

PENGEMBANGAN ANT ALGORITHM DENGAN HYBRIDIZATION CONCEPT UNTUK CLUSTERING DATA

Saiful Bukhori

Program Study of Electrical Engineering Jember University
saiful001@plasa.com

Abstract

Application of ant algorithm for the problem solving with artificial intelligence has grew at full speed. This problem was related to the behavior of ant that was aimed at strove for sustain the ant colony life. Base on the literature survey, the research that was apply the ant algorithm in data mining scope for clustering data has not done

In this research, researcher was design and analysis application of ant algorithm for clustering data . The algorithm, that was used, was modification and improvement for the algorithm that has developed previously. Ant algorithm that was designed was not used four main parameters, ant desirability, ant frequency, heuristic information (α) and pheromone concentration (β), that used in the previous research[2].

The Software that was designed and implemented in operating system Windows has experimented with data from many sources. The result of the experiment was the true classification in 93,48% - 97,00% range, the false classification in 3,00% - 6,52% range and unclassified 0%.

Keyword: *data mining, clustering data, ant algorithm, hibridization concept*

Pendahuluan Latar Belakang

Data Mining adalah kegiatan eksplorasi dan analisa secara otomatis pada suatu basis data yang besar dengan tujuan menemukan pola-pola atau aturan-aturan yang tersembunyi yang mempunyai makna bagi penggunaanya. Data mining memusatkan pada penemuan secara otomatis dari suatu realita baru dalam hubungannya dengan data. Gambaran ini menunjukkan bahwa data mining merupakan bahan baku dalam bisnis data dan algoritma dari

data mining merupakan alat penggantinya.

Definisi *data mining* tersebut mendasari diperlukannya alat penggali dalam hal ini adalah algoritma yang mempunyai kemampuan untuk menghasilkan informasi yang dapat menunjang proses pembuatan keputusan yang dapat dipertanggung jawabkan. Ant algorithm yang didasarkan pada pengamatan terhadap koloni ant yang saat ini terus dikembangkan merupakan pilihan dalam penelitian ini dengan harapan dapat membantu dalam pengembangan data mining

dengan menerapkan *hibridization concept* terutama untuk *clustering data*.

Rumusan Masalah

Rumusan masalah pada penelitian ini ditekankan pada bagaimana mendesain dan menganalisis serta mengimplemen-tasikan *ant algorithm* yang dapat digunakan pada data mining untuk *clustering data* dalam bentuk perangkat lunak yang dapat membantu proses pengambilan keputusan suatu organisasi atau perusahaan

Tujuan

Tujuan utama yang ingin dicapai melalui penelitian ini adalah mendesain dan menganalisis penerapan *ant algorithm* untuk peng-klassifikasi data dengan meng-gunakan konsep *global desirability* dan *global frequency*. Perangkat lunak yang akan dibuat diharapkan dapat dimanfaatkan sebagai sebuah alat yang dapat memberikan informasi tentang *clustering data* dari sebuah basis data yang berukuran besar, sehingga perangkat lunak ini dapat digunakan sebagai alat penunjang proses pengambilan keputusan dalam me-mentukan aturan dari suatu basis data tertentu

Tinjauan Pustaka

Data Mining

Data mining yang juga dikenal dengan *Knowledge Discovery in Database* (KDD) merupakan metoda pencarian terhadap informasi yang bernilai dan tersembunyi dalam suatu basis data yang sulit atau bahkan tidak mungkin untuk ditemukan dengan meng-gunakan mekanisme query standard

atau teknik statistik klasik (Cratochvil, 1999).

Data mining dalam dunia industri disebut sebagai bagian dari *Business Intelligence* (BI) yang merupakan alat bantu bagi perusahaan untuk berkompetisi mendapatkan pasar dan secara cepat dapat memantau kecenderungan pasar. DSS (*Decision Support System*), *Data warehouse*, OLAP (*On-Line Analytical Processing*) dan *Data Mining* merupakan bagian terpenting dari *Business Intelligence* ini (Joseph Bigus P, 1996).

Data mining menggunakan teknik yang berbeda untuk menemukan pola data dan ekstraksi informasi yang umum digunakan adalah *query tools*, teknik statistik, visualisasi, *On-Line Analytical Processing* (OLAP), *case-based learning* (*k-nearest neighbor*), *decision trees*, *association rules*, *neural networks*, *genetic algorithms*.

