

Implementasi Log Analisis Menggunakan Jaringan Saraf Tiruan Untuk Memprediksi Perilaku Pengguna Internet

Nurlindasari Tamsir¹, Nirwana²

STMIK Dipanegara Makassar

Jalan Perintis Kemerdekaan Km. 9 Makassar, Telp. (0411) 587194 – Fax. (0411) 588284

e-mail : stmik14@gmail.com, nirwana_math06@yahoo.com,

Abstrak

Analisis Log didefinisikan sebagai kegiatan menganalisa data yang terkumpul dalam log transaksi untuk menyelidiki informasi sesuai yang kita cari. Hasil analisis log ini dapat membantu kita dalam meningkatkan pelayanan kepuasan konsumen atau user yang terdapat dalam suatu instansi/institusi. Pihak institusi mampu menginformasikan mengenai siapa saja yang memiliki karakter kunjungan web yang cenderung sama berdasarkan histori kunjungan web di internet. Histori kunjungan web ini bisa kita temukan di file squid log. Metode JST ART-2 mampu melakukan clustering terhadap user yang memiliki karakteristik kunjungan web yang sama. Kualitas clustering dari JST ART-2 ini dapat kita hitung menggunakan metode silhouette coefficient. Setelah melakukan penghitungan kualitas clustering dan memilih nilai parameter alpha yang kecil serta nilai rho yang mendekati 1 maka hasilnya yaitu sekitar 0.956025957. Nilai kualitas ini termasuk dalam rentang 0.71-1 pada tabel referensi silhouette coefficient. Ini menunjukkan bahwa struktur cluster yang kuat sudah ditemukan.

Kata kunci: analisis log, clustering, silhouette coefficient

Abstract

Log analysis is defined as the activity of analyzing the data collected in the transaction log to investigate the information according to what we are looking for. The results of this log analysis can help us in improving customer satisfaction services or users contained in an agency / institution. Institutions are able to inform anyone who has a web-visit character that tends to be the same based on the history of web visits on the internet. This web visit history can be found in squid log file. The ART-2 JST method is capable of clustering against users with the same web visit characteristics. The clustering quality of this JST ART-2 can be calculated using silhouette coefficient method. After calculating the quality of clustering and choosing the value of small alpha parameters and rho value close to 1 then the result is about 0.956025957. This quality value is included in the range 0.71-1 in the reference table silhouette coefficient. This shows that a strong cluster structure has been found.

Keywords: analisis log, clustering, silhouette coefficient

1. Pendahuluan

Semakin banyak pengguna internet pada suatu institusi atau perusahaan, maka semakin banyak atau semakin besar pula data yang terekam dalam *log access web server*. Ukuran data log ini tidak hanya dalam ukuran megabyte, tetapi dapat juga dalam terabyte atau bahkan sampai petabyte. Analisa terhadap data *log access web server* diharapkan mampu memberikan informasi sehingga dapat meningkatkan efektifitas sebuah situs, memberikan layanan komunikasi yang lebih baik, dan juga dapat meningkatkan target tertentu terhadap sebuah kelompok pemakai (admin). Sebagai contoh STMIK Dipanegara memiliki satu bagian yang bernama SISFO, SISFO dapat memanfaatkan aplikasi ini untuk mengetahui tren kunjungan web yang diakses oleh mahasiswa. Dengan melihat pentingnya data *log access web server*, maka diperlukan sebuah sistem yang mampu melakukan analisis terhadap data tersebut secara efisien, sehingga informasi yang tersembunyi dibalik data *log access web server* dapat digali sesuai kebutuhan. Kita dapat menampilkan system reporting situs yang dikunjungi baik itu per hari, per bulan atau per tahun (sesuai kebutuhan), frekuensi kunjungan suatu web, dan penggunaan bandwidth suatu jaringan.

Untuk menganalisa data log ini dibutuhkan sebuah metode yang dapat membantu kita dalam memprediksi tingkah laku pengguna internet. Dalam hal ini digunakan metode ART-2 neural network dalam melakukan clustering user yang memiliki tingkah laku kunjungan web yang cenderung sama. Kami menggunakan metode ART2 neural network ini dengan alasan, ART-2 mampu membaca inputan matriks antara 0 dan 1 (desimal).

1.1. Transaction Log Analystis (TLA)

Menurut Bernard, Transaction Log Analysis dapat didefinisikan sebagai kegiatan menganalisa data yang terkumpul dalam log transaksi, untuk menyelidiki informasi sesuai yang kita cari terkait dengan interaksi penggunaan web, browser web yang digunakan, content yang paling sering diakses oleh pengguna internet, dan lain-lain. Penggunaan data yang disimpan dalam log transaksi dari mesin pencari Web, intranet, dan situs web dapat memberikan informasi yang berharga kepada kita, sesuai yang dibutuhkan[1].

