

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:

1. **Data** yaitu Kumpulan Fakta yang terekam atau sebuah entitas yang tidak memiliki arti dan selama ini terabaikan.
2. **Mining** yaitu proses Penambangan

Sehingga *Data Mining* itu dapat diartikan sebagai proses penambangan data yang menghasilkan sebuah output (keluaran) berupa pengetahuan. Selain itu juga Definisi *Data Mining* dapat dikutip dari beberapa sumber yaitu:

1. 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)

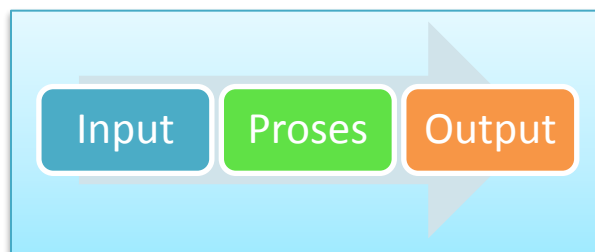
2. Menurut Larose:

Data Mining merupakan analisis dari peninjauan kumpulan data untuk menemukan hubungan yang tidak diduga dan meringkas data 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 musim, 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:

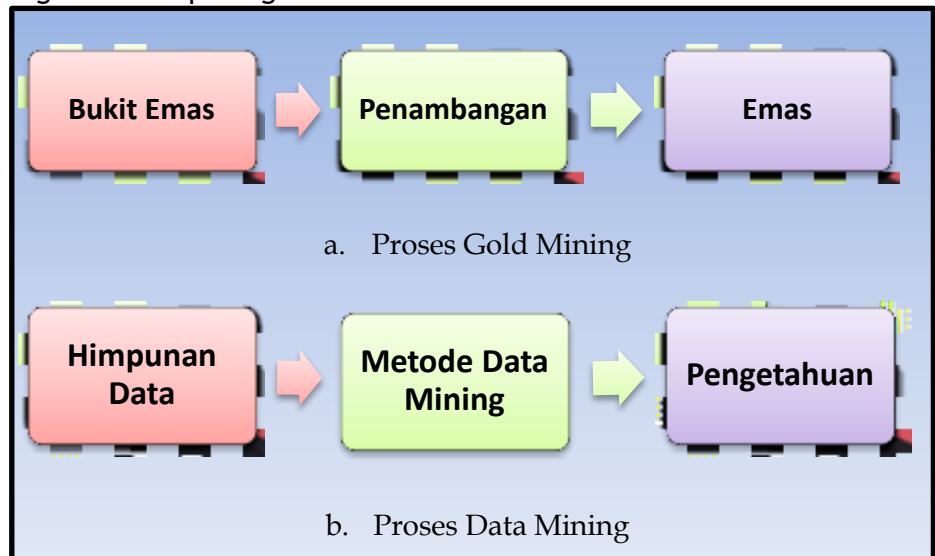
1. *Gold Mining*.

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 *Outputnya*.

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*

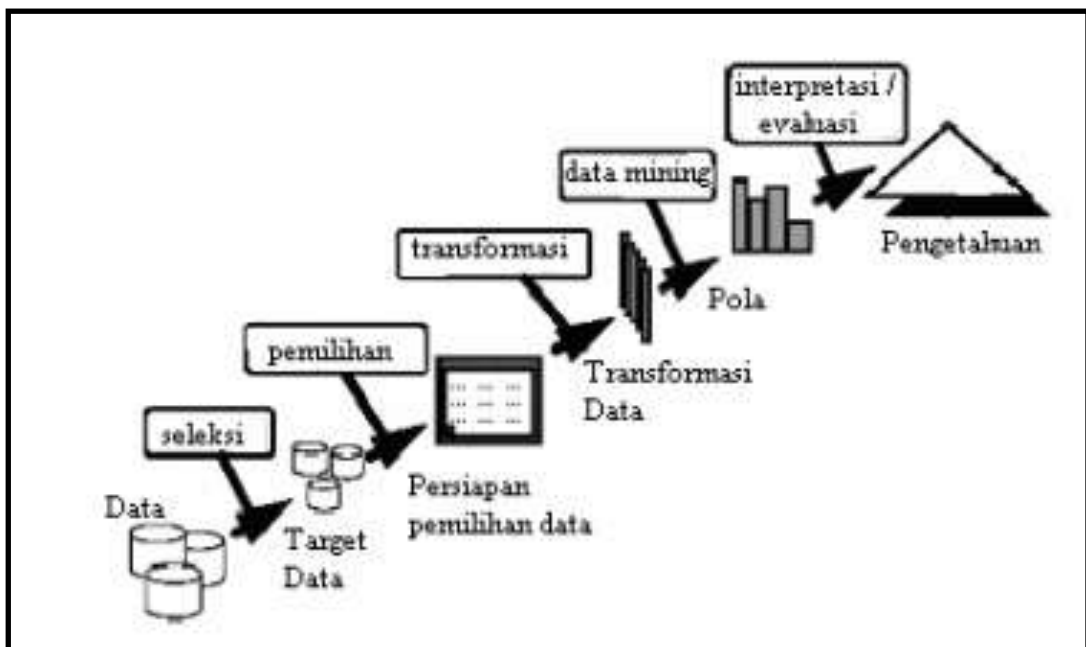
Berdasarkan proses di atas untuk menunjang pemahaman tentang *Data Mining* kita harus memahami beberapa disiplin ilmu lain seperti terlihat pada gambar di bawah ini:



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:

1. *Seleksi Data (Selection)*

Selection (seleksi/pemilihan) data dari merupakan sekumpulan data operasional perlu dilakukan sebelum tahap

penggalan 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 proses *Data Mining*

4. *Data Mining*

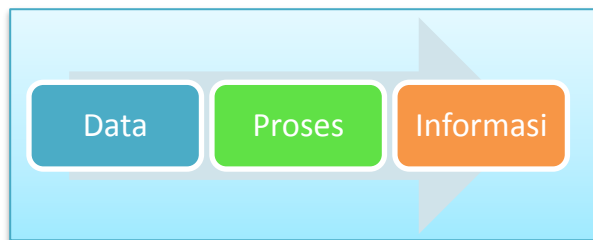
Pada fase ini yang dilakukan adalah menerapkan algoritma atau metode pencarian pengetahuan.

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

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

1.3 Data Dan Himpunan Data

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 diinterpretasikan sebagai Entitas yang tidak memiliki arti yang selam 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 / Variable / Field					Atribut keputusan
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 mengidentifikasi 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, kal^{ikan} 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

Numeric

Nominal

1.4 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 populer 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 menggunakan 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 digunakan 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 atribut 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.

SOAL LATIHAN :

1. Mengapa perlu dilakukan Data mining?
2. Jelaskan setiap tahapan di dalam proses data mining?
3. Apa perbedaan mendasar dari Konsep Data Mining dan Sistem Pendukung Keputusan?



FUNGSI MINOR : METODE ESTIMASI

2.1 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 Algorithm diantaranya yaitu Regresi Linier.

2.1 Metode Regresi Linear Sederhana

1. Pendahuluan

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:

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

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:

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

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

Atau

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

$$a = \bar{Y} - b \cdot \bar{X}$$

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

2. 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 atau 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
28	32	174	238	198

Rumus 1

- Menghitung Nilai Rata-rata

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

- 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,

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

$$a = \frac{5568 - 5544}{870 - 784}$$

$$a = \frac{24}{86}$$

$$a = 0.279$$

dan

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

$$\begin{aligned}b &= \frac{(5)(198)-(28)(32))}{(10(55^2)-(55))^2} \\&= \frac{990-896}{870-784} \\&= 94/86 \\&= 1.093\end{aligned}$$

Maka persamaan regresinya adalah:

$$Y = a+bx$$

$$Y = 0.279 + 1.093x$$

Jadi,

$$Y = 0.279 + 1.093x$$

$$Y = 0.279 + 1.093 (8.5)$$

$$Y = 0.279 + 9.2905$$

$$Y = 9.569 \text{ (ribuan)}$$

$$Y = 9569$$

Rumus 2:

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

$$a = \bar{Y} - b \cdot \bar{X}$$

-Menghitung Nilai a dan b

$$\begin{aligned}b &= \frac{(5)(198)-(28)(32))}{(10(55^2)-(55))^2} \\&= \frac{990-896}{870-784} \\&= 94/86 \\&= 1.093\end{aligned}$$

$$\begin{aligned}a &= 6.4-1.093(5.6) \\&= 6.4-6.1208 \\&= 0.279\end{aligned}$$

Maka persamaan Regresinya adalah:

$$Y = a + bx$$

$$Y = 0.279 + 1.093x$$

Jadi,

$$Y = 0.279 + 1.093x$$

$$Y = 0.279 + 1.093 (8.5)$$

$$Y = 0.279 + 9.2905$$

$$Y = 9.569 \text{ (ribuan)}$$

$$Y = 9569$$

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:

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

$$R^2 = \frac{((5)(198) - (28)(32))^2}{(5(174) - (28)^2)(5(238) - (32)^2)}$$

$$R^2 = \frac{(990 - 896)^2}{(870 - 784)(1190 - 1024)}$$

$$R^2 = \frac{(94)^2}{(86)(166)}$$

$$R^2 = 8.836 / 14.276$$

$$R^2 = 0.6189$$

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 3.1 : Sampel Data Regresi Linier

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

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

3.1 Pendahuluan

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 dalam 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 bentuk “*Decision Tree* (pohon keputusan)”. Dalam pembahasan kali ini saya mencoba untuk membahas tentang Algoritma C4.5.

3.2 Algoritma C4.5

1. Pendahuluan

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

mengeksrak suatu kelas. Entropi digunakan untuk mengukur ketidakaslrian 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:

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

Dimana:

- A : Atribut
- S : Sampel
- n : Jumlah partisis himpunan atribut A
- $|S_i|$: 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:



Gambar4.1 : 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 yang 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 diagram yang biasa disebut pohon keputusan (*decision tree*). Data berikut ini dipergunakan untuk data latihan. Data selengkapnya tampak pada tabel dibawah ini:

Tabel 3.1 : Sampel Data Yang Digunakan

No.	Nama Mahasiswa	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
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 Widiantera	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 *Gainnya*:

1. Nilai *Entropy*

a. $Entropy\ Total = Entropy(S) = \sum_{i=1}^n - p_i \cdot \log_2 p_i$

$$Entropy\ Total = ((-4/14) \cdot \log_2 (4/14) + (-10/14) \cdot \log_2 (10/14))$$

$$= 0.863120569$$

b. *Entropy* Minat Calon Mahasiswa

- Nilai atribut “**Hardware**” = $((-3/5) \cdot \log_2 (3/5) + (-2/5) \cdot \log_2 (2/5))$
= **0.970950594**

- Nilai atribut “**Software**” = $((-1/5) \cdot \log_2 (1/5) + (-4/5) \cdot \log_2 (4/5))$

$$= 0.721928095$$

- Nilai atribut “**Umum**” = $((-0/4) \cdot \log_2 (0/4) + (-4/4) \cdot \log_2 (4/4))$

$$= 0$$

c. *Entropy* Histori Pendidikan (Asal Sekolah)

- Nilai atribut “**SMK Komputer**”

$$= ((-2/4) \cdot \log_2 (2/4) + (-2/4) \cdot \log_2 (2/4)) = 1$$

- Nilai atribut “**SMK Teknik**”

$$= ((-0/4) \cdot \log_2 (0/4) + (-4/4) \cdot \log_2 (4/4)) = 0$$

- Nilai atribut “**SMA Umum**”

$$=((-2/6)*\text{Log } 2(2/6)+(-4/6)*\log 2 (4/6) = \mathbf{0.918295834}$$

d. Entropy Hobi

- Nilai atribut “IT” =

$$((-4/6)*\text{Log } 2(4/6)+(-2/6)*\log 2 (2/6) = \mathbf{0.918295834}$$

- Nilai atribut “Non IT” =

$$((-2/8)*\text{Log } 2(2/8)+(-6/8)*\log 2 (6/8) = \mathbf{0.811278124}$$

e. Entropy Jenis Kelamin

- Nilai atribut “1” = $((-4/7)*\text{Log } 2(4/7)+(-3/7)*\log 2 (3/7)$

$$= \mathbf{0.985228136}$$

- Nilai atribut “0” = $((-0/7)*\text{Log } 2(0/7)+(-7/7)*\log 2 (7/7)$

$$= \mathbf{0}$$

2. Nilai Gain

Berikut ini adalah nilai *Gain* dari setiap kriteria.

- Nilai Gain Minat Calon Mahasiswa
 $= \mathbf{0.863120569 - ((5/14)*0.970950594) + ((5/14)*0.721928095) + ((4/14)*0)} = \mathbf{0.258521037}$

- Nilai Gain Histori Pendidikan

$$= \mathbf{0.863120569 - ((4/14)*1) + ((4/14)*0) + ((6/14)*0.918295834)} = \mathbf{0.183850925}$$

- Nilai Gain Hobi

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

- Nilai Gain Jenis Kelamin

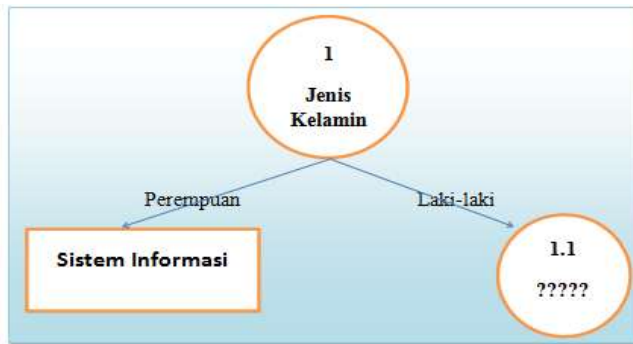
$$= 0.863120569 - ((7/14) * 0.985228136) + ((7/14) * 0) = 0.005977711$$

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

Tabel 3.2 Rekapitulasi Hasil

NODE		Keterangan	Jml Kasus (S)	Sistem Komputer (SK)	Sistem Informasi (SI)	Entropy	Gain
1	TOTAL		14	4	10	0.863120569	
	Minat Calon						0.258521037
		Hardware	5	3	2	0.970950594	
		Software	5	1	4	0.721928095	
		Umum	4	0	4	0	
	Asal Sekolah						0.183850925
		SMK Komputer	4	2	2	1	
		SMK Teknik	4	0	4	0	
		SMA Umum	6	2	4	0.918295834	
	Jenis Kelamin						0.370506501
		Laki-laki	7	4	3	0.985228136	
		Perempuan	7	0	7	0	
	Hobi						0.005977711
		IT	6	4	2	0.918295834	
		Non	8	2	6	0.811278124	

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(kriteria) Jenis Kelamin = Laki-laki.

1. Nilai *Entropy*

Berikut ini adalah tabel penyelesaiannya.

