

## KLASIFIKASI DATA PROSPEKTUS LOKASI WARALABA DENGAN ALGORITMA C4.5

**Ali Khumaidi**

Manajemen Informatika  
Bina Sarana Informatika Jakarta  
Jl. RS. Fatmawati No.24 Jakarta Selatan  
[mr.khumaidi@yahoo.com](mailto:mr.khumaidi@yahoo.com)

### ABSTRACT

*In this study the authors performed data classification prospectus franchise locations derived from franchise consultant MBC (mandiri business consultant) to determine the prospects of a location waralaba. From the data of 186 cases of existing franchise, there were 152 cases are prospects and 34 cases of the remaining are not prospects. This means that 18.27% of potential locations might, could not be maintained or changed into potential locations for the franchise. If possible franchise locations are not known earlier outlook, then the franchise consultant can perform the actions necessary to make the site a prospect. The data are analyzed and classified using the decision tree approach C4.5 algorithm. For the determination of attributes and variables involved in determining the location of the franchise prospectus, the author uses the rules and guidelines of a franchise consultant who has been posted. By enabling attributes or variables not the destination as many as 22 variables obtained by 98 rules. The rules or the rule will become a reference in determining the prospective franchise locations are categorized no outlook or outlook. Of the rules that form, there are two categories of rules. That is the rule that produces a particular classification, prospects and prospects as well as rules that the conclusion is not classified.*

*Keyword: decision tree, C4.5 algorithm, prospectus franchise locations, database design*

### 1. PENDAHULUAN

Hingga saat ini tercatat tidak kurang dari 900 perusahaan franchise dan business opportunity (BO) yang menawarkan usahanya untuk dibeli oleh calon pembeli franchise (Hendry, 2010). Makin banyak suatu pilihan franchise maka akan membuat semakin rumit dalam menentukan pilihan yang bagus, mana yang tidak. Banyak buku yang menjelaskan dan menjabarkan bagaimana kiat sukses dalam membeli suatu franchise, hampir semuanya menyebutkan bahwa pemilihan lokasi menjadi salah satu kriteria. Bahkan banyak sumber juga yang mengatakan bahwa yang penting dalam bisnis (ritel) adalah pertama, lokasi, kedua, lokasi, ketiga, lokasi. Para pebisnis franchise kawakan sekalipun ketika membuka gerai yang ke-10 atau yang ke-100, mereka tetap memperhatikan lokasi (IFBM, 2007). Oleh karena itu, dibutuhkan suatu solusi yang dapat membantu dalam pengambilan keputusan untuk mengatasi hal tersebut.

Penulis sebelumnya telah membuat tulisan tentang “Penerapan Case-Based Reasoning dan Algoritma Nearest Neighbor untuk Penentuan Lokasi Waralaba”. Dari data kasus yang diperoleh dari konsultan waralaba tersebut,

penulis mencoba menggunakan pendekatan klasifikasi data yang lain yaitu, algoritma C4.5 untuk melakukan pencarian pola prospektus lokasi waralaba sehingga dapat dijadikan bahan analisa dalam menentukan prospek atau tidaknya suatu calon lokasi waralaba.

Dari data jumlah 186 kasus franchise yang ada, terdapat 152 kasus yang prospek dan 34 sisanya tidak prospek. Hal ini berarti 18,27 % lokasi yang mungkin potensial, tidak mampu dipertahankan atau diubah menjadi lokasi yang potensial untuk waralaba. Jika kemungkinan lokasi waralaba yang tidak prospek tersebut diketahui lebih dini, maka pihak konsultan waralaba dapat melakukan tindakan-tindakan yang diperlukan untuk menjadikan lokasi tersebut menjadi prospek. Salah satu cara untuk melakukan analisis kemungkinan prospek suatu lokasi adalah dengan melakukan klasifikasi dari kumpulan data calon lokasi yang ada. Salah satu model klasifikasi adalah dengan membuat pohon keputusan.

### 2. LANDASAN TEORI

#### 2.1. Pemilihan Lokasi

Banyak pendapat dari beberapa ahli tentang atribut-atribut untuk menentukan lokasi yang baik untuk mendirikan suatu usaha. Secara garis besar, dalam menilai lokasi terutama masalah akses jalan kita dapat melihat dari aspek makro dan mikro. Yang menyangkut aspek makro, meliputi penilaian terhadap pola jalan, kondisi jalan dan hambatan ke lokasi. Sedangkan dalam aspek mikro, penilaian dilakukan terhadap visibilitas, arus lalu lintas, parkir, keramaian, dan jalan masuk ke lokasi atau jalan keluar dari lokasi (Hendry, 2009).

Menurut Hendry ada tiga tahapan pemilihan lokasi:

1. Pemilihan pasar (market selection)
2. Analisis area (area trading analysis)
3. Analisis tempat (site analysis)

Tahap pertama, yaitu pemilihan pasar (market selection). Pada tahap pemilihan pasar beberapa aspek yang perlu dipertimbangkan dan dinilai meliputi: aspek tingkat perekonomian masyarakat, tingkat persaingan bisnis sekitar, ukuran populasi dan karakteristiknya, serta industri atau bisnis di lingkungan sekitar. Tahap kedua yaitu analisis area (area trading analysis). Pada tahapan analisis area dilakukan analisis terhadap area dari lokasi yang disurvei. Area dari lokasi dibagi menjadi 3 area yaitu area primer, area sekunder dan area tersier. Area primer meliputi area geografis suatu toko atau outlet mendapat sekitar 60% dari pelanggannya. Area sekunder meliputi area suatu toko atau outlet mendapat sekitar 20% dari pelanggannya. Dan area tersier adalah area dimana pelanggan kadang berbelanja di toko atau outlet karena beberapa alasan tertentu, seperti karena suatu toko atau outlet merupakan satu-satunya yang menjual barang tertentu dan tidak ada toko lain yang menjual barang sejenis. Dalam pelaksanaan penentuan area dilakukan penjabaran dari manakah 60% pelanggan di lokasi yang akan dipilih, dan dari manakah 20% pelanggan serta darimanakah pelanggan tersier yang mungkin saja mampir ke toko karena beberapa alasan.