Penggunaan data yang disimpan dalam log transaksi dari mesin pencari Web, intranet, dan situs web dapat memberikan informasi yang berharga kepada kita, sesuai yang dibutuhkan. Sebagai contoh, kita dapat mengetahui seberapa sering user mengakses suatu situs, konten apa yang sering diakses oleh user, serta dapat membantu kita dalam mendesain antarmuka suatu web.

1.2. Bahasa Pemrograman Ruby

Michael menjelaskan bahwa ruby adalah bahasa pemrograman dinamis berbasis skrip yang berorientasi obyek. Tujuan dari ruby adalah menggabungkan kelebihan dari semua bahasa-bahasa pemrograman skrip yang ada di dunia. Ruby ditulis dengan bahasa pemrograman C dengan kemampuan dasar seperti Perl dan Python[2].

1.3. Jaringan Saraf Tiruan (JST)

Suyanto mengatakan Artificial neural network (ANN) atau yang diterjemahkan menjadi jaringan saraf tiruan (JST) merupakan suatu arsitektur jaringan yang digunakan untuk memodelkan cara kerja sistem saraf manusia (otak) dalam melaksanakan tugas-tugas tertentu[3]. Pemodelan ini didasarkan pada cara kerja otak yang dapat mengorganisasikan sel-sel saraf (neuron) secara efektif dalam melaksanakan tugas-tugas tertentu khususnya dalam pengenalan pola. Tentu saja, pemodelan ini jauh lebih sederhana daripada cara kerja sel-sel saraf yang sebenarnya.

Pemodelan JST yang berbasis komputer untuk membantu pemecahan masalah aktivitas manusia telah banyak dimanfaatkan oleh beberapa kalangan, seperti pakar, ilmuwan dan peneliti. Hal ini dikarenakan JST memiliki beberapa kelebihan menurut Bintang, antara lain: nonlinearitas, adaptivitas, pemetaan masukan-keluaran, toleransi atau ambang kesalahan dan Generalisasi[4].

Menurut Suyanto, model neuron ini memiliki tiga elemen dasar, yaitu: Sekumpulan sinapsis atau jalur hubungan, adder untuk menjumlahkan sinyal-sinyal input yang diberi bobot oleh sinapsis neuron yang sesuai dan fungsi aktivasi untuk membatasi amplitud output dari setiap neuron[5].

1.4. JST ART

Menurut Suyanto, pada umumnya JST memiliki kecenderungan untuk melupakan informasi yang lama begitu informasi yang baru diberikan dalam hal pengenalan pola tertentu. Demi mengatasi kekurangan tersebut, pola baru yang akan dipelajari dan ditambah dengan pola lama dilakukan pelatihan[5].

1.5. Adaptive Resonance Theory 2

Menurut Carpenter dan Grossberg, Adaptive Resonance Theory 2 merupakan suatu arsitektur jaringan saraf tiruan yang dibuat dengan beberapa prinsip desain yang muncul dari analisis terhadap jaringan saraf tiruan yang membentuk recognition codes dari input yang masuk secara acak[6]. Yaitu stability-plasticity tradeoff, search-direct access tradeoff, dan match-reset tradeoff, dll. ART-2 mampu mengenali pola data kontinyu sebagai pengembangan dari ART 1 yang hanya bisa mengenali pola data biner. ART-2 termasuk salah satu implementasi dari leader clustering yang menetapkan sebuah objek sebagai representasi sebuah cluster.

1.6 Silhouette Coefficient

Silhouette coefficient merupakan suatu metode evaluasi cluster yang menggabungkan cohesion dan separation[13]. Misal $X = \{X_1, \dots, X_N\}$ merepresentasikan data set-nya dan $C = (C_1, \dots, C_K)$

merepresentasikan Kcluster yang terbentuk, $d(X_k, X_l)$ merupakan jarak diantara X_k dan X_l dan $C_j = \{X_1^j, \dots, X_{m_j}^j\}$ merupakan cluster ke- j dimana $m_j = |C_j|$.

Rentang nilai silhouette coefficient menurut Kauffman dan Rousseeuw memiliki arti tertentu seperti disebutkan pada tabel 1[7].

