

SISTEM PENENTUAN TINGKAT INVESTASI MENGUNAKAN ALGORITMA C 4.5

Yuli Murdianingsih

Manajemen Informatika STMIK Subang, Jawa Barat
Jl.Marsinu no 5 Subang, 40112 telp (0260)417853
e-mail : yuli_murdianingsih@yahoo.com

Abstrak

Selama dua tahun terakhir, di kabupaten Subang telah terjadi peningkatan jumlah investor yang masuk. Jika dibandingkan dengan tahun 2012, nilai investasi sektor industri di kabupaten Subang meningkat hingga mencapai 800 persen pada tahun 2013. Berdasarkan perkembangan investasi di kabupaten Subang diperlukan upaya untuk mengetahui karakteristik tingkat investasi perusahaan berdasarkan pada deskripsi masing-masing perusahaan. seperti dirilis oleh disperindagsar kabupaten Subang setiap perusahaan memiliki jumlah investasi, jumlah tenaga kerja yang berbeda, jumlah produksi, jenis usaha dan status yang berbeda-beda. Dilakukan perhitungan entropy total, entropy atribut jumlah investasi, jumlah tenaga kerja, jumlah produksi, jenis usaha dan status untuk masing-masing nilai atribut tersebut. Dilakukan perhitungan gain untuk masing-masing atribut. Dilakukan perbandingan nilai gain dari masing-masing atribut, sehingga dapat ditentukan root, cabang dan pohon keputusan secara keseluruhan. Setelah diperoleh pohon keputusan lengkap, baru diperoleh aturan sebagai komputasi yang diekstrak dari data penelitian. Aturan direalisasikan dalam view menggunakan DBMS MySQL dan pengantarmukaan sistem menggunakan bahasa pemrograman PHP. Diperoleh sembilan aturan, dimana sistem dapat melakukan penentuan tingkat investasi. Dapat diimplementasikan sistem penentuan tingkat investasi menggunakan algoritma C 4.5 berdasarkan deskripsi perusahaan berupa jumlah investasi, jumlah tenaga kerja, jumlah produksi, jenis usaha dan status.

Kata Kunci : Investasi, Algoritma C 4.5, Entropy, Gain.

1. PENDAHULUAN

Nilai investasi sektor industri meningkat hingga mencapai 800 persen pada tahun 2013 jika dibandingkan dengan tahun 2012. Untuk kategori kota sedang, Kabupaten Subang dalam dua tahun berturut-turut mendapat penghargaan terbaik dalam pelayanan terpadu satu pintu, tingkat Jawa Barat. Tingkat investasi akan terus meningkat seiring telah beroperasinya tol cipali.

Berdasarkan perkembangan investasi di kabupaten Subang diperlukan upaya untuk mengetahui karakteristik tingkat investasi perusahaan berdasarkan pada deskripsi perusahaan yang dimungkinkan. Dalam kenyataannya setiap perusahaan memiliki jumlah investasi, jumlah tenaga kerja yang berbeda, jumlah produksi, jenis usaha dan status yang berbeda-beda. Nurlaela dan Abidin (2013) melakukan pengamatan data investasi perusahaan-perusahaan di kabupaten Subang. Diperoleh deskripsi perusahaan berupa jumlah tenaga kerja, jumlah produksi, jenis usaha dan status perusahaan dalam menentukan tingkat investasi menggunakan metode Naive Bayes.

Diperlukan kajian hubungan jumlah investasi melalui jumlah tenaga kerja yang berbeda, jumlah produksi, jenis usaha dan status yang berbeda-beda menggunakan metode lain selain *Naive Bayes*. Dokumentasi pengetahuan dari data investasi dengan deskripsi perusahaan yang dapat diamati bisa memeberikan nilai tambah jika menggunakan metode data mining.

2. TINJAUAN PUSTAKA

Metode data mining selain *Naive Bayes* terdapat Algoritma C 4.5. Menurut Kusri dkk. dalam Abidin (2011) disampaikan bahwa algoritma C 4.5 lebih cepat dan akurat jika dibandingkan dengan *K-Nearest Neighbour*. Nilai akurasi yang diperoleh melalui percobaan perbandingan kinerja algoritma ID3/C 4.5, *Naive Bayes*, K-NN dan ANN dalam mengklasifikasi dokumen teks berbahasa inggris terhadap 10 partisi data, jumlah atribut 2000, data training 1300 set /partisi, jumlah data testing 700 set/partisi, hasil menunjukkan bahwa algoritma ID3/C4.5 dan *Naive Bayes* merupakan algoritma yang memiliki kehandalan dalam klasifikasi yang ditunjukkan dengan nilai rata-rata akurasi lebih dari 80 %.