Tahapan selanjutnya yaitu analisis tempat (site analysis). Pada tahap ini dilakukan analisis yang terkait langsung dengan lokasi yang ingin disewa, informasi yang dianalisis antara lain: berapa harga sewa, biaya konstruksi atau renovasi, faktor keamanan, kemungkinan mendapatkan tenaga kerja di lokasi tersebut, dll.

**2.2. Algoritma C4.5**

Algoritma C4.5 merupakan algoritma yang digunakan untuk membentuk pohon

keputusan. Banyak algoritma yang dapat dipakai dalam pembentukan pohon keputusan, antara lain ID3, CART, FACT, QUEST, LMDT dan C4.5. Algoritma C4.5 ini merupakan pengembangan dari algoritma ID3. Dibandingkan dengan algoritma yang lain, algoritma C4.5 memiliki kelebihan utama karena dapat menghasilkan model berupa pohon atau aturan yang mudah diinterpretasikan dan dikonversi ke aturan SQL, memiliki tingkat akurasi yang dapat diterima, dapat menangani atribut bertipe diskret dan numerik, serta efisien dalam menangani atribut bertipe diskret (Kamber, 2001).

Data dalam pohon keputusan biasanya dinyatakan dalam bentuk tabel dengan atribut dan record. Atribut menyatakan suatu parameter yang dibuat sebagai kriteria atau variabel dalam pembentukan pohon. Salah satu atribut merupakan atribut yang menyatakan data solusi per item data yang disebut target atribut. Atribut memiliki nilai yang dinamakan instance. Proses pada pohon keputusan adalah mengubah bentuk data(tabel) menjadi model pohon, mengubah model pohon menjadi rule, dan menyederhanakan rule (Kusrini, 2009).

Secara umum algoritma C4.5 untuk membangun pohon keputusan adalah sebagai berikut (Larose, 2005):

1. Pilih atribut sebagai akar
2. Buat cabang untuk tiap-tiap nilai
3. Bagi kasus dalam cabang
4. Ulangi proses untuk tiap cabang sampai semua kasus pada cabang memiliki kelas yang sama.

Untuk memilih atribut sebagai akar, didasarkan pada nilai gain tertinggi dari atribut-atribut yang ada. Untuk menghitung gain digunakan rumus sebagai berikut:

$$\text{Gain}(S, A) = \sum_{i=1}^n \frac{|S_i|}{|S|} * \text{Entropi}(S_i)$$

Keterangan:

- S : Himpunan kasus
- A : Atribut
- n : Jumlah partisi atribut An
- |Si| : Jumlah kasus pada partisi ke-i
- |S| : Jumlah kasus dalam S

Sementara itu, perhitungan nilai entropi dapat dilihat pada persamaan berikut:

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

Keterangan:

S : Himpunan kasus  
 A : Fitur  
 n : Jumlah partisi S  
 pi : Proporsi dari Si terhadap S

Berikut ini penjelasan langkah dalam pembentukan pohon keputusan dengan algoritma C4.5 tentang kasus keputusan bermain tenis:

Tabel 1. Keputusan Bermain Tenis

| No | Outlook | Temperature | Humidity | Windy | Play |
|----|---------|-------------|----------|-------|------|
| 1  | Sunny   | Hot         | High     | False | No   |
| 2  | Sunny   | Hot         | High     | True  | No   |
| 3  | Cloudy  | Hot         | High     | False | Yes  |
| 4  | Rainy   | Mild        | High     | False | Yes  |
| 5  | Rainy   | Cool        | Normal   | False | Yes  |
| 6  | Rainy   | Cool        | Normal   | True  | Yes  |
| 7  | Cloudy  | Cool        | Normal   | True  | Yes  |
| 8  | Sunny   | Mild        | High     | False | No   |
| 9  | Sunny   | Cool        | Normal   | False | Yes  |
| 10 | Rainy   | Mild        | Normal   | False | Yes  |
| 11 | Sunny   | Mild        | Normal   | True  | Yes  |
| 12 | Cloudy  | Mild        | High     | True  | Yes  |
| 13 | Cloudy  | Hot         | Normal   | False | Yes  |
| 14 | Rainy   | Mild        | High     | True  | No   |

Sumber: Kusrini, 2009

Tabel kasus diatas akan dibuat pohon keputusan untuk bermain tenis atau tidak dengan melihat beberapa atribut. Yaitu outlook (cuaca), temperature (suhu), humidity (kelembapan), dan windy (keadaan angin).

Berikut ini penjelasan lebih terperinci tiap langkah dalam pembentukan pohon keputusan dengan menggunakan algoritma C4.5 dalam permasalahan kasus bermain tenis (tabel 1).

