Pemrograman Jaringan. Adapun tahapan-tahapan pengujian yaitu sebagai berikut:

1. Sebelum kita mengimplementasikan sistem terhadap data yang ingin diolah kita harus mempersiapkan terlebih dahulu *Decision System*-nya. Adapun decision system yang digunakan akan di-import terlebih dahulu yang berbasis Microsoft Excel.Setelah kita mentransformasi data sampel di atas ke *File Microsoft Excel* kemudian kita simpan data tersebut dengan nama *file* "DataMahasiswaTesis.xls" pada direktori tertentu seperti terlihat pada gambar di bawah ini.

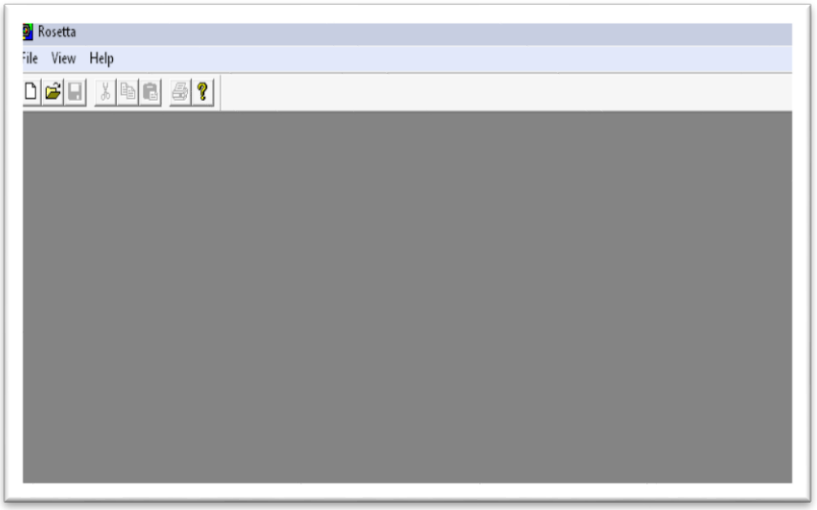


Gambar 13.11: File Yang Akan Di Proses

2. Setelah data tersebut dipersiapkan kemudian kita meng-*installRosetta 1.4.41*. Setelah kita melakukan penginstallan *Rosetta 1.4.41* tersebut, kemudian kita jalankan *tools* tersebut.

a. Menu Utama

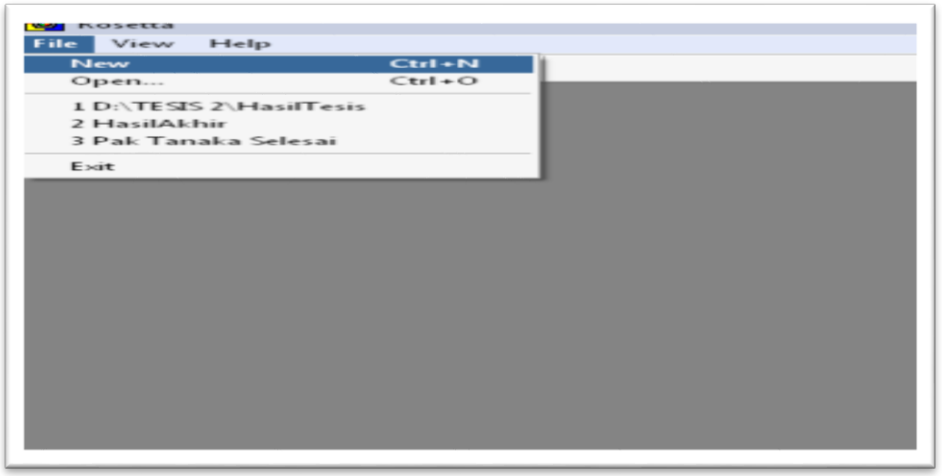
Menu utama dapat dikatakan sebagai antar muka (*user interface*) antara *user* dan program. Menu utama menampilkan pilihan menu yang tersedia pada program. Pada *Rosetta,* tersedia 5 pilihan menu yaitu menu *File, Edit, View, Window* dan *Help.* Gambar untuk tampilan menu utama dapat dilihat pada gambar 6.12.



Gambar 13.12 Tampilan Menu Utama

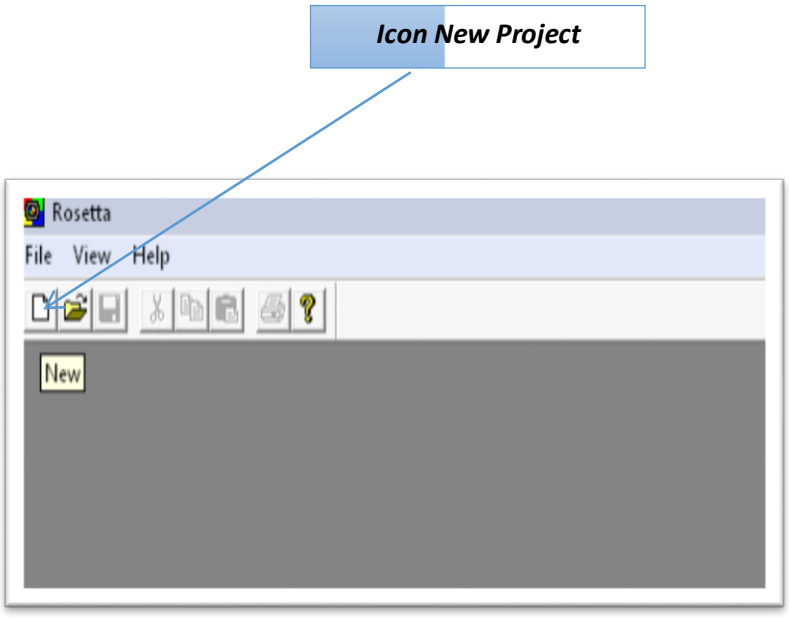
b. *FormNew Project*

Form New Project merupakan tempat di mana *project* akan dikerjakan. Di sinilah dimulai langkah-langkah pengerjaan untuk memperoleh *rules* yang dibutuhkan. Berikut adalah tampilan *Form New Project.*



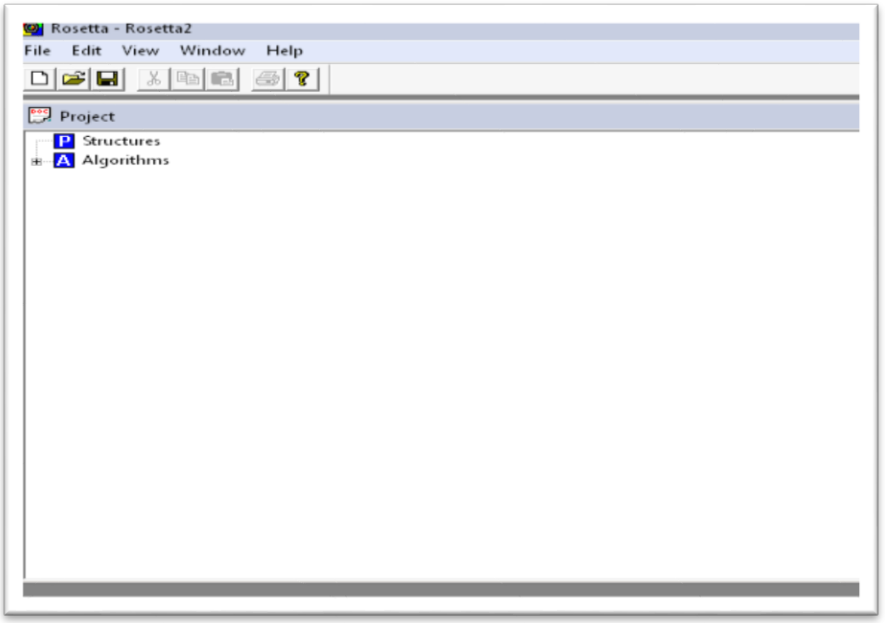
Gambar 13.13 *Create New Project*

Dengan cara lain yaitu dengan memilih *Icon Menu New Pada Toolbar Standard Rosetta 1.4.41* seperti terlihat pada gambar 13.13



Gambar 13.14 *Create New Project 2nd Way*

Setelah menambahkan *file* baru maka akan terlihat *window* awal dari Halaman Proses *Rosetta*-nya seperti terlihat pada gambar 6.15.



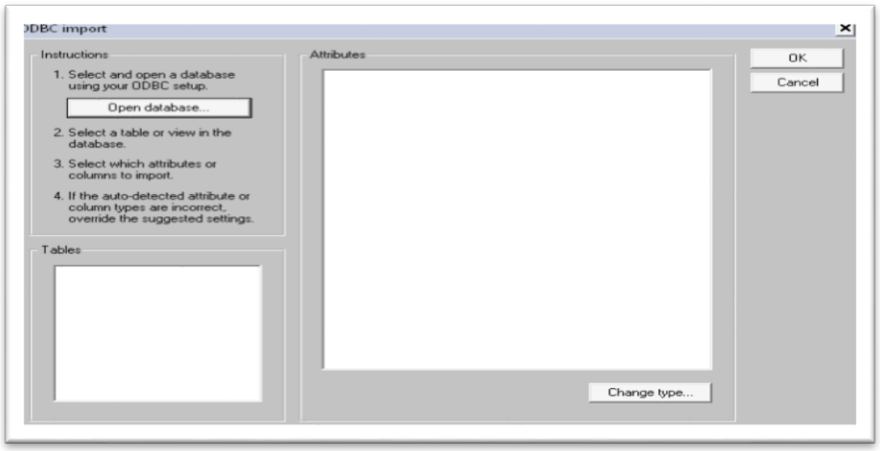
Gambar 13.15 Halaman *Project Rosetta1.4.41*

Di dalam pemrosesan awal kita terlebih dahulu harus menambahkan *Driver* sesuai dengan format *database* yang kita gunakan yaitu dengan cara "*Right-Click*" menu *Structures* kemudian pilih *ODBC* seperti terlihat pada gambar di 13.15



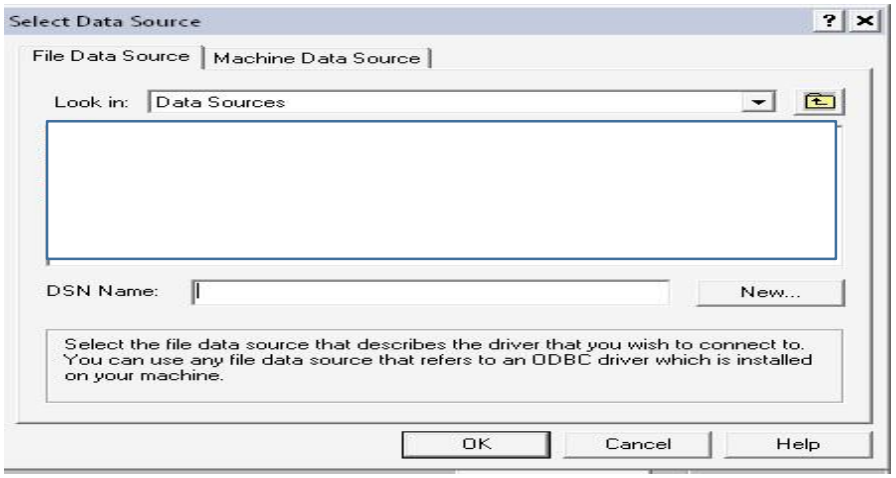
Gambar 13.16Langkah Awal Men-*Create ODBC*

Setelah kita memilih menu *ODBC* kemudian kita melakukan proses Impor *ODBC* seperti pada gambar13.7.