Investasi merupakan penggunaan modal pada suatu masa dengan harapan terjadi perolehan keuntungan pada masa lainnya (Irma, 2014). Pengambilan keputusan investasi pada umumnya didasarkan pada pertimbangan investor terhadap besarnya return (pengembalian) yang diharapkan serta risiko yang diperkirakan akan dihadapi.

Terdapat banyak pilihan investasi oleh para investor, baik secara individu maupun secara kelompok. Berdasarkan bentuknya, investasi dapat dibedakan menjadi dua jenis, yaitu:

- Investasi pada sektor riil, seperti peralatan, properti, dll.
- Investasi pada sektor finansial, baik instrumen investasi pada pasar uang maupun instrumen investasi pada pasar modal, seperti obligasi dan saham maupun derivatifnya.

Disperindagsar kabupaten Subang (2013) merilis deskripsi perusahaan yang berhubungan dengan nilai investasi yaitu jumlah tenaga kerja, status, jenis komoditas, kelompok industri, kapasitas produksi dan nilai ekspor.

Craw (2005) menyampaikan pohon keputusan menggunakan algoritma C 4.5 dalam menentukan keputusan bermain tenis berdasarkan pada outlook, temperature, humidity dan windy. Untuk memperoleh root pada pohon keputusan dan seluruh node yang lain yang tertera pada pohon keputusan, Craw melakukan penghitungan Entropy, seperti pada Rumus 1, melakukan penghitungan Gain informasi seperti pada Rumus 2.

$$Entropy(S) = \sum_{i=1}^n - p_i * \log_2 p_i \dots 1)$$

Dimana:

- S : himpunan kasus
- n : jumlah partisi S
- pi : proporsi dari Si terhadap S

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i) \dots 2)$$

Dimana:

- S : himpunan kasus
- A : atribut
- n : jumlah partisi atribut A
- |Si| : jumlah kasus pada partisi ke-i
- |S| : jumlah kasus dalam S

Seleksi atribut melalui teori entropi dilakukan dengan menentukan gain informasi (*Information gain*), gain informasi yang tertinggi menjadi acuan dalam menentukan sebuah atribut n sebagai root atau node-node lainnya dalam sebuah pohon keputusan. Pada model pohon keputusan C4.5 proses learning dan klasifikasinya sederhana dan cepat. Secara umum model C4.5 memiliki tingkat akurasi yang tinggi (Han dan Kamber, 2006).

3. METODE PENELITIAN

Tabel 1 menunjukkan data penelitian yang merupakan data sekunder, data bernilai numerik. Data pada Tabel 1 diubah menjadi data kategorikal dan hasilnya diperlihatkan pada Tabel 2. Terdapat 18 baris data sebagai data penelitian tingkat investasi di kabupaten Subang dengan menggunakan deskripsi tingkat investasi berupa status perusahaan, jumlah tenaga kerja, kapasitas produksi dan jenis komoditas.

Tabel 1 Data perusahaan dan aktivitas investasi (disperindagsar, 2013)

No	Nama Perusahaan	Status	Jenis Komoditi	Jumlah Tenaga Kerja	Kapasitas Produksi/thn	Nilai Investasi
1	PTPN VIII Ciater	BUM N	Teh Hitam	2573	9.600 Ton	17.061.446
2	PT. Sinkona Indonesia Lestari	PMD N	Bahan Baku Farmasi	248	150 Ton	21.310.000