Pertama, menghitung jumlah kasus untuk keputusan yes dan kasus untuk keputusan no, serta Entropi dari semua kasus dan kasus yang dibagi berdasarkan atribut outlook, temperature, humidity, dan windy. Kemudian lakukan perhitungan Gain untuk tiap atribut. Hasil perhitungan sebagai berikut:

Tabel 2. Perhitungan node 1

| N           | S  | S1 | S2 | E    | G    |
|-------------|----|----|----|------|------|
| 1 Total     | 14 | 4  | 10 | 0,86 |      |
| Outlook     |    |    |    |      | 0,26 |
| Cloudy      | 4  | 0  | 4  | 0    |      |
| Rainy       | 5  | 1  | 4  | 0,72 |      |
| Sunny       | 5  | 3  | 2  | 0,97 |      |
| Temperature |    |    |    |      | 0,18 |
| Cool        | 4  | 0  | 4  | 0    |      |
| Hot         | 4  | 2  | 2  | 1    |      |
| Mild        | 6  | 2  | 4  | 0,92 |      |
| Humidity    |    |    |    |      | 0,37 |
| High        | 7  | 4  | 3  | 0,98 |      |
| Normal      | 7  | 0  | 7  | 0    |      |
| Windy       |    |    |    |      | 0,01 |
| False       | 8  | 2  | 6  | 0,81 |      |
| True        | 6  | 4  | 2  | 0,92 |      |

Sumber: Kusrini, 2009

Keterangan : N = Node, S = Jumlah kasus, S1 = No (tidak bermain tenis), S2 = Yes (jadi bermain tenis), E = entropi, G= gain.

$$\text{Entropi(Total)} = \sum_{n=1}^n \frac{|\text{Outlook}|}{|\text{Total}|} * \text{Entropi(Outlook1)}$$

Berikut contoh dalam menghitung nilai entropi dan Gain. Baris Total pada kolom entropi (tabel 2) dihitung dengan persamaan berikut:

$$\text{Gain(Total, Outlook)} = 0,86 - ((4/14*0) + (5/14*0,72) + (5/14*0,97))$$

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

$$\text{Gain(Total, Outlook)} = 0,26$$

$$\text{Entropi(Total)} =$$

Demikian selanjutnya cara untuk menghitung nilai Gain dari temperature, humidity, dan windy menggunakan cara yang sama.

$$(-4/14 * \log_2(4/14)) + (-10/14 * \log_2(10/14))$$

Dari hasil pada tabel 2 dapat diketahui bahwa atribut dengan Gain tertinggi adalah humidity, yakni sebesar 0,37. Dengan demikian humidity dapat menjadi node akar. Ada dua nilai atribut dari humidity yaitu high dan normal. Dari kedua nilai atribut tersebut, nilai atribut normal sudah mengklasifikasikan kasus menjadi 1, yaitu keputusan yes (karena kasus yang tidak bermain tenis jumlahnya 0). Sehingga tidak perlu dilakukan perhitungan lebih lanjut, tetapi untuk nilai atribut high masih perlu dilakukan perhitungan lagi.

$$\text{Entropi(Total)} = 0,86$$

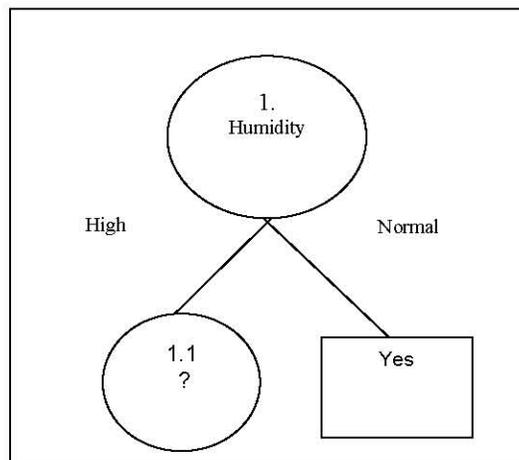
Demikian selanjutnya cara untuk menghitung nilai entropi dari cloudy, rainy, sunny, cool, hot, mild, high, normal, false dan true menggunakan cara yang sama.

Dari hasil tersebut dapat digambarkan pohon keputusan sementara seperti gambar berikut:

Sementara itu, nilai gain pada baris Outlook dihitung sebagai berikut:

$$\text{Gain(S, A)} = \sum_{i=1}^n \frac{|S_i|}{|S|} * \text{Entropi(S}_i)$$

$$\text{Gain(Total, Outlook)} =$$



Gambar 1. Pohon keputusan perhitungan node1  
Sumber: Kusriani, 2009

Kedua, menghitung semua jumlah kasus, jumlah kasus untuk keputusan No dan jumlah kasus untuk keputusan Yes. Serta Entropi dari semua kasus dan kasus yang dibagi berdasarkan atribut Outlook, Temperature, dan

Windy yang dapat menjadi node akar dari nilai atribut High. Setelah itu, lakukan penghitungan Gain untuk tiap atribut. Berikut hasil perhitungan untuk node 1.1.

Tabel 3. Perhitungan node 1.1

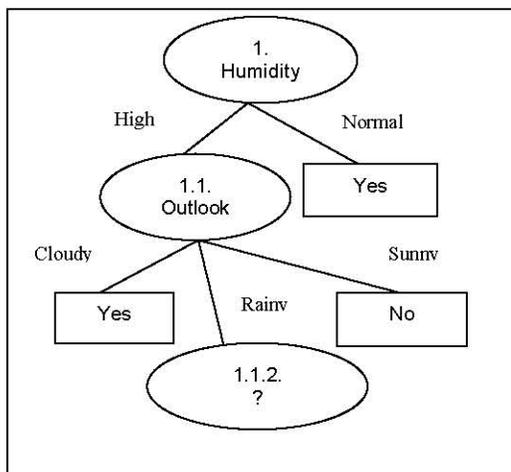
| N  |               | S | S1 | S2 | E    | G    |
|----|---------------|---|----|----|------|------|
| 1. | Humidity-High | 7 | 4  | 3  | 0,98 |      |
| 1  | Outlook       |   |    |    |      | 0,70 |
|    | Cloudy        | 2 | 0  | 2  | 0    |      |
|    | Rainy         | 2 | 1  | 1  | 1    |      |
|    | Sunny         | 3 | 3  | 0  | 0    |      |
|    | Tempe Rature  |   |    |    |      | 0,02 |
|    | Cool          | 0 | 0  | 0  | 0    |      |
|    | Hot           | 3 | 2  | 1  | 0,92 |      |
|    | Mild          | 4 | 2  | 2  | 1    |      |
|    | Windy         |   |    |    |      | 0,02 |
|    | False         | 4 | 2  | 2  | 1    |      |
|    | True          | 3 | 2  | 1  | 0,92 |      |