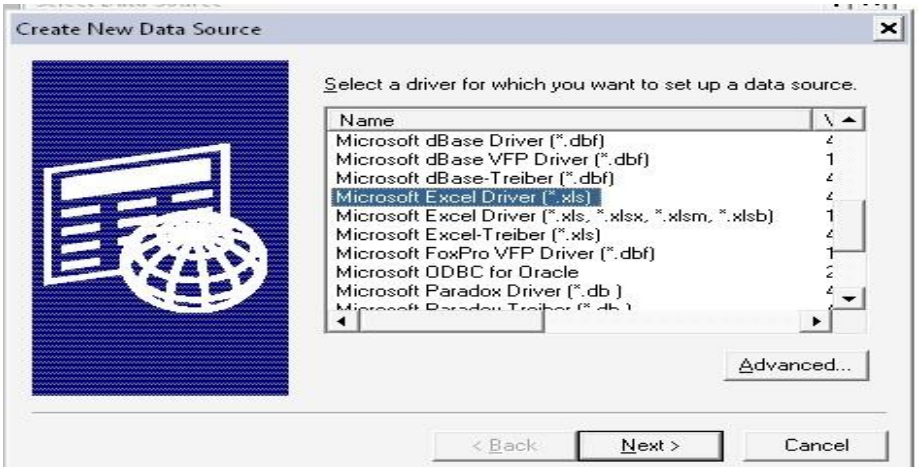
Gambar 13.17 Impor Drive ODBC

Pada gambar di atas menjelaskan mengenai proses Impor *Driver* ODBC-nya dengan meng-*ClickOpen Database*. Kemudian langkah selanjutnya kita akan memilihi *Data Source* yang akan kita gunakan. Apabila *Data Source*-nya telah ada kita *finding data source* tersebut pada direktori dimana kita *saving data source*tersebut. Sedangkan apabila *data source* tersebut belum pernah dilakukan pemrosesan, maka kita harus men-*Create* terlebih dahulu *Data Source*-nya. Adapun caranya yaitu dengan meng-*Click* tombol *New* pada *Window Select Data Source* seperti terlihat pada gambar 13.17.



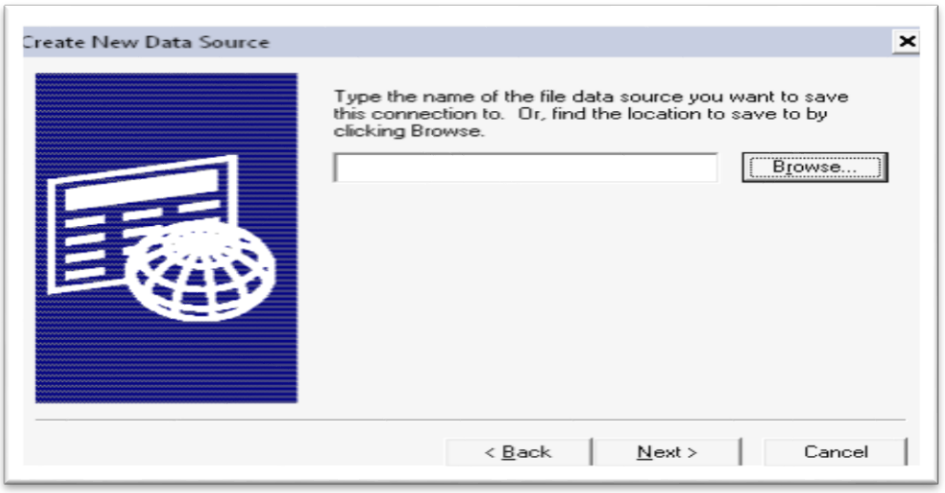
Gambar 13.18 *Window Select Data Source*

Setelah kita meng-*Create Data Source* yang akan kita proses, tahapan selanjutnya adalah memilih *driver* mana yang akan kita gunakan untuk men-*setupdata source*-nya. Ada beberapa *Driver Data Source* yaitu seperti *Microsoft ODBC for Oracle* yaitu untuk *database* yang berbasis *DBMS (Database Management System) Oracle, Microsoft Paradox Driver(*.db)* berfungsi untuk *database* yang berbasis *DBMS(Database Management System)Paradox* dan juga *Microsoft Excel Driver(*.xls)* dan *Microsoft Excel Driver(*.xls,*.xlsx, *.xlsm, *xlsb)* yaitu untuk *DBMS(Database Management System)* berbasis *Microsoft Excel* seperti terlihat pada gambar 13.18



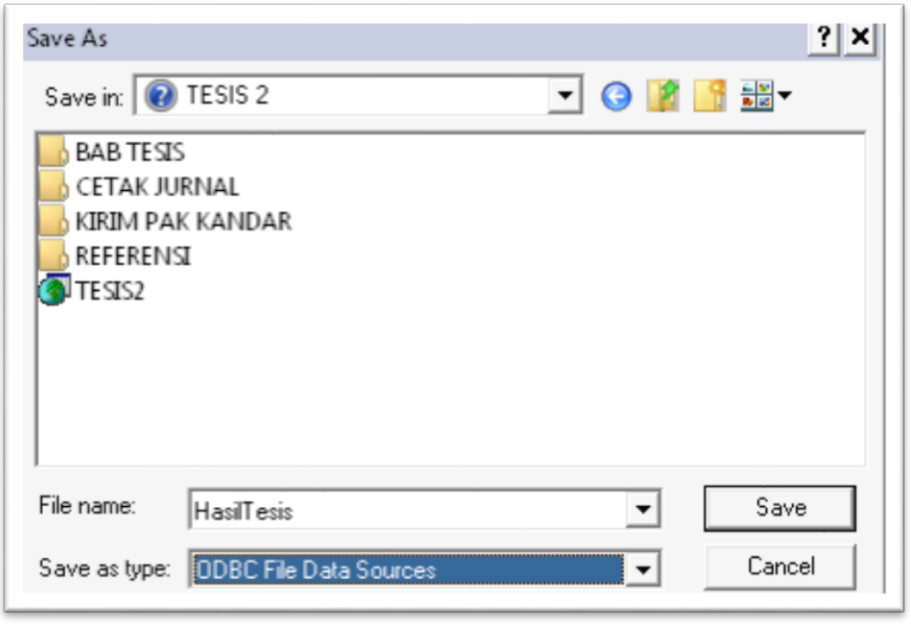
Gambar 13.19 *Window Of Create New Data Source*

Setelah kita memilih *Driver* dari *data source* yang kita gunakan selanjutnya kita tekan tombol *"Next"*. Setelah kita menekan tombol *Next* maka akan terlihat gambar 6.19



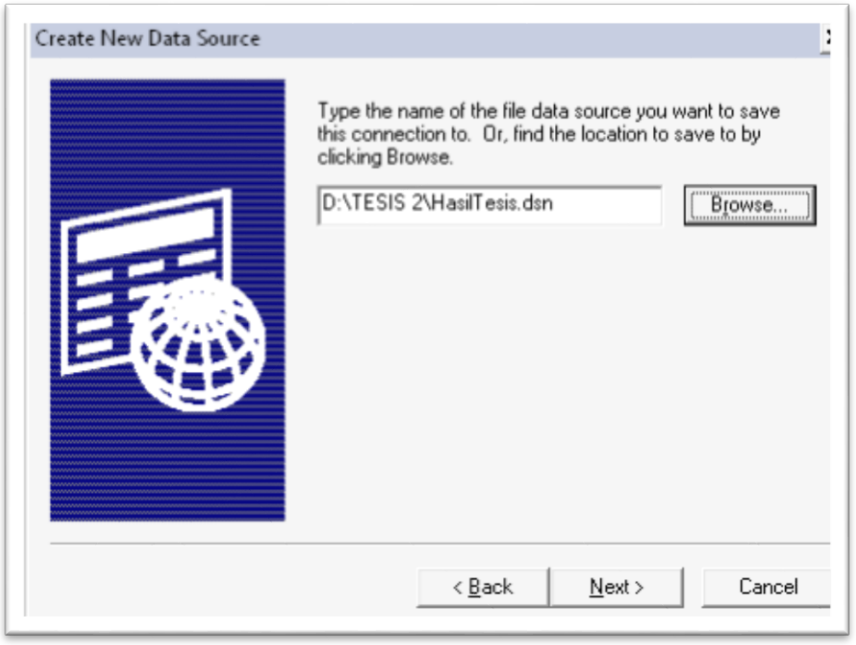
Gambar 13.20 *Window Of Browse New Data Source*

Setelah terlihat *window* seperti di atas kemudian langkah selanjutnya menekan tombol *"Next"* setelah kita menekan tombol *Next* maka akan muncul jendela seperti gambar 13.21.



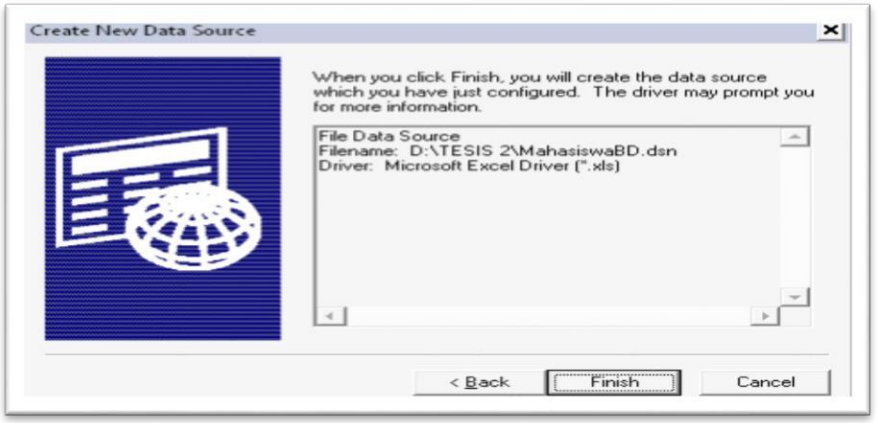
Gambar 13.21 Direktori Penyimpanan *Data Source*

Gambar di atas menggambarkan tentang lokasi penyimpanan *file rosetta*. Setelah melakukan proses penyimpanan langkah selanjutnya melakukan *Create New Data Source* seperti terlihat pada gambar 13.22