Tabel 1. Rentang Nilai Silhouette Coefficient

No	Rentang Nilai	Keterangan
1	0.71-1	Struktur cluster yang kuat sudah ditemukan
2	0.51-0.7	Struktur cluster yang cukup bagus sudah ditemukan
3	0.26-0.5	Struktur cluster yang terbentuk lemah
4	0.25	Tidak terlihat adanya struktur cluster

2. Metode Penelitian

Jenis penelitian yang dilakukan meliputi penelitian lapangan, penelitian pustaka dan penelitian laboratorium. Penelitian lapangan dilakukan dengan cara mengambil data squid log kemudian dianalisis agar siap digunakan sebagai bahan penelitian. Implementasi dilakukan dengan menggunakan ruby dan menganalisa data squid log dengan metode ART-2. terdapat 2 tugas utama yang dilakukan pada pengujian yaitu tahap training pada metode ART-2 dan tahap testing pada metode ART-2.

2.1. Spesifikasi Perangkat Keras dan Perangkat Lunak

Perangkat Keras (Hardware) yang digunakan dalam penelitian ini memiliki spesifikasi sebagai berikut: Processor Intel Core 2 Duo 2 GHz, Memory DDR2 2 GB, Monitor 14,1 WXGA resolusi 1280 x 800 px dan Graphic Card IntelGMA4500MHD. Adapun Perangkat Lunak (Software): sistem operasi Windows 7 Professional 32 bit, LTF (Large Text File) Viewer 5.2u, Matlab R2009a version 7.8, Ruby 1.9.2-p290.

2.2. Bagan Alir

Pada penelitian ini akan dibangun sebuah sistem untuk mengenali pola tingkah laku pengguna internet berdasarkan data squid log dengan metode ART-2. Sistem ini terbagi menjadi 3 modul utama, yaitu *preprocessing*, *processing* dan *reporting*.



Gambar 1. Diagram Alir Umum Sistem

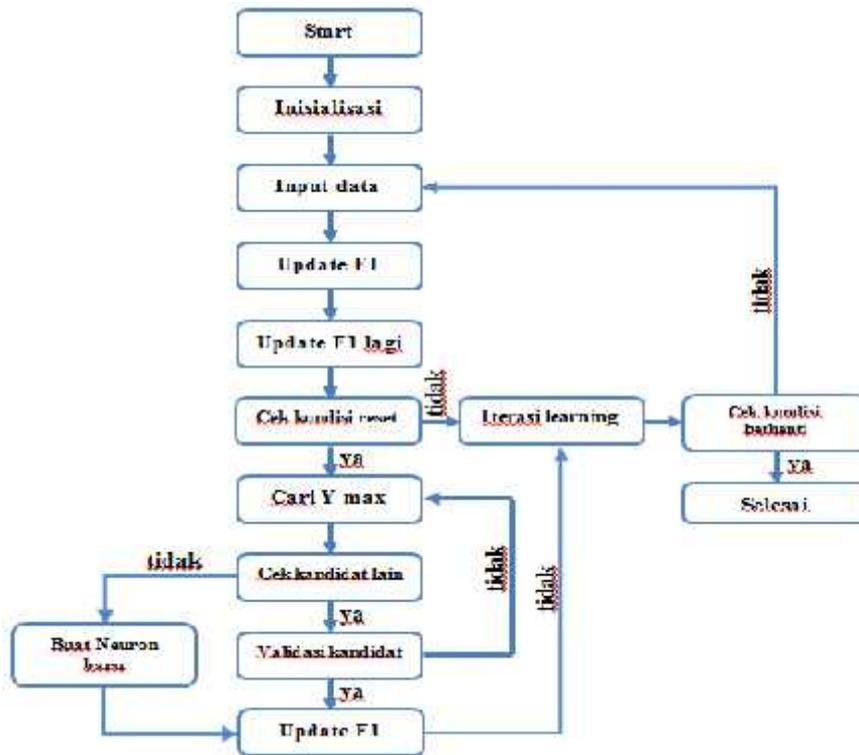
Pada gambar 1 dapat kita lihat bahwa kegiatan dimulai pada apa yang akan kami analisa adalah data squid log yang diperoleh dari server STMIK Diponegara. Preprocessing data yang diperoleh untuk penelitian ini masih berupa data mentah ataupun data yang belum terstruktur dengan jelas. Oleh sebab itu diperlukan data preprocessing agar data yang diperoleh dapat mendukung penelitian. Setelah data preprocessing selesai, data akan diproses agar mendapatkan hasil cluster IP-Address user yang memiliki karakter kunjungan web yang cenderung sama. Setelah semuanya selesai, hasil akhir dari sistem ini berupa tampilan IP-Address user yang berada dalam satu cluster.



Gambar 2. Diagram Alir perluasan Modul Preprocessing

Pada gambar 2 dapat kita lihat bahwa kegiatan pada modul preprocessing mempunyai tahapan yaitu: Data Cleaning, Sorting IP, Split URL, Stopword dan Pembobotan.