Sumber: Kusrini, 2009

Keterangan : N = Node, S = Jumlah kasus, S1 = No (tidak bermain tenis), S2 = Yes (jadi bermain tenis), E = entropi, G= gain.

Dari hasil pada tabel 3 dapat diketahui bahwa atribut dengan nilai gain tertinggi yaitu outlook dengan nilai sebesar 0,70. Dengan demikian outlook yang akan menjadi pilihan menjadi node cabang dari nilai atribut high. Ada tiga nilai atribut dari outlook, yaitu cloudy, rainy, dan sunny. Dari ketiga nilai atribut tersebut, nilai atribut cloudy sudah

mengklasifikasikan kasus menjadi satu dengan keputusan yes (karena kasus yang tidak jadi bermain tenis jumlahnya 0). Dan nilai atribut sunny sudah mengklasifikasikan kasus menjadi satu dengan keputusan no (karena kasus yang jadi bermain tenis jumlahnya 0), sehingga tidak perlu lagi perhitungan lebih lanjut. Tetapi untuk nilai atribut rainy masih perlu dilakukan perhitungan lagi. Adapun pohon keputusan sementara dapat digambarkan seperti gambar berikut (gambar 2):



Gambar 2. Pohon keputusan perhitungan node 1.1

Sumber: Kusrini, 2009

Ketiga, menghitung semua jumlah kasus, jumlah kasus untuk keputusan No dan jumlah kasus untuk keputusan Yes. Serta Entropi dari semua kasus dan kasus yang dibagi berdasarkan atribut Temperature, dan Windy

yang dapat menjadi node akar dari nilai atribut Rainy. Setelah itu, lakukan penghitungan Gain untuk tiap atribut. Berikut hasil perhitungan untuk node 1.1.2.

Tabel 4. Perhitungan Node 1.1.2.

| N   | S | S1 | S2 | E | G |
|---|---|----|----|---|---|
| 1.1. Humidity-High dan<br>2 Outlook-Rainy | 2 | 1  | 1  | 1 |   |
| Temperature                               |   |    |    |   | 0 |
| Cool                                      | 0 | 0  | 0  | 0 |   |
| Hot                                       | 0 | 0  | 0  | 0 |   |
| Mild                                      | 2 | 1  | 1  | 1 |   |
| Windy                                     |   |    |    |   | 1 |
| False                                     | 1 | 0  | 1  | 0 |   |
| True                                      | 1 | 1  | 0  | 0 |   |

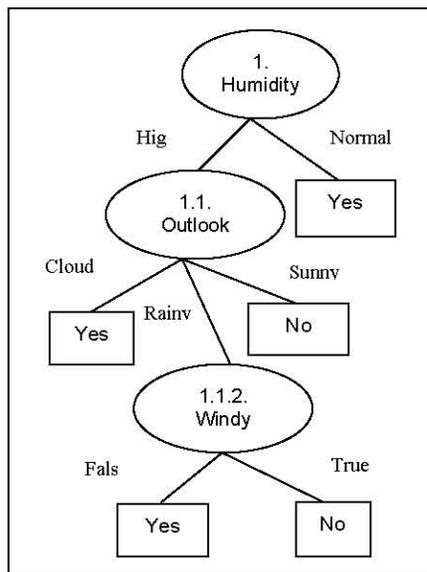
Sumber: Kusrini, 2009

Keterangan :

N = Node, S = Jumlah kasus, S1 = No (tidak bermain tenis), S2 = Yes (jadi bermain tenis), E = entropi, G= gain.

Dari hasil pada tabel 4 dapat diketahui bahwa atribut dengan nilai gain tertinggi yaitu windy dengan nilai sebesar 1. Dengan demikian windy yang akan menjadi pilihan menjadi node cabang dari nilai atribut rainy. Ada dua nilai atribut dari windy, yaitu false dan

true. Dari kedua nilai atribut tersebut, nilai atribut false sudah mengklasifikasikan kasus menjadi satu dengan keputusan yes (karena kasus yang tidak jadi bermain tenis jumlahnya 0). Dan nilai atribut true sudah mengklasifikasikan kasus menjadi satu dengan keputusan no (karena kasus yang jadi bermain tenis jumlahnya 0), sehingga tidak perlu lagi perhitungan lebih lanjut untuk nilai atribut ini. Adapun pohon keputusan dapat digambarkan seperti gambar berikut:



Gambar 3. Pohon keputusan perhitungan node1.1.2.

Sumber: Kusrini, 2009

Pohon keputusan yang terbentuk sampai tahap ini seperti pada gambar 3. Dengan memperhatikan pohon keputusan tersebut diketahui bahwa semua kasus sudah masuk

dalam kelas dan pohon keputusan gambar 3 merupakan pohon keputusan terakhir yang terbentuk.