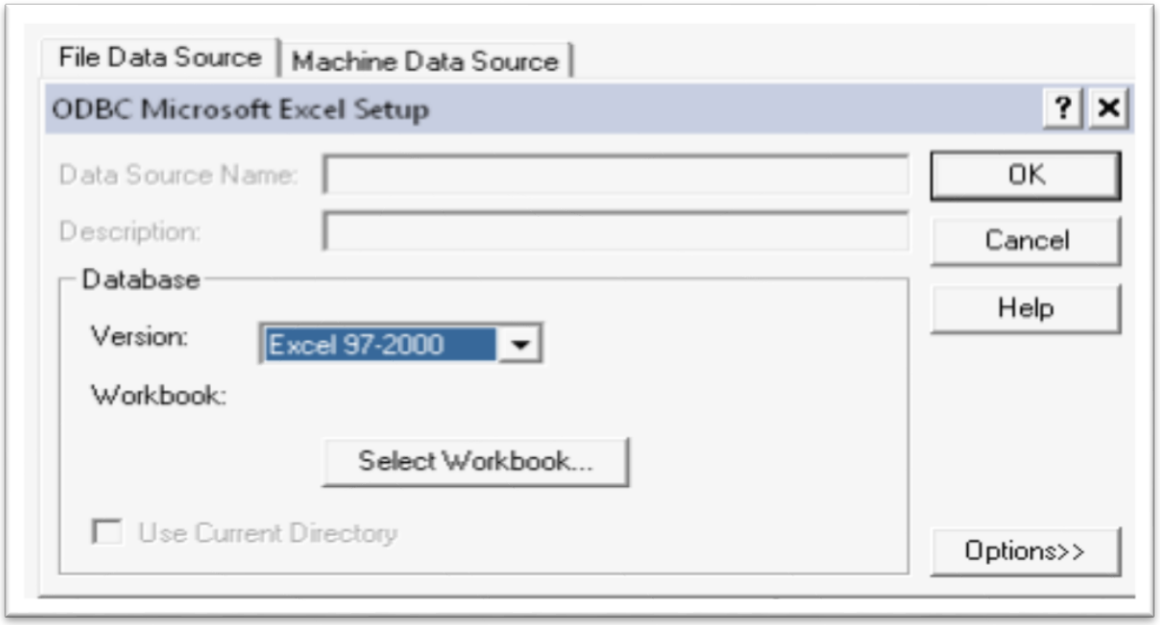
Gambar 13.22 Pemilihan *Data Source*

Setelah kita men-*create New Data Source* seperti terlihat pada gambar diatas, kemudian kita melanjutkan proses selanjutnya dengan menekan tombol "*Next"*. Setelah kita menekan tombol *"Next"* berarti kita telah menyelesaikan Konfigurasi *New Data Source*-nya, seperti terlihat pada gambar 13.23



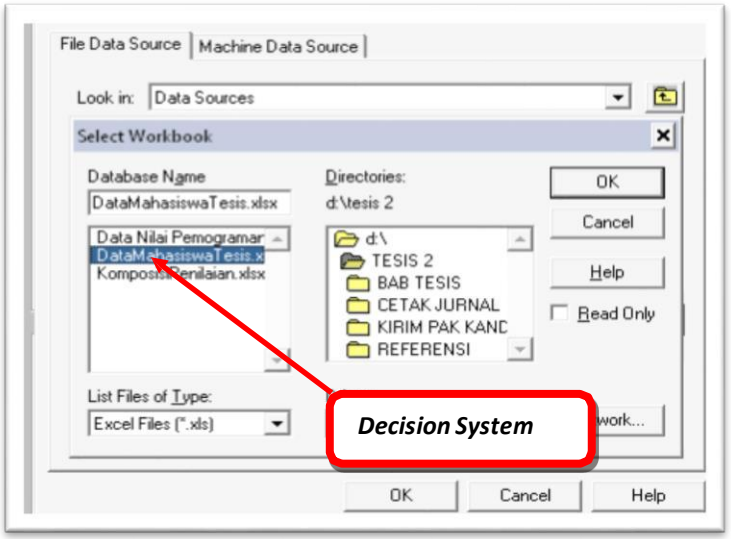
Gambar 13.23 Tahapan Akhir Proses Pemilihan *Data Source*

Gambar di atas menjelaskan bahwasanya kita telah menyelesaikan rangkaian proses *Create Data Source*. Tahap selanjutnya adalah kita Men-*Set-up ODBC Driver* kita. Penulis di sini menggunakan *DBMS* berbasis *Microsoft Excel 2007* maka window yang terlihat seperti gambar 13.24



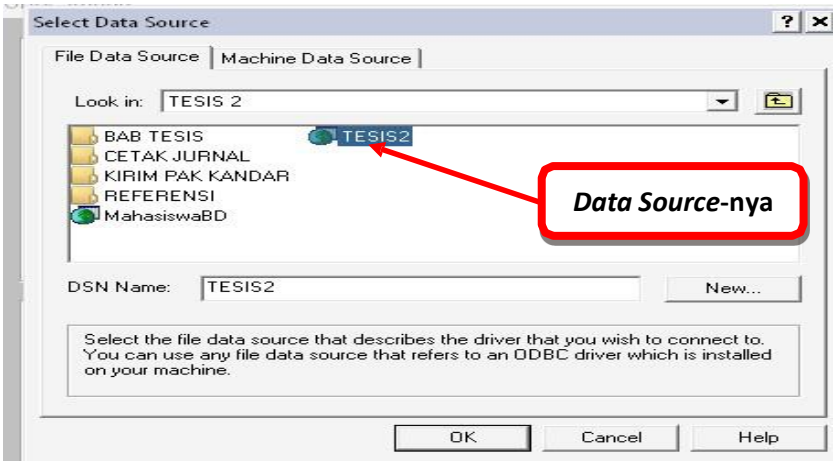
Gambar 13.24 *ODBC Microsoft Excel Setup*

Selanjutnya kita menekan tombol “OK” untuk melakukan proses selanjutnya yaitu mencari di mana letak file *DBMS* yang akan kita gunakan pada *Directory* penyimpanan *file*-nya, seperti terlihat pada gambar 6.25.



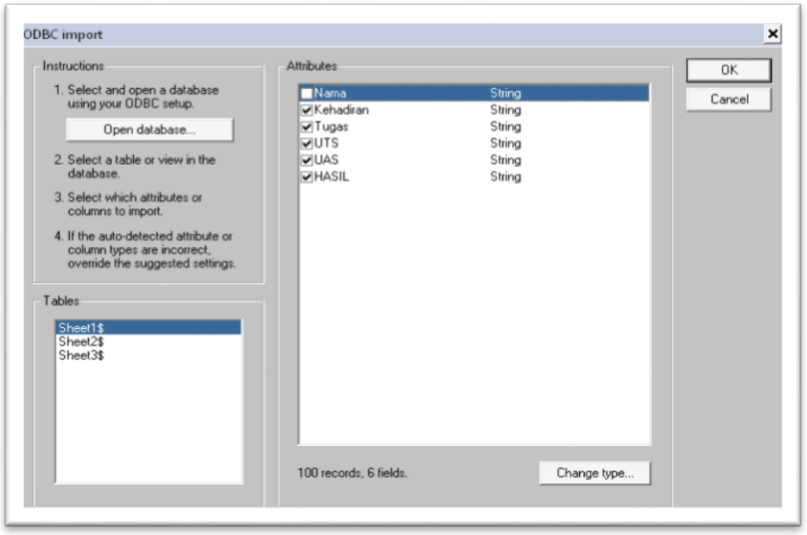
Gambar 13.25 Pemilihan *File Decision System*

Pada penelitian ini, penulis menggunakan "DataMahasiswaTesis.xls" sebagai *data processing*. Untuk tahapan selanjutnya kita mencari dimana letak *Data Source* yang telah kita *Create* pada tahapan sebelumnya, seperti terlihat pada gambar 6.26.



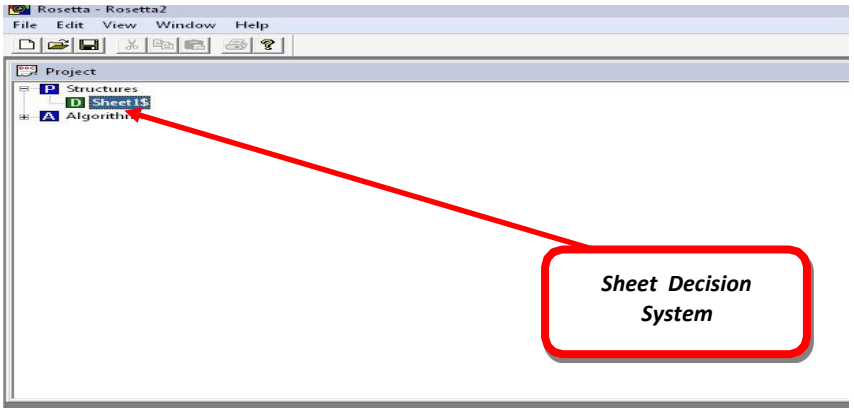
Gambar 13.26 Pemilihan *Data Source* Yang Telah Di *Create*

Data Source yang digunakan sesuai dengan gambar di atas yaitu dengan file name "TESIS2". Setelah rangkaian proses diatas selesai untuk selanjutnya adalah tahapan *Data Processing* berdasarkan metode yang digunakan di dalam menganalisa pemecahan masalah. Setelah selesai proses *Import ODBC* akan terlihat semua atribut yang digunakan di dalam proses *Rough Set*, seperti terlihat pada gambar 13.27.



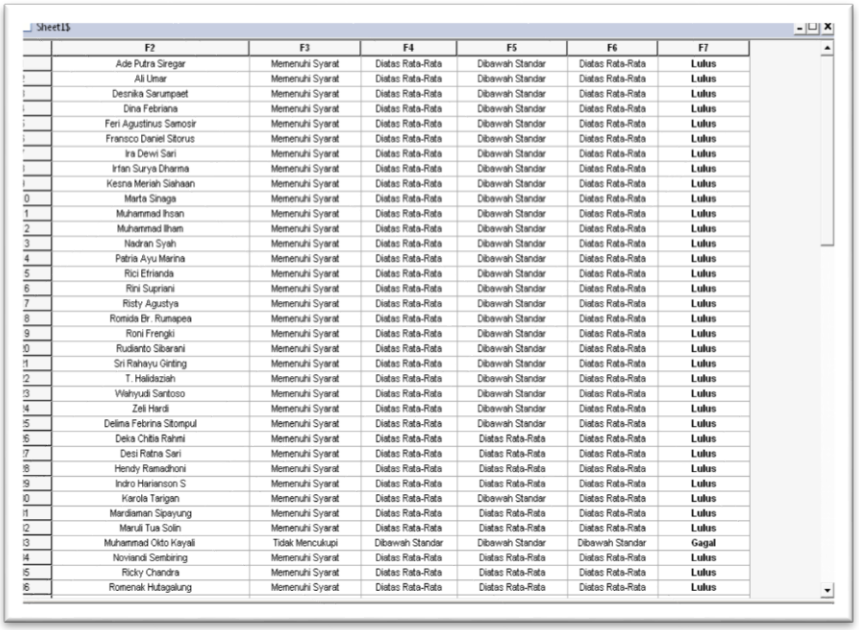
Gambar 13.27 Atribut *Decision System*

Gambar di atas menjelaskan tentang attribute yang digunakan dan berapa banyak sheet atau lembar kerja yang akan di proses. Kemudian langkah selanjutnya adalah menekan tombol “OK” dan akan muncul *window project* baru seperti terlihat pada gambar 13.27



