ISSN : 2355-9365 e-Proceeding of Engineering : Vol.2, No.1 April 2015 | Page 1762

KOMBINASI ALGORITMA AGGLOMERATIVE CLUSTERING DAN K-MEANS UNTUK SEGMENTASI PENGUNJUNG WEBSITE

Yudha Agung Wirawan, Dra.Indwiarti ,M.Si, Yuliant Sibaroni,S.SI., M,T
Program Studi Ilmu Komputasi Fakultas Informatika Universitas Telkom
Jl. Telekomunikasi Terusan Buah Batu Bandung 40257

Abstrak

Clustering merupakan salah satu bagian penting dalam penggunaan web mining untuk segmentasi pengunjung web. Dalam tulisan ini, kami melakukan pengelompokan pengunjung web menggunakan kombinasi metode *clustering* hirarki dan non-hirarki terhadap data log web*site* akademik. Metode pengelompokan hirarki dan non-hirarki yang digunakan dalam tugas akhir ini yaitu *Agglomerative Clustering* dan *K-Means*. *Agglomerative Clustering* digunakan untuk menentukan jumlah *cluster, K-Means* digunakan untuk membentuk segmentasi. Dari pengujian yang dilakukan pada data log website *akademik*, beberapa kelompok pengunjung web dihasilkan. Terdapat beberapa hari dimana banyak menu yang diakses oleh *user*. Pada minggu 1, lebih cenderung pada menu yang diakses adalah tentang menu Registrasi, baik itu dalam tagihan pembayaran, keterlambatan registrasi dan proses registrasi. Pada minggu 2, *user* cenderung dominan pada menu silabus dan akademik mahasiswa baik kehadiran maupun jadwal mahasiswa. Sementara pada minggu 3 *user* tidak cenderung pada beberapa menu saja melainkan banyak menu yang di kunjungi, namun pada minggu 3 ini hal yang paling diperhatikan adalah pada menu tentang Tugas Akhir/Proyek Akhir. Pada minggu 4 yang paling diperhatikan adalah pada menu tentang akademik mahasiswa baik kehadiran, presensi maupun jadwal.

Kata kunci : Clustering, Data Mining, Data log web, Segmentasi.

I. PENDAHULUAN

Perkembangan teknologi internet saat ini telah memacu pesatnya pertumbuhan dan pertukaran informasi yang mencakup semua aspek kehidupan. Seiring dengan perkembangan ini, aktivitas *user* semakin meningkat dalam mengakses *(World Wide Web)* atau website. Ini menandakan bahwa peran *(World Wide Web)* sangat penting. Untuk menjamin kepuasan *user* dalam mengakses *(World Wide Web)* dalam bentuk website, perlu diperhatikan performansi dan kualitasnya. Salah satu tolak ukurnya adalah kecenderungan *user* dalam mengakses website. Website merupakan bagian yang terpenting dalam era informasi saat ini. Hampir 80% [9] layanan di Internet tersedia dalam bentuk *(World Wide Web)* atau website sebagai media dalam menyebarkan informasi dalam teks, gambar, video, atau suara dan multimedia. Tingginya jumlah pengunjung sebuah website, mengakibatkan website tersebut mempunyai data log yang sangat besar dan data tersebut perlu diolah lebih lanjut (*Web Usage Mining*) guna mendapat pola pengunjung website untuk berbagai keperluan. Data *log* yang besar tersebut mengandung hal-hal yang tidak diperlukan (irrelevant data) dalam proses mining sehingga perlu upaya untuk memperbaiki kualitasnya.