### 3. METODE PENELITIAN

Langkah-langkah yang dilakukan dalam penelitian ini, sebagai berikut:

Studi literatur

Pada tahap ini penulis mempelajari beberapa literatur yang membahas metode klasifikasi data, terutama algoritma C4.5.

Pengumpulan Data

Data yang penulis kumpulkan berasal dari MBC (Mandiri Business Consultans), yang merupakan konsultan waralaba. Adapun data yang penulis dapatkan sejumlah 186 data franchise, yang terdiri dari dua kategori yaitu makanan dan retail.

Analisis

Menganalisis penerapan algoritma C4.5 terkait dengan studi kasus yaitu membentuk pohon keputusan prospektus lokasi waralaba. Dalam analisis ini, dicari atribut-atribut yang ada dalam data prospektus lokasi waralaba yang mungkin mempengaruhi atribut tujuan yaitu atribut prospektus.

Perancangan Sistem

Pada tahap perancangan ini penulis membangun arsitektur sistem, perancangan basis data dan perancangan antar muka.

Implementasi Sistem

Dalam tahap ini, penulis membangun aplikasi yang didasarkan pada perancangan. Adapun

aplikasinya menggunakan bahasa pemrograman Borland Delphi dan database Interbase.

Uji Coba

Untuk melakukan uji coba ini menggunakan data prospektus lokasi waralaba yang sudah penulis peroleh dengan membagi menjadi dua bagian yaitu sebagai pembuat pohon keputusan dan sebagai penguji kasus.

### 4. PEMBAHASAN

#### 4.1. Pemilihan Kriteria atau Atribut

Berdasarkan data kasus yang diperoleh dari konsultan waralaba terdapat 26 kriteria atau atribut dalam penentuan lokasi, yaitu: Waralaba, alamat, investasi, kategori, tingkat perekonomian, tingkat persaingan, ukuran populasi, industri atau Bisnis sekitar, calon konsumen, Posisi lokasi, pola jalan, kondisi jalan, hambatan ke lokasi, visibilitas, parkir, keramaian, renovasi, arus lalu lintas, keamanan, tenaga kerja, bahan baku, transportasi, daya beli, luas area, usia penduduk, prospektus. Berikut data kasus yang penulis peroleh, karena datanya sangat banyak dalam tulisan ini hanya diberikan satu sampel saja (tabel 5).

Tabel 5. Sampel Kasus

|    |                |               |
|----|----------------|---------------|
| 1  | Waralaba       | Bakmi Gila    |
| 2  | Alamat         | Bakti, Medan  |
| 3  | Investasi      | 250 Juta      |
| 4  | Kategori       | Makanan       |
| 5  | Perekonomian   | Size A1       |
| 6  | Persaingan     | Tinggi        |
| 7  | Populasi       | Tinggi        |
| 8  | Bisnis sekitar | Tinggi        |
| 9  | Konsumen       | Sekunder      |
| 10 | Posisi         | Pinggir jalan |
| 11 | Pola jalan     | 2 arah        |
| 12 | Kondisi jalan  | Baik          |
| 13 | Hambatan       | Rendah        |
| 14 | Visibilitas    | Mudah         |
| 15 | Parkir         | Luas          |
| 16 | Keramaian      | Tinggi        |
| 17 | Renovasi       | Rendah        |
| 18 | Lalu lintas    | Tidak macet   |
| 19 | Keamanan       | Aman          |
| 20 | Tenaga kerja   | Mudah         |
| 21 | Bahan baku     | Mudah         |
| 22 | Transportasi   | Mudah         |
| 23 | Daya beli      | Tinggi        |
| 24 | Luas area      | Besar         |
| 25 | Usia penduduk  | Tinggi        |
| 26 | Prospektus     | Prospek       |

Namun dari data tersebut yang diambil sebagai atribut atau variabel keputusan atau tujuan adalah prospektus, sedangkan yang diambil sebagai variabel penentu dalam pembentukan pohon keputusan adalah 22 kriteria atau atribut yaitu :

Tabel 6. Penentuan pembentukan pohon keputusan

|                     |                   |
|---------------------|-------------------|
| 1. Kategori         | 12. Parkir        |
| 2. Perekonomian     | 13. Keramaian     |
| 3. Persaingan       | 14. Renovasi      |
| 4. Populasi         | 15. Lalu lintas   |
| 5. Industri sekitar | 16. Daya beli     |
| 6. Calon konsumen   | 17. Luas area     |
| 7. Posisi Lokasi    | 18. Usia penduduk |
| 8. Pola jalan       | 19. Keamanan      |
| 9. Kondisi jalan    | 20. Tenaga kerja  |
| 10. Hambatan        | 21. Bahan baku    |
| 11. Visibilitas     | 22. Transportasi  |

**4.2. Penentuan Instance atau Nilai Suatu Atribut**

Setelah ditentukan atribut yang pembentukan pohon keputusan kemudian

dilakukan penentuan instance atau fitur masing-masing atribut, sebagai berikut:

Tabel 7. Pengelompokan nilai atribut kategori

|                         |                                       |
|-------------------------|---------------------------------------|
| Makanan                 | produk makanan                        |
| Minuman                 | produk minuman                        |
| Kesehatan               | produk dan jasa kesehatan             |
| Kecantikan              | produk dan jasa kecantikan            |
| Anak                    | produk anak                           |
| Pakaian                 | produk pakaian                        |
| Laundry&jasa kebersihan | jasa laundry dan kebersihan           |
| Otomotif                | produk dan jasa otomotif              |
| Penginapan              | jasa penginapan                       |
| Travel                  | jasa perjalanan                       |
| Furniture               | produk furniture                      |
| Komputer dan internet   | produk dan jasa komputer dan internet |
| Hiburan&Hobi            | produk hiburan dan hobi               |
| Retail                  | produk retail                         |
| Lainnya                 | produk/jasa selain kategori           |