No	Nama Perusahaan	Status	Jenis Komoditi	Jumlah Tenaga Kerja	Kapasitas Produksi/thn	Nilai Investasi
3	PT.Perkasa Rekadaya Nusantara	PMD N	Komponen Suku Cadang	2065	1,5 juta	3.000.000
4	PT.Tirta Investama	PMA	Air Minum Dalam Kemasan	519	900.000.000 L	4.000.000
5	PT. Kondobo Texindo	PMA	Benang Nilon	725	60900 bale	66.000.000
6	PT.Papertech Indonesia	PMD N	Kertas karton	162	28.800 Ton	14.811,73
7	PTPN VIII Jalupang	PMD N	Pengolahan Karet	2.506	1.680 Ton	3.222.360
8	PT.Ado Internasional	PMA	Gordyn/ Vitrage	116	1500jt/thn	64.500.000
9	PT. Anugrah Setia Lestari	PMD N	Kecap	775	125 Ton	10.000
10	PT. Youtex Garment	PMA	Pakaian Jadi	1100	60jt pc/thn	765.000.000
11	PT. Crevis Tex Jaya	PMA	Pakaian Jadi	2500	3000 Ton	4.500.000
12	PTPN VIII Wangunreja Dawuan	PMD N	Pengolahan Karet	1.394	2.430.000 Ton	5.438.870
13	PT. Dunia Hibar	NON FAS	PMDK	20	3.750 ekor	850.000
14	PT. Quty Karuna (Young Do Kim)	PMA	pakaian jadi	1425	3000000 buah	10.000.000
15	PT. Inawan Chemtex Sukses Abadi	PMD N	Industri Damar Buatan	220	2000/ tahun	5.000.000
16	PT. Kwang Lim YH Indah	PMA	Pakaian Jadi	1100	1.000.000 lusin/thn	1.750.000
17	PT. Budi Makmur Perkasa	PMD N	Bihun, Tepung dan Beras Ketan	456	3.783 Ton	6.103.000
18	PT. C-Site Texpia	PMA	Pakaian Jadi	516	86.000 lusin	6.355.152

Tabel 2 Data kategorikal deskripsi investasi perusahaan

No	Nama Perusahaan	Status	Jenis Komoditi	Jumlah Tenaga Kerja	Kapasitas Produksi/th N	Class
1	PTPN VIII Ciater	BUMN	Pangan	Besar	Besar	Tinggi
2	PT. Sinkona Indonesia Lestari	PMDN	Pangan	Kecil	Sedang	Tinggi
3	PT.Perkasa Rekadaya Nusantara	PMDN	Papan	Besar	Kecil	rendah
4	PT.Tirta Investama	PMA	Pangan	Kecil	Besar	rendah
5	PT. Kondobo Texindo	PMA	Sandang	Kecil	Sedang	Tinggi
6	PT.Papertech Indonesia	PMDN	Papan	Kecil	Sedang	Tinggi
7	PTPN VIII Jalupang	PMDN	Papan	Besar	Besar	Rendah
8	PT.Ado Internasional	PMA	Sandang	Kecil	Besar	Tinggi
9	PT. Anugrah Setia Lestari	PMDN	Pangan	Kecil	Sedang	Rendah
10	PT. Youtex Garment	PMA	Sandang	Sedang	Sedang	Tinggi
11	PT. Crevis Tex Jaya	PMA	Sandang	Besar	Sedang	Rendah
12	PTPN VIII Wangunreja Dawuan	PMDN	Papan	Kecil	Besar	Tinggi

No	Nama Perusahaan	Status	Jenis Komoditi	Jumlah Tenaga Kerja	Kapasitas Produksi/th N	Class
13	PT. Dunia Hibar	NON FAS	Pangan	Kecil	Kecil	Rendah
14	PT. Quty Karuna (Young Do Kim)	PMA	Sandang	Sedang	Sedang	Tinggi
15	PT. Inawan Chemtex Sukses Abadi	PMDN	Papan	Kecil	Kecil	Tinggi
16	PT. Kwang Lim YH Indah	PMA	Sandang	Sedang	Sedang	Rendah
17	PT. Budi Makmur Perkasa	PMDN	Pangan	Kecil	Sedang	Tinggi
18	PT. C-Site Texpia	PMA	Sandang	Kecil	Sedang	Tinggi

Dilakukan penghitungan entropy dan gain. Dalam paparan ini disampaikan penghitungan entropy berupa entropy total, penghitungan entropy status dengan nilai Non Fas, PMA, PMDN dan BUMN. Selain status, dilakukan penghitungan entropy untuk semua nilai atribut dari semua atribut yang ada.