Web Usage Mining adalah salah satu kategori di bidang pertambangan web, yang merupakan penambangan yang dilakukan diweb berdasarkan data log web. Secara khusus, *Web Usage Mining* adalah penerapan teknik data mining untuk menemukan interaksi antara pengunjung website melalui data log web. Salah satu teknik yang dikenal dalam data mining yaitu teknik *clustering* [3]. Pengertian *clustering* keilmuan dalam data mining adalah pengelompokan sejumlah data atau obyek ke dalam *cluster* (group) sehingga setiap dalam *cluster* tersebut akan berisi data yang semirip mungkin dan berbeda dengan obyek dalam *cluster* yang lainnya. Sampai saat ini, para ilmuwan masih terus melakukan berbagai usaha untuk melakukan perbaikan model *cluster* dan menghitung jumlah *cluster* yang optimal sehingga dapat dihasilkan *cluster* yang paling baik. Ada dua metode *clustering* yang kita kenal, yaitu *hierarchical clustering* dan *partitioning*. Dalam membentuk segmentasi *clustering* ada beberapa metode yang dapat digunakan, tetapi pada umumnya metode yang sering digunakan yaitu: metode *Hierarchical Agglomerative Clustering* yang merupakan salah satu bagian dari metode hirarki. Sedangkan metode partitioning sendiri yang sering digunakan yaitu: *K-Means*. *K-Means* merupakan metode *clustering* yang paling sederhana dan umum [9]. Hal ini dikarenakan K-Means mempunyai kemampuan mengelompokkan data dalam jumlah yang cukup besar dengan waktu komputasi yang relatif cepat dan efisien [4]. Namun, *K-Means* mempunyai mempunyai kelemahan yang diakibatkan oleh penentuan pusat awal *cluster*. Hasil *cluster* yang terbentuk dari metode *K-Means* ini sangatlah tergantung pada inisiasi nilai pusat awal *cluster* yang diberikan [9]. Hal ini menyebabkan hasil *cluster*nya berupa solusi yang sifatnya *local optimal*. Untuk itu, maka *K-Means* dikolaborasikan oleh metode hirarki untuk penentuan pusat awal *cluster*. Metode hirarki yang akan dicoba diterapkan dalam tugas akhir ini adalah metode *Hierarchical Agglomerative Clustering* (*HAC*) yang diharapkan dapat memberikan hasil pengelompokan yang lebih baik. Dari proses pengelompokan ini nantinya diharapkan akan diketahui kemiripan atau kedekatan antar data sehingga dapat dikelompokkan ke dalam beberapa *cluster*, dimana antar anggota *cluster* memiliki tingkat kemiripan yang tinggi. Berdasarkan hal tersebut, maka penulis akan mencoba melakukan pengelompokan data log

ISSN : 2355-9365 e-Proceeding of Engineering : Vol.2, No.1 April 2015 | Page 1763

pengunjung web berdasarkan kombinasi metode hirarki dan metode non-hierarki yang akan diimplementasikan pada website *akademik*.

II. LANDASAN TEORI

II.1 *Web Usage Mining*

Definisi yang banyak diterima mengenai *web usage mining* adalah definisi yang dikemukakan dalam [2] yaitu " *the application of data mining techniques to large web datarepositories in order to extract usage patterns* ", penerapan teknik data mining untuk data repositori web yang besar untuk mengetahui pola akses *user*. Seperti diketahui bahwa web sangat berkaitan erat dengan sebuah *web server*, yaitu suatu *software server* yang memiliki tugas utama melayani dan memenuhi permintaan halaman *web* oleh *client* (pengguna). Selain itu, *web server* juga akan mencatat setiap aktivitas yang dilakukan oleh *client* (pengguna) tersebut ke dalam sebuah *file* yang sering disebut *web access log*. Hasil catatan aktivitas tersebut yang menjadi sumber data utama dalam *web usage mining*. Dari sebuah *web access log*, dapat diketahui beberapa informasi mengenai pola akses dan kelakuan (*behaviour*) pengguna dalam mengakses halaman website.