Berikut ini akan diperlihatkan proses yang terjadi dalam ART-2 :



Gambar 3. Proses ART-2

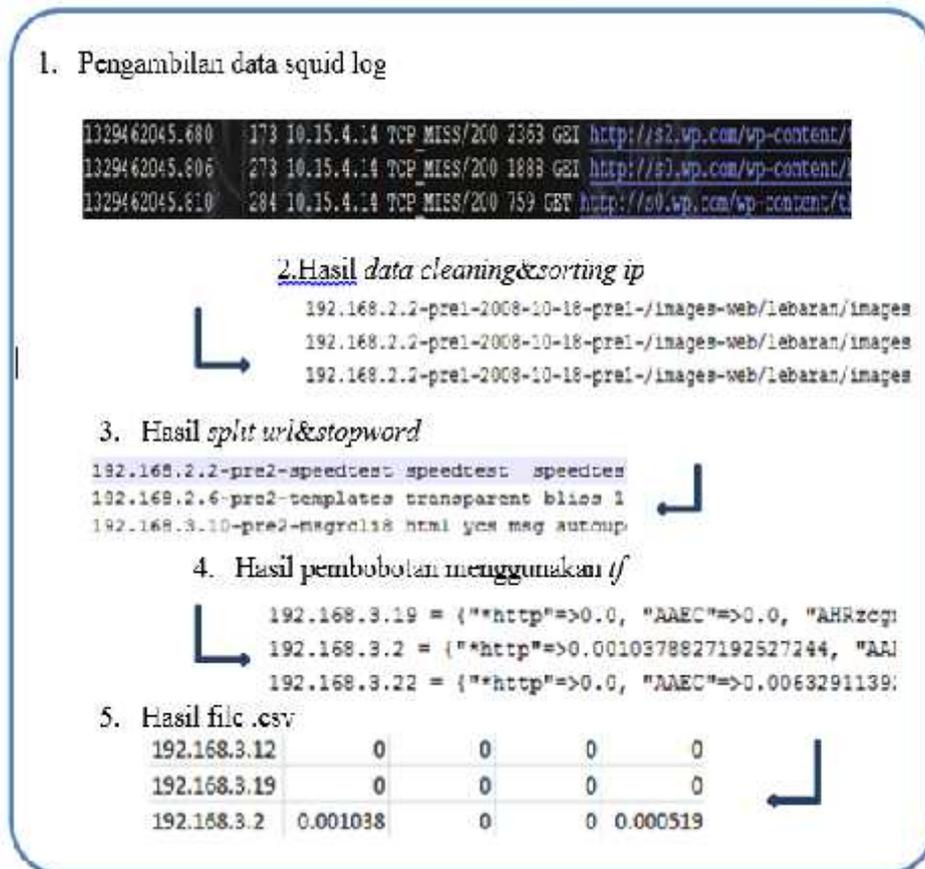
2.3. Inisialisasi Parameter

- $(\rho) = 0.99$
dipilih antara 0–1. Namun, hanya nilai antara 0.7–1 menghasilkan peranan penting dalam mengontrol jumlah cluster. Nilai di bawah 0.7 mempunyai efek yang sama dengan mengubah menjadi nol.
- $(\alpha) = 0.2$
Nilai yang lebih kecil akan memperlambat proses belajar pada mode fastlearning atau slow learning. Nilai yang lebih kecil meyakinkan bobot mencapai kesetimbangan (equilibrium).
- $e = 0.0000000001$
Nilai yang lebih kecil digunakan untuk mencegah pembagian dengan nol.
- $(\theta) = 0$
Parameter penekan derau ini digunakan untuk mengganti nilai yang berada di bawah menjadi nol.

2.4. Ilustrasi Perancangan Sistem

Untuk memperjelas perencanaan sistem kami, maka kami akan membuat ilustrasi dari modul *preprocessing*, *processing* dan modul *reporting*.