Gambar 13.28 *Sheet Of Decision System*

Gambar di atas mendeskripsikan *Project* Baru dari *Rosetta*-nya dan menyatakan dengan benar data yang di impor telah masuk ke dalam sistem. Untuk melihat hasil data dari proses *import*nya, seperti terlihat pada gambar 6.29


Sheet1$

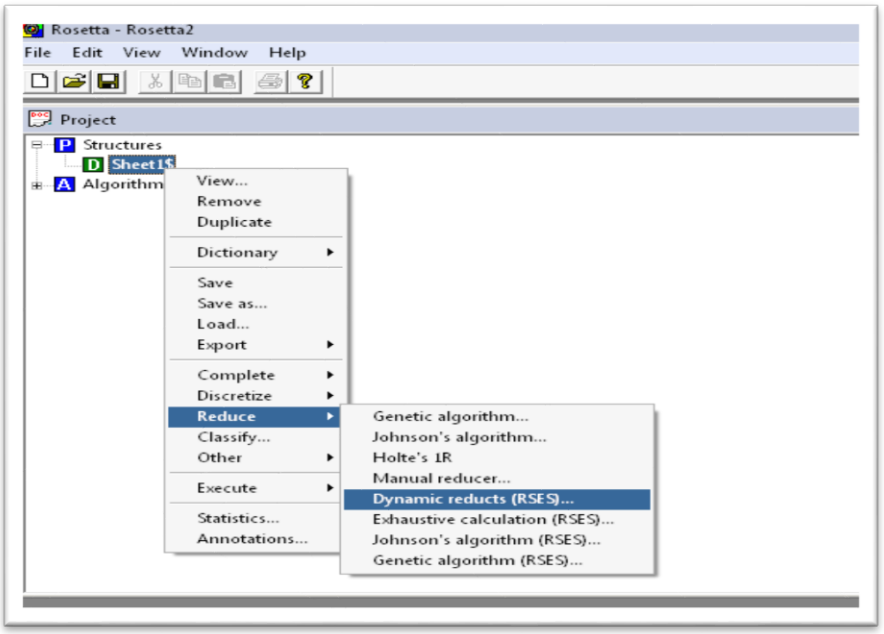
	F2	F3	F4	F5	F6	F7
	Ade Putra Siregar	Memenuhi Syarat	Diatas Rata-Rata	Dibawah Standar	Diatas Rata-Rata	Lulus
	Ali Umar	Memenuhi Syarat	Diatas Rata-Rata	Dibawah Standar	Diatas Rata-Rata	Lulus
	Desnika Sarumpaet	Memenuhi Syarat	Diatas Rata-Rata	Dibawah Standar	Diatas Rata-Rata	Lulus
	Dina Febriana	Memenuhi Syarat	Diatas Rata-Rata	Dibawah Standar	Diatas Rata-Rata	Lulus
	Feri Agustinus Samosir	Memenuhi Syarat	Diatas Rata-Rata	Dibawah Standar	Diatas Rata-Rata	Lulus
	Fransco Daniel Sitorus	Memenuhi Syarat	Diatas Rata-Rata	Dibawah Standar	Diatas Rata-Rata	Lulus
	Ira Dewi Sari	Memenuhi Syarat	Diatas Rata-Rata	Dibawah Standar	Diatas Rata-Rata	Lulus
	Irfan Surya Dharma	Memenuhi Syarat	Diatas Rata-Rata	Dibawah Standar	Diatas Rata-Rata	Lulus
	Kesna Meriah Siahaan	Memenuhi Syarat	Diatas Rata-Rata	Dibawah Standar	Diatas Rata-Rata	Lulus
0	Marta Sinaga	Memenuhi Syarat	Diatas Rata-Rata	Dibawah Standar	Diatas Rata-Rata	Lulus
1	Muhammad Ihsan	Memenuhi Syarat	Diatas Rata-Rata	Dibawah Standar	Diatas Rata-Rata	Lulus
2	Muhammad Ilham	Memenuhi Syarat	Diatas Rata-Rata	Dibawah Standar	Diatas Rata-Rata	Lulus
3	Nadran Syah	Memenuhi Syarat	Diatas Rata-Rata	Dibawah Standar	Diatas Rata-Rata	Lulus
4	Patria Ayu Marina	Memenuhi Syarat	Diatas Rata-Rata	Dibawah Standar	Diatas Rata-Rata	Lulus
5	Rici Efrianda	Memenuhi Syarat	Diatas Rata-Rata	Dibawah Standar	Diatas Rata-Rata	Lulus
6	Rini Supriani	Memenuhi Syarat	Diatas Rata-Rata	Dibawah Standar	Diatas Rata-Rata	Lulus
7	Risty Agustya	Memenuhi Syarat	Diatas Rata-Rata	Dibawah Standar	Diatas Rata-Rata	Lulus
8	Romida Br. Rumapea	Memenuhi Syarat	Diatas Rata-Rata	Dibawah Standar	Diatas Rata-Rata	Lulus
9	Roni Frengki	Memenuhi Syarat	Diatas Rata-Rata	Dibawah Standar	Diatas Rata-Rata	Lulus
0	Rudianto Sibarani	Memenuhi Syarat	Diatas Rata-Rata	Dibawah Standar	Diatas Rata-Rata	Lulus
1	Sri Rahayu Ginting	Memenuhi Syarat	Diatas Rata-Rata	Dibawah Standar	Diatas Rata-Rata	Lulus
2	T. Halidaziah	Memenuhi Syarat	Diatas Rata-Rata	Dibawah Standar	Diatas Rata-Rata	Lulus
3	Wahyudi Santoso	Memenuhi Syarat	Diatas Rata-Rata	Dibawah Standar	Diatas Rata-Rata	Lulus
4	Zeli Hardi	Memenuhi Syarat	Diatas Rata-Rata	Dibawah Standar	Diatas Rata-Rata	Lulus
5	Delima Febrina Sitompul	Memenuhi Syarat	Diatas Rata-Rata	Dibawah Standar	Diatas Rata-Rata	Lulus
6	Deka Chitia Rahmi	Memenuhi Syarat	Diatas Rata-Rata	Diatas Rata-Rata	Diatas Rata-Rata	Lulus
7	Desi Ratna Sari	Memenuhi Syarat	Diatas Rata-Rata	Diatas Rata-Rata	Diatas Rata-Rata	Lulus
8	Hendy Ramadhoni	Memenuhi Syarat	Diatas Rata-Rata	Diatas Rata-Rata	Diatas Rata-Rata	Lulus
9	Indro Hariansson S	Memenuhi Syarat	Diatas Rata-Rata	Diatas Rata-Rata	Diatas Rata-Rata	Lulus
0	Karola Tarigan	Memenuhi Syarat	Diatas Rata-Rata	Dibawah Standar	Diatas Rata-Rata	Lulus
1	Mardiaman Sipayung	Memenuhi Syarat	Diatas Rata-Rata	Diatas Rata-Rata	Diatas Rata-Rata	Lulus
2	Maruli Tua Solin	Memenuhi Syarat	Diatas Rata-Rata	Diatas Rata-Rata	Diatas Rata-Rata	Lulus
3	Muhammad Okto Kayali	Tidak Mencukupi	Dibawah Standar	Dibawah Standar	Dibawah Standar	Gagal
4	Novisandi Sembiring	Memenuhi Syarat	Diatas Rata-Rata	Diatas Rata-Rata	Diatas Rata-Rata	Lulus
5	Ricky Chandra	Memenuhi Syarat	Diatas Rata-Rata	Diatas Rata-Rata	Diatas Rata-Rata	Lulus
6	Romenak Hutagalung	Memenuhi Syarat	Diatas Rata-Rata	Diatas Rata-Rata	Diatas Rata-Rata	Lulus

Gambar 13.29 *View Of Data Decision System*

Gambar di atas mendeskripsikan *Decision System* yang akan kita proses pada tahapan selanjutnya. Terlihat pada gambar di atas yaitu attribut-attribut beserta *record-record* dari data mahasiswa STMIK XXXX sebagai sampel data di dalam pembuktian hasil analisa data pada bab sebelumnya.

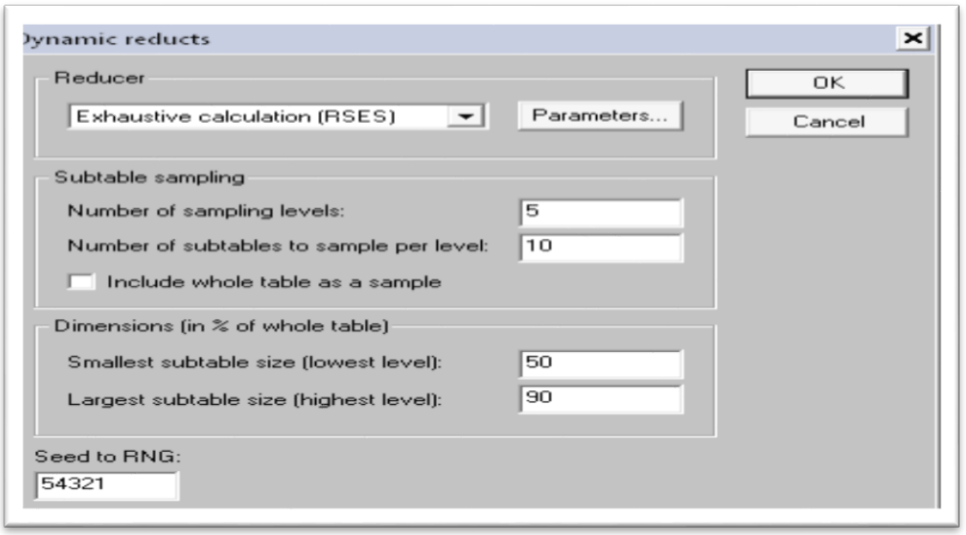
c. Reduction Process

Untuk tahapan selanjutnya di dalam rangkaian proses pemecahan masalah menggunakan metode *Rough Set*, sama halnya dengan analisa data manual, Rosetta juga melakukan proses *Reduction* seperti terlihat pada gambar 13.30



Gambar 13.30 *Process Of Dynamic Reduct*

Gambar di atas menjelaskan tentang proses *Reduction* dengan langkah-langkah yaitu "*Right Click*" pada *Data Sheet* atau *Decision System* kemudian Pilih *Reduce*☆*Dinamic Reduct (RSES)* dan akan terlihat *window* baru seperti gambar 13.31.