Setelah diperoleh entropy masing-masing nilai atribut status, jenis komoditas, jumlah tenaga kerja dan kapasitas produksi, kemudian dilakukan penghitungan gain untuk atribut status, jumlah tenaga kerja, kapasitas produksi dan jenis komoditas.

$$\begin{aligned}
 Entropy(S) &= \sum_{i=1}^n - p_i * \log_2 p_i \\
 &= - p_1 * \log_2 p_1 - p_2 * \log_2 p_2 \\
 &= - \left(\frac{s1}{S} \right) * \log_2 \left(\frac{s1}{S} \right) - \left(\frac{s2}{S} \right) * \log_2 \left(\frac{s2}{S} \right) \\
 &= - \frac{11}{18} * \log_2 \left(\frac{11}{18} \right) - \frac{7}{18} * \log_2 \left(\frac{7}{18} \right) \\
 &= 0.43419 + 0.529888 \\
 &= 0.964079
 \end{aligned}$$

Menghitung Entropy dan Gain Status

$$\begin{aligned}
 Entropy(Status_{BUMN}) &= \sum_{i=1}^n - p_i * \log_2 p_i \\
 &= - \frac{1}{1} * \log_2 \left(\frac{1}{1} \right) - \frac{0}{1} * \log_2 \left(\frac{0}{1} \right) \\
 &= 0 + 0 \\
 &= 0
 \end{aligned}$$

$$\begin{aligned}
 Entropy(Status_{PMDN}) &= \sum_{i=1}^n - p_i * \log_2 p_i \\
 &= - \frac{5}{8} * \log_2 \left(\frac{5}{8} \right) - \frac{3}{8} * \log_2 \left(\frac{3}{8} \right) \\
 &= 0.038404 + 0.530639 \\
 &= 0.569043
 \end{aligned}$$

$$Entropy(Status_{PMA}) = \sum_{i=1}^n - p_i * \log_2 p_i$$

$$\begin{aligned}
 &= -\frac{5}{8} * \log_2\left(\frac{5}{8}\right) - \frac{3}{8} * \log_2\left(\frac{3}{8}\right) \\
 &= 0.038404 + 0.530639 \\
 &= 0.569043 \\
 Entropy(Status_{NONFAS}) &= \sum_{i=1}^n -p_i * \log_2 p_i \\
 &= -\frac{0}{1} * \log_2\left(\frac{0}{1}\right) - \frac{1}{1} * \log_2\left(\frac{1}{1}\right) \\
 &= 0 + 0 \\
 &= 0
 \end{aligned}$$

Gain Status

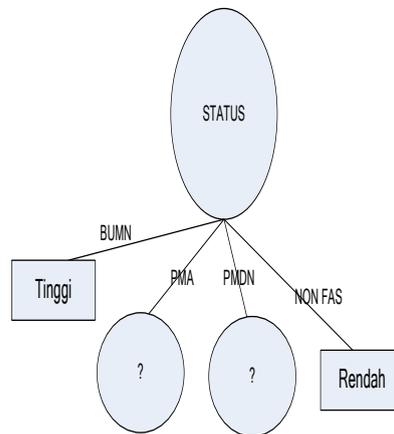
$$\begin{aligned}
 Gain(S, A) &= Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i) \\
 &= Entropy(total) - \sum_{i=1}^2 \frac{S_{Status}}{S_{total}} * Entropy(Status) \\
 &= 0.964079 - \left(\frac{1}{18}\right) * 0 - \left(\frac{8}{18}\right) * 0.569043 - \left(\frac{8}{18}\right) * 0.569043 - \left(\frac{1}{18}\right) * 0 \\
 &= 0.458263
 \end{aligned}$$

Hasil perhitungan diperlihatkan dalam Tabel 3, dimana nilai gain untuk atribut status adalah terbesar yaitu 0,458 sehingga status sebagai root dari pohon keputusan. Perolehan root dari rangkuman perhitungan pada Tabel 3 diperlihatkan pada Gambar 1, demikian seterusnya seperti pada Tabel dan 4, serta pohon keputusnnya pada Gambar 2 dan 3.