Tabel: 3.3 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 3.4 Atribut Jenis Kelamin

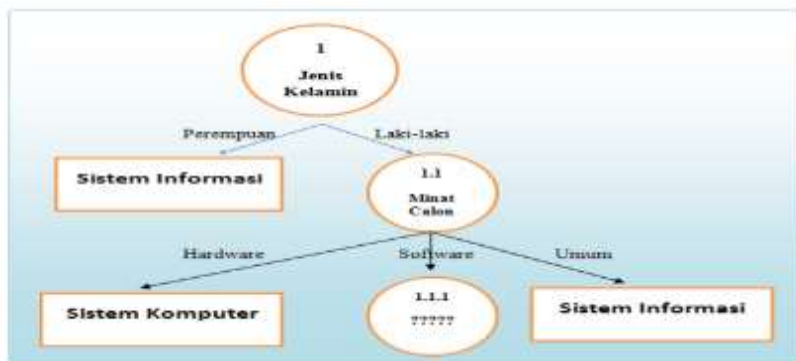
No.	Nama Mahasiswa	Minat Calon Mahasiswa	Asal Sekolah	Jenis Kelamin	Hobi	Hasil
1	Novita Devi Batu Bara	Hardware	SMK Komputer	Laki-Laki	Non	SK
2	Ahmad Riyandi	Hardware	SMK Komputer	Laki-Laki	IT	SK
3	Reza Adriansyah	Umum	SMK Komputer	Laki-Laki	Non	SI
4	Gafar Dwi Satrio	Software	SMA UMUM	Laki-Laki	Non	SI
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 Atribut = 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 3.6 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 3.7 Atribut Minat Calon Siswa

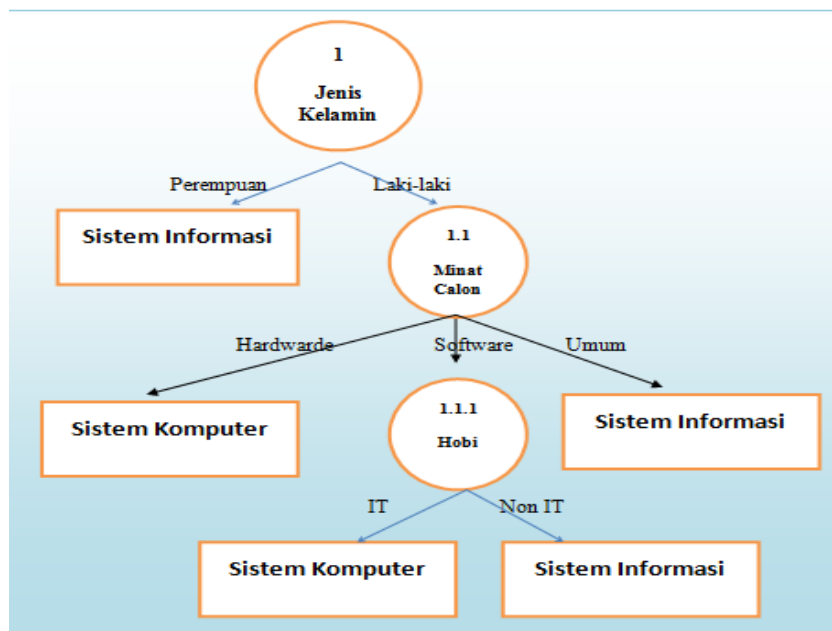
No.	Nama Mahasiswa	Minat Calon Mahasiswa	Asal Sekolah	Hobi	Hasil
1	Gafar Dwi Satrio	Software	SMA UMUM	Non	SI
2	Akbar Widiantera	Software	SMA UMUM	IT	SK

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

Tabel 3.8 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 3.4: Hasil Dari Pohon Keputusan

Maka basis pengetahuan atau rule yang terbentuk yaitu:

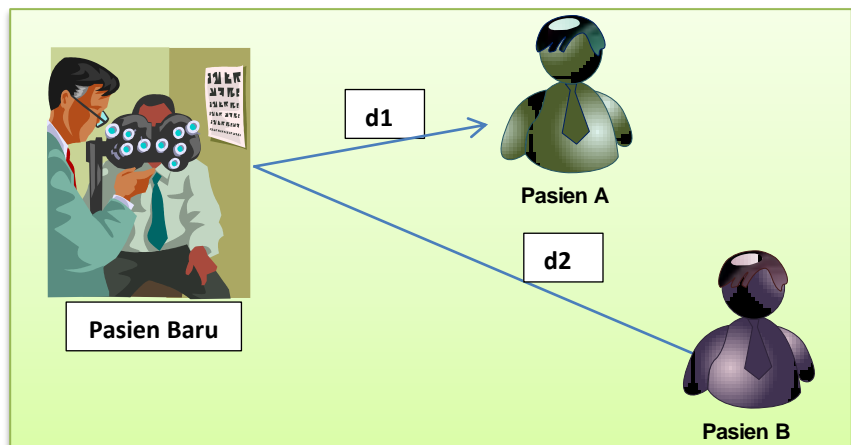
1. Jika **Jenis Kelamin** = Perempuan maka **Hasil**= Sistem Informasi
2. Jika **Jenis Kelamin** = Laki-laki dan **Minat Calon**=Hardware maka **Hasil**= Sistem Komputer
3. Jika **Jenis Kelamin** = Laki-laki dan **Minat Calon**=Umum maka **Hasil**= Sistem Informasi
4. Jika **Jenis Kelamin** = Laki-laki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

3.3 Algoritma Nearest Neighbor

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.



Gambar 3.5: 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:

$$\text{Similarity (T,S)} = \frac{(\sum_{i=1}^n f(T_i, S_i) * w_i)}{w_i}$$

Keterangan:

T : Kasus Baru

S : Kasus yang ada dalam penyimpanan

n : jumlah attribut dalam setiap kasus

- i : atribut 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 3.9: Himpunan Data Algoritma C4.5

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

Atribut Kondisi

Atribut Tujuan

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 3.10 : 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 3.11 : Sampel Kasus

Atribut	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 3.12: 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 3.13: 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 3.14 : 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 mengajukan 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 penyelesaiannya

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

Tabel 3.15 : 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 3.16: Bobot Kasus 1

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

Hitung:

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

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

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

$$\text{Jarak} = 3.75 / 2.25$$

$$\text{Jarak} = 1.667$$

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

Tabel 3.17 : 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 3.18 : Bobot Kasus 2

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

Hitung:

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

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

$$\text{Jarak} = \frac{0.5 + 0.375 + 1}{2.25}$$

$$\text{Jarak} = 1.875 / 2.25$$

$$\text{Jarak} = 0.833333$$

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

Tabel 3.19: 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 3.20 : Bobot Kasus 2

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

Hitung:

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

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

$$\text{Jarak} = \frac{0.5 + 0.28125 + 0.75}{2.25}$$

$$\text{Jarak} = 1.53125 / 2.25$$

$$\text{Jarak} = 0.68055$$

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

Tabel 3.21: Rekapitulasi Nilai Kedekatan

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

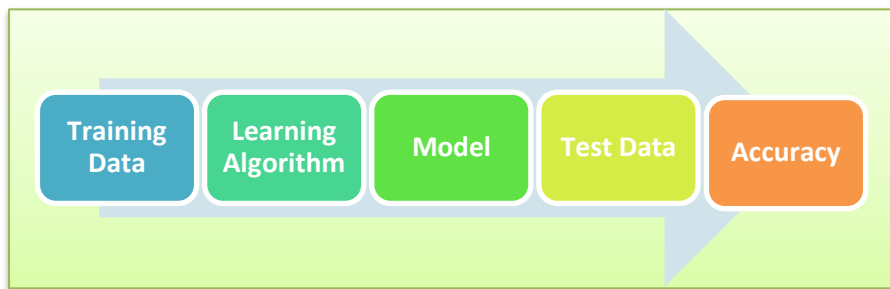
Berdasarkan tabel diatas untuk nilai Kedekatan tertinggi adalah pada nilai Kedekatan dengan Kasus 1. Jadi, untuk soal di atas maka nasabah atas nama “Dian” pada nilai atribut keterangannya bernilai “Ya”.

3.4 Algoritma Naïve Bayesian Clasifier

1. Pendahuluan

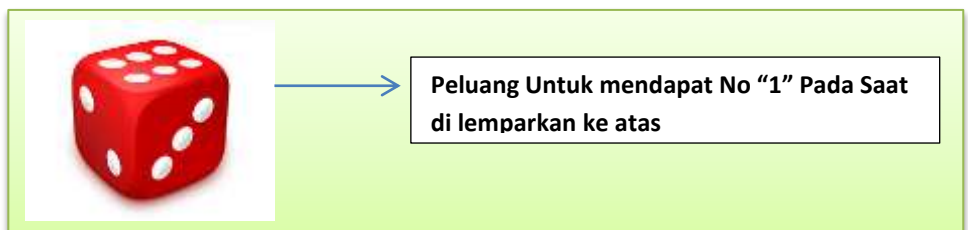
Naïve Bayesian Clasifier merupakan salah satu algoritma pemecahan masalah yang termasuk kedalam Metode Klasifikasi pada Data Mining. Naïve Bayesian Clasifier mengadopsi ilmu statistika yaitu dengan menggunakan teori kemungkinan (Probabilitas) untuk menyelesaikan sebuah kasus *Supervised Learning*, artinya 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 3. 6: 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 3.7 : Ilustrasi Peluang

Dari gambar di atas dapat kita mengetahui secara sederhana bahwasanya peluang untuk mendapatkan no.1 yaitu : $\frac{1}{6}$. Dengan asumsi jumlah yang bernilai dadu no.1 ada 1 sedangkan total keseluruhan dadu 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

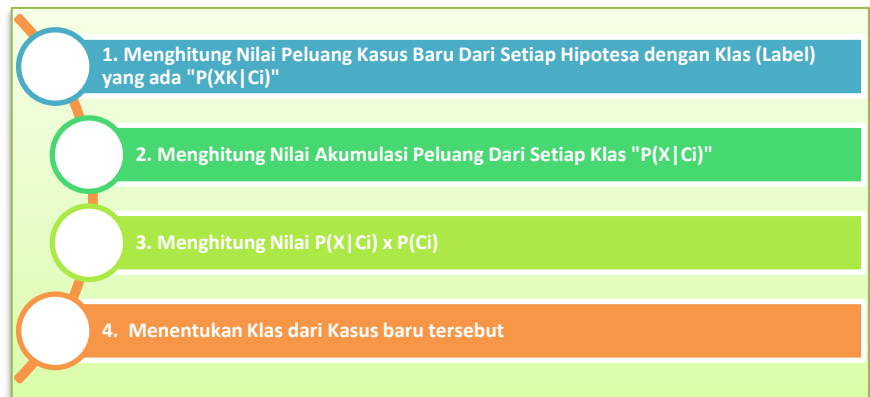
H : hipotesa bahwa x adalah data kelas (label)

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 3.8: Algoritma Naïve Bayesian Clasifier

2. 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 3.22 : 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 mengajukan Kredit Perumahan Rakyat (KPR) dengan keterangan di bawah ini:

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

Penyelesaian:

- Hitung nilai $P(XK|Ci)$ untuk setiap class i
 - $P(\text{Kriteria 1} = \text{"Di atas 30"} \mid \text{Keterangan} = \text{"Ya"})$
 $P(\text{Kriteria 1} = 2/4 = 0.5$
 - $P(\text{Kriteria 1} = \text{"Di atas 30"} \mid \text{Keterangan} = \text{"Tidak"})$
 $P(\text{Kriteria 1} = 1/2 = 0.5$
 - $P(\text{Kriteria 2} = \text{"Sedang"} \mid \text{Keterangan} = \text{"Ya"})$
 $P(\text{Kriteria 2} = 2/4 = 0.5$
 - $P(\text{Kriteria 2} = \text{"Sedang"} \mid \text{Keterangan} = \text{"Tidak"})$
 $P(\text{Kriteria 2} = 0/2 = 0$
 - $P(\text{Kriteria 3} = \text{"Baik"} \mid \text{Keterangan} = \text{"Ya"})$
 $P(\text{Kriteria 2} = 4/4 = 1$
 - $P(\text{Kriteria 3} = \text{"Baik"} \mid \text{Keterangan} = \text{"Tidak"})$
 $P(\text{Kriteria 2} = 1/2 = 0.5$
- Hitung nilai $P(X|Ci)$ untuk setiap Kelas (label)
 - $P(X \mid \text{Keterangan} = \text{"Ya"})$
 $= 0.5 \times 0.5 \times 1 = 1.25$ 
 - $P(X \mid \text{Keterangan} = \text{"Tidak"})$
 $= 0.5 \times 0 \times 0.5 = 0$
- Hitung nilai $P(X|Ci) * P(Ci)$
 - $-(P(X \mid \text{Keterangan} = \text{"Ya"}) \times P(\text{Keterangan} = \text{"Ya"}))$
 $= 1.25 \times 4/6 = 0.8333$
 - $-(P(X \mid \text{Keterangan} = \text{"Ya"}) \times P(\text{Keterangan} = \text{"Ya"}))$
 $= 0 \times 2/6 = 0$
- 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 3.22 : Sampel Kasus *Naïve Bayesian Classifier*

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 Classifier*)??

4.1 Pendahuluan

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

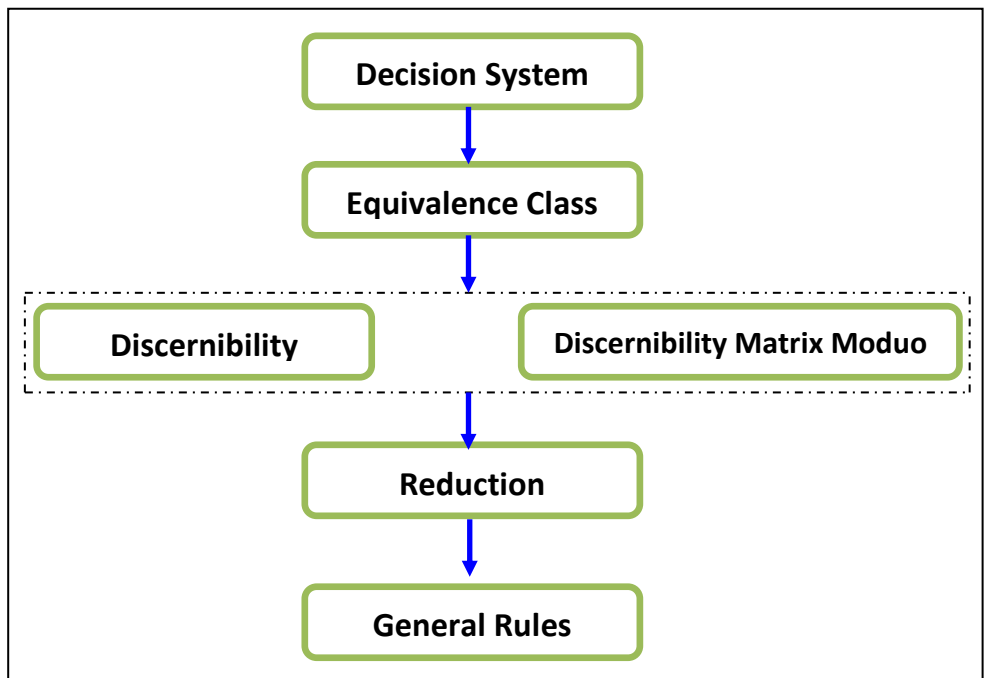
Melalui jurnal tersebut Metode *Rough Set*, metodologi yang elegan dan kuat dalam penggalian dan meminimalkan aturan dari tabel keputusan dan sistem informasi Pawlak. Gagasan utamanya adalah inti, mengecil, dan ketergantungan pengetahuan.

Pada jurnal internasional lainnya bahwasanya *Rough Set* di definisikan *An approach first forwarded by mathematician Zdzislaw Pawlak at the beginning of the eighties; 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 this approach is expressed by a boundary region of a set, and not by a partial membership as in Fuzzy Set Theory. Rough Set concept can be defined quite generally by means of interior and 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 4.1 Algoritma Penyelesaian Dengan Metode Rough Set

Keterangan:

1. *Decision System*

Decision system adalah *information system* dengan atribut tambahan yang dinamakan dengan *decision attribute*, 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 *Information Systems (IS)* menjadi

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

Keterangan

$U= \{x_1, x_2, \dots, x_m\}$ yang merupakan sekumpulan *example*.

$A= \{a_1, a_2, \dots, a_n\}$ 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 *Discernibility Matrix*: Diberikan sebuah IS $A=(U,A)$ and B Gabung A , *discernibility matrix* dari A adalah MB , dimana tiap-tiap entry $MB(i,j)$ terdiri dari sekumpulan attribute yang berbeda antara objek X_i dan X_j .

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 , MB_d , didefinisikan seperti berikut dimana $MB(i,j)$ adalah sekumpulan attribute yang berbeda antara objek X_i dan X_j 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 = $(2^n - 1)$. 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 minimal 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.

4.2 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 teknik 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 4.1 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 4.2 Sampel Data Yang Akan Di Proses

NO	Objek	Nilai				Hasil
		Kehadiran	Tugas	UTS	UAS	
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 XXXXXX yang 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 4.3 Decision System

NO	Objek	Nilai				Hasil
		Kehadiran	Tugas/Quiz	UTS	UAS	
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 3.3 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 *DecisionAttribute*.

2. Equivalen Class

Equivalence Class adalah mengelompokan objek-objek yang sama untuk *attribute* tertentu. Diberikan *DecisionSystems* pada

tabel 3.4, maka dapat diperoleh *equivalence class* (*Equivalence Class* 1 sampai dengan *Equivalence Class* 7) seperti digambarkan pada Tabel 3.4

Tabel 4.4 *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 4.5 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 4.6 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 4.7 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 = $(2^n - 1)$. 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 minimal 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.