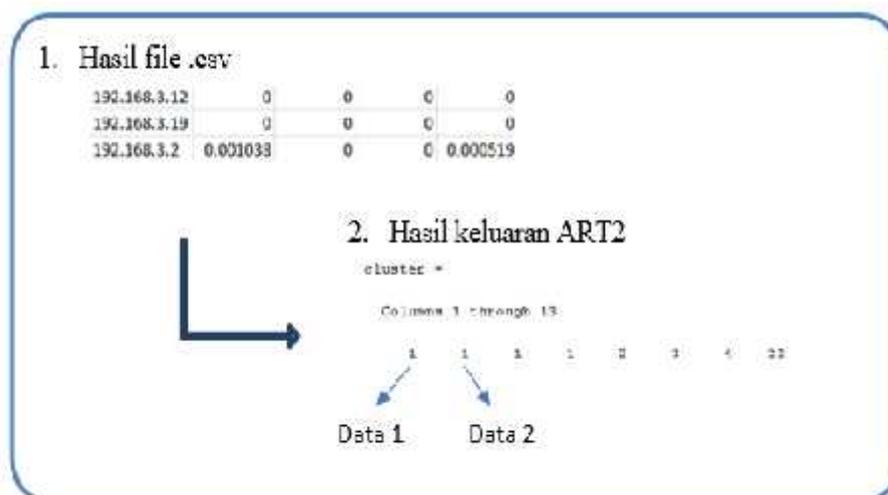
2.4.1 Modul Preprocessing



Gambar 4. Ilustrasi Modul Preprocessing

Penjelasan hasil keluaran ART2: Untuk data pertama, kedua, ketiga dan keempat berada dalam 1 cluster, yaitu cluster 1. Untuk data ke 5 berada pada cluster 2, data ke 6 pada cluster 3, data ke 7 pada cluster ke 4, dan data ke 8 pada cluster ke 22, dan seterusnya.

2.3.1 Modul Processing



Gambar 5. Ilustrasi Modul Preprocessing

3. Hasil dan Pembahasan

3.1. Implementasi Sistem

Sistem ini merupakan perangkat lunak yang berbasis desktop pada lingkungan sistem operasi Windows 7 Professional 32 bit. Sistem ini dibangun dengan menggunakan Ruby 1.9.2-p290 untuk melakukan parser dengan kata lain modul preprocessingnya, Matlab R2009 version 7.8 untuk melakukan komputasi fungsi matematis dari ART2 dengan kata lain modul processingnya. Perangkat lunak LTF Viewer kami gunakan untuk membuka file squid log yang ukurannya sangat besar yaitu 18 gigabyte.

3.2. Pengujian White Box

Pengujian ini dilakukan dengan tujuan untuk mengetahui pengaruh parameter alpha (*learning rate*) dan parameter rho (*vigilance* parameter) terhadap kualitas *clustering* dan jumlah cluster yang terbentuk. Pada pengujian ini, digunakan data uji berjumlah 10000 baris data yang terdapat pada squid log. Adapun parameter yang akan dianalisa adalah parameter alpha dan parameter rho dari ART-2. Pengujian ini dikelompokkan kedalam 7 skenario utama berikut.

3.2.1. Skenario 1

Hasil yang didapat dari pengujian skenario 1 adalah sebagai berikut :

Tabel 2. Hasil Clustering Skenario 1

No	alpha	jumlah cluster	silhoutte global
1	0.01	1	0
2	0.1	1	0
3	0.3	1	0
4	0.5	1	0
5	0.7	1	0
6	0.9	1	0
7	0.999	1	0

Berdasarkan hasil pengujian pada tabel 2 maka dapat diperlihatkan bahwa, hasil clustering dengan menggunakan skenario 1 tidak menunjukkan adanya struktur cluster. Hal ini dapat kita lihat dari nilai silhouette coefficient globalnya yang masuk dalam kategori 2.5 (tabel 1). Nilai alpha dan rho dari skenario 1 ini sangat mempengaruhi kualitas clustering yang dihasilkan. Nilai alpha yang kecil akan mempengaruhi kecepatan learning, sedangkan nilai rho yang kecil akan mempengaruhi jumlah cluster yang akan terbentuk. Untuk nilai rho 0.7 tidak akan membentuk cluster.

3.2.2. Skenario 2, 3, 4 dan 5

Hasil yang didapat dari pengujian skenario 2, 3, 4 dan 5 adalah sebagai berikut :

Tabel 3. Hasil Clustering Skenario 2, 3,4 dan 5

No	alpha	jumlah cluster	silhoutte global
1	0.01	1	0
2	0.1	1	0
3	0.3	1	0
4	0.5	1	0
5	0.7	1	0
6	0.9	1	0
7	0.999	1	0

Berdasarkan hasil pengujian pada tabel 3 maka dapat diperlihatkan bahwa, hasil clustering dengan menggunakan skenario 2, 3, 4 dan 5 tidak menunjukkan adanya struktur cluster. Hal ini dapat kita lihat dari nilai silhouette coefficient globalnya yang masuk dalam kategori 2.5 (tabel 1). Nilai alpha dan rho dari skenario 1 ini sangat mempengaruhi kualitas clustering yang dihasilkan. Nilai alpha yang kecil akan mempengaruhi kecepatan learning, sedangkan nilai rho yang kecil akan mempengaruhi jumlah cluster yang akan terbentuk. Untuk nilai rho 0.7 tidak akan membentuk cluster. Jika tidak ada cluster maka nilai dari silhouette coefficient juga tidak ada.