Menurut Srivastava, *web usage mining* merupakan teknik data mining yang berusaha mengungkap pola penggunaan dari halaman *web*, untuk memahami dan meningkatkan pelayanan kebutuhan dari aplikasi berbasis *web* [7]. Jadi *web usage mining* sedikit berbeda *web structure mining* dan *web content mining*. Pada jenis *structure* dan *content mining*, yang dianalisa atau digali adalah data didalam *web* itu sendiri, namun pada *web usage mining* yang dianalisa adalah pengguna atau pengunjung dari halaman web. Sehingga yang akan dianalisa adalah tingkah laku dari pengunjung (pengguna) dari *web* maka hasil dari *web usage mining* banyak digunakan dalam *e-marketing* dan *e-commerce*.Hasil dari analisa *web usage mining* antara lain informasi mengenai segmentasi pengunjung dari situs (aplikasi *web*). Segmentasi dapat dilihat berdasarkan *user* yang menjadi anggota *cluster* pada *cluster* yang dihasilkan.

II.2 *Hierarchical Agglomerative Clustering*

Hierarchical Agglomerative Clustering (HAC) adalah suatu metode *clustering* yang bersifat *bottom-up* yaitu menggabungkan n buah *cluster* menjadi satu *cluster* tunggal. Metode ini dimulai dengan meletakkan setiap obyek data sebagai sebuah *cluster* tersendiri dan selanjutnya menggabungkan *cluster-cluster* tersebut menjadi *cluster* yang lebih besar dan lebih besar lagi sampai akhirnya semua objek data menyatu dalam sebuah *cluster* tunggal. Secara logika semua obyek pada akhirnya hanya akan membentuk sebuah *cluster* [6]. Metode ini dimulai membentuk *cluster* pada setiap obyek. Kemudian dua obyek dengan memiliki jarak terdekat digabungkan. Selanjutnya obyek ketiga digabung dengan obyek lain memiliki jarak terdekat dan membentuk *cluster* baru. Proses akan berlanjut hingga akhirnya terbentuk satu cluster yang terdiri dari keseluruhan obyek.

Ada beberapa teknik dalam *HAC* [5], ada yang menggunakan *war's linkage, centroid linkage, single linkage, complete linkage, average linkage, median linkage,* dan lain-lainnya. Dalam Tugas Akhir ini teknik yang digunakan adalan single linkage*(nearest neighbor methods)*. Metode *Single linkage (nearest neighbor methods)* menggunakan prinsip jarak minimum yang diawali dengan mencari dua obyek terdekat dan keduanya membentuk *cluster* yang pertama. Pada langkah selanjutnya terdapat dua kemungkinan, yaitu : Obyek ketiga akan bergabung dengan *cluster* yang telah terbentuk, atau Dua obyek lainnya akan membentuk *cluster* baru. Proses ini akan berlanjut sampai akhirnya terbentuk jumlah *cluster* yang digunakan sebagai penentuan jumlah *cluster* awal pada *K-Means*. Pada metode ini jarak antar *cluster* didefinisikan sebagai jarak terdekat antar anggotanya.

II.3 *K-Means Clustering*

K-Means Clustering adalah Metode *clustering* berbasis jarak yang membagi data ke dalam sejumlah *cluster* dan algoritma ini hanya bekerja pada attribut numerik. Berikut ini adalah contoh dataset numerik.

No	Obyek	Menu 1	Menu 2
1	M1	27	3
2	M2	38	14
3	M3	24	8
4	M4	43	15
5	M5	28	5
6	M6	32	11
7	M7	30	7
8	M8	17	5
9	M9	20	6
10	M10	29	8

Langkah-langkah algoritma *K-Means* sebagai berikut,
Algoritma *K-Means* [7] :

1. Partisi *item* menjadi K initial *cluster*.
2. Lakukan proses perhitungan dari daftar item, tandai item untuk kelompok yang mana berdasarkan pusat (*mean*) yang terdekat (dengan menggunakan *distance* dapat digunakan *Euclidean distance*). Hitung kembali pusat centroid untuk item baru yang diterima pada *cluster* tersebut dari *cluster* yang kehilangan item.
3. Ulangi step 2 hingga tidak ada lagi tempat yang akan ditandai sebagai *cluster* baru atau tidak ada perubahan pada centroidnya. Seperti pada gambar 2.1 dibawah ini.