Tabel 3 Menentukan root (akar)

Node 1	jml kasus (S)	Tinggi (S1)	Rendah(S2)	Entropy	Gain
Total	8	5	3	0.964079	
Status					0.458263
BUMN	1	1	0	0	
PMDN	8	5	3	0.569043	
PMA	8	5	3	0.569043	
Non Fas	1	0	1	0	
Jenis Komoditi					0.025379
Pangan	6	3	3	1	
papan	5	2	3	0.970951	
Sandang	7	5	2	0.863121	
Jml Tenaga Kerja					0.114142
Besar	4	1	3	0.811278	
Sedang	3	2	1	0.918296	
Kecil	11	8	3	0.845351	

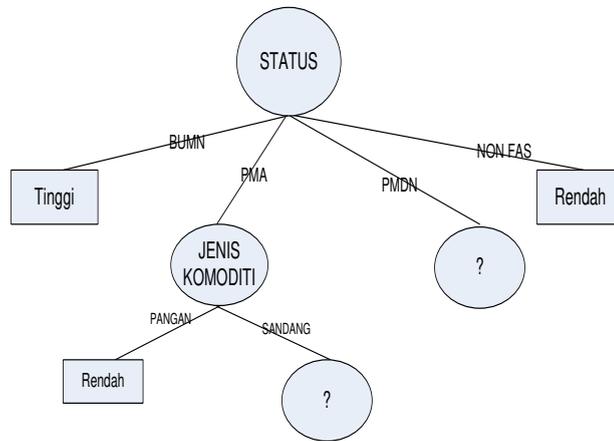
Node 1	jml kasus (S)	Tinggi (S1)	Rendah(S2)	Entropy	Gain
Kapasitas Produksi					0.051715
Besar	5	3	2	0.970951	
Sedang	10	7	3	0.881291	
Kecil	3	1	2	0.918296	



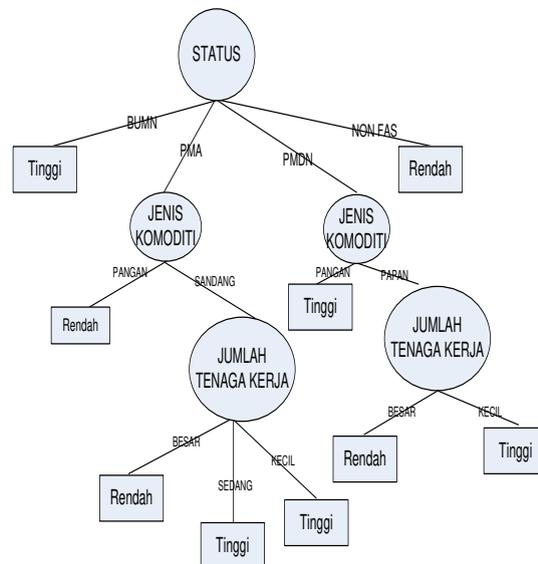
Gambar 1 Root pohon keputusan

Tabel 4 Menentukan cabang setelah root

	jml kasus	Tinggi	rendah	entrophy	Gain
Total	8	5	3	0.954434	
Jenis Komoditi					0.502595
pangan	1	0	1	0	
papan	0	0	0	0	
sandang	7	5	2	0.516387	
Jml Tenaga Kerja					0.360073
besar	1	0	1	0	
sedang	3	2	1	0.918296	
kecil	4	3	1	0.5	
Kapasitas Produksi					0.140712
besar	2	1	1	0.5	
sedang	6	4	2	0.918296	
kecil	0	0	0	0	