3.2.3. Skenario 6

Hasil yang didapat dari pengujian skenario 6 adalah sebagai berikut :

Tabel 4. Hasil Clustering Skenario 6

No	alpha	jumlah cluster	silhouette global
1	0.01	2	0.112538372867996
2	0.1	7	0.277669142755548
3	0.3	13	0.723427584232302
4	0.5	16	0.875793101782478
5	0.7	16	0.88326122616409
6	0.9	16	0.88326122616409
7	0.999	16	0.88326122616409

Berdasarkan hasil pengujian pada tabel 4 maka dapat diperlihatkan bahwa, hasil clustering dengan menggunakan skenario 6 menunjukkan adanya struktur cluster yang bervariasi. Hal ini dapat kita lihat dari nilai silhouette coefficient globalnya yang bervariasi pula. Untuk alpha 0.01 dan 0.1 tidak memperlihatkan adanya struktur cluster, untuk alpha 0.3 memperlihatkan struktur cluster yang terbentuk lemah, dan untuk alpha 0.5, 0.7, 0.9 serta 0.999 memperlihatkan struktur cluster yang kuat sudah ditemukan. Variasi nilai kualitas clustering ini sangat dipengaruhi dengan nilai alpha dan rho yang dipilih. Pemilihan nilai alpha akan mempengaruhi learning rate dan pemilihan nilai rho akan mempengaruhi seberapa banyak cluster yang akan terbentuk.

3.2.4. Skenario 7

Hasil yang didapat dari pengujian skenario 7 adalah sebagai berikut :

Tabel 5. Hasil Clustering Skenario 7

No	alpha	jumlah cluster	silhouette global
1	0.01	8	0.51176822570633
2	0.1	26	1
3	0.3	26	1
4	0.5	26	1
5	0.7	26	1
6	0.9	26	1
7	0.999	26	1

Berdasarkan hasil pengujian pada tabel 5 maka dapat diperlihatkan bahwa, hasil clustering dengan menggunakan skenario 7 menunjukkan adanya struktur cluster yang bervariasi. Hal ini dapat kita lihat dari nilai silhouette coefficient globalnya yang bervariasi pula. Untuk alpha 0.01 memperlihatkan struktur cluster yang terbentuk sudah cukup bagus, untuk alpha 0.1, 0.3, 0.5, 0.7, 0.9, 0.999 memperlihatkan struktur cluster yang kuat sudah ditemukan. Nilai kualitas clustering ini dipengaruhi oleh nilai alpha dan rho yang dipilih. Pemilihan nilai alpha akan mempengaruhi learning rate, sedangkan pemilihan nilai rho > 0.7 akan mempengaruhi jumlah cluster yang akan terbentuk.

Setelah melakukan beberapa macam skenario maka dapat kita lihat, hasil yang terbaik kita peroleh pada skenario ke-7 (alpha = 0,1 rho = 0,999 epoch = 3). Dengan nilai rho yang tidak terlalu tinggi dan nilai alpha yang tinggi metode ART-2 dapat mencapai hasil yang cukup baik dalam memprediksi tingkah laku user pengguna internet. Hal ini terlihat dari nilai silhouette coefficient globalnya yang mencapai 1, berdasarkan tabel 1 maka skenario 7 ini sudah mendapatkan struktur cluster yang kuat. Adapun pemilihan

parameter epoch berdasarkan hasil percobaan yang dilakukan, nilai shilloutte dengan epoch di atas 3 menghasilkan nilai shilloutte yang sama yaitu 1.

Dengan melihat hasil keluaran dari ART-2 yaitu bahan analisa squid log, maka kita dapat melakukan analisa terhadap pengunjung web khususnya di ittelkom yang terekam dalam 10000 baris data dari squid log yang diproses. Dari 10000 baris data tersebut, terdiri dari 33 ip dan menghasilkan 26 cluster. Ini menunjukkan karakter kunjungan web user yang terdapat dalam data squid log yang diproses sangatlah beragam. Hal ini dapat kita buktikan dengan melihat history kunjungan web masing-masing cluster.