ISSN : 2355-9365 e-Proceeding of Engineering : Vol.2, No.1 April 2015 | Page 1764

No	Obyek	Menu1	Menu 2	Anggota Cluster		
				C1	C2	C3
1	M1	27	3		*	
2	M2	38	14			*
3	M3	24	8		*	
4	M4	43	15			*
5	M5	28	5		*	
6	M6	32	11			*
7	M7	30	7		*	
8	M8	17	5	*		
9	M9	20	6	*		
10	M10	29	8			*

No	Obyek	Menu1	Menu 2	Anggota Cluster		
				C1	C2	C3
1	M1	27	3		*	
2	M2	38	14			*
3	M3	24	8		*	
4	M4	43	15			*
5	M5	28	5		*	
6	M6	32	11			*
7	M7	30	7		*	
8	M8	17	5	*		
9	M9	20	6	*		
10	M10	29	8			*

Gambar 2.1 Centroid tidak berubah.

Gambar 2.1 menunjukkan posisi data sudah tidak mengalami perubahan, sehingga tidak ada perubahan pada centroidnya.

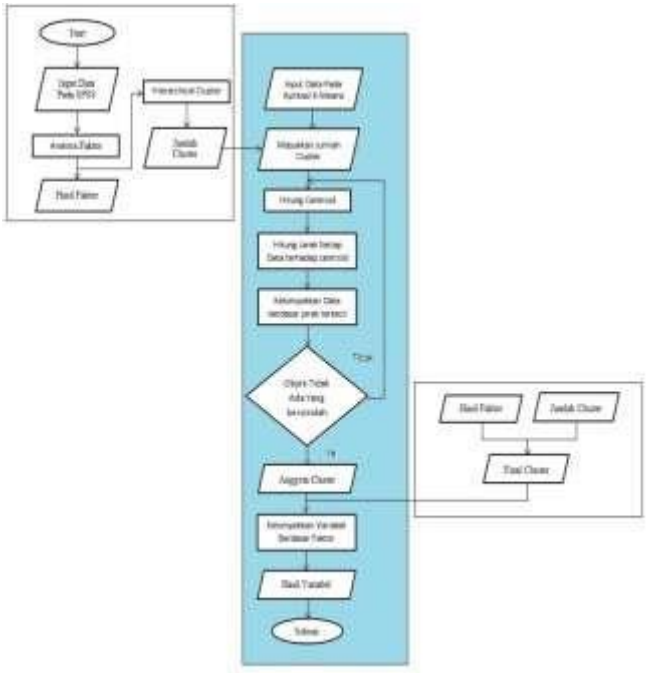
III. PERANCANGAN SISTEM

III.1 Deskripsi Sistem Secara Umum

Sistem yang dibangun pada Tugas Akhir ini adalah sistem untuk menentukan pola akses *user* pada halaman website akademik Universitas Telkom yang bernama *i-gracias* menggunakan metode *Hierarchical Agglomerative Clustering* dan *Non-Hierarchical K-Means Clustering*. Tahapan awal yang dilakukan adalah *preprocessing* dan *factor analysis* dari *data log server* yang di gunakan. Pada tahap *preprocessing* ini, dilakukan beberapa tahapan terhadap data *log* untuk memisahkan data yang tidak dibutuhkan serta menyesuaikan dengan kebutuhan sistem.
Setelah data yang diperoleh sesuai dengan kebutuhan, dilakukan proses *faktor analysis*, selanjutnya dengan menggunakan teknik *clustering* dengan algoritma *Hierarchical Agglomerative Clusterng* dan dilanjutkan dengan algoritma *Non – Hierarchical K-Means Clustering*. Keluaran dari proses ini akan menghasilkan sebuah segmentasi *user* yang menggambarkan kecenderungan akses *user* terhadap halaman web *akademik*.