Pengelompokan nilai atribut tingkat perekonomian (dalam radius 1-5 Km)

|    |  |
|----|--|
| A1 | > 500 KK, rumah tipe 90/70, mobil 3000cc |
| A2 | > 500 KK, rumah tipe 70/45, mobil 2400cc |
| B  | > 500 KK, rumah tipe 60/36, mobil 1500cc |
| C  | < 500 KK, rumah tipe 60/36, mobil 1500cc |

Pengelompokan nilai atribut tingkat persaingan (radius 1-5km, bisnis yang sama)

|          |               |
|----------|---------------|
| Tinggi   | Lebih dari 10 |
| Menengah | Antara 5 – 10 |
| Rendah   | Kurang dari 5 |

Pengelompokan nilai atribut ukuran populasi (dalam radius 1-5 Km)

|          |                      |
|----------|----------------------|
| Tinggi   | Lebih dari 1000 KK   |
| Menengah | Antara 500 – 1000 KK |
| Rendah   | Kurang dari 500 KK   |

Pengelompokan nilai atribut industri sekitar (dalam radius 1-5 Km)

|          |                |
|----------|----------------|
| Tinggi   | Lebih dari 50  |
| Menengah | Antara 30 – 50 |
| Rendah   | Kurang dari 30 |

Pengelompokan nilai atribut calon konsumen (Area calon pelanggan yang berbelanja)

|          |                  |
|----------|------------------|
| Primer   | Lebih dari 60%   |
| Sekunder | Antara 20% - 60% |
| Tersier  | Kurang dari 10%  |

Pengelompokan nilai atribut posisi lokasi

|                 |  |
|-----------------|--|
| Pinggir jalan   | Lalu lintas ramai                              |
| Pusat keramaian | Dekat pasar, pusat perbelanjaan, perkantoran   |
| Lain-lain       | Selain pinggir jalan besar dan pusat keramaian |

Pengelompokan nilai atribut pola jalan

|        |                     |
|--------|---------------------|
| 1 arah | Jalan hanya 1 arah  |
| 2 arah | Jalan dengan 2 arah |

Pengelompokan nilai atribut kondisi jalan

|        |                                  |
|--------|----------------------------------|
| Baik   | Jalan baik dan tidak ada kendala |
| Sedang | Jalan baik, air bisat menggenang |
| Rusak  | Jalan rusak                      |

Pengelompokan nilai hambatan

|        |   |
|--------|---|
| Tinggi | Jalan rusak, transportasi sulit, jauh dari perumahan        |
| Rendah | Jalan tidak rusak, transportasi mudah, dekat dari perumahan |

Pengelompokan nilai atribut visibilitas

|       |                |
|-------|----------------|
| Mudah | Mudah terlihat |
| Sulit | Sulit terlihat |

Pengelompokan nilai atribut parkir

|       |                       |
|-------|-----------------------|
| Luas  | Lebih dari 5 mobil    |
| Cukup | Kapasitas 1 - 5 mobil |
| Tidak | Tidak ada             |

Pengelompokan nilai atribut keramaian (jumlah orang lewat tiap hari)

|        |                   |
|--------|-------------------|
| Tinggi | Lebih dari 1000   |
| Sedang | Antara 500 – 1000 |
| Rendah | Kurang dari 500   |

Pengelompokan nilai atribut renovasi

|        |                                   |
|--------|-----------------------------------|
| Tinggi | Perlu perbaikan besar / banyak    |
| Rendah | Hanya cat dinding/perbaikan kecil |

Pengelompokan nilai atribut lalu lintas

|       |                                     |
|-------|-------------------------------------|
| Macet | Sering terjadi macet sekitar lokasi |
| Tidak | Tidak terjadi macet sekitar lokasi  |

Pengelompokan nilai atribut daya beli

|        |   |
|--------|---|
| Tinggi | Sebagian besar usia produktif dan bekerja serta konsumtif             |
| Sedang | Sebagian besar bekerja dan agak konsumtif                             |
| Rendah | Sebagian besar usia tak produktif, tidak bekerja, dan tidak konsumtif |

Pengelompokan nilai atribut luas area

|        |                                  |
|--------|----------------------------------|
| Tinggi | Pilihan jenis yang paling besar  |
| Sedang | Pilihan jenis yang paling Sedang |
| Kecil  | Pilihan jenis yang paling kecil  |

Pengelompokan nilai atribut usia penduduk (Usia target radius 1-5Km)

|        |                            |
|--------|----------------------------|
| Tinggi | > jumlah setengah populasi |
| Rendah | < jumlah setengah populasi |

Pengelompokan nilai atribut keamanan

|        |                                      |
|--------|--------------------------------------|
| Aman   | tak pernah ada pencurian, ada satpam |
| Kurang | Pernah ada pencurian, tak ada satpam |

Pengelompokan nilai atribut tenaga kerja

|       |                                |
|-------|--------------------------------|
| Sulit | Sulit, harus dari luar wilayah |
| Mudah | Mudah, ada di wilayah sekitar  |

Pengelompokan nilai atribut bahan baku

|       |                                     |
|-------|-------------------------------------|
| Sulit | Sulit, tidak ada di wilayah sekitar |
| Mudah | Mudah, ada di wilayah sekitar       |

Pengelompokan nilai atribut transportasi

|        |                             |
|--------|-----------------------------|
| Sulit  | Tak ada angkutan ke lokasi  |
| Sedang | 1 angkutan umum ke lokasi   |
| Mudah  | > 2 angkutan umum ke lokasi |