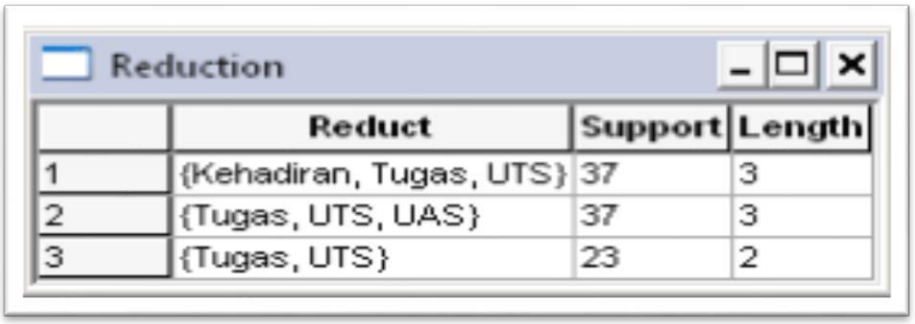
Gambar 13.32 *Window Of Dynamic Reduct*

Gambar di atas mendeskripsikan tentang proses *dynamic reducts*. Setelah itu tekan tombol "OK" dan akan terlihat penambahan 1(satu) proses pada *Project Rosetta*-nya yaitu *Reduction*, seperti terlihat pada gambar 13.33.



Gambar 13.33 *Icon Of Reduct Processing*

Untuk melihat hasil dari proses *Reduct* yaitu dengan cara "*Double Click*" *icon Reduction* dan akan terlihat hasil dari *Data Reduct* seperti terlihat pada gambar 13.34.

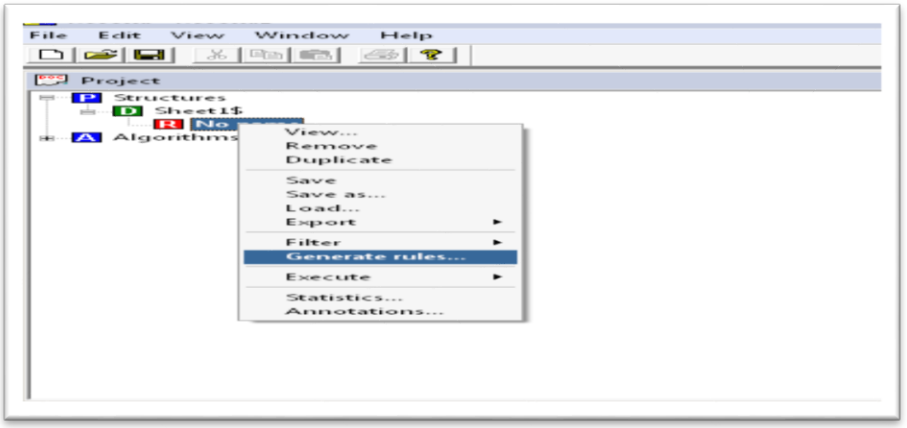


	Reduct	Support	Length
1	{Kehadiran, Tugas, UTS}	37	3
2	{Tugas, UTS, UAS}	37	3
3	{Tugas, UTS}	23	2

Gambar 13.34 *Result Of Reduction*

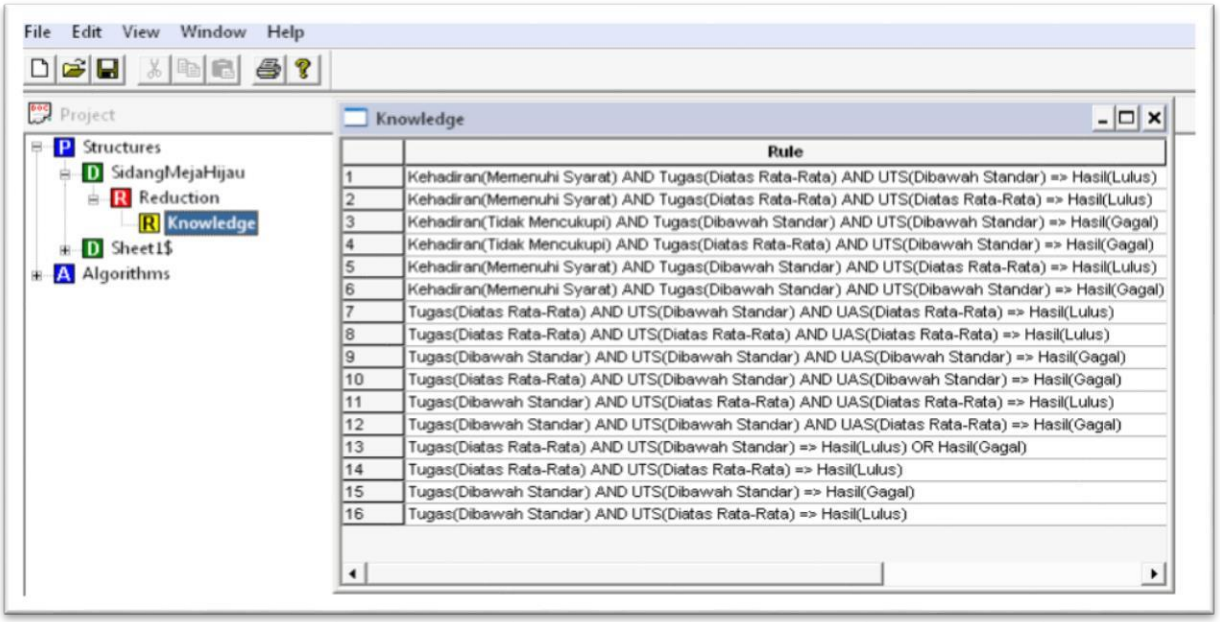
d. *General Rules*

Setelah melakukan proses *Reduction*, langkah terakhir adalah mencari *"General Rules"* atau keputusan dari *Information System* yang di proses dengan cara *"Right-Click" Icon Reduct*☆ Pilih *General Rules*, seperti terlihat pada gambar 13.35



Gambar 13.35 *Result Of Reduction*

Setelah melakukan proses di atas, pada *Project Rosetta*-nya akan terlihat penambahan *Icon"General Rules*-nya" dan untuk melihat hasil atau dari *Decision System* yang di proses, terlihat pada gambar 13.36



	Rule
1	Kehadiran(Memenuhi Syarat) AND Tugas(Diatas Rata-Rata) AND UTS(Dibawah Standar) => Hasil(Lulus)
2	Kehadiran(Memenuhi Syarat) AND Tugas(Diatas Rata-Rata) AND UTS(Diatas Rata-Rata) => Hasil(Lulus)
3	Kehadiran(Tidak Mencukupi) AND Tugas(Dibawah Standar) AND UTS(Dibawah Standar) => Hasil(Gagal)
4	Kehadiran(Tidak Mencukupi) AND Tugas(Diatas Rata-Rata) AND UTS(Dibawah Standar) => Hasil(Gagal)
5	Kehadiran(Memenuhi Syarat) AND Tugas(Dibawah Standar) AND UTS(Diatas Rata-Rata) => Hasil(Lulus)
6	Kehadiran(Memenuhi Syarat) AND Tugas(Dibawah Standar) AND UTS(Dibawah Standar) => Hasil(Gagal)
7	Tugas(Diatas Rata-Rata) AND UTS(Dibawah Standar) AND UAS(Diatas Rata-Rata) => Hasil(Lulus)
8	Tugas(Diatas Rata-Rata) AND UTS(Diatas Rata-Rata) AND UAS(Diatas Rata-Rata) => Hasil(Lulus)
9	Tugas(Dibawah Standar) AND UTS(Dibawah Standar) AND UAS(Dibawah Standar) => Hasil(Gagal)
10	Tugas(Diatas Rata-Rata) AND UTS(Dibawah Standar) AND UAS(Dibawah Standar) => Hasil(Gagal)
11	Tugas(Dibawah Standar) AND UTS(Diatas Rata-Rata) AND UAS(Diatas Rata-Rata) => Hasil(Lulus)
12	Tugas(Dibawah Standar) AND UTS(Dibawah Standar) AND UAS(Diatas Rata-Rata) => Hasil(Gagal)
13	Tugas(Diatas Rata-Rata) AND UTS(Dibawah Standar) => Hasil(Lulus) OR Hasil(Gagal)
14	Tugas(Diatas Rata-Rata) AND UTS(Diatas Rata-Rata) => Hasil(Lulus)
15	Tugas(Dibawah Standar) AND UTS(Dibawah Standar) => Hasil(Gagal)
16	Tugas(Dibawah Standar) AND UTS(Diatas Rata-Rata) => Hasil(Lulus)

Gambar 13.36 Hasil Atau *Rule* dari Implementasi *Rough Set* Pada *Rosseta 1.4.41*

Setelah kita menyelesaikan keseluruhan tahapan, dari proses penginputan *Decision System* sampai terbentuknya *knowledge* baru yang dapat kita lihat pada output *General Rules*. Berikut ini adalah *output* yang dihasilkan pada proes di atas.

BAB 14

BAB XIV : PENGUJIAN ALGORITMA LANJUTAN

a. Pengujian Teknik Single Lingkage

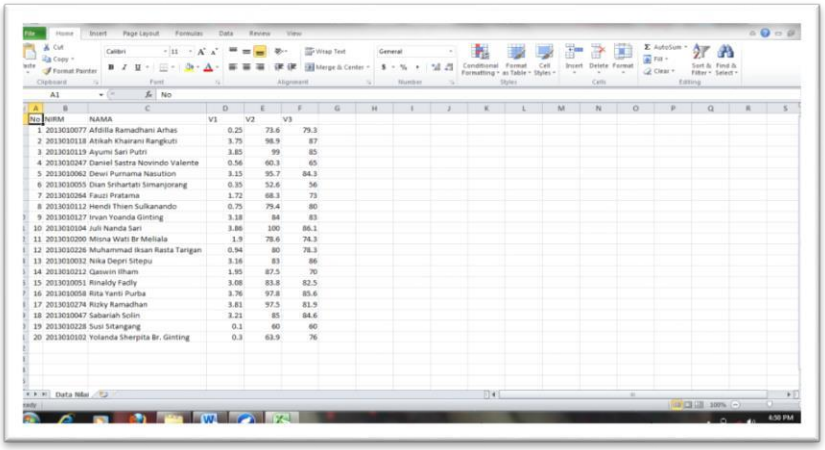
Dalam pengujian kasus ini penulis menggunakan Perangkat Lunak Weka 3.6.7. Tujuan dari pengujian ini apakah hasil analisa tersebut diatas sesuai dengan hasil pengujian perangkat lunaknya.

a. Masukan (*Input*) Data

Input data sangat berpengaruh di dalam pengisian data sehingga data yang dikehendaki dapat diterima oleh sistem dan juga diproses untuk pembuatan laporan yang dikehendaki. Adapun bentuk input tersebut adalah sebagai berikut:

1. Sumber Data

Proses *data mining* secara aplikatif dimana proses *data mining* yang akan dilakukan menggunakan *software data mining WEKA*. Langkah yang dilakukan adalah input data nilai IPK, nilai absensi dan nilai etika yang disimpan dengan *type Microsoft Excel Comma Separated Values File* (.csv). Tampilan data nilai IPK, nilai absensi dan nilai etika dapat dilihat pada gambar 14.1