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

Tabel 4.8 Proses Penyeleksian

Class	CNF of Boolean Function	Prime Implicant	Reduct
EC1	$C \wedge (A \vee B \vee D) \wedge (A \vee D) \wedge B \wedge (B \vee C) \wedge (B \vee D)$	$C \wedge (A \vee D) \wedge B$	{C}, {A}, {B}
EC2	$C \wedge (A \vee B \vee C \vee D) \wedge (A \vee C \vee D) \wedge (B \vee C) \wedge B \wedge (B \vee C \vee D)$	$C \wedge B$	{C}, {B}
EC3	$(A \vee B \vee D) \wedge (A \vee B \vee C \vee D) \wedge B \wedge (A \vee D) \wedge (A \vee C \vee D) \wedge A$	$B \wedge A$	{B}, {A}
EC4	$(A \vee D) \wedge (A \vee C \vee D) \wedge D \wedge (A \vee B \vee D) \wedge (A \vee B \vee C \vee D) \wedge (A \vee B)$	$(A \vee D) \wedge D$	{D}, {A}, {B}
EC5	$(B \vee C) \wedge B \wedge (A \vee C \vee D) \wedge (ABCD) \wedge D \wedge (C \vee D)$	$B \wedge (C \vee D)$	{B}, {C}, {D}
EC6	$(B \vee D) \wedge (B \vee C \vee D) \wedge A \wedge (A \vee B \vee D) \wedge C \wedge (C \vee D)$	$(B \vee D)$	{B}, {D}
EC7	$(B \vee D) \wedge (B \vee C \vee D) \wedge A \wedge (A \vee B \vee D) \wedge C \wedge (C \vee D)$	$(B \vee D)$	{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 4.9 Keputusan Atau Knowledge Baru

N o	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 = Lulus
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?



5.1 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 analisis dan interpretasi 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 segmentasi 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

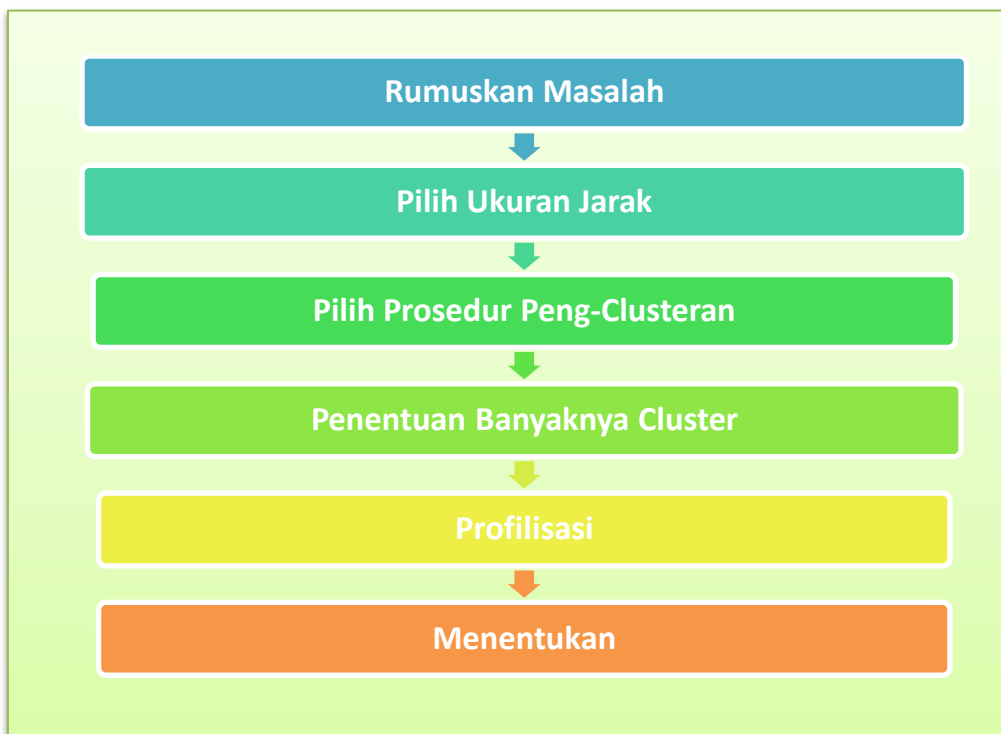
Analisis *cluster* digunakan untuk mengenali/mengidentifikasi kelompok pembeli yang homogen/relatif homogen. Kemudian perilaku untuk dalam setiap kelompok perlu dikaji secara terpisah. Responden (pembeli) dikelompokkan didasarkan pada *self-reported importance* yang terkait pada setiap faktor pilihan 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 5.1: Prosedur Analisis *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 $m \times p$. Matrik tersebut ditransformasikan ke dalam bentuk matriks similaritas (kemiripan) berupa $n \times n$ yang dihitung berdasarkan pasangan-pasangan obyek p peubah. Konsep dasar pengukuran analisis *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 analisis

cluster didasarkan pada kedekatan. Pengukuran 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.

$$D(X, Y) = \sqrt{\sum (X_i - Y_i)^2}$$

2. *Squared Euclidean Distance*

Merupakan ukuran jarak antara dua item X dan Y.

$$D(X, Y) = \sum (X_i - Y_i)^2$$

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*

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

5. *Block*

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

6. *Minkowski*

7. *Chi-Square*

8. *Phi-Square*

9. *Hamming*

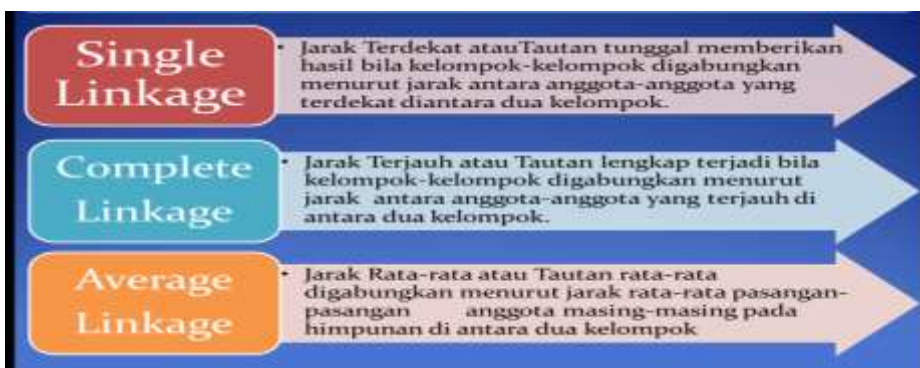
Teknik hirarki (*hierarchical methods*) adalah teknik *clustering* membentuk konstruksi hirarki atau berdasarkan tingkatan tertentu seperti struktur pohon (struktur pertandingan). Dengan demikian proses pengelompokannya dilakukan secara bertingkat atau bertahap. Hasil dari pengelompokan ini dapat

disajikan dalam bentuk dendogram. Metode-metode yang digunakan dalam teknik hirarki :

- 1 • Agglomerative Method
- 2 • Divisive Method

5.2 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 5.2 : Jenis-jenis Hierarchy Agglomerathip Cluster

Pada pembahasan kali ini penulis mencoba untuk menjelaskan tentang penyelesaian Teknik Single Linkage di dalam proses Clustering dengan menggunakan pendekatan Euclidian

Distance. Berikut ini adalah proses penyelesaian Clustering dengan Teknik Single Linkage dengan pendekatan Euclidian Distance:



Gambar 5.3 : 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 menyelesaikan metode clustering dengan Pendekatan Teknik Single Linkage.

1. Menghitung Nilai Rata-rata Variabel (V1)

Rumus :

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n}$$

Keterangan:

\bar{X} = Nilai Rata-rata Variabel

X_i = Nilai Variabel

N = Jumlah atau kuantitas dari objek

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

Rumus :

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

Keterangan:

Std (X) = Nilai Rata-rata Variabel

X_i = Nilai Variabel

N = Jumlah atau kuantitas dari objek

3. Menghitung Nilai Zero Standar (Zi)

Rumus:

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

Keterangan:

Z_i = Nilai Zero Standar

\bar{X} = Nilai Rata-rata Variabel

Std (X) = Nilai Rata-rata Variabel

X_i = Nilai Variabel

Contoh Kasus:

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 menggunakan Teknik Single Linkage. Adapun sampel data yang akan di proses terlihat pada tabel di bawah ini:

Tabel 5.1 : 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 (preprocessing) yang harus dilakukan adalah membuat tabel asumsi terhadap objek. Berikut ini adalah tabel asumsi objek dari sampel data yang digunakan.

Tabel 5.2: 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 menggunakan 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 sebagai berikut:

Rumus :

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n}$$

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 $\bar{V1}$

2. Menghitung Nilai Standar Deviasi Variabel

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

Maka,

$$\text{Std}(V1) = \sqrt{\frac{(\overline{V1}_1 - \overline{V1})^2 + (\overline{V1}_2 - \overline{V1})^2 + (\overline{V1}_3 - \overline{V1})^2 + (\overline{V1}_4 - \overline{V1})^2 + (\overline{V1}_5 - \overline{V1})^2 + (\overline{V1}_6 - \overline{V1})^2 + (\overline{V1}_7 - \overline{V1})^2 + (\overline{V1}_8 - \overline{V1})^2 + (\overline{V1}_9 - \overline{V1})^2 + (\overline{V1}_{10} - \overline{V1})^2 + (\overline{V1}_{11} - \overline{V1})^2 + (\overline{V1}_{12} - \overline{V1})^2 + (\overline{V1}_{13} - \overline{V1})^2 + (\overline{V1}_{14} - \overline{V1})^2 + (\overline{V1}_{15} - \overline{V1})^2 + (\overline{V1}_{16} - \overline{V1})^2 + (\overline{V1}_{17} - \overline{V1})^2 + (\overline{V1}_{18} - \overline{V1})^2 + (\overline{V1}_{19} - \overline{V1})^2 + (\overline{V1}_{20} - \overline{V1})^2}{N-1}}$$

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

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

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

$$= 1.442$$

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

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

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

Maka,

$$Z(V1_1) = \frac{V1_1 - \bar{V1}}{\text{Std}(V1)} = (0.25 - 2.182) / 1.442 = -1.340$$

Untuk Nilai

- Z(V1₂) sampai dengan Z(V1₂₀)
- Z(V2₁) sampai dengan Z(V2₂₀)
- Z(V3₁) sampai dengan Z(V3₂₀)

Silahkan dihitung dengan konsep perhitungan Z(V1₁). Setelah semua data distandarkan dengan cara yang sama, maka didapatkan tabel baru yang berisi data yang telah standar, yaitu:

Tabel 5.3: 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 5.4 : 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	0.639	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	2.013	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	4.546	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	1.761	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	0.940	1.067	1.030	0.191	3.709	2.947
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	0.639	2.531	2.514	2.013	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	0.940	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	0.735	2.227	2.443	1.097	2.349	1.123	0	0.412	0.965	4.605	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	0.994	4.371	3.414
R	2.274	1.058	1.058	3.018	0.735	4.546	2.007	1.822	0.191	1.135	1.523	1.756	0.210	1.845	0.263	0.965	0.994	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	4.605	4.371	3.874	0	1.797
T	0.759	3.601	3.584	1.256	3.083	2.346	1.082	1.192	2.947	3.670	1.511	1.216	2.622	2.090	2.469	3.505	3.414	2.659	1.797	0

Nilai Distance Terkecil

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 5.5 : 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	2.551	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	2.175	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	1.175	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	1.151	2.121	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	2.461	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	2.551	2.514	2.015	1.987	3.129	1.151	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	2.517	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	1.175	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	2.461	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

Nilai Distance Terkecil

- 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 5.6 : 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	2.443	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	0.357	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	1.516	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	0.357	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	1.516	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	2.019	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	2.589	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	2.094	0.805	0	1.772	1.265	2.358	1.756	2.514	1.216
M	0.339	0.097	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	2.580	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	2.498	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

Nilai Distance Terkecil

- 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 5.7 : 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) $d(\text{CPJB})\text{IO} = \min \{d\text{CPJIO}, d\text{BIO}\} = d\text{CPJIO} = 1.067$
- 2) $d(\text{CPJB})\text{A} = \min \{d\text{CPJA}, d\text{BA}\} = a = 3.027$
- 3) $d(\text{CPJB})\text{D} = \min \{d\text{CPJD}, d\text{BD}\} = d\text{CPJD} = 4.091$
- 4) $d(\text{CPJB})\text{E} = \min \{d\text{CPJE}, d\text{BE}\} = d\text{CPJE} = 0.470$
- 5) $d(\text{CPJB})\text{F} = \min \{d\text{CPJF}, d\text{BF}\} = d\text{CPJF} = 5.092$
- 6) $d(\text{CPJB})\text{G} = \min \{d\text{CPJG}, d\text{BG}\} = d\text{CPJG} = 2.835$
- 7) $d(\text{CPJB})\text{H} = \min \{d\text{CPJH}, d\text{BH}\} = d\text{CPJH} = 2.517$
- 8) $d(\text{CPJB})\text{K} = \min \{d\text{CPJK}, d\text{BK}\} = d\text{CPJK} = 2.227$
- 9) $d(\text{CPJB})\text{L} = \min \{d\text{CPJL}, d\text{BL}\} = d\text{CPJL} = 2.443$
- 10) $d(\text{CPJB})\text{M} = \min \{d\text{CPJM}, d\text{BM}\} = d\text{CPJM} = 1.097$
- 11) $d(\text{CPJB})\text{N} = \min \{d\text{CPJN}, d\text{BN}\} = d\text{CPJN} = 2.249$
- 12) $d(\text{CPJB})\text{Q} = \min \{d\text{CPJQ}, d\text{BQ}\} = d\text{CPJQ} = 0.359$
- 13) $d(\text{CPJB})\text{R} = \min \{d\text{CPJR}, d\text{BR}\} = d\text{CPJR} = 0.963$

$$14) d(\text{CPJB})S = \min \{d(\text{CPJS}, dBS)\} = d(\text{CPJS}) = 4.603$$

$$15) d(\text{CPJB})T = \min \{d(\text{CPJT}, dBT)\} = d(\text{CPJT}) = 3.505$$

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

Tabel 5.8 : 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*.

$$1) d(\text{IOR})\text{CPBJ} = \min \{d(\text{IOCPBJ}, d\text{RCPBJ})\} = d\text{RCPBJ} = 0.963$$

$$2) d(\text{IOR})A = \min \{d(\text{IOA}, d\text{RA})\} = d\text{IOA} = 2.114$$

$$3) d(\text{IOR})D = \min \{d(\text{IOD}, d\text{RD})\} = d\text{RD} = 3.018$$

$$4) d(\text{IOR})E = \min \{d(\text{IOE}, d\text{RE})\} = d\text{RE} = 0.735$$

$$5) d(\text{IOR})F = \min \{d(\text{IOF}, d\text{RF})\} = d\text{IOF} = 4.095$$

$$6) d(\text{IOR})G = \min \{d(\text{IOG}, d\text{RG})\} = d\text{IOG} = 1.768$$

$$7) d(\text{IOR})H = \min \{d(\text{IOH}, d\text{RH})\} = d\text{IOH} = 1.667$$

$$8) d(\text{IOR})K = \min \{d(\text{IOK}, d\text{RK})\} = d\text{IOK} = 1.273$$

$$9) d(\text{IOR})L = \min \{d(\text{IOL}, d\text{RL})\} = d\text{IOL} = 1.577$$

$$10) d(\text{IOR})M = \min \{d(\text{IOM}, d\text{RM})\} = d\text{RM} = 0.210$$

- 11) $d(\text{IOR})N = \min \{d(\text{ION}), d(\text{RN})\} = d(\text{ION}) = 1.610$
- 12) $d(\text{IOR})Q = \min \{d(\text{IOQ}), d(\text{RQ})\} = d(\text{RQ}) = 0.998$
- 13) $d(\text{IOR})S = \min \{d(\text{IOS}), d(\text{RS})\} = d(\text{IOS}) = 3.625$
- 14) $d(\text{IOR})T = \min \{d(\text{IOT}), d(\text{RT})\} = d(\text{IOT}) = 2.469$

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

Tabel 5.9: Matriks Jarak Turunan 5

	CPBJ	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) $d(\text{IORM})\text{CPBJ} = \min \{d(\text{IORCPBJ}), d(\text{MCPBJ})\} = d(\text{IORCPBJ}) = 0.963$
 - 2) $d(\text{IORM})A = \min \{d(\text{IORA}), d(\text{MA})\} = d(\text{IORA}) = 2.114$
 - 3) $d(\text{IORM})D = \min \{d(\text{IORD}), d(\text{MD})\} = d(\text{IORD}) = 3.018$
 - 4) $d(\text{IORM})E = \min \{d(\text{IORE}), d(\text{ME})\} = d(\text{IORE}) = 0.735$
 - 5) $d(\text{IORM})F = \min \{d(\text{IORF}), d(\text{MF})\} = d(\text{IORF}) = 4.095$
 - 6) $d(\text{IORM})G = \min \{d(\text{IORG}), d(\text{MG})\} = d(\text{IORG}) = 1.768$
 - 7) $d(\text{IORM})H = \min \{d(\text{IORH}), d(\text{MH})\} = d(\text{IORH}) = 1.667$
 - 8) $d(\text{IORM})K = \min \{d(\text{IORK}), d(\text{MK})\} = d(\text{IORK}) = 1.273$
 - 9) $d(\text{IORM})L = \min \{d(\text{IORL}), d(\text{ML})\} = d(\text{IORL}) = 1.577$
 - 10) $d(\text{IORM})N = \min \{d(\text{IORN}), d(\text{MN})\} = d(\text{IORN}) = 1.610$

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

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

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

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

Tabel 5.10 : Matrik Perbandingan Turunan 6 (Cluster 6)

	IORM	CPBU	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
CPBU	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 5.11 : 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 5.12 : 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 5.13 : 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 5.14 : 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 5.15 : 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 5.16 : 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 5.17 : 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 5.18 : 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 5.19 : 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 5.20 : 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 5.21 : 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 5.22: Matriks Perbandingan Turunan 18 (Cluster 18)

	DSFGKNHALT	EQCPBJIORM
DSFGKNHALT	0	1.273
EQCPBJIORM	1.273	0

Tabel 5.23 : Matriks Perbandingan Turunan 18 (Cluster 18)

0	DSFGKHALT-EQCPBJIORM
DSFGKHALT-EQCPBJIORM	0

Tabel 5.24 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 5.24 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- GKNHL AT	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 5.25 Hasil *Cluster*

Nomor <i>Cluster</i>	Nirm	Nama
<i>Cluster 0</i>	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
<i>Cluster 1</i>	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?