Gambar 2 Perolehan cabang pohon keputusan untuk jenis komoditas

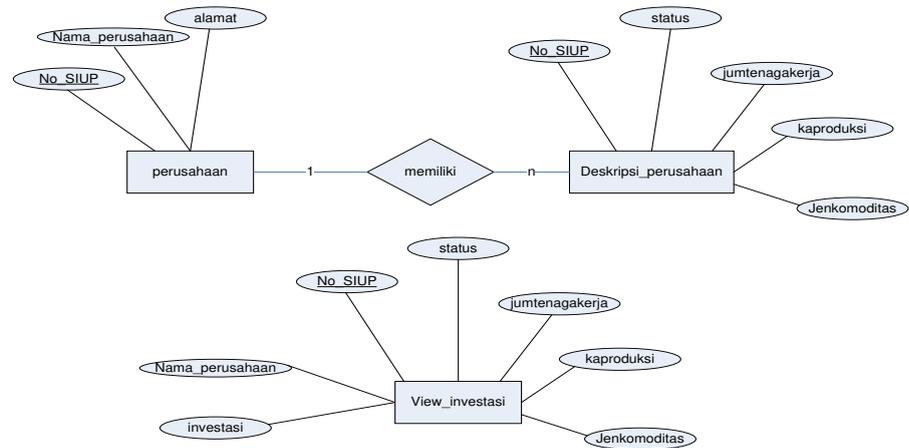


Gambar 3 Pohon keputusan lengkap

Dari pohon keputusan pada Gambar 3 diperoleh 9 aturan sebagai berikut:

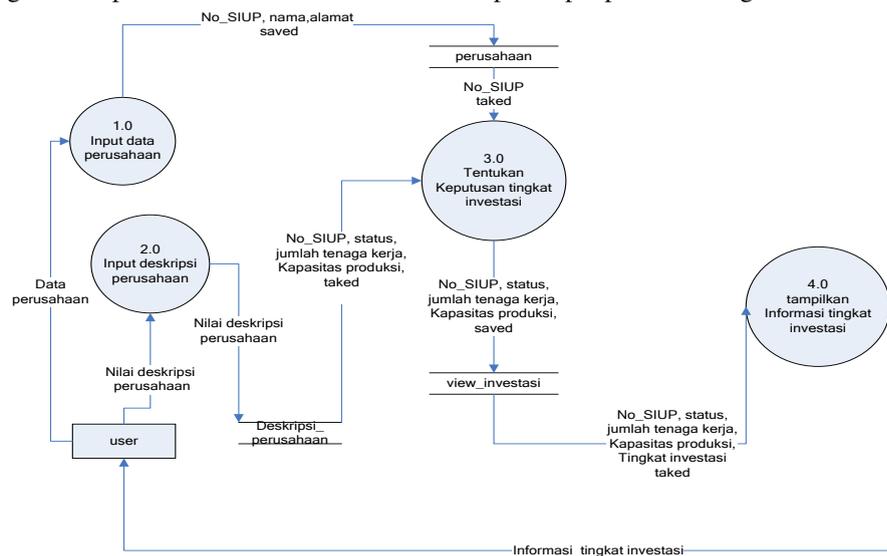
1. Jika status BUMN maka Investasi Tinggi
2. Jika status PMA dan jenis komoditas pangan maka investasi Rendah
3. Jika status PMA dan jenis komoditi sandang dan jumlah tenaga kerja besar maka investasi rendah
4. Jika status PMA dan jenis komoditi sandang dan jumlah tenaga kerja sedang maka investasi tinggi
5. Jika status PMA dan jenis komoditi sandang dan jumlah tenaga kerja kecil maka investasi tinggi
6. Jika status PMDN dan jenis komoditi pangan maka investasi tinggi
7. Jika status PMDN dan jenis komoditi papan dan jumlah tenaga kerja besar maka investasi rendah
8. Jika status PMA dan jenis komoditi papan dan jumlah tenaga kerja kecil maka investasi tinggi
9. Jika status NonFAS maka investasi rendah

Untuk memperoleh prototipe sistem dimodelkan data dengan tiga entitas masing-masing adalah perusahaan, deskripsi perusahaan dan view investasi, seperti diperlihatkan dalam Gambar 4.



Gambar 4 Model data system

Dan model fungsional diperlihatkan dalam Gambar 5. Terdapat empat proses dan tiga data store.



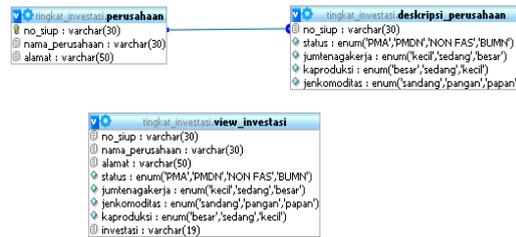
Gambar 5 Model fungsional (DFD level 1)