3.3 Analisis Cluster

Dengan mengacu pada skenario ke-7 yaitu skenario dengan nilai shiloutte yang terbaik, maka cluster yang terbentuk ada sebanyak 26 cluster. Dari 26 cluster ini hanya beberapa cluster saja yang akan kita analisis yaitu cluster 1, 4, 9, 10, 13, dan 14. Hal ini dilakukan karena hanya cluster ini yang memiliki anggota lebih dari 1 ip, sehingga analisis cluster dapat kita lakukan.

3.3.1 Analisis Cluster 4

Setelah melalui proses data cleaning, sorting ip, split url, stopword dan pembobotan, maka data squid log yang kita gunakan dapat kita analisa. Pada cluster 4 terdapat 4 ip. Ip address yang terdapat pada cluster 4 ini memiliki kemiripan tingkah laku dalam mengakses situs di dunia maya. Hal ini dapat kita lihat dari hasil tahap preprocessing yang menunjukkan nilai term frekuensi tiap ip address sama. Jumlah ip pada cluster 4 ini adalah 4 user. Nilai tf tiap ip address pada cluster 4 adalah sama. Nilai tfnya dapat kita lihat pada gambar dibawah ini.

Karakteristik dari user yang berada dalam cluster 4 adalah mengakses google dan yahoo. Hal ini terlihat dari link-link pada squid log yang diakses oleh ke 4 ip dalam cluster 4 memiliki kesamaan path. Dengan kata lain kata-kata yang dicari/ diakses oleh ke 4 ip ini memiliki kesamaan/ kemiripan yang dicari melalui google dan yahoo.

```
1 172.16.102.31 = {"en"=>0.2, "US"=>0.2, "firefox"=>0.2, "headlines"=>0.2, "xml"=>0.2}
2 172.16.142.178 = {"en"=>0.2, "US"=>0.2, "firefox"=>0.2, "headlines"=>0.2, "xml"=>0.2}
3 172.16.146.221 = {"en"=>0.2, "US"=>0.2, "firefox"=>0.2, "headlines"=>0.2, "xml"=>0.2}
4 172.16.150.13 = {"en"=>0.2, "US"=>0.2, "firefox"=>0.2, "headlines"=>0.2, "xml"=>0.2}
```

Gambar 6. Nilai tf cluster 4

3.3.2 Analisis Cluster 9

Jumlah ip pada cluster 9 ini adalah 2 user. Nilai tf tiap ip address pada cluster 9 adalah sama, sehingga tidak terdapat adanya pencilan. Nilai tfnya dapat kita lihat pada gambar dibawah ini.

Karakteristik dari user yang berada dalam cluster 9 adalah mengakses google. Hal ini terlihat dari link-link pada squid log yang diakses oleh ke 2 ip dalam cluster 9 memiliki kesamaan path. Dengan kata lain kata-kata yang dicari/ diakses oleh ke 2 ip ini memiliki kesamaan/ kemiripan yang dicari melalui google.

```
1 172.16.129.32 = {"complete"=>0.5, "search"=>0.5}
2 172.16.183.47 = {"complete"=>0.5, "search"=>0.5}
```

Gambar 7. Nilai tf cluster 9

3.3.3 Analisis Cluster 10

Jumlah ip pada cluster 10 ini adalah 2 user. Nilai tf tiap ip address pada cluster 10 adalah sama, sehingga tidak terdapat adanya pencilan. Nilai tfnya dapat kita lihat pada gambar dibawah ini.

Karakteristik dari user yang berada dalam cluster 10 adalah mengakses google dan facebook. Hal ini terlihat dari link-link pada squid log yang diakses oleh ke 2 ip dalam cluster 10 memiliki kesamaan path. Dengan kata lain kata-kata yang dicari/ diakses oleh ke 2 ip ini memiliki kesamaan/ kemiripan yang mengacu ke google dan facebook.

```
1 172.16.66.155 = {"search"=>1.0}
2 172.16.137.46 = {"search"=>1.0}
```

Gambar 8. Nilai tf cluster 10

3.3.4 Analisis Cluster 13

Jumlah ip pada cluster 13 ini adalah 2 user. Nilai tf tiap ip address pada cluster 13 adalah sama. Nilai tfnya dapat kita lihat pada gambar dibawah ini.

Karakteristik dari user yang berada dalam cluster 13 adalah mengakses igracias. Path yang diakses ke 2 ip pada cluster ini mengacu kepada igracias.

```
1 172.16.62.163 = {"safebrowsing"=>0.5, "downloads"=>0.5}
2 172.16.171.222 = {"safebrowsing"=>0.5, "downloads"=>0.5}
```

Gambar 9. Nilai tf cluster 11

3.3.5 Analisis Cluster 14

Jumlah ip pada cluster 14 ini adalah 3 user. Nilai tf tiap ip address pada cluster 14 adalah sama. Nilai tfnya dapat kita lihat pada gambar dibawah ini.