Data mining dapat menimbulkan kinerja perusahaan lebih kompetitif karena data *mining* dapat membantu perusahaan lebih mengerti pelanggan dan klien mereka. Salah satu contoh informasi yang didapat dari data mining adalah perilaku pelanggan. Dengan mengetahui perilaku pelanggan maka perusahaan dimungkinkan untuk memutuskan perubahan strategi untuk meningkatkan keuntungan dan untuk memenuhi kepuasan pelanggan yang didasarkan dari informasi berharga yang diperoleh dari kumpulan data yang dimilikinya

Jenis data mining didasarkan pada kegunaannya dapat digolongkan menjadi enam golongan, dimana permasalahan-permasalahan intelektual, ekonomi dan kepentingan-

kepentingan bisnis dapat diselesaikan dengan keenam golongan ini (Joseph Bigus P, 1996).

- a. **Klasifikasi (*Classification*)**
Classification adalah jenis data mining yang paling umum di gunakan. Contoh jenis data mining ini adalah dalam pengklasifikasi resiko pemberian kredit atau investasi pada suatu perusahaan, apakah berisiko rendah, sedang atau tinggi
- b. **Estimasi (*Estimation*)**
Estimation adalah bentuk lain dari klasifikasi tetapi berbeda dalam bentuk *outputnya*. Kalau dalam klasifikasi *outputnya* berupa data diskrit, tetapi kalau estimasi digunakan untuk memberikan nilai pada variable-variabel yang bersifat *continous*.
- c. **Prediksi (*Prediction*)**
Prediction atau peramalan pada intinya sama dengan klasifikasi atau estimasi tetapi lebih mengarah pada nilai-nilai pada masa yang akan datang..
- d. ***Affinity Grouping***
Affinity Grouping disebut juga *Market Basket Analysis* adalah suatu analisis yang mendefinisikan benda atau sesuatu yang berhubungan dengan benda atau sesuatu lainnya. Contoh data mining jenis ini adalah untuk mencari keterkaitan berapa kemungkinan sabun akan dibeli jika seseorang sedang membeli *shampoo*.
- e. ***Clustering***
Clustering adalah jenis data mining yang membuat segmen-segmen atau bagian-bagian dari sesuatu populasi yang heterogen

ke suatu sub group yang lebih homogen.

f. ***Description***

Data mining kadang digunakan untuk mendeskripsikan apa yang ada pada suatu basis data yang rumit dan berskala besar dengan tujuan untuk meningkatkan pengetahuan atas orang, produk atau proses yang mempunyai pengaruh besar terhadap data.

Konsep Dasar *Ant algorithm*

Perilaku serangga sosial lebih diarahkan kepada mempertahankan kehidupan keseluruhan koloninya daripada individu-individu tunggal dari koloni tersebut. Perilaku ini menangkap perhatian banyak ilmuwan karena tingginya tingkat struktur koloni yang dapat dicapai, terutama ketika dibandingkan dengan kesederhanaan individu-individu dari koloni. Perilaku yang penting dan menarik dari ant adalah kemampuannya untuk menemukan jalur terpendek antara sumber makanan dan sarang mereka (Dorigi Marco, Roli Andrea, Blum Christian, 2001). Selama *ant* berjalan dari tempat sumber pangan ke sarang dan sebaliknya aktifitas ant antara lain adalah sebagai berikut :

- a. *Ant* meninggalkan di atas tanah suatu zat yang disebut pheromone, dengan demikian membentuk jejak pheromone.
- b. *Ant* dapat mencium bau pheromone, ketika memilih jalan, mereka cenderung memilih jalur yang ditandai dengan konsentrasi pheromone yang kuat.
- c. Jejak *pheromone* membuat ant mampu menemukan jalan kembali ke sumber pangan atau sarangnya. Jejak itu juga dapat

digunakan oleh ant yang lain untuk menemukan lokasi sumber pangan yang ditemukan oleh rekan satu sarangnya.

- d. Pergerakan ant tidak terlalu banyak, tergantung keadaannya dan apakah dia melihat makanan pada arah di depannya atau tidak, ant hanya akan melangkah kedepan, belok kiri, belok kanan atau diam.
- e. Memori ant berisi kondisinya atau disebut chromosome nya, dan chromosome terdiri dari gen-gen yang membawa informasi berupa apa yang ia kerjakan dan apa yang akan dikerjakan selanjutnya.
- f. Secara eksperimental telah terbukti bahwa perilaku mengikuti jejak pheromone ini, apabila diterapkan oleh sebuah koloni ant, mengarah kepada munculnya shortest path. Apabila ada banyak jalur yang tersedia dari sarang menuju ke sebuah sumber pangan, sebuah koloni ant dapat mengeksploitasi jejak pheromone yang ditinggalkan oleh individu-individu ant untuk menemukan lintasan terpendek dari sarang menuju ke sumber pangan dan kembali lagi.

