

BAB I Pendahuluan

I.1 Latar Belakang

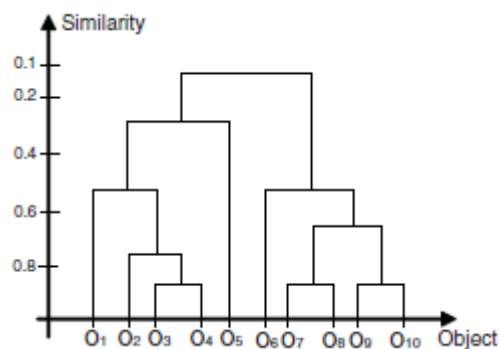
Perkembangan era internet saat ini meningkat dengan sangat pesat. Pemanfaatan teknologi internet (*World Wide Web*) saat ini telah mencakup semua aspek di kehidupan. Seiring perkembangan teknologi internet ini, aktivitas *user* dalam penggunaannya ikut meningkat. Menurut survei yang dilakukan Netcraft, tercatat sebanyak 628.170.204 situs yang aktif. Dengan memanfaatkan media internet ini, *user* dapat lebih mudah dalam mendapatkan informasi ataupun berita melalui berbagai aplikasi internet seperti e-commerce, e-news, online shop, e-learning dan lain-lain.

Untuk menjamin kepuasan *user* dalam menggunakan suatu website, perlu diperhatikan performansi dan kualitasnya. Kecenderungan *user* dalam mengakses suatu website dapat dijadikan tolok ukur untuk meningkatkan dua hal tersebut. Analisis ini dapat dilakukan dengan pemanfaatan ilmu *Web Mining*. *Web mining* adalah ilmu yang mempelajari penggalian data yang berhubungan dengan World Wide Web (WWW).

Web Mining sendiri dapat dibedakan atas tiga kategori yang berbeda, yaitu *web content mining*, *web structure mining*, dan *web usage mining*. Ketiga kategori ini memiliki fungsi masing-masing dalam pengolahan data. *Web usage mining* sendiri merupakan proses analisis terhadap pola akses user dan segala aktivitasnya pada suatu website. Web server akan menyimpan semua data tentang interaksi user dengan website pada sebuah log file. Log file ini kemudian akan diolah melalui beberapa tahap, yaitu *preprocessing*, *pattern discovery*, dan *pattern analysis*. Sangat banyak metode yang digunakan untuk melakukan penerapan *web usage mining* ini, salah satunya adalah metode *clustering*. *Clustering* merupakan proses mengelompokkan sekumpulan *object* fisik maupun abstrak ke dalam kelas tertentu berdasarkan kesamaannya [8]. Dalam kaitannya dengan *web usage mining*, teknik *clustering* sering digunakan untuk menentukan segmentasi pasar pengunjung suatu situs e-commerce berdasarkan kesamaan pola akses maupun demografinya.

Metode *hierarchical agglomerative clustering* adalah salah satu teknik *clustering* yang dapat diterapkan untuk mencari kesamaan pola akses suatu website. Metode ini adalah teknik *clustering* yang membentuk konstruksi hirarki berdasarkan tingkatan tertentu seperti struktur pohon. Algoritma ini bekerja dengan menggabungkan N *cluster* menjadi satu *cluster* dengan melihat jarak antar *cluster* yang ada. Penggabungan ini akan dilakukan secara bertahap dengan melihat jarak antar *cluster*. Dengan demikian proses pengelompokannya dilakukan secara bertingkat atau bertahap. Untuk beberapa penelitian sebelumnya, HAC juga dapat digunakan untuk pengelompokan data seperti dokumen bahasa Indonesia.

Dengan metode seperti ini, HAC dinilai cocok dengan kebutuhan untuk pengolahan dan pengelompokan data log file. Hasil dari pengelompokan ini dapat disajikan dalam bentuk dendrogram seperti pada Gambar 1.1 :



Gambar I.1 Contoh Dendrogram

I.2 Rumusan Masalah

Dari latar belakang yang dijelaskan sebelumnya, rumusan masalah yang diberikan adalah :

1. Bagaimana menerapkan metode *clustering* untuk menentukan pola akses user pada suatu website dengan memperhatikan kemiripan aktivitas user.
2. Bagaimana menghasilkan rekomendasi peningkatan kualitas suatu website berdasarkan analisis terhadap pola akses dan segmentasi user.
3. Bagaimana segmentasi pengguna *icaring* untuk dijadikan informasi evaluasi website.

I.3 Tujuan

Tujuan yang ingin dicapai dari penelitian tugas akhir ini adalah :

1. Menerapkan metode *clustering* sebagai salah satu teknik *pattern discovery* untuk menentukan pola akses user pada suatu *icaring*.
2. Memberikan rekomendasi perbaikan website berdasarkan keterkaitan pola akses user.
3. Melihat segmentasi pengguna user *icaring* sebagai informasi evaluasi website.

I.4 Batasan Masalah

Batasan masalah pada penelitian ini adalah :

1. Data yang diambil adalah data aktivitas user dalam mengakses situs <http://icaring.ittelkom.ac.id/>
2. Data yang digunakan dari bulan Februari sampai dengan bulan Mei (dari awal i-caring digunakan).
3. Sistem yang dibuat masih bersifat *offline*.
4. Format data web server log yang digunakan adalah NCSA *common* yang mengandung IP address, waktu, request, status, size.

I.5 Hipotesa

Hierarchical Agglomerative Clustering dapat menjadi salah satu metode *web usage mining* untuk pencarian pola akses user melalui *cluster* yang terbentuk yang berguna untuk informasi pengembangan website.

I.6 Metodologi Penelitian

Penelitian ini akan dikerjakan dengan metodologi sebagai berikut:

1. Studi Literatur
Pada tahap ini dilakukan dengan mempelajari buku-buku referensi atau sumber-sumber yang berkaitan dengan skripsi ini, baik dari *text book* maupun *internet*.
2. Pengumpulan Data
Pada tahap ini dilakukan pengumpulan data log server situs yang telah ditentukan sebelumnya.
3. Analisis Data

- a. Melakukan data selection dari server log, dengan mengambil data aktivitas akses user dari server aplikasi web
- b. Melakukan preprocessing terhadap data mentah berupa server log, dengan melakukan tahapan: *parsing data* yaitu memisahkan data log menjadi bagian-bagian tertentu yang lebih kecil, *cleaning data* yaitu membuang data yang tidak diperlukan saat proses pencarian pola (sehingga menghasilkan atribut yang akan dianalisis seperti User IP dan halaman yang diakses) dan *data transformation* yaitu mengubah data menjadi data bertipe *discrete sequence* sebagai inputan pada tahap selanjutnya.
- c. Melakukan pencarian pola terhadap data , dalam tahap ini akan dilakukan pencarian pola dengan metode *Hierarchical Agglomerative Clustering*.

4. Perancangan Sistem

Membuat pemodelan sistem yang berorientasi objek agar dapat diimplementasikan menggunakan metode *Hierarchical Agglomerative Clustering*

5. Implementasi Program(*Coding*)

Pada tahap ini dilakukan pengkodean program untuk mengimplementasikan perancangan sistem dengan menggunakan bahasa pemrograman *Java*.

6. Pengujian

Melakukan analisis terhadap hasil dari program dengan inputan *raw log* yang telah di *preprocessing* sebelumnya sehingga menghasilkan output berupa pola kesamaan akses user.

7. Pembuatan Laporan

Pembuatan laporan skripsi bertujuan untuk dijadikan sebagai dokumentasi hasil penelitian.