III.2 Desain Sistem

Data yang digunakan adalah data yang sudah di *preprocessing* sebagai inputan awal pada sistem ini. Sistem ini dilakukan secara sekuensial dengan menghasilkan jumlah *cluster* terlebih dahulu menggunakan algoritma *Hierarchical Agglomerative Clustering*, kemudian jumlah *cluster* tersebut digunakan sebagai inputan pada penentuan jumlah awal *cluster* algoritma *K-Means* dan setelah itu dilakukan analisis dengan *final cluster* yang dihasilkan, sehingga menghasilkan suatu kumpulan menu. Jumlah *cluster* dengan algoritma *Hierarchical Agglomerative Clustering* dan *final cluster* dilakukan dengan *tool* SPSS versi windows 9.0. Penentuan jumlah awal *cluster* algoritma *K-Means* dilakukan menggunakan aplikasi *K-Means* dengan PHP. Desain sistem yang ditunjukkan pada gambar 3.1 di bawah ini.

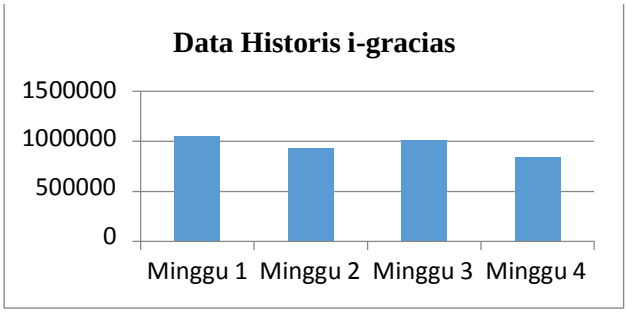


Gambar 3.1 Desain Sistem

Secara umum, proses yang akan dilakukan pada Tugas Akhir ini terdiri dari beberapa tahapan . Tahapan proses yang dilakukan pada penelitian ini yaitu : *Web logs data selection*, *Preprocessing*, *Factor analysis*, *Hierarchical Agglomerative Clustering*, *Non- Hirarchical K-Means Clustering*, Hasil dan analisis.

III.2.1 *Web logs data selection*

Data historis yang di gunakan untuk menjadi dataset dalam proses mining adalah data *access log*, *field* yang akan di-*clustering*kan adalah, *ip adress*, *page id*, *request*, *access time*, yang tercatat untuk menghasilkan data yang optimal. Dalam mendapatkan *record* yang sesuai karakteristik yang di inginkan maka akan di lakukan filtering terhadap data yang ada. Data ini dimulai dari tanggal 24 Agustus 2014 sampai dengan 20 September 2014. Data disajikan perhari dalam setiap minggunya. Berikut adalah data yang disajikan dalam bentuk histogram.



Gambar 3.2 Jumlah Data Historis *I-gracias*.

III.2.2 *Pre-Processing*

- ***Parsing Data.***

Proses ini bertujuan untuk mendapatkan bagian-bagian data yang diinginkan. Proses ini dilakukan dengan mengelompokkan baris-baris data menjadi beberapa bagian bagian data yang diinginkan.

- ***Cleaning Data.***

Setelah data selesai pada tahap parsing data, data yang terkelompok tersebut dibersihkan dari bagian-bagian yang tidak perlu seperti data berekstensi, .jpg, .gif, ukuran byte, dan status. Hasil dari cleaning data ini adalah informasi yang dibutuhkan untuk tugas akhir ini.

- ***Page User Identification***

Proses ini bertujuan untuk mengidentifikasi *user* yang melakukan akses terhadap website. Proses ini dilakukan setiap sistem menemukan baris data "*ip address*".