2. Pembahasan

Desain *ant algorithm* pada data mining untuk *clustering* data pada penelitian ini didasarkan pada konsep *hybridization*. Konsep *hybridization* adalah konsep yang membangkitkan segmen-segmen atau bagian-bagian dari sesuatu populasi tanpa menggunakan keempat parameter yang dibutuhkan pada penelitian sebelumnya (Saiful Bukhori, 2003).

Struktur Ant algorithm pada data mining untuk pengclustering data pada penelitian ini dirancang terdiri dari 4 proses utama yaitu : proses pemasukan data, proses pengkodean, proses pembentukan jejaring pheromone dan proses Segmentasi. global *pseudocode* untuk *ant algorithm* pada data mining dalam penelitian ini seperti dalam Gambar 1.

```

input = Dataset(),Kelas = Jumlah Kelas;
THO = 0,FREKTHO = 0,PROBTOT = 0,INFOTHOK = 0,HEUR = 0,PROB = 0,PTUR =0;
FOR (j =0; j <= FieldCount; j++)
//Aturan berdasar konsep hybridization
BEGIN
    value = Table1->Fields[j]-
>AsString;
    FOR (i = 0 ; i <= RecordCount;
i++)
        BEGIN
            IF (ARRAY [i][j] == value )
            THEN // Menentukan Nilai
            Pheromone
                THO = THO + 1;
                LENK= (( k ** 0.001)-
1)/0.001; //Menentukan Nilai
                LOG2(K);
                LOGKLS[k] =
                LENK/LENB;
                namarule = Table1-
>FieldByName("RULE") ;//
                Nilai INFO  $\tau_{ij}$ 
                IF (ARRAY [i][j] == value
                && RECORD[i] == namarule
                )THEN
                    FREKTHO = FREKTHO +
                    1;
                    FREKW = FREKTHO /
                    THO;
                    LENA = (( FREKW **
                    0.001)-1)/0.001;
                    LENB = 0.693;
                    LOGINFO =
                    LENA/LENB;

```

```

        INFOTHO = FREKW *
    LOGINFO;
        INFOTHOK =
    INFOTHOK + INFOTHO;
        HEUR =
        INFOTHO/INFOTHOK
        ; //Menentukan Nilai
        Heuristik
        PROB1 = (HEUR**HIB1)
    * (THO * HIB2);
        PROB[i] = PROB1;
    //Menentukan Nilai Probabilitas
    dari term
        IF (PROB[i][J] >=
    PROB{i-1}[j])
        PROB = PROB[i][j];
    END
    PROBATURAN [i] =
    PROBATURAN + PROB;
    PROBATURAN =
    PROBATURAN[i];
    END
    FOR (k = 0 ; k <= Jmlrule; k++)
    BEGIN
        FOR (i = 0 ; i <= RecordCount;
    i++)
        BEGIN
            IF (PROBATURAN{i} >=
        PROBATURAN{i-1}) THEN
            //penentuan cluster
                PROBATURANCLUS =
        PROBATURAN[i];
            END
            PROBABILITASRULECLUS[k]
    =(( jmlrule - k )/jmlrule) *
    PROBATURANCLUS;
        END
        KELAS[k] =
    PROBABILITASRULECLUS[k];
    
```

Sumber: Hasil Olahan

Gambar 1. *Global Pseudocode Ant algorithm Untuk Penemuan Clustering*

Proses Pemasukan Data

Proses pemasukan data terdiri dari dua bagian yaitu pembacaan data set yang akan ditemukan segmennya dan masukan dari pemakai. Proses pembacaan data set

akan menghasilkan informasi tentang field-field data, informasi tentang record-record data, informasi tentang value (nilai) dari record-record tersebut pada field-field tertentu, jumlah dan nama kelas yang akan dicari segmennya dan dimensi data. Pseudocode proses pemasukan data dapat di rancang seperti dalam Gambar 2.

```

    FOR (0 ≤ i ≤ Data_Set →
    RecordCount)
    BEGIN
        IF (nama_field == "Class") THEN
        Namakelas = Data_Set → Field
        dengan nama("Class");
        int k = 0;
        Bool next=true;
        WHILE ((next==true) and
        (k<jumlah kelas))
        BEGIN
            IF (anamaclus[k] ==
        namaclus)THEN
            Next= false;
            k++;
            END
            FOR (0<j< DataSet→FieldCount -
        1)
            BEGIN
                Temp = 0;
                IF (nama_field !="Class) THEN
                nonclus =
                Dataset→Field→field[i][j];
                Temp++;
                END
            END
        
```