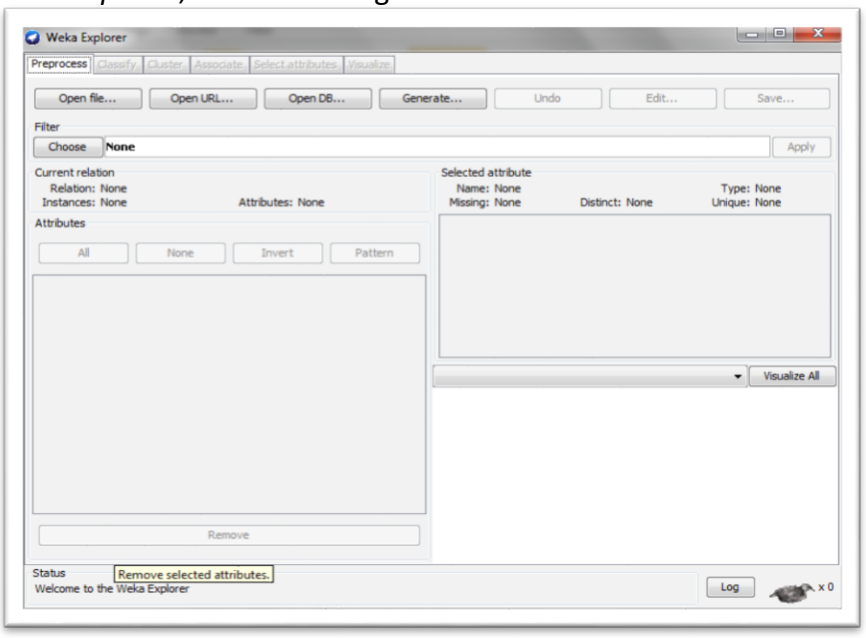
Gambar 14.1 Data Nilai Awal Mahasiswa

a. Buka program *WEKA* yang telah diinstal, akan muncul seperti gambar 6.38



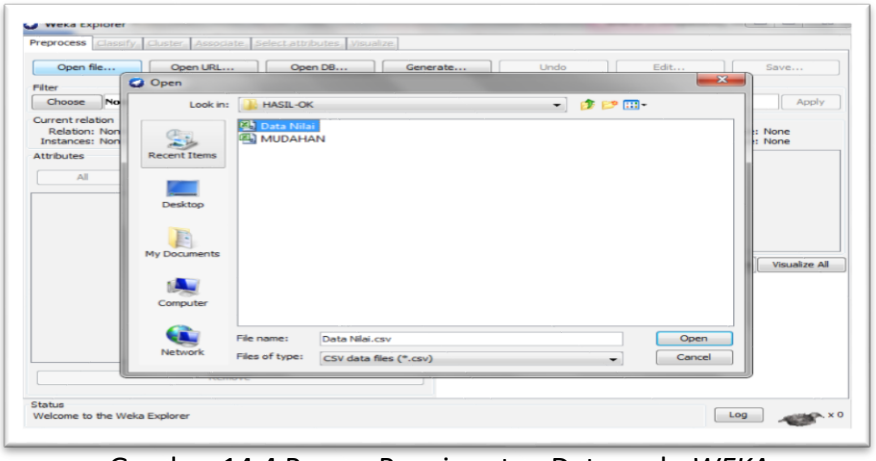
Gambar 14.2 Tampilan Awal Aplikasi *WEKA*

Kemudian klik *Explorer*, akan muncul gambar 6.39.



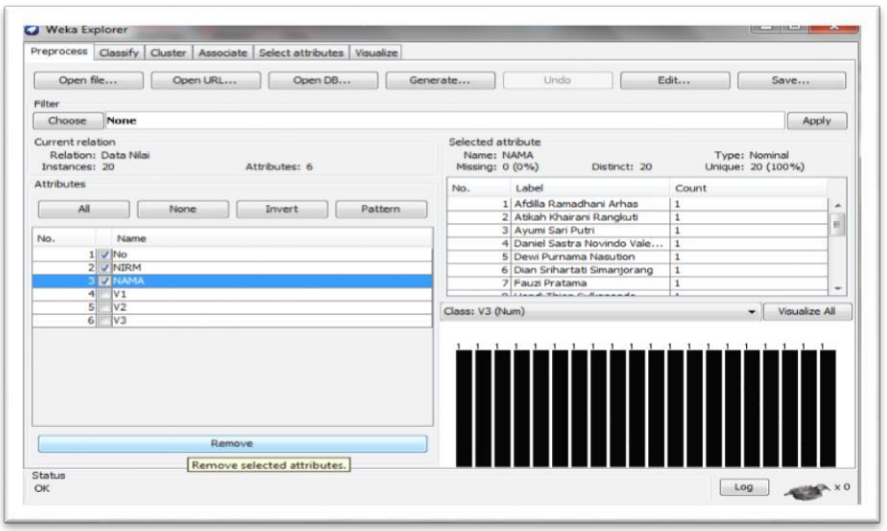
Gambar 14.3 Tampilan *WEKA Explorer*

b. Kemudian buka file data nilai mahasiswa yang sudah di save tadi dengan *typeMicrosoft ExcelCSV* untuk dimasukkan ke dalam aplikasi *WEKA* seperti pada gambar 14.4



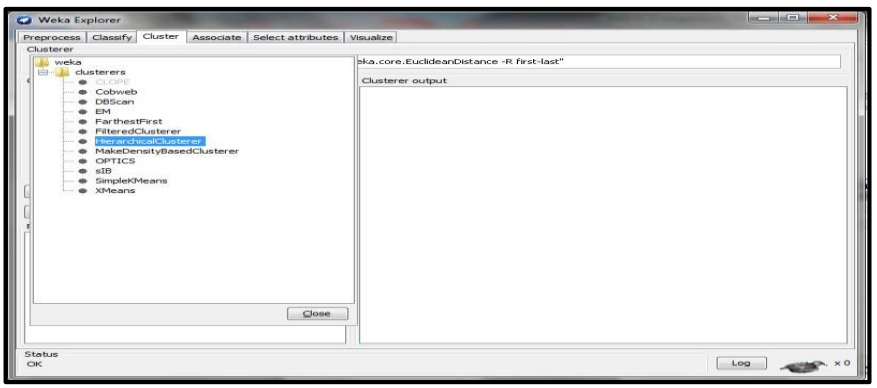
Gambar 14.4 Proses Penginputan Data pada *WEKA*

c. Kemudian pilih attributes V1, V2, V2, selanjutnya klik *remove*, seperti pada gambar 14.5



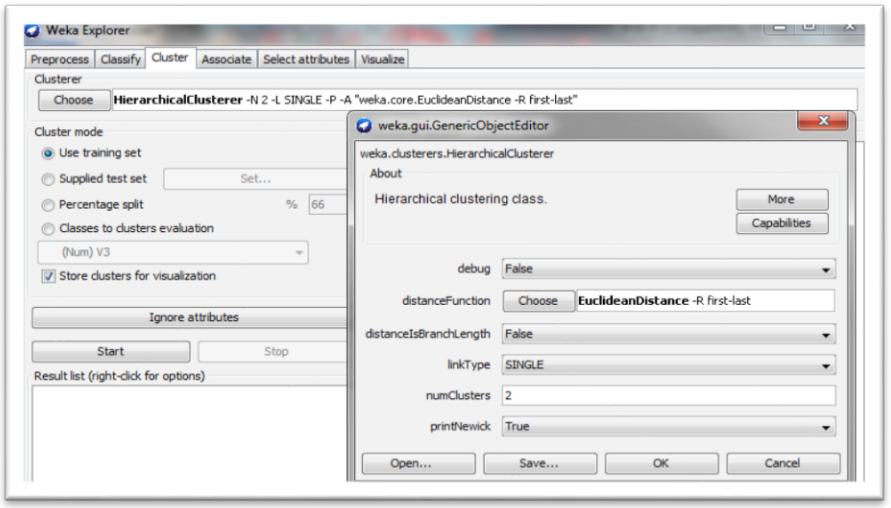
Gambar 14.5 Proses *Selected Attributes* pada *WEKA*

d. Kemudian klik menu *cluster*, pilih *choose* dan pilih *hierarchicalclusterer* seperti pada gambar 14.6



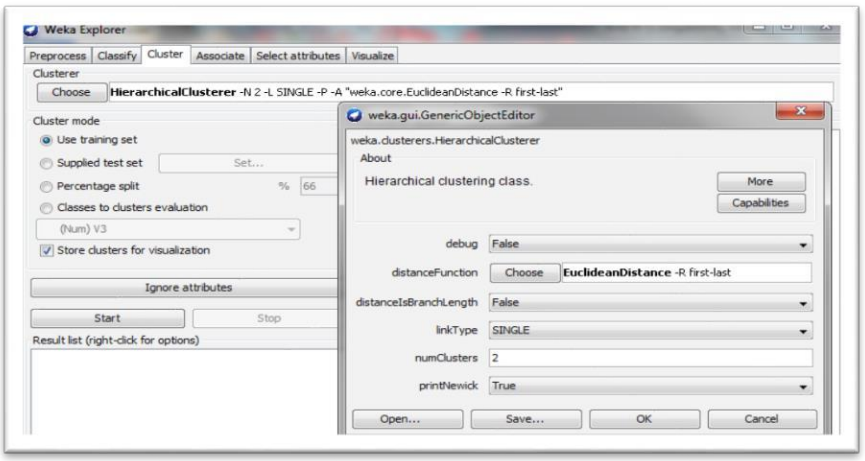
Gambar 14.6 : Proses Menentukan Metode Pada *Weka*

e. Maka langkah yang dilakukan selanjutnya yaitu *double* klik *hierarchicalclusterer* yang sudah dipilih, kemudian akan muncul seperti pada gambar 14.7



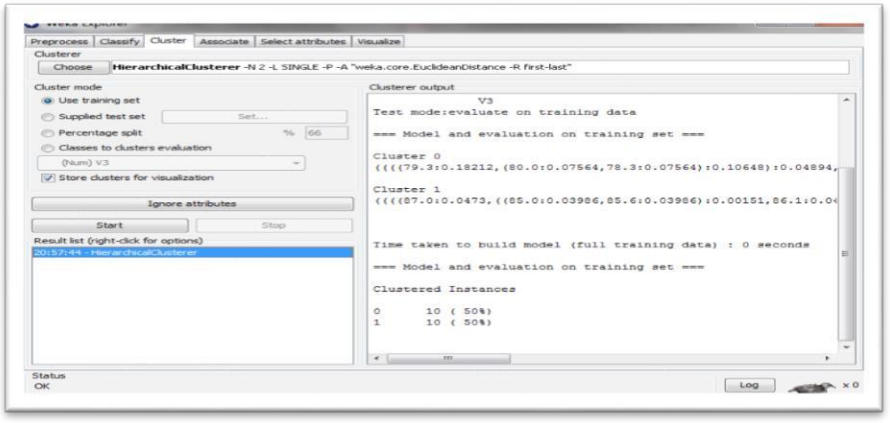
Gambar 14.7 Tampilan *WEKA* untuk Menentukan *EuclideanDistance*

f. Kemudian untuk menentukan jumlah *cluster* yang diinginkan yaitu dengan menentukan 2 *cluster* karena jumlah *cluster* menentukan hasil pembagian *cluster* dalam mengelompokkan mahasiswa yang masuk ke kelas unggulan dan mana yang tidak dan kemudian pilih *linkType single linkage* sesuai dengan teknik yang digunakan dalam analisis skripsi ini. Selanjutnya klik “OK”, seperti yang dilihat pada gambar 6.44