Karakteristik dari user yang berada dalam cluster 14 adalah mengakses igracias dan www.msftncsi.com. Hal ini terlihat dari link-link yang diakses oleh ke 3 ip dalam cluster 14 memiliki kesamaan path. Dengan kata lain kata-kata yang dicari/ diakses oleh ke 4 ip ini memiliki kesamaan/kemiripan yang mengacu ke igracias dan msftncsi.

```
1 172.16.99.225 = {"ncsi"=>0.5, "txt"=>0.5}
2 172.16.196.142 = {"ncsi"=>0.5, "txt"=>0.5}
3 172.16.197.225 = {"ncsi"=>0.5, "txt"=>0.5}
```

Gambar 10. Nilai tf cluster 14

3.4 Analisis Menggunakan Data Dummy

Pada analisis menggunakan data dummy ini, penulis mencoba untuk mensimulasikan beberapa ip yang mengunjungi situs yang cenderung sama. Misalkan ada 4 ip baru yang terekam pada squid log, pada kasus ini ip 172.16.73.160 (situs bola), 172.16.72.161 (situs politik dan bola), 172.16.74.162 (situs bola) dan 172.16.74.163 (sama dengan kunjungan web cluster 9). Maka jumlah cluster yang sebelumnya ada 26 cluster, sekarang menjadi 28 cluster. Hal ini terjadi di karenakan ip 172.16.73.160 dan 172.16.74.162 berada dalam 1 cluster baru dan ip 172.16.72.161 ini juga membentuk cluster baru sedangkan untuk ip 172.16.74.163 bergabung dengan ip yang berada pada cluster 9. Pada kasus ini Ip user 172.16.72.161 dibedakan dengan ip 172.16.73.160 dan 172.16.74.162 karena memiliki path yang cukup berbeda walaupun web yang dikunjungi terdapat beberapa situs bola.

Adapun kualitas hasil cluster data dummy ini mencapai 0.959274194044996. Dengan mengacu pada tabel 2.1, maka kualitas hasil cluster ini sudah membentuk struktur cluster yang kuat.

3.5 Analisis Karakter Kunjungan Web

Dengan melihat hasil keluaran dari ART-2 yaitu bahan analisa squid log, maka kita dapat melakukan analisa terhadap pengunjung web khususnya di ittelkom yang terekam dalam 10000 baris data dari squid log yang diproses. Dari 10000 baris data tersebut, terdiri dari 33 ip dan menghasilkan 26 cluster. Ini menunjukkan karakter kunjungan web user yang terdapat dalam data squid log yang diproses sangatlah beragam. Hal ini dapat kita buktikan dengan melihat history kunjungan web masing-masing cluster.

4. Kesimpulan

Berdasarkan pemaparan di atas, penulis dapat tarik kesimpulan sebagai berikut :

1. Kualitas Clustering ART-2 dengan nilai alpha yang kecil dan rho yang mendekati 1 memberikan hasil yang cukup bagus, yaitu masuk ke dalam rentang 0,7-1. Sesuai dengan tabel 2.1 nilai yang terdapat pada rentang ini sudah menemukan struktur cluster yang kuat.
2. Metode ART-2 dapat digunakan untuk menghasilkan bahan analisis data log access web server dalam memprediksi tingkah laku kunjungan pengguna internet.

Daftar Pustaka

- [1] Bernard J. Jansen, 2006, *Search log analysis: "What it is, what's been done, how to do it"*.
- [2] Michael Stal. Senior Principal Engineer. 2005. *The Programming Language Ruby*. Siemens Corporate Technology, CT SE 2.
- [3] Suyanto, ST., MSc., 2008. *Soft Computing: Membangun Mesin Ber-IQ Tinggi*. Bandung: Penerbit Informatika Bandung.
- [4] Bintang, Samiaji. 2002. *Analisa Peramalan Menggunakan Jaringan Syaraf Tiruan Adaptive Resonance Theory: Studi Kasus Pelanggan Telepon Seluler di Indonesia*. Tugas Akhir. Bandung: STT Telkom.
- [5] Suyanto, ST., MSc., 2007. "Artificial Intelligence: Searching, Reasoning, and Learning". Bandung: Penerbit Informatika Bandung.
- [6] Carpenter, G. A., & Grossberg, S. (1987b). *ART-2: Self-organization of stable category recognition codes for analog input patterns*. *Applied Optics*, 26, 4919-4930.
- [7] L. Kaufman, and P.J. Rousseeuw. (1999). "Finding Groups in Data: an Introduction to Cluster Analysis". John Wiley & Sons.