Sumber: Hasil Olahan

Gambar 2. *Pseudocode Proses Pemasukan Data*

Proses Pengkodean

Masukan dari proses pengkodean ini adalah berasal dari *array non clus* yang dihasilkan pada proses pemasukan data. Elemen-elemen *array non clus* yang merupakan nilai dari dari atribut-atribut untuk masing-masing record akan

menentukan terbentuknya simpul dan busur pada jejaring pheromone. Pembentukan simpul dan busur pertama ditentukan oleh agen yang berfungsi sebagai pembentuk segmen yang menunjukkan adanya aturan dengan tipe sesuai yang ditunjukkan oleh simpul. Busur yang telah dibentuk oleh agen yang berfungsi sebagai pembentuk segmen akan diikuti oleh agen yang bekerja sebagai pembentuk cluster berdasarkan bobotnya dengan meletakkan pheromone pada busur tersebut. Konsentrasi pheromone ini akan menentukan kualitas aturan yang telah dibentuk.

Proses pembentukan aturan dilakukan dengan melakukan analisa atribut-atribut berdasarkan ada atau tidak adanya hubungan antara atribut dan nilai yang dimodelkan dengan vektor biner. Atribut ke- i yang disimbolkan dengan A_i menyatakan simpul ke i dan Nilai ke- j dari Atribut ke- i yang disimbolkan dengan V_{ij} menyatakan simpul ke- j dari atribut ke- i . Jika ada busur dari simpul i ke simpul j , maka elemen (A_i, V_{ij}) ditandai dengan "1", dan sebaliknya apabila tidak ada busur dari simpul i ke simpul j , maka elemen (A_i, V_{ij}) ditandai dengan "0". Pseudocode untuk proses pengkodean seperti dalam Gambar 3.

```
FOR (j=0; j <= FieldCount - 1; j++)
BEGIN
FOR (i=0; i <= RecordCount; i++)
BEGIN //Pembentukan termij
value = Table1->Fields[j]->AsString;
term[i][j] = value;
IF(termi,j == termi-1,j) THEN code = 1;
//Pengkodean
ELSE IF (termi,j != termi-1,j) THEN code
= 0;
END
```

END

Sumber: Hasil Olahan

Gambar 3. Pseudocode Proses Pengkodean

Proses Pembentukan Jejaring Pheromone

Masukan dari proses pembentukan jejaring pheromone ini adalah pola data biner seperti yang dijelaskan sebelumnya. Dari pola data tersebut, dapat dibentuk model himpunan solusi yang mungkin yang merupakan subset dari himpunan solusi yang dimodelkan dengan ruang berdimensi n . Pembentukan ruang berdimensi n dilakukan dengan diawali dari ruang dengan dimensi 2 kemudian dimensi 3, dimensi 4 sampai dengan dimensi n dengan cara menarik sudut-sudut dari dimensi sebelumnya. Kode biner '0' ditambahkan pada prefik dari kode bagian sudut luar dan kode '1' ditambahkan pada prefik dari kode sudut bagian dalam.

Simpul pada himpunan solusi yang dimodelkan dengan ruang berdimensi n tersebut menunjukkan adanya aturan yang telah dibentuk oleh agen ant yang berfungsi membentuk segmen, sedangkan busur antara simpul-simpul tersebut menunjukkan lintasan yang diikuti oleh agen ant yang berfungsi membentuk bobot *clustering* dengan meninggalkan *pheromone* pada lintasan tersebut. Konsentrasi *pheromone* yang ada pada lintasan tersebut menunjukkan adanya pembobotan dari aturan yang dibentuk yang nantinya akan menentukan kualitas dari aturan tersebut sebagai pembentuk segmen.