Gambar 14.8 Proses Menentukan *EuclideanDistance*

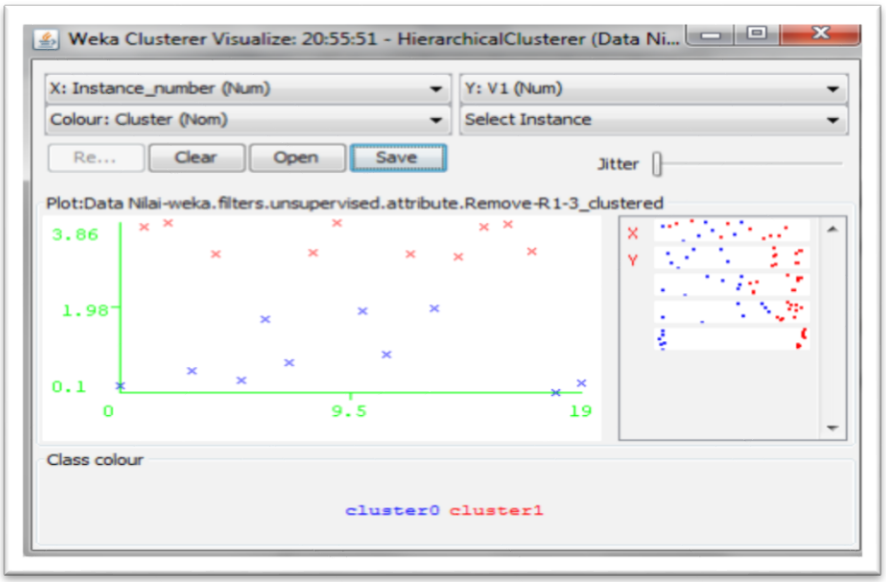
g. Kemudian klik *start*, dan kemudian akan muncul seperti yang dilihat pada gambar 14.9.



Gambar 14.9 Tampilan Hasil Analisa

Temuan di atas menghasilkan 10 mahasiswa yang masuk ke dalam *cluster* 0 (50%) dan 10 mahasiswa masuk ke dalam *cluster* 1 (50%).

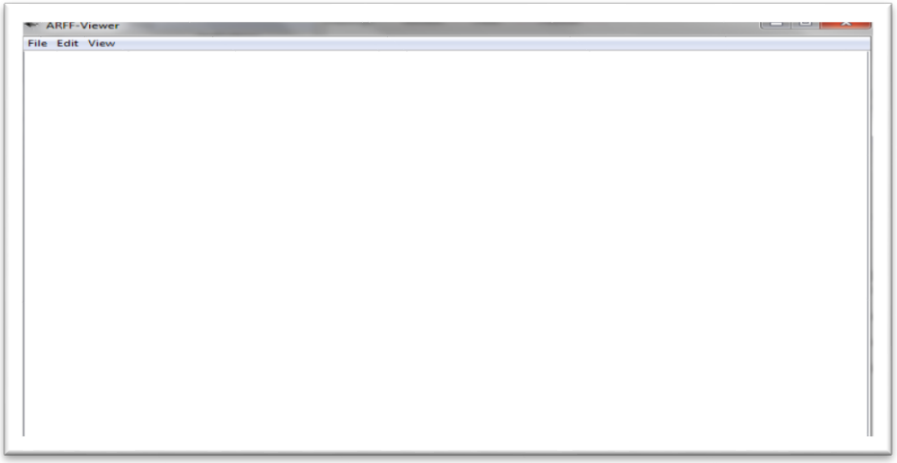
h. Selanjutnya klik kanan di bawah *result list* pilih *visualize cluster assignments*, dan kemudian akan muncul seperti pada gambar 14.10



Gambar 14.11 : Prosesuntuk Menyimpan Hasil *Cluster* pada *Weka*

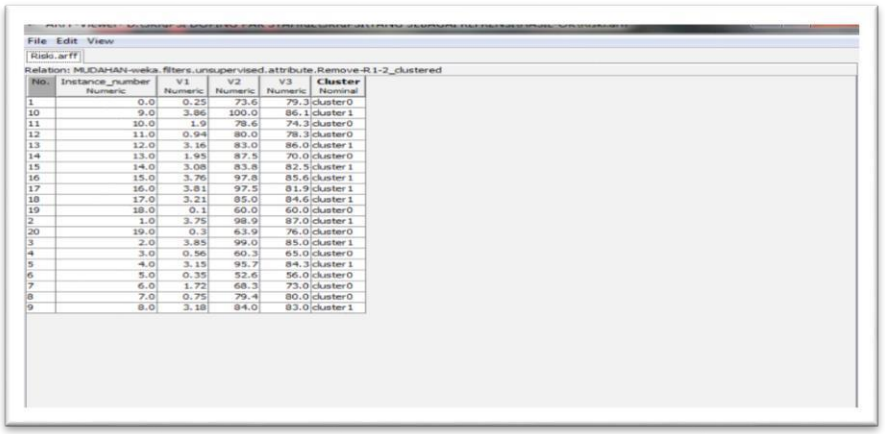
i. Selanjutnya dari gambar di atas klik *save* dan simpan hasil *cluster* dengan *type ArffViewer*.

j. Kemudian buka tampilan awal *WEKA* dan pilih *tools*, selanjutnya klik *ArffViewer* (CTRL+A), kemudian akan muncul seperti yang dilihat pada gambar 14.12



Gambar 14.12 Tampilan Awal *WEKA* untuk Menampilkan Hasil *Cluster*

k. Kemudian pilih menu *file*, klik *open* (CTRL+O), cari data yang sudah disimpan sebelumnya, dan selanjutnya akan mucul hasil *cluster* yang diinginkan seperti pada gambar 14.13



Gambar 14.13Tampilan Hasil *Cluster*

Berdasarkan hasil pengujian di atas dapat terlihat hasil pengujian dan hasil analisa dengan metode pengujiannya sesuai.

Pengujian Algoritma Apriori

Uji coba sistem bertujuan untuk membuktikan bahwa input, proses, output yang dihasilkan oleh sistem aplikasi Xl Miner telah benar dan sesuai dengan yang diinginkan.

Pengujian sistem dengan cara memasukkan data training ke dalam sistem dan memperhatikan output yang dihasilkan. Jika input, proses dan output telah sesuai, maka sistem telah benar. Berikut merupakan tahapan untuk pengujian sistem yaitu:

1. Melakukan penginputan dari data analisa bisnis yaitu data *historis* yang tersimpan didatabase transaksi penjualan atau database kasir.
2. Melakukan transformasi data dari jenis file *.xls menjadi *.csv
3. Lakukan analisa data Xl miner *Association Rule*.

Penggunaan program sebagai penerapan data mining ini, agar dapat berjalan dengan baik di dalam aplikasi Xl Miner harus ditempatkan pada sebuah folder, dan dilengkapi dengan input data dari analisa bisnis. Lokasi folder yang telah ditentukan adalah tempat untuk menyimpan file-file yang telah dikumpulkan, untuk menghindari kesalahan sebaiknya data tidak diletakan kedalam folder yang berbeda. Selanjutnya untuk menerapkan data mining aturan asosiasi untuk menganalisa keranjang pasar atau kebiasaan Toko dalam membeli barang, maka data dari bukti transaksi yang ada didatabase kasir diinput ke aplikasi Microsoft Excel lalu simpan dan ubah jenis file data tersebut menjadi *.csv. Jalankan aplikasi Xl Miner yang telah terinstall dikomputer.

Input data merupakan pengaruh di dalam pengisian data sehingga data yang dikehendaki dapat diterima oleh sisem dan juga diproses untuk pembuatan laporan yang di inginkan. Adapun bentuk input tersebut adalah sebagai berikut:

1. Tampilan Data Transaksi

Sebelum malukukan analisa keranjang pasar maka dilakukan input data transaksi terlebih dahulu yang diperoleh dari database kasir ke Microsoft Excel dan data diubah kejenis *.Csv. Tampilan input transaksi dapat dilihat pada gambar 14.14

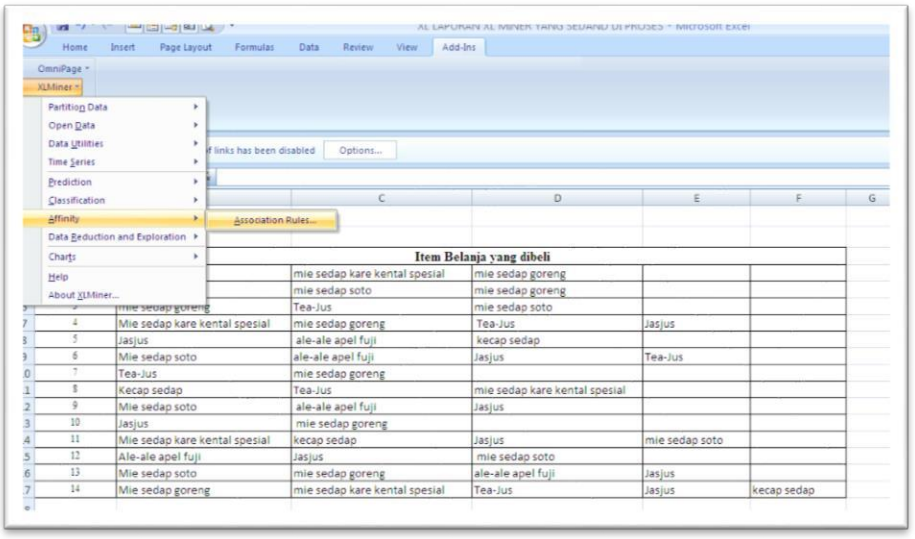
	A	B	C	D	E	F
2						
3	**Transaksi ke-**	**Item Belanja yang dibeli**				
4	1	Kecap sedap	mie sedap kare kental spesial	mie sedap goreng		
5	2	Ale-ale apel fuji,	mie sedap soto	mie sedap goreng		
6	3	mie sedap goreng	Tea-Jus	mie sedap soto		
7	4	Mie sedap kare kental spesial	mie sedap goreng	Tea-Jus	Jasjus	
8	5	Jasjus	ale-ale apel fuji	kecap sedap		
9	6	Mie sedap soto	ale-ale apel fuji	Jasjus	Tea-Jus	
10	7	Tea-Jus	mie sedap goreng			
11	8	Kecap sedap	Tea-Jus	mie sedap kare kental spesial		
12	9	Mie sedap soto	ale-ale apel fuji	Jasjus		
13	10	Jasjus	mie sedap goreng			
14	11	Mie sedap kare kental spesial	kecap sedap	Jasjus	mie sedap soto	
15	12	Ale-ale apel fuji	Jasjus	mie sedap soto		
16	13	Mie sedap soto	mie sedap goreng	ale-ale apel fuji	Jasjus	
17	14	Mie sedap goreng	mie sedap kare kental spesial	Tea-Jus	Jasjus	kecap sedap
18						