Pengelompokan nilai atribut prospektus

|               |  |
|---------------|--|
| Propek        | target omset harian dan BEP investasi tercapai       |
| Tidak Prospek | target omset harian dan BEP investasi tidak tercapai |

**4.3. Perancangan Basis Data**

Dalam pembuatan aplikasi dibuat 5 tabel. Adapun tabel saat pengembangan sistem sebagai berikut:

**Tabel Kasus**

Script pembuatan tabel Kasus, yaitu:  
 Create Table Kasus (kategori varchar(50), perekonomian varchar(50), persaingan varchar(50), populasi varchar(50), Bisnis\_sekitar varchar(50), konsumen varchar(50), lokasi varchar(50), pola\_jalan varchar(50), kondisi\_jalan varchar(50), hambatan varchar(50), visibilitas varchar(50), parker varchar(50), keramaian varchar(50), renovasi varchar(50), lalulintas varchar(50), keamanan varchar(50), tenaga\_kerja varchar(50), bahan\_baku varchar(50), transportasi varchar(50), daya\_beli varchar(50), luas\_area varchar(50), usia\_penduduk varchar(50), prospectus varchar(50) );

**Tabel Atribut**

Script pembuatan tabel Atribut, yaitu:  
 Create Table Atribut ( nama\_atribut varchar(50) not null, aktif char(1) default 'Y', hasil char(1) default 'T', primary key(nama\_atribut) );

**Tabel Pohon**

Script pembuatan tabel Pohon, yaitu:

Create Table Pohon ( id\_node integer, node varchar(50), nilai varchar(50), induk varchar(50), atribut char(1) default 'Y' );

**Tabel Gain**

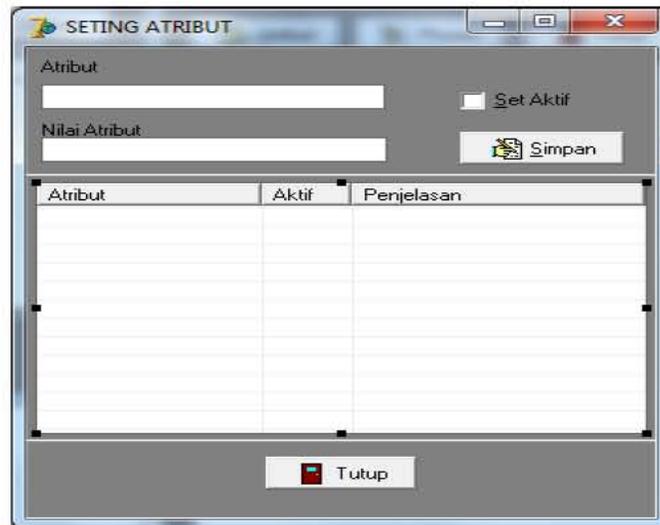
Script pembuatan tabel Gain, yaitu:  
 Create Table Gain ( nama\_atribut varchar(50), gain numeric(10,2) );

**Tabel Entropi**

Script pembuatan tabel Entropi, yaitu:  
 Create Table Entropi ( nama\_atribut varchar(50), nilai varchar(255), entropi numeric(10,2), hasil1 varchar(50), hasil2 varchar(50), jumlahkasus integer );

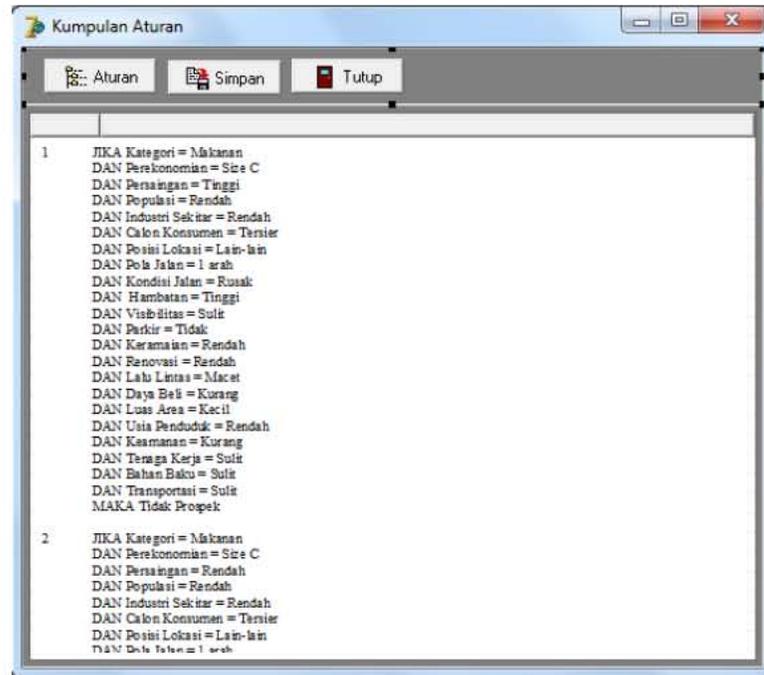
**5. HASIL DAN KESIMPULAN**

Hasil penerapan klasifikasi data dengan menggunakan algoritma C4.5 untuk prospektus lokasi waralaba mampu mengklasifikasikan data dengan atribut atau variabel yang masuk kategori prospektif dan tidak prospektif. Hanya sedikit atribut yang tidak digunakan dalam klasifikasi. Pemilihan atribut cukup penting dalam proses klasifikasi untuk itu sebaiknya atribut yang tidak begitu penting dan berpengaruh jangan dimasukkan dalam klasifikasi. Berikut beberapa tampilan dari aplikasi yang dibangun dengan Delphi Borland 7 dan Interbase 2009.