Proses pembentukan segmen dimulai dengan pembentukan

struktur dalam bentuk aturan IF (*condition*) THEN (segmen). Masing-masing ant mulai dengan aturan tanpa *term* dalam *antecedent* dan menambah satu term pada setiap waktu untuk aturan yang sedang dibentuk. Aturan yang sedang dibentuk, dibangun oleh ant sesuai dengan bagian yang diikuti oleh ant tersebut, demikian juga pemilihan term untuk ditambahkan pada bagian aturan yang sedang dibentuk sesuai dengan pemilihan pada bagian yang berlangsung dan akan diperluas pada semua bagian yang memungkinkan. Pemilihan term (pasangan atribut – nilai) untuk ditambahkan tergantung pada fungsi heuristik dan pada jumlah kumpulan *pheromone* pada masing-masing term. Berdasarkan konsep tersebut di atas, maka dapat dibuat pseudocode untuk proses klasifikasi seperti dalam Gambar 4.

```

FOR(j=0; j<= RecordCount; j++)
BEGIN
FOR (i=0; i<= FieldCount – 1; i++)
BEGIN // Rule Construction
value = Table1->Fields[j]; //
Menentukan Nilai Pheromone
IF (ARRAY [i][j] == value ) THEN
THO = THO + 1; //Menentukan
Nilai LOG2 (Kelas)
LENK= (( k ** 0.001)-1)/0.001;
LOGKLS[k] = LENK/LENB; //
Menentukan Nilai INFO τij
namakelas = Table1-
>FieldByName("CLASS");
IF (ARRAY [i][j] == value &&
RECORD[i] == namakelas)THEN
FREKTHO = FREKTHO + 1;
FREKW = FREKTHO / THO;
LENA = (( FREKW ** 0.001)-
1)/0.001;
LENB = 0.693;
LOGINFO = LENA/LENB;
INFOTHO = FREKW *
LOGINFO;

```

```

INFOTHOK = INFOTHOK +
INFOTHO;
HEUR =
INFOTHO/INFOTHOK;
//Menentukan Nilai Heuristik
PROB1=(HEUR**Hibri1)*(THO
* Hibri2);//Tentukan Prob. term
PROB[i] = PROB1;
IF (PROB[i][J] >= PROB{i-1}[j])
THEN
PROB = PROB[i][j];
PROBATURAN [i] =
PROBATURAN + PROB;
PROBATURAN =
PROBATURAN[i];
Aturan = Term[i][j];//Rule
berdasar penambahan termij
END
END
FOR(j=0; j<= RecordCount; j++)
BEGIN
IF (PROBATURAN{j} >=
PROBATURAN{j-1})THEN
PROBATURANKLAS =
PROBATURAN[j];
END

```

Sumber: Hasil Olahan

Gambar 4. Pseudocode Proses Pembentukan Jejaring *Pheromone*

Proses Segmentasi

Proses klasifikasi digunakan untuk menentukan segmen yang sesuai dengan aturan yang terbaik yang ditemukan oleh algoritma, sedangkan jumlah segmen yang telah ditemukan dalam proses pemasukan data digunakan sebagai acuan untuk menentukan jumlah aturan segmen yang diambil dari proses pembentukan jejaring pheromone yang dilakukan sebelumnya. Proses pengupdate-an segmen dilakukan dengan cara menarik semua aturan yang berada pada dimensi yang terluar menjadi dimensi di bawahnya satu

demisatu sampai mencapai dimensi pada aturan terbaik.

Dari konsep tersebut, maka dapat dibuat *pseudo code* seperti dalam Gambar 5.

```

FOR (k = 0; k< Jmlclus; k++)
BEGIN
FOR (j = 0; j<= RecordCount; j++)
BEGIN
IF (PRATURAN{j} >=
PRATURAN{j-1})
PRATURANKLAS = PRATURAN{j}
PRRULECLASS[k]=((jmlclus-
k)/jmlclus)*PRATURANKLS
END
END
KELAS[k] =
PRAILITASRULECLASS[k]
    
```

Sumber: Hasil Olahan

Gambar 5. Pseudocode Proses Segmentasi

Uji Coba Perangkat Lunak

Uji coba perangkat lunak dilakukan dalam lingkungan sistem operasi Windows yang dijalankan pada komputer PC Pentium III 667 MHz dengan RAM berkapasitas 128 MB dan Harddisk sebesar 10 GB. Dalam penelitian ini, terdapat lima jenis data yang digunakan dalam uji coba. Statistik dari kelima jenis data uji coba yang digunakan dapat dilihat dalam tabel 1.