- ***Page Access Identification.***

Proses ini sama seperti proses sebelumnya. Disini yang diidentifikasi adalah *pageId* yang diakses *user*. Sistem mengidentifikasi *pages* tersebut jika menemukan "view.php?pageid=", "index.php?pageid=" , "category.php?pageid=" pada baris data. Item data dengan kode selain 200 dihapus [14] .

Hasil akhir dari tahap *pra*-pengolahan dalam bentuk vektor matriks pada tabel 4.1 dibawah ini.

Tabel 3.1 Matriks Vektor.

User	p1	p2	p3	p4	p5	p6	..	pn
u1	6	9	0	0	0	0		X_{1n}
u2	0	0	0	35	0	35		X_{2n}
u3	0	11	0	0	0	0		X_{3n}
u4	0	1	4	3	2	3		X_{4n}
u5	0	84	0	0	0	0		X_{5n}
u6	5	5	5	0	1	5		X_{6n}
u7	1	37	0	0	3	0		X_{7n}
u8	2	21	3	0	4	0		X_{8n}
u9	6	0	1	4	7	0		X_{9n}
...								
um	X_{m1}	X_{m2}	X_{m3}	X_{m4}	X_{m5}	X_{m6}		X_{nn}

Dengan p1, p2, p3, pn adalah variabel untuk halaman *web*, misalnya, p1 adalah halaman *web* dengan nama index.php. u1, u2, u3, um adalah variabel untuk pengunjung web, untuk contoh u1 adalah *web* pengunjung dengan alamat IP, 72.233.234.xxx. Dari Tabel 4.1, dapat menyimpulkan bahwa pengunjung dengan variabel u1 yang telah diakses berapa kali halaman *web* p16, halaman web n kali p29 dan seterusnya. Data ini dalam bentuk *vektor matriks* ini yang diproses lebih lanjut.

III.2.3 *Factor Analysis*

Analisis faktor digunakan untuk mengidentifikasi dimensi yang mendasari sekelompok variabel kemudian membangun struktur pengelompokkan baru yang lebih sederhana berdasarkan sifat dasar tersebut. Proses analisis faktor mencoba menemukan hubungan (*inter relationship*) antar sejumlah variabel-variabel yang saling independen satu dengan yang lain, sehingga bisa dibuat satu atau beberapa kumpulan variabel yang lebih sedikit dibandingkan dengan jumlah variabel awal tanpa kehilangan sebagian besar informasi penting yang terkandung didalamnya. Sebagai contoh, jika ada 16 variabel yang independen satu dengan yang lain, dengan analisis faktor mungkin bisa diringkas hanya menjadi 3 kumpulan variabel baru yang disebut *faktor*. Hasil dari analisis faktor adalah kumpulan variabel-variabel apa saja yang termasuk ke dalam faktor, dari faktor tersebut kita dapat mengetahui variabel-variabel pada setiap *cluster* yang dihasilkan yang dilakukan dengan algoritma *K-Means clustering*.

III.2.4 *Hierarchical Agglomerative Clustering*

Cluster hirarki yang digunakan adalah metode *agglomerative clustering*. Tahap pertama dari *cluster* hirarki adalah menghitung jarak antara obyek dengan metode jarak *manhattan distance* dan pembentukan *cluster* menggunakan metode *single linkage*. Berdasarkan *Agglomeration Schedule* dari metode ini, jumlah *cluster* berdasarkan aturan siku ditentukan, seperti yang ditunjukkan pada Gambar 3.3 dibawah ini.