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

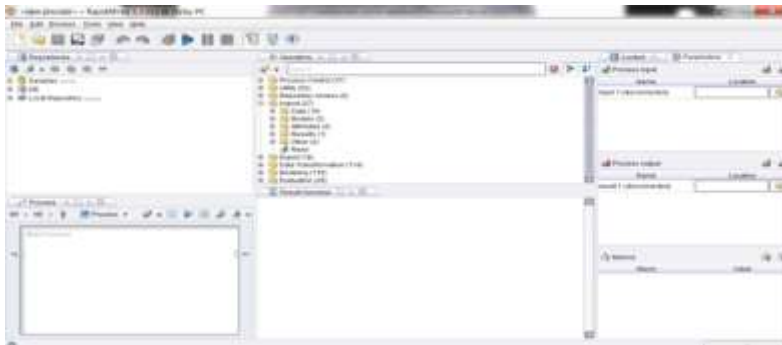
6.1 Pengujian Algoritma C4.5

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 6.1 : 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.



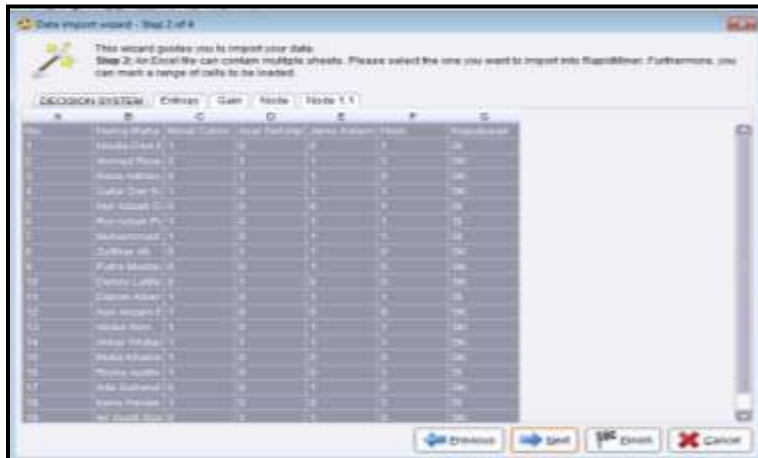
Gambar 6.2 : Tampilan Menu Utama Rapid Miner

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.



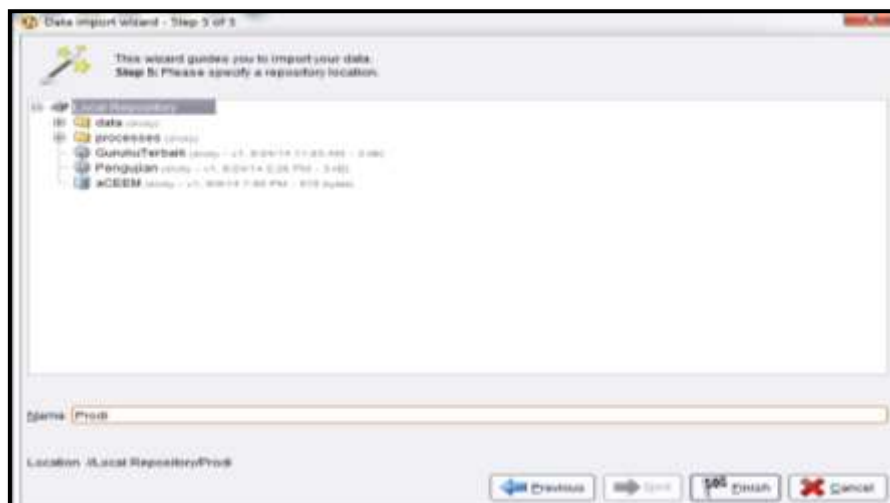
Gambar 6.3 : Proses Awal Untuk Mengimpor Data

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



Gambar 6.6: Decision System

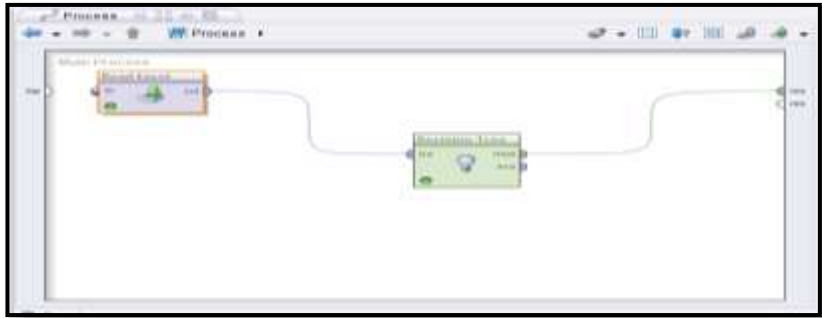
Gambar di atas menjelaskan tentang *Decision Sitem* yang akan diberikan operator untuk pengolahan selanjutnya. Setelah itu *tools* akan memberikan informasi untuk memasukkan *file name repository* dari datanya seperti terlihat pada gambar di bawah ini.



Gambar 6.7: Repositori Data

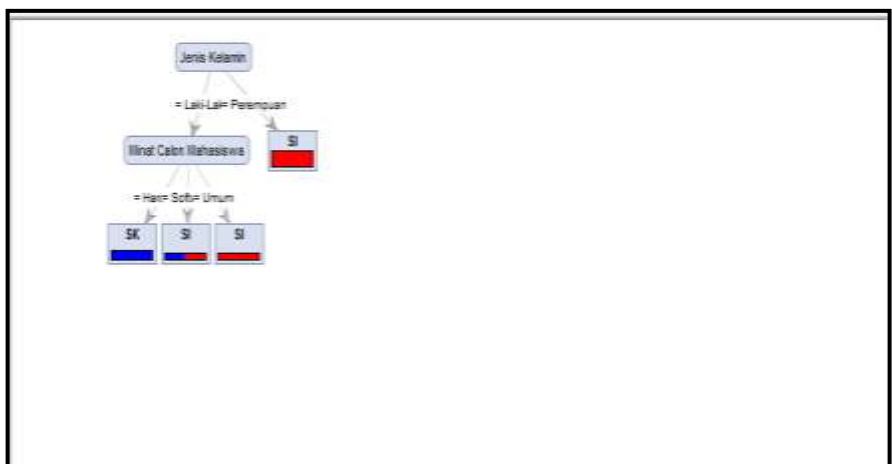
3. Menambahkan Operator

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



Gambar 6.8: Pemrosesan Dan Penambahan Tree

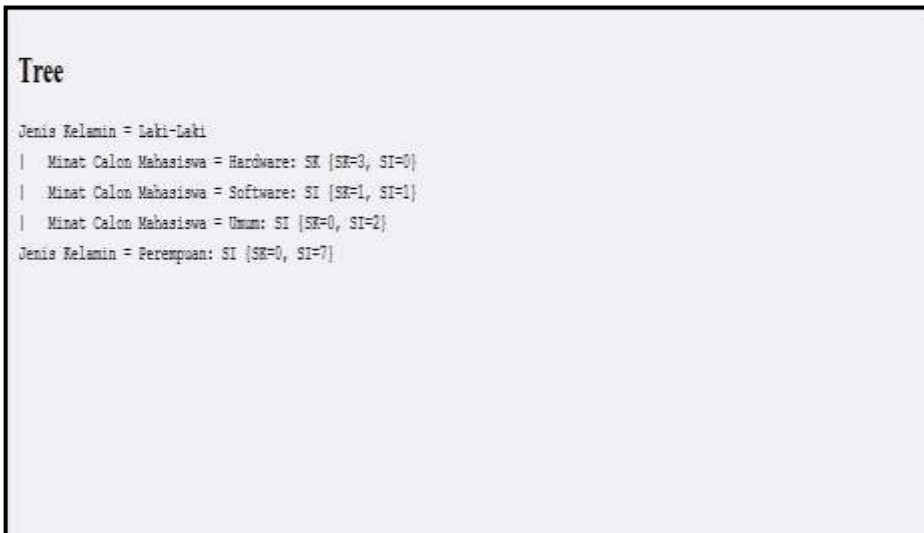
Setelah di bentuk diagram dari proses *decision tree* yang terlihat pada gambar di atas langkah selanjutnya adalah melihat *tree view* atau pohon keputusan dari hasil penelitian seperti terlihat pada gambar di bawah ini:



Gambar 6.9 :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:



Gambar 6.10 :Hasil Pengujian

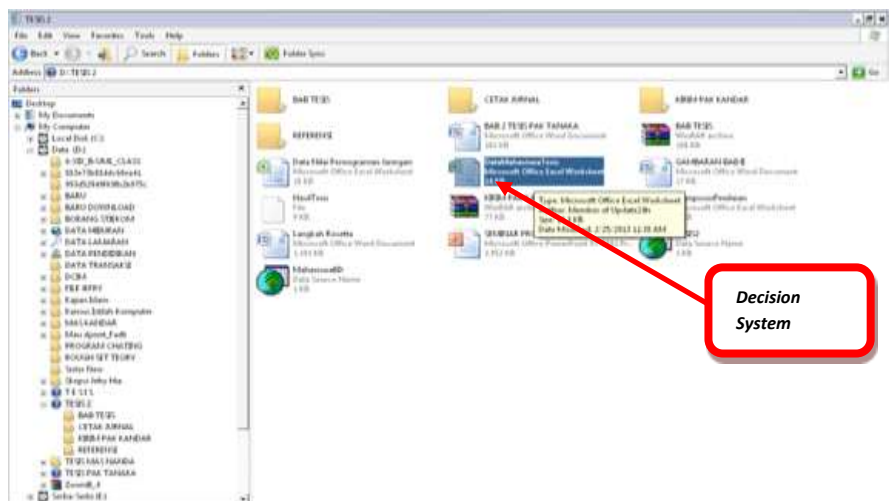
6.2 Pengujian Metode *Rough Set*

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:

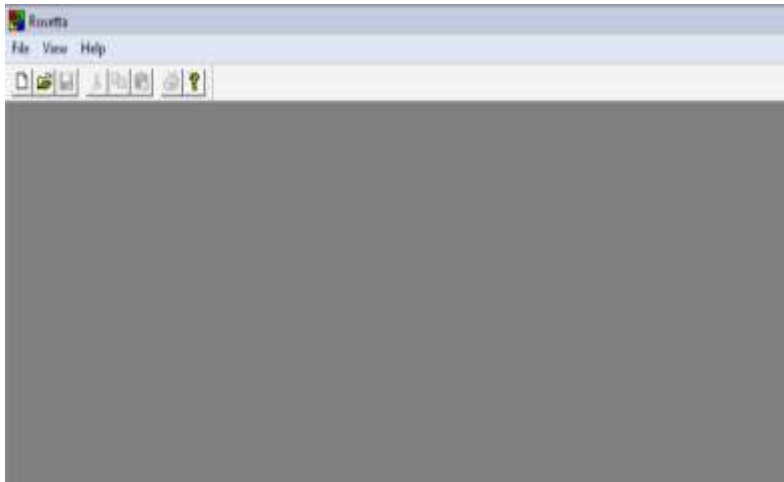
1. Sebelum kita mengimplementasikan sistem terhadap data yang ingin diolah kita harus mempersiapkan terlebih dahulu *Decision System*-nya. Adapun decision system yang digunakan akan di-import terlebih dahulu yang berbasis Microsoft Excel. Setelah kita mentransformasi data sampel di atas ke *File Microsoft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



Gambar 6.11: File Yang Akan Di Proses

2. Setelah data tersebut dipersiapkan kemudian kita meng-*install Rosetta 1.4.41*. Setelah kita melakukan penginstallan *Rosetta 1.4.41* tersebut, kemudian kita jalankan *tools* tersebut.
 - a. Menu Utama

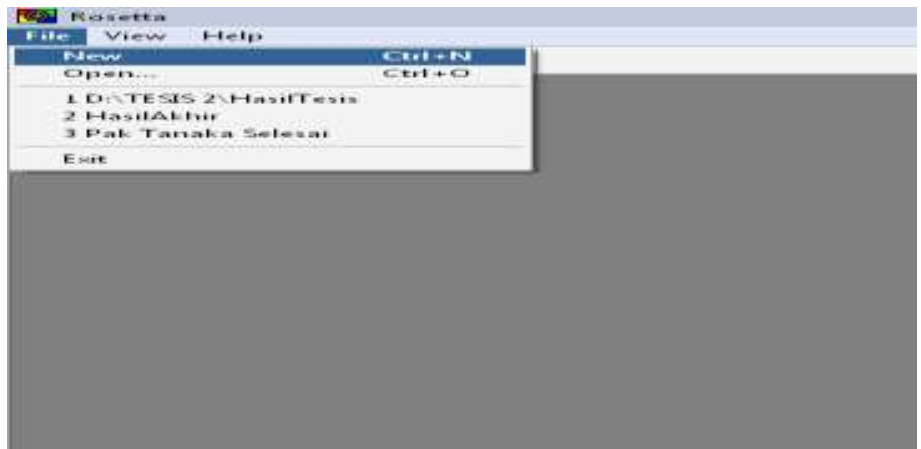
Menu utama dapat dikatakan sebagai antar muka (*user interface*) antara *user* dan program. Menu utama menampilkan pilihan menu yang tersedia pada program. Pada *Rosetta*, tersedia 5 pilihan menu yaitu menu *File*, *Edit*, *View*, *Window* dan *Help*. Gambar untuk tampilan menu utama dapat dilihat pada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



Gambar 6.12 Tampilan Menu Utama

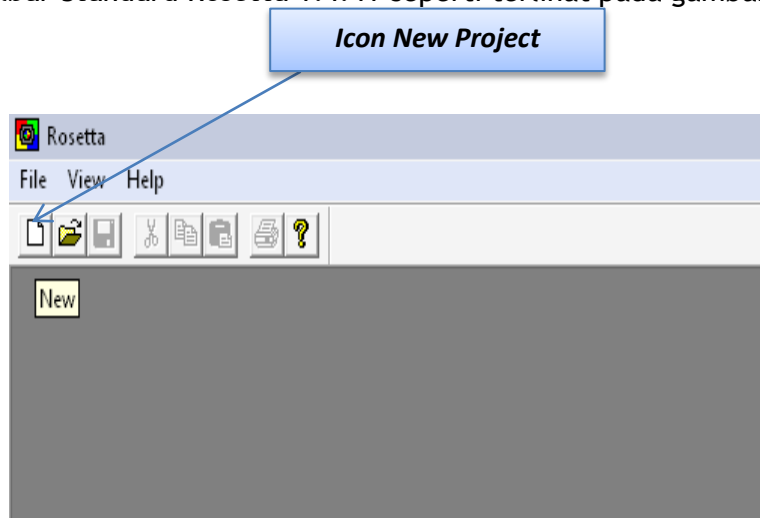
b. *FormNew Project*

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*.