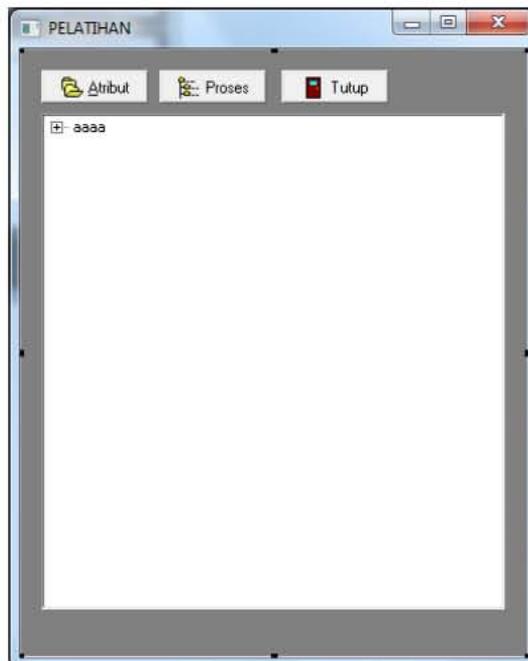
Gambar 4. Tampilan Setting Atribut

Pada menu ini dapat dilakukan setting atribut yang terdiri dari 22 atribut.



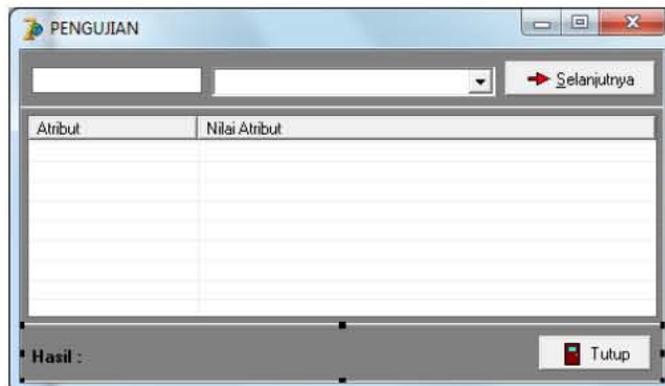
Gambar 5. Menu Kumpulan Aturan

Kumpulan aturan ini akan berisi beberapa waralaba yaitu terdiri 98 aturan. aturan yang menyatakan prospek dari lokasi



Gambar 6. Menu Pelatihan

Menu pelatihan ini digunakan untuk mengetes apakah aturan yang diperoleh dari menu kumpulan aturan sudah sesuai atau tidak. Untuk mengetes kebenaran aturan dibandingkan dengan kasus keputusan bermain tenis (tabel 1)



Gambar 7. Tampilan Menu Pengujian

Pada menu pengujian ini jika ada lokasi yang baru maka tinggal memilih atribut dan nilai atribut yang sesuai kemudian akan memberikan hasil bahwa lokasi dengan kriteria yang ada merupakan lokasi yang prospek atau tidak.

Dalam penerapan aplikasi pada kasus prospektus lokasi waralaba, data yang dimasukkan dalam database sejumlah 186 record kasus waralaba. Dari keseluruhan data tersebut terdapat 152 kasus yang prospek dan sisanya tidak prospek. Dengan mengaktifkan atribut atau variabel bukan tujuan yaitu kategori, tingkat perekonomian, tingkat persaingan, ukuran populasi, industri atau Bisnis sekitar, calon konsumen, Posisi lokasi, pola jalan, kondisi jalan, hambatan ke lokasi, visibilitas, parkir, keramaian, renovasi, arus lalu lintas, keamanan, tenaga kerja, bahan baku, transportasi, daya beli, luas area, dan usia penduduk diperoleh sebanyak 98 aturan. Aturan-aturan atau rule tersebut akan menjadi acuan dalam menentukan calon suatu lokasi waralaba apakah masuk kategori prospek atau tidak.

Dari aturan-aturan yang terbentuk, ada dua kategori aturan. Yaitu aturan yang menghasilkan klasifikasi tertentu, yakni prospek dan tidak prospek. Serta aturan yang kesimpulannya adalah tidak terklasifikasi. Aturan-aturan yang hasilnya tidak terklasifikasi disebabkan karena semua atribut atau variabel yang terdefiniskan, seperti: kategori, tingkat perekonomian, tingkat persaingan, ukuran

populasi, industri atau Bisnis sekitar, calon konsumen, Posisi lokasi, pola jalan, kondisi jalan, hambatan ke lokasi, visibilitas, parkir, keramaian, renovasi, arus lalu lintas, keamanan, tenaga kerja, bahan baku, transportasi, daya beli, luas area, dan usia penduduk sudah muncul dalam tree untuk suatu cabang, tetapi kasus belum bisa mengelompok dalam satu klasifikasi, yakni prospek dan tidak prospek.

#### DAFTAR PUSTAKA

- Han, J. Kamber, M. 2001. Data Mining: Concept and technique. San Fransisco: Morgan Kaufmann Publishers.
- IFBM. 2007. Franchise Manual From Small Drops to Profit. Jakarta: IFBM
- Kusrini & Luthfi Emha T. 2009. Algoritma Data Mining. Jakarta: Penerbit Andi
- Larose, Daniel T. 2005. Discovering Knowledge in Data An Introduction to Data Mining, John Wiley & Sons, Inc., Hoboken, New Jersey.
- Ramadhan E. Hendry. 2009. Franchise untuk Orang Awam. Jakarta: Gramedia Pustaka Utama
- Ramadhan E. Hendry. 2010. Jitu Membeli Franchise. Jakarta: Penebar Plus