4. HASIL DAN PEMBAHASAN

Dilakukan implementasi basis data dengan cara membuat tabel perusahaan dan tabel deskripsi perusahaan seperti diperlihatkan dalam Gambar 7. Gambar 6 menunjukkan skrip SQL untuk pembuatan view investasi dalam rangka mengimplementasikan at Hasil dan pembahasan berisi hasil analisis fenomena di wilayah ran dari sembilan aturan yang diperoleh. Gambar 8 merupakan implementasi antar muka untuk menginput data deskripsi perusahaan.

```
create view view_investasi as select
perusahaan.no_siup,perusahaan.nama_perusahaan,perusahaan.alamat,deskripsi_perusahaan.status,de
skripsi_perusahaan.jumtenagakerja,deskripsi_perusahaan.jenkomoditas,deskripsi_perusahaan.kapro
duksi,CASE WHEN status='BUMN' THEN 'tinggi' WHEN status='PMA' AND
jenkomoditas='pangan' THEN 'rendah' WHEN status='PMA' AND jenkomoditas='sandang' AND
jumtenagakerja='besar' THEN 'rendah' WHEN status='PMA' AND jenkomoditas='sandang' AND
jumtenagakerja='sedang' THEN 'tinggi' WHEN status='PMA' AND jenkomoditas='sandang' AND
jumtenagakerja='kecil' THEN 'tinggi' WHEN status='PMDN' AND jenkomoditas='pangan' THEN
'tinggi' WHEN status='PMDN' AND jenkomoditas='papan' AND jumtenagakerja='besar' THEN
'rendah' WHEN status='PMA' AND jenkomoditas='papan' AND jumtenagakerja='kecil' THEN
'tinggi' WHEN status='NON FAS' THEN 'rendah' ELSE 'Belum Ada keputusan' end as investasi
from perusahaan,deskripsi_perusahaan WHERE perusahaan.no_siup=deskripsi_perusahaan.no_siup
```

Gambar 6 implementasi pohon keputusan



Gambar 7 Relasi antar tabel

Gambar 8 Proses input data deskripsi perusahaan

5. KESIMPULAN

Berdasarkan hasil pembahasan pada bab sebelumnya diperoleh kesimpulan sebagai berikut:

1. Dapat diimplementasikan prototipe perangkat lunak penentuan tingkat investasi menggunakan algoritma C 4.5 dalam DBMS MySQL dan PHP.
2. Diperoleh sembilan aturan untuk komputasi pengambilan keputusan tingkat investasi perusahaan di kabupaten Subang menggunakan deskripsi perusahaan status perusahaan, jumlah tenaga kerja, kapasitas produksi dan jenis komoditas.
3. Untuk penyempurnaan sebaiknya diperbanyak data training penelitian kalau bisa tidak sebatas data dari kabupaten Subang akantetapi juga dari seluruh Indonesia.

DAFTAR PUSTAKA

- Abidin, A.Z.Z., 2011, Implementasi Algoritma C 4.5 dalam menentukan tingkat bahaya Tsunami, Semnasif 2011, dapat diakses pada: <http://repository.upnyk.ac.id/620/1/A-4.pdf>
- Adji, Y., 2014, Investasi Di Subang tahun 2013 naik 800 persen, media online pikiran rakyat, dapat diakses pada: <http://www.pikiran-rakyat.com/jawa-barat/2014/09/18/297466/investasi-di-subang-tahun-2013-naik-800-persen>
- Craw, 2005, Case Base Reasoning : Lecture 3: CBR Case-Base Indexing, dapat diakses pada <http://www.comp.rgu.ac.uk/staff/smc/teaching/kbp3/>
- Dinas Periindustrian perdagangan dan pasar, 2013, Daftar Industri Besar di Kabupaten Subang Sampai Dengan September 2013, dapat diakses pada: <http://disperindagsar.subang.go.id/files/Industri%20Besar%202013.pdf>
- Han, J., & Kamber, M (2006) Data Mining Concept and Tehniques San Fransisco: Morgan Kauffman.
- Irma, 2014, Analisis Investasi Perusahaan, dapat diakses pada: <http://irmajhe.blogspot.co.id/2014/04/analisis-investasi-perusahaan.html>
- Nurlaela,S., Abidin, A.Z.Z, 2013, Implementasi Naive Bayes dalam menentukan tingkat investasi perusahaan, Tugas Akhir STMIK Subang
- Ramadan, R., 2007, Penerapan Pohon Untuk Klasifikasi Dokumen Teks Berbahasa Inggris, publikasi ilmiah