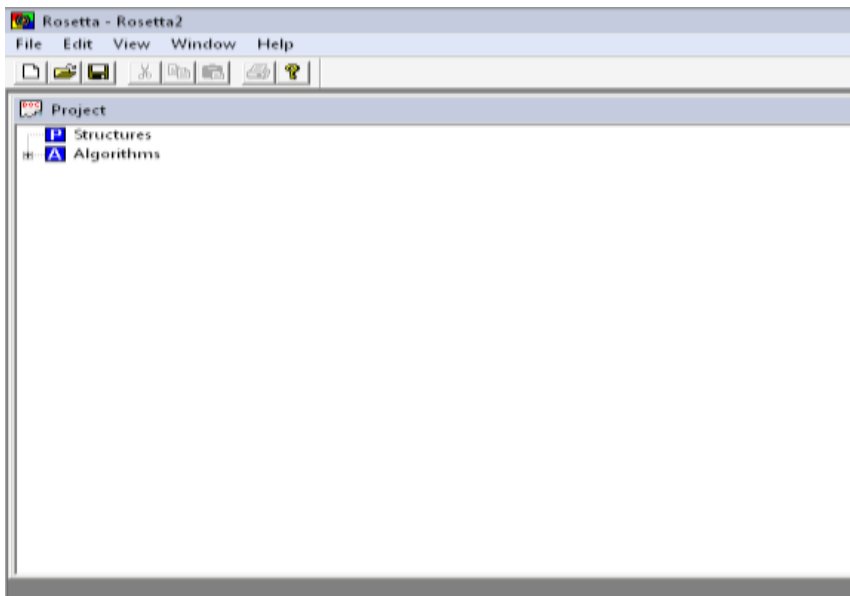
Gambar 6.13 *Create New Project*

Dengan cara lain yaitu dengan memilih *Icon Menu New Pada Toolbar Standard Rosetta 1.4.41* seperti terlihat pada gambar 6.13



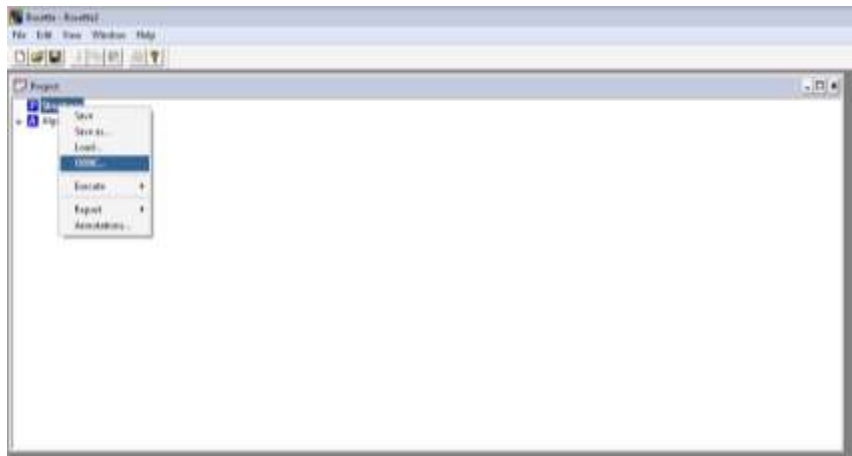
Gambar 6.14 *Create New Project 2nd Way*

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



Gambar 6.15 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 6.15



Gambar 6.16 Langkah Awal Men-Create ODBC

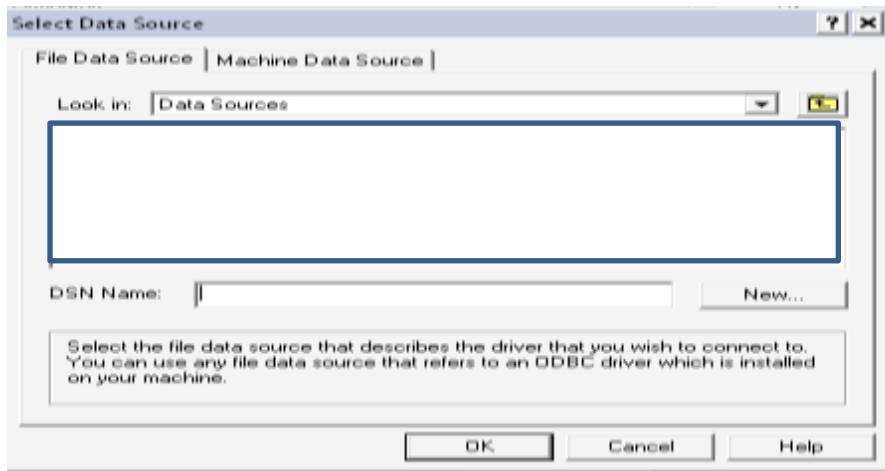
Setelah kita memilih menu *ODBC* kemudian kita melakukan proses Impor *ODBC* seperti pada gambar 5.7.



Gambar 6.17 Impor Drive ODBC

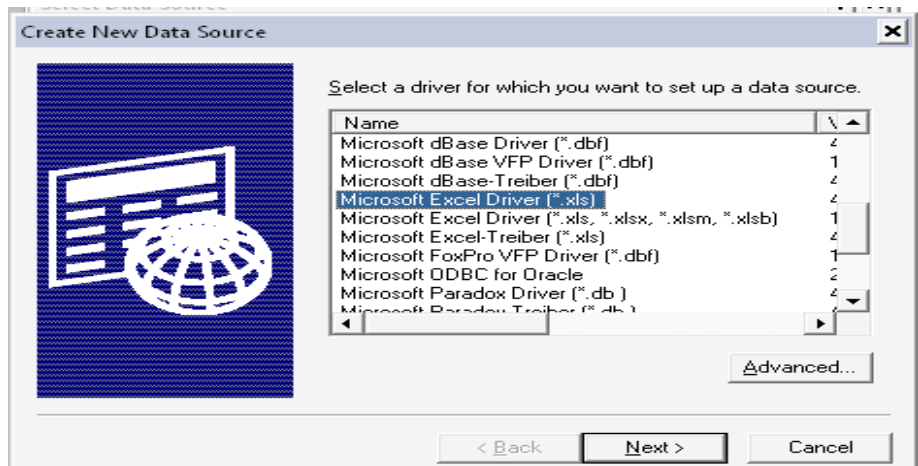
Pada gambar di atas menjelaskan mengenai proses Impor *Driver* ODBC-nya dengan meng-Click *Open Database*. Kemudian langkah selanjutnya kita akan memilih *Data Source* yang akan kita gunakan. Apabila *Data Source*-nya telah ada kita *finding data*

source tersebut pada direktori dimana kita *saving data source* tersebut. Sedangkan apabila *data source* tersebut belum pernah dilakukan pemrosesan, maka kita harus *Create* terlebih dahulu *Data Source*-nya. Adapun caranya yaitu dengan meng-*Click* tombol *New* pada *Window Select Data Source* seperti terlihat pada gambar 6.17.



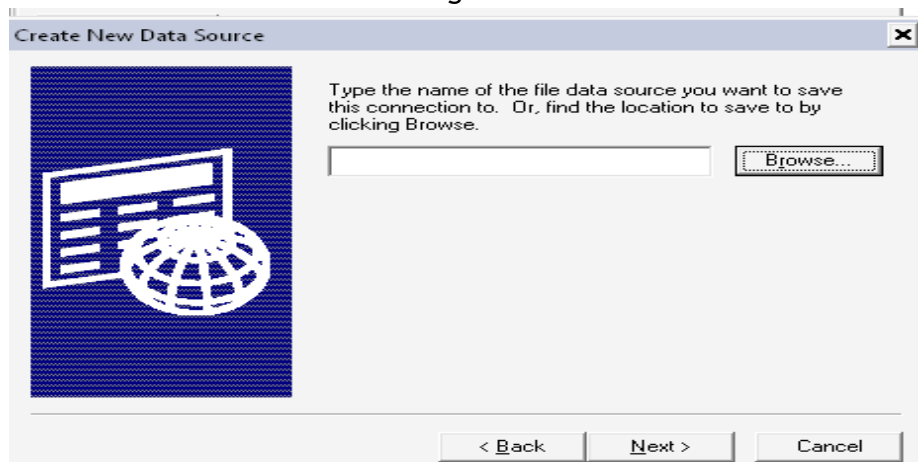
Gambar 6.18 *Window Select Data Source*

Setelah kita meng-*Create Data Source* yang akan kita proses, tahapan selanjutnya adalah memilih *driver* mana yang akan kita gunakan untuk men-*setup data 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 6.18



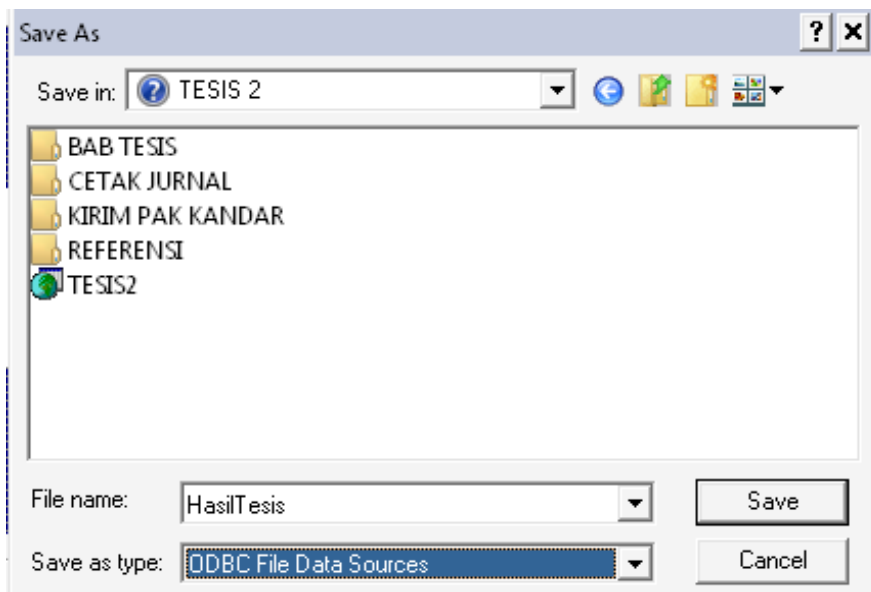
Gambar 6.19 Window Of Create New Data Source

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



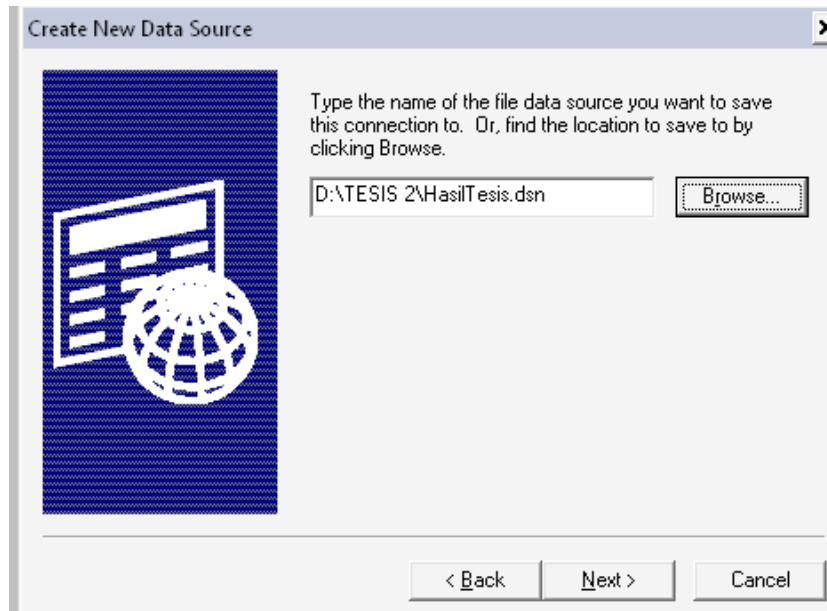
Gambar 6.20 Window Of Browse New Data Source

Setelah terlihat *window* seperti di atas kemudian langkah selanjutnya menekan tombol “Next” setelah kita menekan tombol *Next* maka akan muncul jendela seperti gambar 6.21.



Gambar 6.21 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 6.22



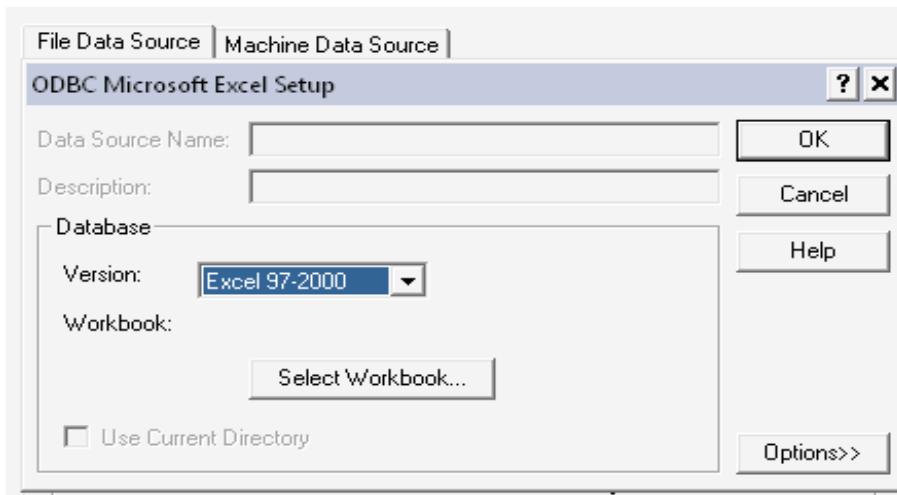
Gambar 6.22 Pemilihan *Data Source*

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 6.23



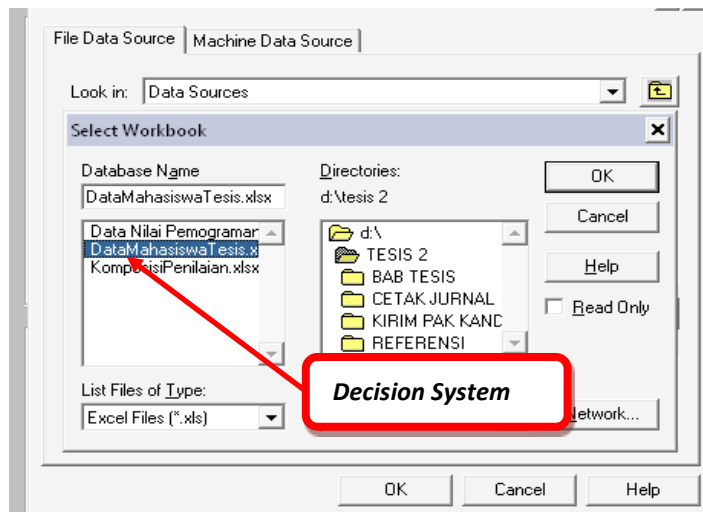
Gambar 6.23 Tahapan Akhir Proses Pemilihan *Data Source*

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 6.24



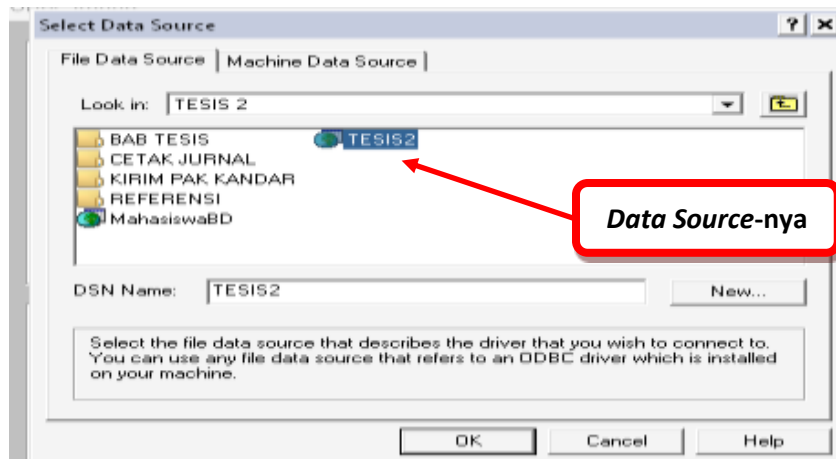
Gambar 6.24 *ODBC Microsoft Excel Setup*

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.