Gambar 14.14 : Datasheet Pengujian

2. Tampilan *Aftinity Association Rule* Xl Miner

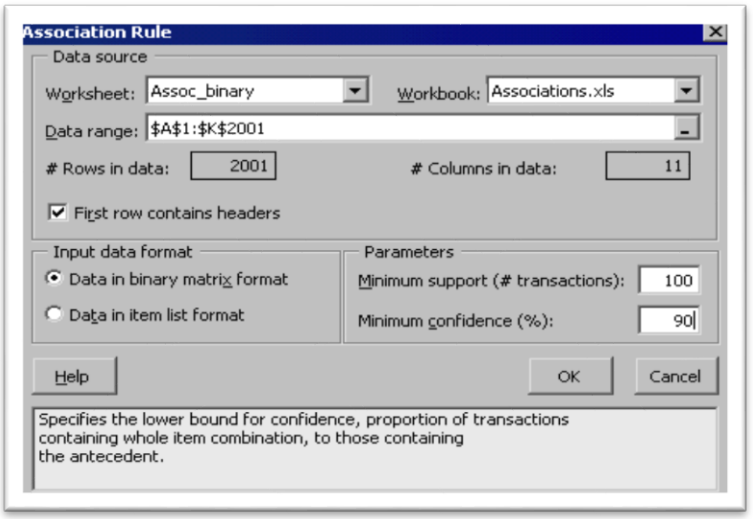
Data transaksi yang diinput dari database kasir telah selesai maka dilakukan langkah sebagai berikut:

a. Pilih menu tab Add-Ins, klik xl miner, aftinity kemudian klik *Association Rules* atau dapat dilihat pada gambar 14.15. berikut:



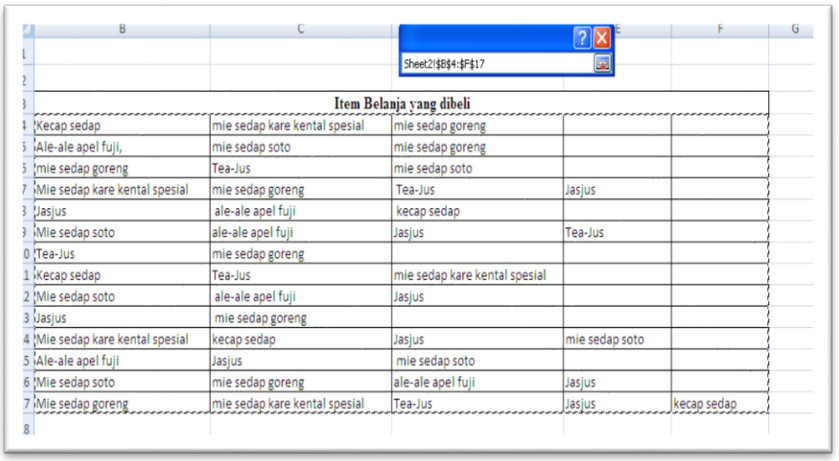
Gambar 14.15. Tampilan Aftinity Association Rule Xl Miner

Nilai *Min Support* dan *Min Confidence* merupakan data masukan yang digunakan untuk mencari item apa sajakah yang sering muncul dan seberapa kuat kaitannya dengan item lain yang juga dibeli, kemunculan sebuah item dan seberapa kuat kaitanya dengan item dapat dilihat pada Gambar 14.16



Gambar 14.16 Tampilan Minimum Support Dan Minimum Confidence

b. Input data yang akan dijadikan analisa dengan mengklik tanda yang ada di data range, kemudian seleksi data maka akan muncul tampilan berikut.

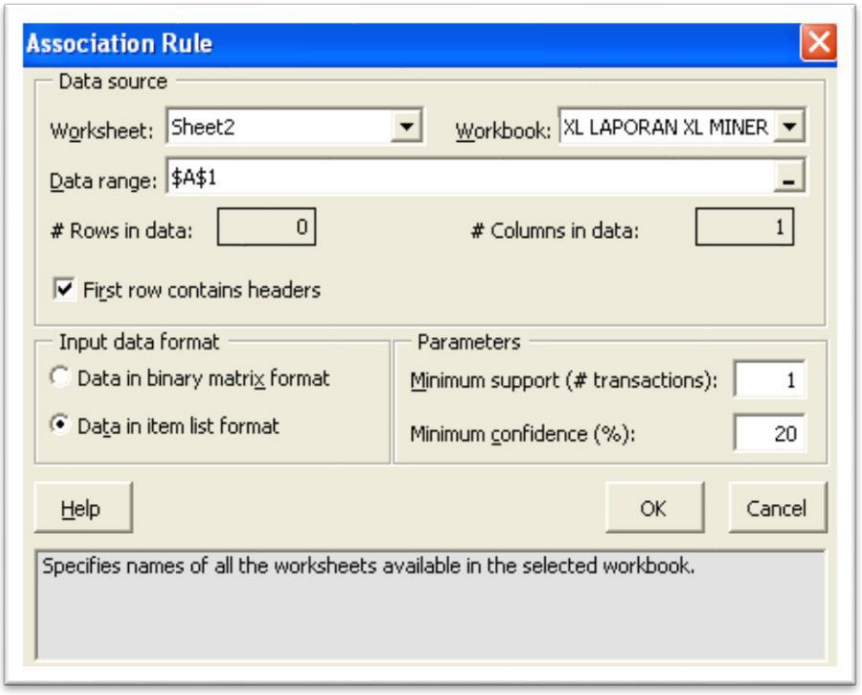


Item Belanja yang dibeli				
Kecap sedap	mie sedap kare kental spesial	mie sedap goreng		
Ale-ale apel fuji,	mie sedap soto	mie sedap goreng		
mie sedap goreng	Tea-Jus	mie sedap soto		
Mie sedap kare kental spesial	mie sedap goreng	Tea-Jus	Jasjus	
Jasjus	ale-ale apel fuji	kecap sedap		
Mie sedap soto	ale-ale apel fuji	Jasjus	Tea-Jus	
Tea-Jus	mie sedap goreng			
Kecap sedap	Tea-Jus	mie sedap kare kental spesial		
Mie sedap soto	ale-ale apel fuji	Jasjus		
Jasjus	mie sedap goreng			
Mie sedap kare kental spesial	kecap sedap	Jasjus	mie sedap soto	
Ale-ale apel fuji	Jasjus	mie sedap soto		
Mie sedap soto	mie sedap goreng	ale-ale apel fuji	Jasjus	
Mie sedap goreng	mie sedap kare kental spesial	Tea-Jus	Jasjus	kecap sedap

Gambar 14.17 Tampilan Seleksi Data Range

c. Lalu klik tanda yang ada pada tampilan diatas.

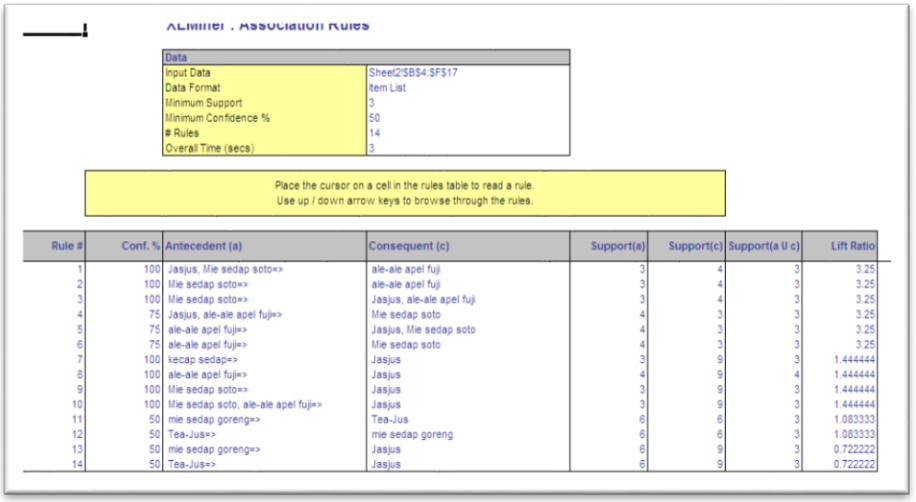
Pada pilihan input data format aktifkan data in item list format, pada pilihan parameters masukan nilai *min support* dan *min confindence* yang diinginkan. Lihat gambar dibawah ini:



Gambar 14.18 Tampilan Input *Minimum Support* Dan *Confidence*

d. Klik OK jika telah selesai maka akan muncul layout hasil analisa tampak pada gambar 14.19 berikutnya.

Hasil keluaran dari data mining aturan asosiasi untuk analisa keranjang pasar pada aplikasi Xl Miner ini menunjukkan laporan yang diharapkan mampu membantu memberikan keputusan bagi seorang yang berkepentingan untuk kelangsungan bisnisnya. Hasil dari keluaran aturan asosiasi dapat dilihat pada Gambar 14.19 berikut:



XLMiner : Association Rules

Data	
Input Data	Sheet2!$B$4:$F$17
Data Format	Item List
Minimum Support	3
Minimum Confidence %	50
# Rules	14
Overall Time (secs)	3

Place the cursor on a cell in the rules table to read a rule.
Use up / down arrow keys to browse through the rules.

Rule #	Conf. %	Antecedent (a)	Consequent (c)	Support(a)	Support(c)	Support(a U c)	Lift Ratio
1	100	Jasjus, Mie sedap soto=>	ale-ale apel fuji	3	4	3	3.25
2	100	Mie sedap soto=>	ale-ale apel fuji	3	4	3	3.25
3	100	Mie sedap soto=>	Jasjus, ale-ale apel fuji	3	4	3	3.25
4	75	Jasjus, ale-ale apel fuji=>	Mie sedap soto	4	3	3	3.25
5	75	ale-ale apel fuji=>	Jasjus, Mie sedap soto	4	3	3	3.25
6	75	ale-ale apel fuji=>	Mie sedap soto	4	3	3	3.25
7	100	kecap sedap=>	Jasjus	3	9	3	1.444444
8	100	ale-ale apel fuji=>	Jasjus	4	9	4	1.444444
9	100	Mie sedap soto=>	Jasjus	3	9	3	1.444444
10	100	Mie sedap soto, ale-ale apel fuji=>	Jasjus	3	9	3	1.444444
11	50	mie sedap goreng=>	Tea-Jus	6	6	3	1.083333
12	50	Tea-Jus=>	mie sedap goreng	6	6	3	1.083333
13	50	mie sedap goreng=>	Jasjus	6	9	3	0.722222
14	50	Tea-Jus=>	Jasjus	6	9	3	0.722222

Gambar 14.19 Keluaran Dari Algoritma Aprio