Tabel 1. Statistik Data Uji Coba

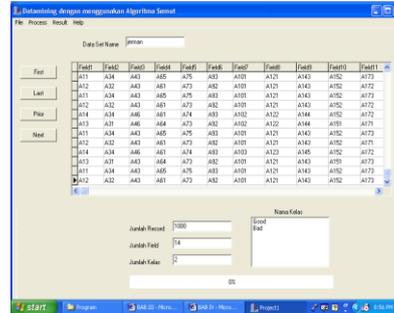
| Data Set | Jumlah Record | Jumlah Atribut Kategorikal | Jumlah Kelas |
|---------------|---------------|----------------------------|--------------|
| Dermatologi | 366 | 33 | 6 |
| German Credit | 1000 | 13 | 2 |
| Car | 1.210 | 6 | 4 |
| Nursery | 12.960 | 8 | 5 |

Sumber: Hasil Olahan

Pelaksanaan Uji Coba

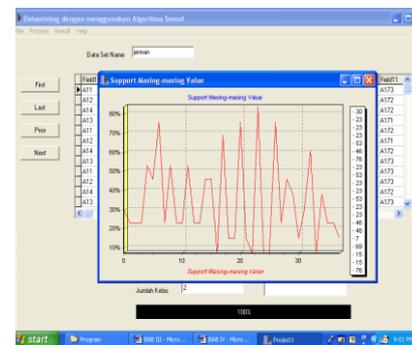
Untuk mengevaluasi kinerja dari perangkat lunak, pelaksanaan uji coba dilakukan dengan mencoba

semua dataset dengan metodologi seperti dalam Gambar 6, Gambar 7 dan Gambar 8 seperti di berikut:



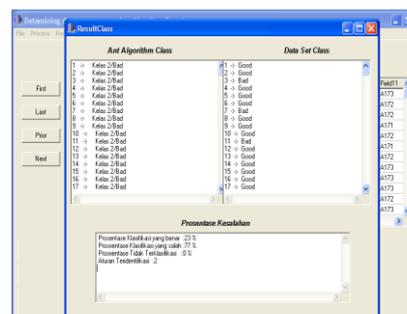
Sumber: Hasil Olahan

Gambar 5.6 Pembacaan Dataset



Sumber: Hasil Olahan

Gambar 5.7 Grafik Bobot masing-masing Segmen



Sumber: Hasil Olahan

Gambar 5.8 Segmen yang ditemukan

Kesimpulan

Di bawah ini diberikan beberapa simpulan yang dapat diperoleh dari hasil penelitian yang dilakukan

sehubungan dengan desain dan analisis penggunaan *ant algorithm* pada data mining untuk clustering data

- a) Dari hasil desain dan analisis algoritma yang digunakan dalam perangkat lunak *clustering* data dengan menggunakan *ant algorithm* dapat disimpulkan bahwa algoritma yang dikembangkan dalam penelitian ini kompleksitas waktu terjeleknya adalah $O(mn)$ dengan m dan n berturut-turut menyatakan jumlah atribut dan jumlah *record* dari data.
- b) Pada Uji coba data set German Credit diperoleh persentase nilai tersegmentasi benar terbesar yaitu 97,00%, sedangkan pada data set dermatology persentase nilai tersegmentasi benar yaitu 93,47%, pada data set Car prosentase nilai tersegmentasi benar yaitu 94,28% dan pada data set Nursery persentase nilai tersegmentasi benar yaitu 95,38%
- c) Berdasarkan persentase jumlah segmen yang dapat diklusterisasi secara benar, sesuai dengan kategori segmen yang ada dalam data uji coba, hasil uji coba menunjukkan bahwa hasil yang diperoleh persentase keberhasilan segmentasi (97,00% untuk data uji coba German credit, 93,47% untuk data uji coba Dermatology, 94,28 % untuk data uji coba Car dan 95,38% untuk data uji coba Nursery)

Daftar Pustaka

- Bigus P. Joseph, "Data mining with Neural Networks :Solving business problem form Application Development to Decision Support", McGraw-Hill, United States of America, 1996.
- Bukhori, Saiful, "Developing of Ant Algortihm with Global Desirability and Global Frequency Concept in Data Mining for Data Classification", Jurnal Rekayasa, 2003
- Cratochvil Anda, "Data Mining Techniques in Supporting Decision Making", Master Thesis, Universiteit Leiden, <http://www.Ainet-sp.si.vti.bin.shtml.dll/education.html>, 1999.
- Cover, T. M, Thomas, J. A, "Element of Information Theory", Hohn Wiley & Sons, New York, 1991.
- Dorigo Marco, Roli Andrea, Blum Christian, "HC-ACO : The Hyper-Cube Framework for Ant Colony Optimization", the 4th Metaheuristic International Conference (MIC'2001), Porto, Portugal, 2001.