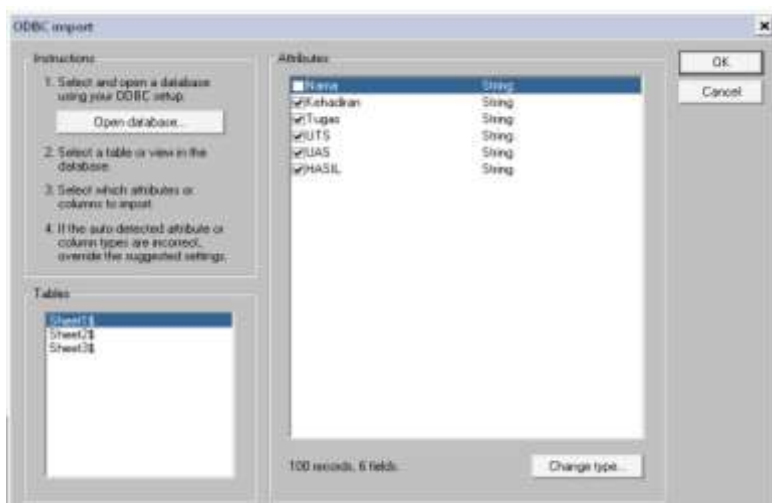
Gambar 6.25 Pemilihan *File Decision System*

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.



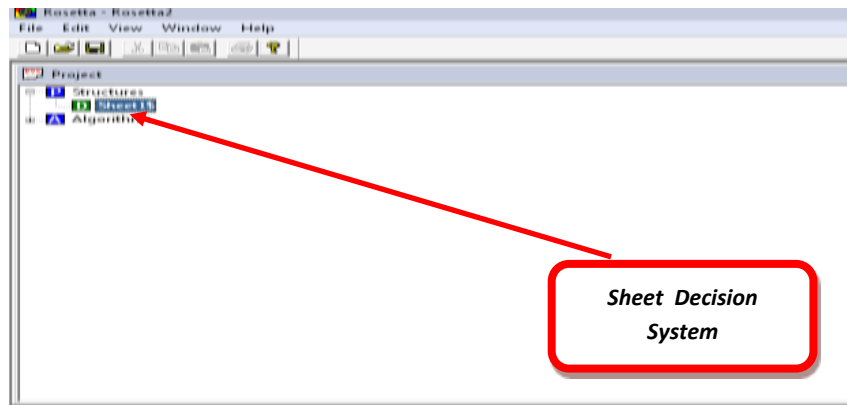
Gambar 6.26 Pemilihan *Data Source* Yang Telah Di *Create*

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 6.27.



Gambar 6.27 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 6.27



Gambar 6.28 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

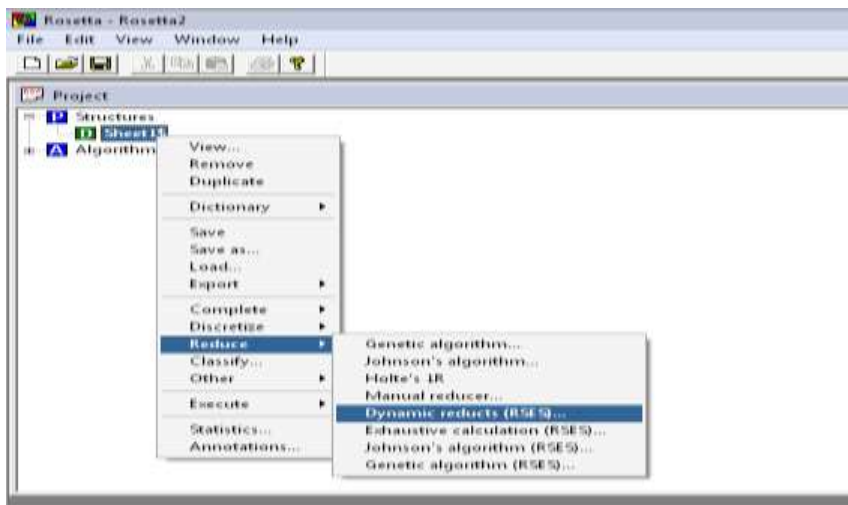
	E2	E3	E4	E5	E6	E7
1	Ade Prita Sengul	Memenuhi Syarat	Salah Prita Prita	Salah Prita Prita	Salah Prita Prita	Salah
2	Adi Prita	Memenuhi Syarat	Salah Prita Prita	Salah Prita Prita	Salah Prita Prita	Salah
3	Adinda Sengul	Memenuhi Syarat	Salah Prita Prita	Salah Prita Prita	Salah Prita Prita	Salah
4	Dani Prita	Memenuhi Syarat	Salah Prita Prita	Salah Prita Prita	Salah Prita Prita	Salah
5	Pati Agustina Sengul	Memenuhi Syarat	Salah Prita Prita	Salah Prita Prita	Salah Prita Prita	Salah
6	Prasasti Sengul Sengul	Memenuhi Syarat	Salah Prita Prita	Salah Prita Prita	Salah Prita Prita	Salah
7	Ira Dewi Sengul	Memenuhi Syarat	Salah Prita Prita	Salah Prita Prita	Salah Prita Prita	Salah
8	Yana Surya Sengul	Memenuhi Syarat	Salah Prita Prita	Salah Prita Prita	Salah Prita Prita	Salah
9	Yana Sengul Sengul	Memenuhi Syarat	Salah Prita Prita	Salah Prita Prita	Salah Prita Prita	Salah
10	Marta Sengul	Memenuhi Syarat	Salah Prita Prita	Salah Prita Prita	Salah Prita Prita	Salah
11	Mahmud Prita	Memenuhi Syarat	Salah Prita Prita	Salah Prita Prita	Salah Prita Prita	Salah
12	Mahmud Prita	Memenuhi Syarat	Salah Prita Prita	Salah Prita Prita	Salah Prita Prita	Salah
13	Marta Sengul	Memenuhi Syarat	Salah Prita Prita	Salah Prita Prita	Salah Prita Prita	Salah
14	Marta Sengul	Memenuhi Syarat	Salah Prita Prita	Salah Prita Prita	Salah Prita Prita	Salah
15	Marta Sengul	Memenuhi Syarat	Salah Prita Prita	Salah Prita Prita	Salah Prita Prita	Salah
16	Marta Sengul	Memenuhi Syarat	Salah Prita Prita	Salah Prita Prita	Salah Prita Prita	Salah
17	Marta Sengul	Memenuhi Syarat	Salah Prita Prita	Salah Prita Prita	Salah Prita Prita	Salah
18	Marta Sengul	Memenuhi Syarat	Salah Prita Prita	Salah Prita Prita	Salah Prita Prita	Salah
19	Marta Sengul	Memenuhi Syarat	Salah Prita Prita	Salah Prita Prita	Salah Prita Prita	Salah
20	Marta Sengul	Memenuhi Syarat	Salah Prita Prita	Salah Prita Prita	Salah Prita Prita	Salah
21	Marta Sengul	Memenuhi Syarat	Salah Prita Prita	Salah Prita Prita	Salah Prita Prita	Salah
22	Marta Sengul	Memenuhi Syarat	Salah Prita Prita	Salah Prita Prita	Salah Prita Prita	Salah
23	Marta Sengul	Memenuhi Syarat	Salah Prita Prita	Salah Prita Prita	Salah Prita Prita	Salah
24	Marta Sengul	Memenuhi Syarat	Salah Prita Prita	Salah Prita Prita	Salah Prita Prita	Salah
25	Marta Sengul	Memenuhi Syarat	Salah Prita Prita	Salah Prita Prita	Salah Prita Prita	Salah
26	Marta Sengul	Memenuhi Syarat	Salah Prita Prita	Salah Prita Prita	Salah Prita Prita	Salah
27	Marta Sengul	Memenuhi Syarat	Salah Prita Prita	Salah Prita Prita	Salah Prita Prita	Salah
28	Marta Sengul	Memenuhi Syarat	Salah Prita Prita	Salah Prita Prita	Salah Prita Prita	Salah
29	Marta Sengul	Memenuhi Syarat	Salah Prita Prita	Salah Prita Prita	Salah Prita Prita	Salah
30	Marta Sengul	Memenuhi Syarat	Salah Prita Prita	Salah Prita Prita	Salah Prita Prita	Salah
31	Marta Sengul	Memenuhi Syarat	Salah Prita Prita	Salah Prita Prita	Salah Prita Prita	Salah
32	Marta Sengul	Memenuhi Syarat	Salah Prita Prita	Salah Prita Prita	Salah Prita Prita	Salah
33	Marta Sengul	Memenuhi Syarat	Salah Prita Prita	Salah Prita Prita	Salah Prita Prita	Salah
34	Marta Sengul	Memenuhi Syarat	Salah Prita Prita	Salah Prita Prita	Salah Prita Prita	Salah
35	Marta Sengul	Memenuhi Syarat	Salah Prita Prita	Salah Prita Prita	Salah Prita Prita	Salah
36	Marta Sengul	Memenuhi Syarat	Salah Prita Prita	Salah Prita Prita	Salah Prita Prita	Salah
37	Marta Sengul	Memenuhi Syarat	Salah Prita Prita	Salah Prita Prita	Salah Prita Prita	Salah
38	Marta Sengul	Memenuhi Syarat	Salah Prita Prita	Salah Prita Prita	Salah Prita Prita	Salah
39	Marta Sengul	Memenuhi Syarat	Salah Prita Prita	Salah Prita Prita	Salah Prita Prita	Salah
40	Marta Sengul	Memenuhi Syarat	Salah Prita Prita	Salah Prita Prita	Salah Prita Prita	Salah
41	Marta Sengul	Memenuhi Syarat	Salah Prita Prita	Salah Prita Prita	Salah Prita Prita	Salah
42	Marta Sengul	Memenuhi Syarat	Salah Prita Prita	Salah Prita Prita	Salah Prita Prita	Salah
43	Marta Sengul	Memenuhi Syarat	Salah Prita Prita	Salah Prita Prita	Salah Prita Prita	Salah
44	Marta Sengul	Memenuhi Syarat	Salah Prita Prita	Salah Prita Prita	Salah Prita Prita	Salah
45	Marta Sengul	Memenuhi Syarat	Salah Prita Prita	Salah Prita Prita	Salah Prita Prita	Salah
46	Marta Sengul	Memenuhi Syarat	Salah Prita Prita	Salah Prita Prita	Salah Prita Prita	Salah
47	Marta Sengul	Memenuhi Syarat	Salah Prita Prita	Salah Prita Prita	Salah Prita Prita	Salah
48	Marta Sengul	Memenuhi Syarat	Salah Prita Prita	Salah Prita Prita	Salah Prita Prita	Salah
49	Marta Sengul	Memenuhi Syarat	Salah Prita Prita	Salah Prita Prita	Salah Prita Prita	Salah
50	Marta Sengul	Memenuhi Syarat	Salah Prita Prita	Salah Prita Prita	Salah Prita Prita	Salah

Gambar 6.29 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.

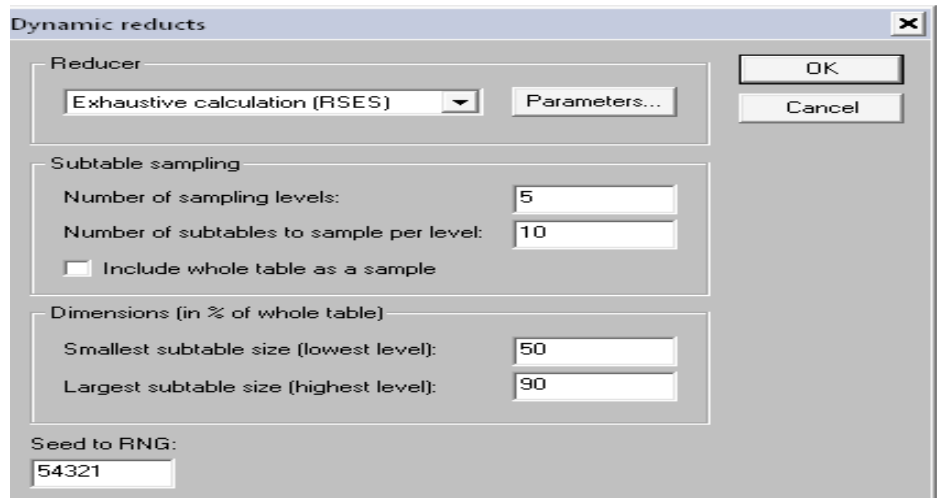
c. Reduction Process

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 6.30



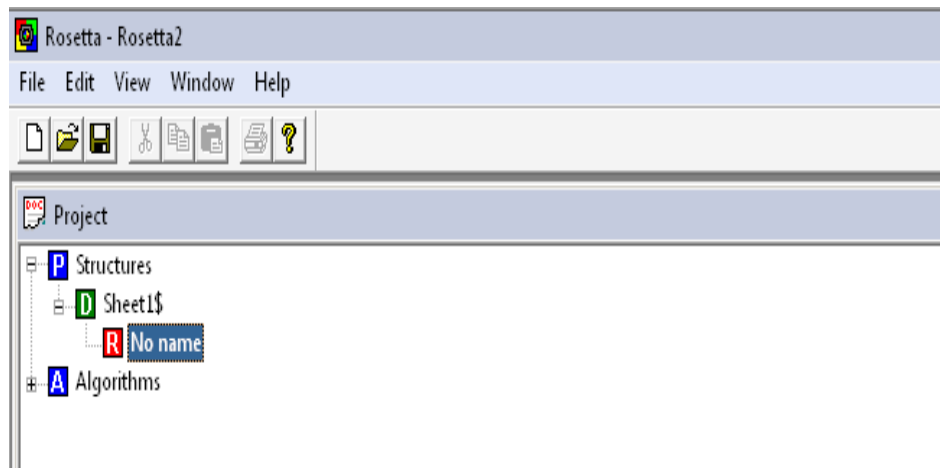
Gambar 6.31 Process Of Dynamic Reduct

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 6.31.



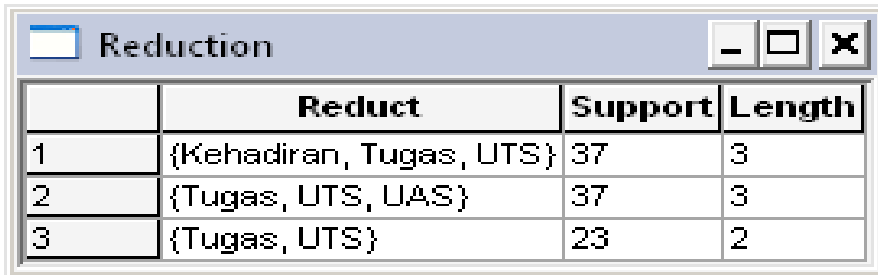
Gambar 6.32 Window Of Dynamic Reduct

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 6.33.



Gambar 6.33 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 6.34.

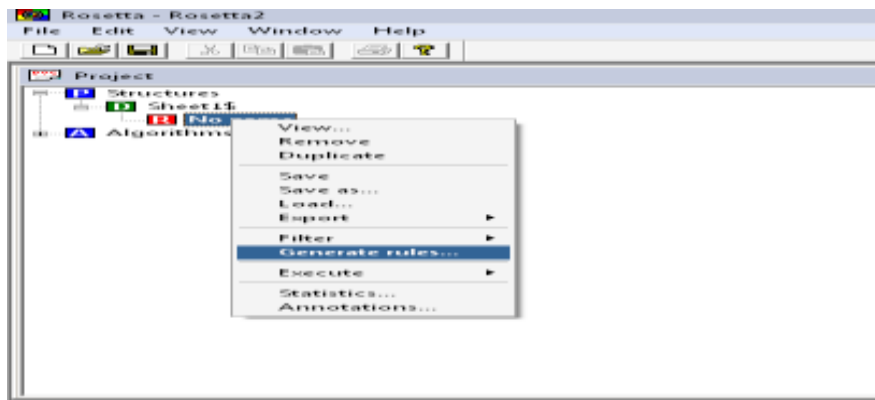


	Reduct	Support	Length
1	{Kehadiran, Tugas, UTS}	37	3
2	{Tugas, UTS, UAS}	37	3
3	{Tugas, UTS}	23	2

Gambar 6.34 Result Of Reduction

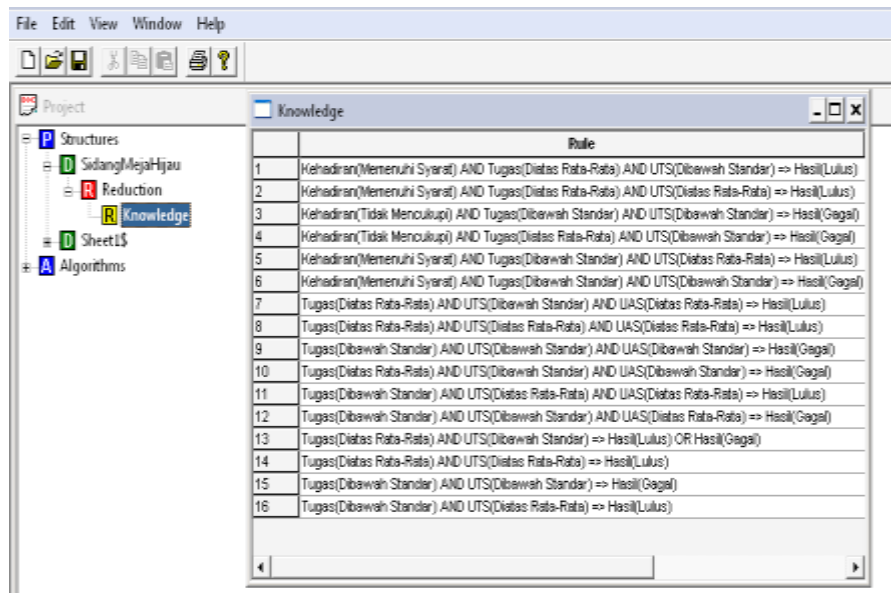
d. General Rules

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 6.35



Gambar 6.35 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 6.36



Gambar 6.36 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 proses di atas.

6.3 Pengujian Teknik Single Linkage

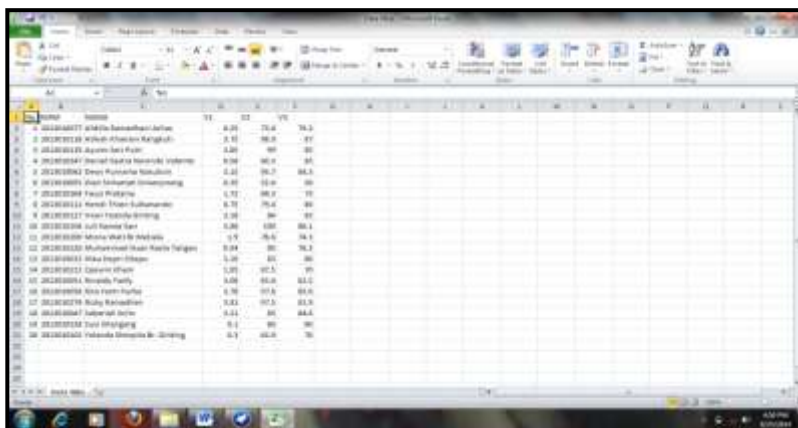
Dalam pengujian kasus ini penulis menggunakan Perangkat Lunak Weka 3.6.7. Tujuan dari pengujian ini apakah hasil analisa tersebut diatas sesuai dengan hasil pengujian perangkat lunaknya.

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 6.37.



Nama	IPK	Absensi	Etika
Wahid Husein	8,20	75,0	75,0
Agus	8,10	80,0	80,0
Fauzi	8,00	80,0	80,0
...
...
...
...
...
...
...
...
...
...
...
...
...
...
...
...
...
...
...

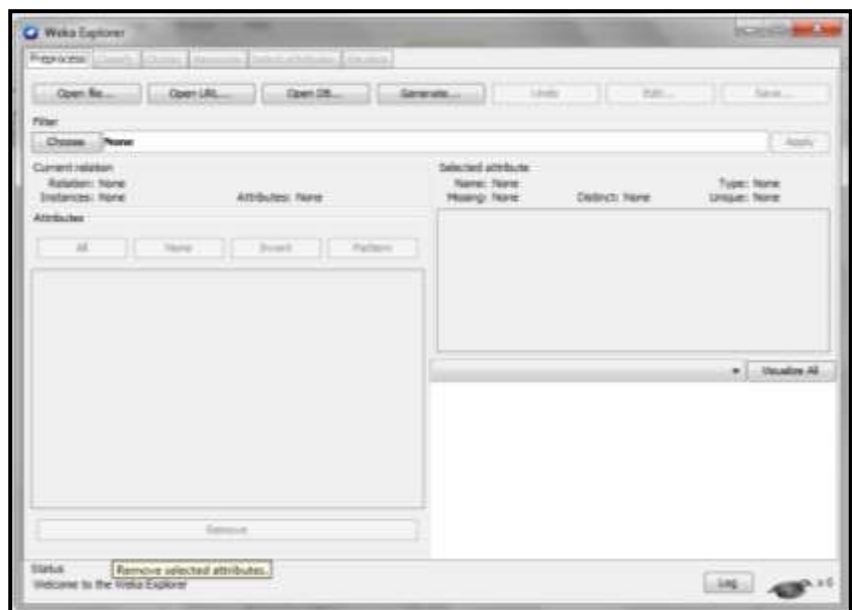
Gambar 6.37 Data Nilai Awal Mahasiswa

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



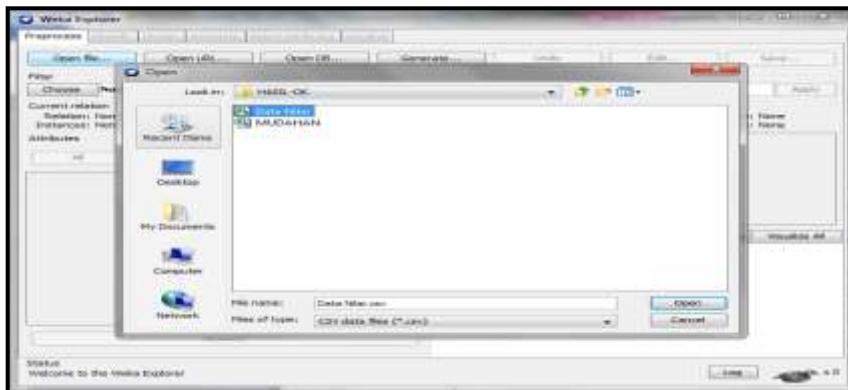
Gambar 6.38 Tampilan Awal Aplikasi WEKA

Kemudian klik *Explorer*, akan muncul gambar 6.39.



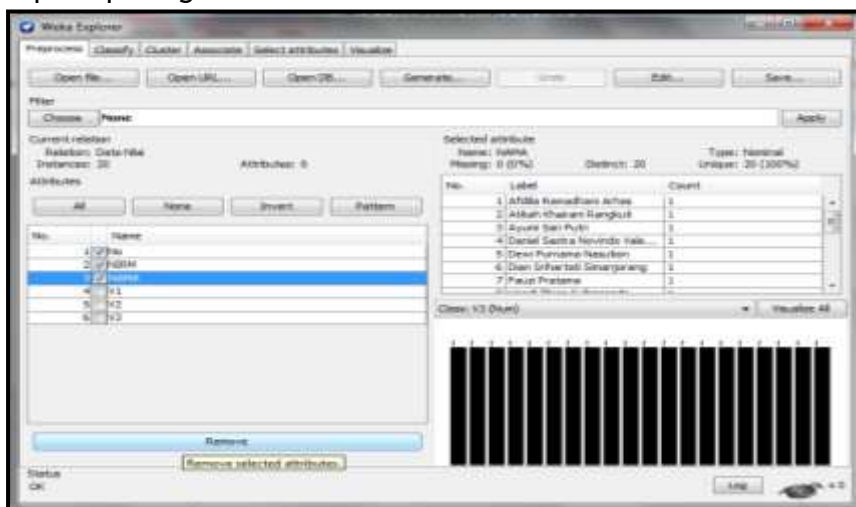
Gambar 6.39 Tampilan WEKA Explorer

- b. Kemudian buka file data nilai mahasiswa yang sudah di save tadi dengan *typeMicrosoft ExcelCSV* untuk dimasukkan ke dalam aplikasi WEKA seperti pada gambar 6.40



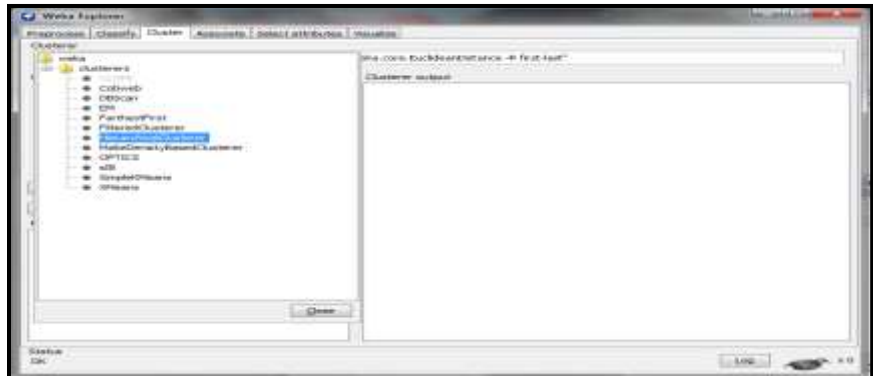
Gambar 6.40 Proses Penginputan Data pada WEKA

- c. Kemudian pilih attributes V1, V2, V2, selanjutnya klik *remove*, seperti pada gambar 6.41



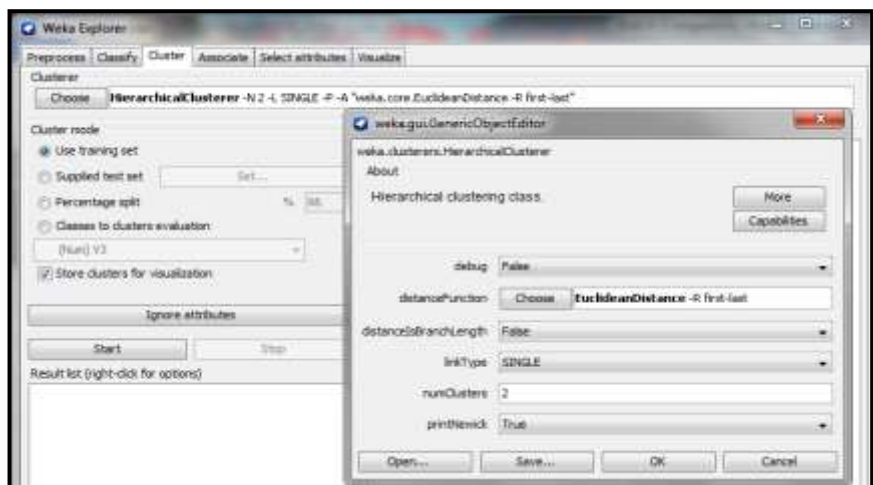
Gambar 6.41 Proses Selected Attributes pada WEKA

- d. Kemudian klik menu *cluster*, pilih *choose* dan pilih *hierarchicalclusterer* seperti pada gambar 6.42



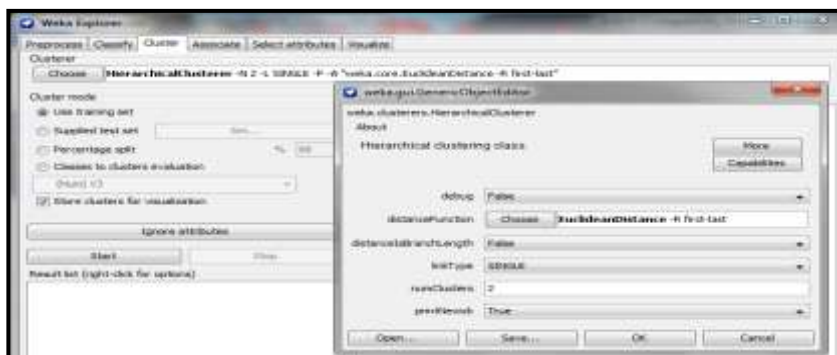
Gambar 6.42 : Proses Menentukan Metode Pada Weka

- e. Maka langkah yang dilakukan selanjutnya yaitu *double* klik *hierarchicalclusterer* yang sudah dipilih, kemudian akan muncul seperti pada gambar 6.43



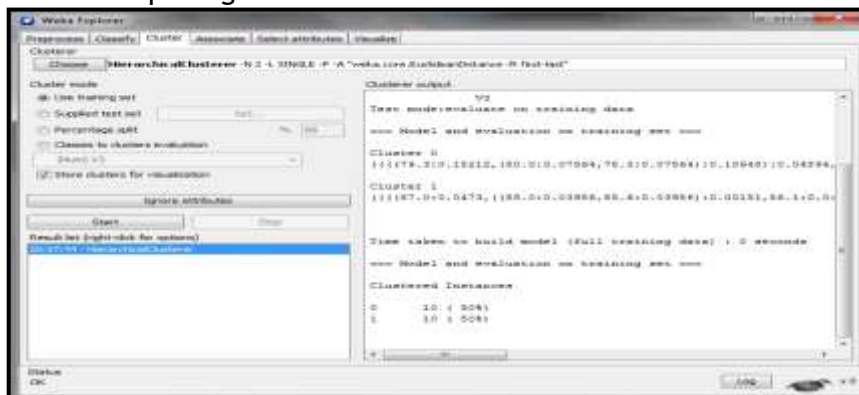
Gambar 6.43 Tampilan WEKA untuk Menentukan *EuclideanDistance*

- 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 6.44 Proses Menentukan *EuclideanDistance*

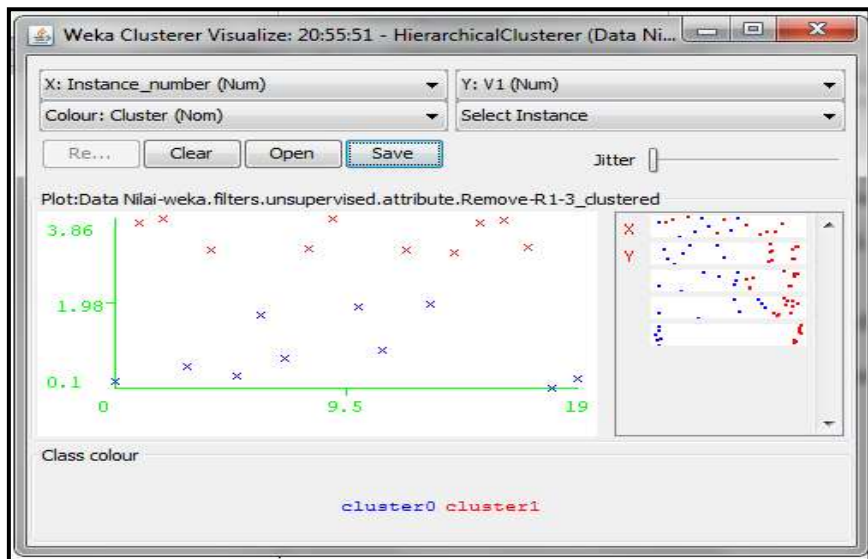
- g. Kemudian klik *start*, dan kemudian akan muncul seperti yang dilihat pada gambar 6.45.



Gambar 6.45 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 6.46



Gambar 6.46 : Proses untuk 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 6.47.



Gambar 6.47 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 muncul hasil *cluster* yang diinginkan seperti pada gambar 6.48.

id	kecepatan_kemudi	Waktu	Waktu	Waktu	Waktu
1	10.0	10.0	10.0	10.0	10.0
2	10.0	10.0	10.0	10.0	10.0
3	10.0	10.0	10.0	10.0	10.0
4	10.0	10.0	10.0	10.0	10.0
5	10.0	10.0	10.0	10.0	10.0
6	10.0	10.0	10.0	10.0	10.0
7	10.0	10.0	10.0	10.0	10.0
8	10.0	10.0	10.0	10.0	10.0
9	10.0	10.0	10.0	10.0	10.0
10	10.0	10.0	10.0	10.0	10.0
11	10.0	10.0	10.0	10.0	10.0
12	10.0	10.0	10.0	10.0	10.0
13	10.0	10.0	10.0	10.0	10.0
14	10.0	10.0	10.0	10.0	10.0
15	10.0	10.0	10.0	10.0	10.0
16	10.0	10.0	10.0	10.0	10.0
17	10.0	10.0	10.0	10.0	10.0
18	10.0	10.0	10.0	10.0	10.0
19	10.0	10.0	10.0	10.0	10.0
20	10.0	10.0	10.0	10.0	10.0
21	10.0	10.0	10.0	10.0	10.0
22	10.0	10.0	10.0	10.0	10.0
23	10.0	10.0	10.0	10.0	10.0
24	10.0	10.0	10.0	10.0	10.0
25	10.0	10.0	10.0	10.0	10.0
26	10.0	10.0	10.0	10.0	10.0
27	10.0	10.0	10.0	10.0	10.0
28	10.0	10.0	10.0	10.0	10.0
29	10.0	10.0	10.0	10.0	10.0
30	10.0	10.0	10.0	10.0	10.0
31	10.0	10.0	10.0	10.0	10.0
32	10.0	10.0	10.0	10.0	10.0
33	10.0	10.0	10.0	10.0	10.0
34	10.0	10.0	10.0	10.0	10.0
35	10.0	10.0	10.0	10.0	10.0
36	10.0	10.0	10.0	10.0	10.0
37	10.0	10.0	10.0	10.0	10.0
38	10.0	10.0	10.0	10.0	10.0
39	10.0	10.0	10.0	10.0	10.0
40	10.0	10.0	10.0	10.0	10.0
41	10.0	10.0	10.0	10.0	10.0
42	10.0	10.0	10.0	10.0	10.0
43	10.0	10.0	10.0	10.0	10.0
44	10.0	10.0	10.0	10.0	10.0
45	10.0	10.0	10.0	10.0	10.0
46	10.0	10.0	10.0	10.0	10.0
47	10.0	10.0	10.0	10.0	10.0
48	10.0	10.0	10.0	10.0	10.0
49	10.0	10.0	10.0	10.0	10.0
50	10.0	10.0	10.0	10.0	10.0
51	10.0	10.0	10.0	10.0	10.0
52	10.0	10.0	10.0	10.0	10.0
53	10.0	10.0	10.0	10.0	10.0
54	10.0	10.0	10.0	10.0	10.0
55	10.0	10.0	10.0	10.0	10.0
56	10.0	10.0	10.0	10.0	10.0
57	10.0	10.0	10.0	10.0	10.0
58	10.0	10.0	10.0	10.0	10.0
59	10.0	10.0	10.0	10.0	10.0
60	10.0	10.0	10.0	10.0	10.0
61	10.0	10.0	10.0	10.0	10.0
62	10.0	10.0	10.0	10.0	10.0
63	10.0	10.0	10.0	10.0	10.0
64	10.0	10.0	10.0	10.0	10.0
65	10.0	10.0	10.0	10.0	10.0
66	10.0	10.0	10.0	10.0	10.0
67	10.0	10.0	10.0	10.0	10.0
68	10.0	10.0	10.0	10.0	10.0
69	10.0	10.0	10.0	10.0	10.0
70	10.0	10.0	10.0	10.0	10.0
71	10.0	10.0	10.0	10.0	10.0
72	10.0	10.0	10.0	10.0	10.0
73	10.0	10.0	10.0	10.0	10.0
74	10.0	10.0	10.0	10.0	10.0
75	10.0	10.0	10.0	10.0	10.0
76	10.0	10.0	10.0	10.0	10.0
77	10.0	10.0	10.0	10.0	10.0
78	10.0	10.0	10.0	10.0	10.0
79	10.0	10.0	10.0	10.0	10.0
80	10.0	10.0	10.0	10.0	10.0
81	10.0	10.0	10.0	10.0	10.0
82	10.0	10.0	10.0	10.0	10.0
83	10.0	10.0	10.0	10.0	10.0
84	10.0	10.0	10.0	10.0	10.0
85	10.0	10.0	10.0	10.0	10.0
86	10.0	10.0	10.0	10.0	10.0
87	10.0	10.0	10.0	10.0	10.0
88	10.0	10.0	10.0	10.0	10.0
89	10.0	10.0	10.0	10.0	10.0
90	10.0	10.0	10.0	10.0	10.0
91	10.0	10.0	10.0	10.0	10.0
92	10.0	10.0	10.0	10.0	10.0
93	10.0	10.0	10.0	10.0	10.0
94	10.0	10.0	10.0	10.0	10.0
95	10.0	10.0	10.0	10.0	10.0
96	10.0	10.0	10.0	10.0	10.0
97	10.0	10.0	10.0	10.0	10.0
98	10.0	10.0	10.0	10.0	10.0
99	10.0	10.0	10.0	10.0	10.0
100	10.0	10.0	10.0	10.0	10.0

Gambar 6.49 Tampilan Hasil Cluster

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

6.4 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 6.50

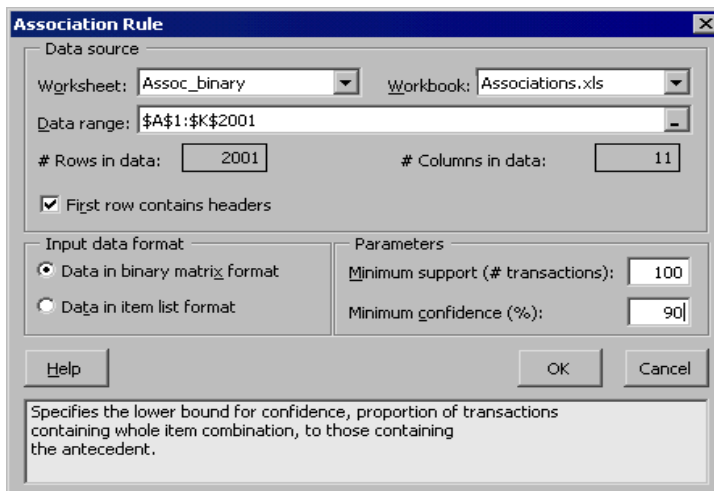
	A	B	C	D	E	F
2						
3	Transaksi ke-	Item Belanja yang dibeli				
4	1	Kecap sedap	mie sedap kare kental spesial	mie sedap goreng		
5	2	Ale-ale apel fuji	mie sedap soto	mie sedap goreng		
6	3	mie sedap goreng	Tea-Jus	mie sedap soto		
7	4	Mie sedap kare kental spesial	mie sedap goreng	Tea-Jus	JasJus	
8	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						

2. Tampilan Aftinity Association Rule XI Miner

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

- [illegible]

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 6.52




Gambar 6.52 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.

	B	C	D	E	F	G
1						
2						
3						
4						
5						
6						
7						
8						
9						
10						
11						
12						
13						
14						
15						
16						
17						
18						

Gambar 6.53 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 confidence* yang diinginkan. Lihat gambar dibawah ini:

Gambar 6.54 Tampilan Input *Minimum Support* Dan *Confidence*

d. Klik OK jika telah selesai maka akan muncul layout hasil analisa tampak pada gambar 6.55 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 6.55 berikut:

XLMiner : Association Rules

Data	
Input Data	Grawo20054 SP517
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 (x)	Consequent (y)	Support(x)	Support(y)	Support (x,y)	Lift Ratio
1	100	jaksa, Ma sedang sakit	Ma-ak apri fuj	3	4	3	3.25
2	100	Ma sedang sakit	Ma-ak apri fuj	3	4	3	3.25
3	100	Ma sedang sakit	jaksa, Ma-ak apri fuj	3	4	3	3.25
4	75	jaksa, Ma-ak apri fuj	Ma sedang sakit	4	3	3	3.25
5	75	Ma-ak apri fuj	jaksa, Ma sedang sakit	4	3	3	3.25
6	75	Ma-ak apri fuj	Ma sedang sakit	4	3	3	3.25
7	100	jaksa sedang	jaksa	3	3	3	1.444444
8	100	Ma-ak apri fuj	jaksa	4	3	4	1.444444
9	100	Ma sedang sakit	jaksa	3	3	3	1.444444
10	100	Ma sedang sakit, Ma-ak apri fuj	jaksa	3	3	3	1.444444
11	50	Ma sedang garing	Ma-Jak	4	3	2	1.333333
12	50	Ma-Jak	Ma sedang garing	4	3	2	1.333333
13	50	Ma sedang garing	jaksa	4	3	2	0.722222
14	50	Ma-Jak	jaksa	4	3	2	0.722222

Gambar 6.55 Keluaran Dari Algoritma Apriori

LATIHAN:

1. Lakukanlah pengujian untuk Sampel Data Pada Tabel 3.1 dan 3.22 menggunakan Aplikasi Rapid Miner 6.3
2. Berdasarkan Hasil Pengujian Apakah perbedaan mendasarkan antara Aplikasi Rapid Miner 6.3, Weka dan Rough Set dalam proses Pengujian sebuah sampel kasus.

DAFTAR PUSTAKA

Bonczek, dkk., 1980. *Decision Support Systems And Intelligent Systems*

Larose, Daniel T. 2005, *Discovering Knowledge In Data: An Introduction To Data Mining*. John Willey & Sons. Inc.

Fajar Astuti Hermawati, 2013, *Data Mining*. Yogyakarta : Andi Offset

Kusrini, dkk., 2009. *Algoritma Data Mining*. Yogyakarta: Andi Offset

Sani Susanto, dkk., 2010. *Pengantar Data Mining Menggali Pengetahuan dari Bongkahan Data*. Yogyakarta : Andi Offset

GLOSARIUM

Agglomerative Method adalah Metode Pendekatan Untuk Pembentukan Klaster sendiri dan berpasangan sebagai salah satu gerak naik Hirarki

Algoritma adalah Prosedur langkah-demi-langkah untuk penghitungan dan metode efektif diekspresikan sebagai rangkaian terbatas

Asosiasi adalah Salah satu fungsi matematika yang mendasari aturan berkelempok

Atribut adalah sifat yg menjadi ciri khas (suatu benda atau orang)

Basis Data adalah Kumpulan data yang salah berhubungan satu dengan yang lainnya

Bobot adalah Kadar atau nilai sebuah objek terhadap sebuah kasus

Data adalah Kumpulan Fakta yang terekam atau sebuah entitas yang tidak memiliki arti dan selama ini terabaikan

Data Mining adalah Metode penggalian dari bongkahan data sehingga menemukan

Definitif adalah Arti atau makna dari sebuah kata

Diagnosis adalah penentuan jenis penyakit dengan cara meneliti (memeriksa) gejala-gejalanya

Ekstrak adalah Proses transformasi sebuah data

Entropy adalah jumlah bit yang diperkirakan dibutuhkan untuk dapat mengekstrak suatu kelas (+ atau -) dari sejumlah data acak pada ruang sampel S

Estimasi adalah Perkiraan, Penilaian dan Pendapat sebuah kasus

Gain adalah Perolehan informasi dari atribut A relative terhadap output data S

Geodesi adalah cabang geologi yang menyelidiki ukuran dan bangun bumi dan ilmu mengukur tanah

Geologi adalah ilmu tentang komposisi, struktur, dan sejarah bumi

Hirarki adalah Tingkatan dari sebuah objek

Input adalah Sebuah masukan dari sebuah permasalahan

Klasifikasi adalah Proses pengelompokan sebuah data

Klaster adalah Pengelompokan dari sebuah objek berdasarkan kriteria dan keluaran

Koefisien adalah bagian suku yg berupa bilangan atau konstan, biasanya dituliskan sebelum lambang peubah seperti angka

Konstanta adalah lambang untuk menyatakan objek yg sama dalam keseluruhan operasi matematika

Mayor adalah Faktor yang utama dan yang besar

Minor adalah Objek yang menjadi alternatif

Nominal adalah menurut yg tercatat atau apa yg tertulis saja yang bersifat nilai dan harga

Numerik adalah Yang bersifat angka atau sistem angka

Operator adalah orang yg bertugas menjaga, melayani, dan menjalankan suatu peralatan, mesin, telepon, radio, dsb

Peubah adalah simbol yg digunakan untuk menyatakan unsur yg tidak tentu dl suatu himpunan atau besaran yg bervariasi atau besaran yg dapat mengambil salah satu dr suatu himpunan nilai tertentu (dl matematika)

Pola adalah bentuk (struktur) yg tetap:

Prediksi adalah ramalan; prakiraan:

Sampel adalah sesuatu yg digunakan untuk menunjukkan sifat suatu kelompok yg lebih besar atau bagian kecil yg mewakili kelompok atau keseluruhan yg lebih besar; percontoh

Seismologi adalah ilmu tentang gempa bumi

Siklus adalah putaran waktu yg di dalamnya terdapat rangkaian kejadian yg berulang-ulang secara tetap dan teratur; daur;

Standar Deviasi adalah Nilai Standarisasi dari kumpulan data angka

Tipografi adalah ilmu cetak; seni percetakan

Variabel adalah sesuatu yg dapat berubah; faktor atau unsur yg ikut menentukan perubahan

INDEKS

Analytical Hierarchy Clustering, 12
Asosiasi, 10
Block, 68
CART (Classification And Regression Tree), 20
Chebychev, 68
Chi-Square, 68
Class, 8
Cleaning, 6
Cluster, 65
Cuaca, 2
Data, 1
Data Mining, 1
Data Reduction, 24
Data Transformation, 24
Decision System, 50
Determinasi Koefisien, 18
Direktori, 100
Discernibility Matrix, 50
Discernibility Matrix Modulo D, 50
Emas, 3
Entitas, 7
Entropy, 21
Estimasi, 10
Euclidean Distance, 68
Euclidean Single Linkage, 78
Gain, 21
General Rules, 51
Gold Mining, 3
Hamming, 68
Hipotesa, 44
Homogen, 66
ID3, 20
Klasifikasi, 10
Label, 8
Minimum Confidence, 131

Minimum Support, 131
Mining, 1
Minkowski, 68
Minor, 13
Naïve Bayesian Clasification, 20
Otomatis, 1
Output, 1
Peluang, 45
Person Correlation, 68
Phi-Square, 68
Pola, 1
Prediksi, 49
Preprocessing, 6
Rapid Miner, 98
Reduksi Data, 66
Regresi Linier, 12
Rough Set, 49
Segmen, 65
Seleksi Data, 5
Squared Euclidean Distance, 68
Standar Deviasi, 71
Statistik, 2
Supervised Learning, 11
Target, 8
Transformasi, 6
Unsupervised Learning, 11
variabel bebas, 14
variabel terikat, 14
Visualisasi, 2
Weka, 120
XL Miner, 129



Nama Saya : Dicky Nofriansyah, S.Kom., M.Kom Saya dilahirkan di Medan Tanggal 31 Oktober 1989 anak ke-4 dari 4 bersaudara, dari Seorang Ibu yang bernama : Hj.Rodiah dan Ayah : H.Syamsul Bachri (alm). Dan telah menikah dengan seorang wanita bernama: Febriani Sartika, S.Kom dan telah dikarunia Putera: Assyadil Dzikri Nofriansyah. Saya Lulus dari Program Studi Strata -1 (S1) yaitu tahun 2011 di STMIK Budidarma Medan, dan Pascasarjana Magister Ilmu Komputer (S2) di Universitas Putera Indonesia YPTK Padang. Pekerjaan saya saat ini adalah Dosen Tetap di STMIK

Triguna Dharma Medan sekaligus menjabat sebagai Ketua Program Studi (Ka.Prodi) S1-Sistem Komputer sejak 8 September 2013. Motivator terbesar saya saat ini adalah Ibu saya dan Istri saya yang saat ini sedang mengandung Anak Pertama saya. Secara Akademik pencapaian terbesar saya yaitu menjadi Wisudawan Terbaik Program Pascasarjana di Universitas Putera Indonesia YPTK Padang.

Dengan adanya buku ini mudah-mudahan dapat menambah ilmu pengetahuan kepada pembaca dan saya sebagai seorang penulis dapat meningkatkan kualitas bahan bacaan yang saya tulis. Aamiin ya Robbal Alamin. Setelah ini akan ada edisi selanjutnya berkenaan tentang Sistem Pendukung Keputusan secara kompleks.

Kata Mutiara :

“Engkau tak dapat meraih ilmu kecuali dengan enam hal yaitu cerdas, selalu ingin tahu, tabah, punya bekal dalam menuntut ilmu, bimbingan dari guru dan dalam waktu yang lama. Maka, Ikatlah ilmu dengan menuliskannya.

(Ali bin Abi Thalib)