Stage	*Combined Cluster*		*Coefficients*	*Stage Cluster First Appears*		*Next Stage*
	Cluster 1	*Cluster 2*		*Cluster 1*	*Cluster 2*	
158	3	24	61.64	157	0	159
159	3	10	100.27	158	0	160
160	3	6	115.04	159	0	161
161	3	4	116.71	160	0	162
162	3	78	133.67	161	0	163
163	1	3	147.53	162	0	164
164	1	2	175.52	163	0	0

Gambar 3.3 *Agglomeration Schedule.*

Gambar 3.3 menunjukkan perbedaan dalam koefisien di mana koefisien dalam tahap 158 memiliki selisih atau peningkatan terbesar pada *stage*. Dengan demikian, berdasarkan aturan siku dengan jumlah data sebagai 165, 165-158 = 7 (menghasilkan 7 kelompok). Hasil ini digunakan sebagai *input* untuk analisis *cluster* non-hirarki. Berikut adalah jumlah *cluster* yang dihasilkan dari *agglomeration schedule*.

III.2.5 *Non - Hierarchical Clustering*

Cluster non-hirarkis yang digunakan adalah metode K-Means. Metode tersebut digunakan untuk menentukan segmentasi pengunjung Web. Berikut ini contoh hasil segmentasi yang ditunjukkan pada tabel 3.2 dibawah ini.

Tabel 3.2 Jumlah Anggota *Cluster.*

Anggota / *User*	*Cluster*					
	1	2	3	4	5	6
Jumlah Anggota / *User*	2	1	143	13	3	3

III.2.6 Hasil dan Analisis

Pada tahap ini, kita dapat menganalisis dari penelitian yang telah dilakukan. Sehingga, kita mendapatkan suatu hasil dari segmentasi dari pengunjung website.

ISSN : 2355-9365 e-Proceeding of Engineering : Vol.2, No.1 April 2015 | Page 1766

IV. PENGUJIAN DAN ANALISIS

Berdasarkan hasil uji coba yang telah dilakukan pada tugas akhir ini, kombinasi *Agglomerative Clustering* dan *K-Means* menghasilkan jumlah *cluster* dan segmentasi pengunjung web yang ditunjukkan pada tabel-tabel berikut.

Tabel 4.1 Jumlah *Cluster Agglomerative Clustering*.

Tanggal	Jumlah Cluster
24-Agust-14	2 Cluster
25-Agust-14	2 Cluster
26-Agust-14	2 Cluster
27-Agust-14	2 Cluster
31-Agust-14	2 Cluster
01-Sep-14	2 Cluster
02-Sep-14	2 Cluster
03-Sep-14	2 Cluster
04-Sep-14	5 Cluster
05-Sep-14	2 Cluster
06-Sep-14	2 Cluster
08-Sep-14	3 Cluster
09-Sep-14	3 Cluster
10-Sep-14	2 Cluster
11-Sep-14	2 Cluster
12-Sep-14	3 Cluster
13-Sep-14	3 Cluster
15-Sep-14	4 Cluster
16-Sep-14	2 Cluster
17-Sep-14	3 Cluster
18-Sep-14	2 Cluster
19-Sep-14	2 Cluster
20-Sep-14	2 Cluster

Tabel 4.1 menunjukkan jumlah *cluster* yang dihasilkan dengan menggunakan *agglomerative clustering.* Jumlah *cluster* tersebut digunakan untuk penentuan jumlah awal *K-Means*. *K-Means* menghasilkan segmentasi dari pengunjung web. Segmentasi pengunjung web di tunjukkan pada tabel x.x di bawah ini.

Tabel 4.2 Segmentasi Pengunjung Web

Tanggal	Cluster					
	1	2	3	4	5	6
24-Agust-14	312	449	-	-	-	761
25-Agust-14	2429	125	-	-	-	3680
26-Agust-14	308	453	-	-	-	761
27-Agust-14	312	449	-	-	-	761
31-Agust-14	453	308	-	-	-	761
01-Sep-14	308	453	-	-	-	761
02-Sep-14	343	318	-	-	-	761
03-Sep-14	337	424	-	-	-	761
04-Sep-14	384	148	224	160	893	1809
05-Sep-14	425	336	-	-	-	761
08-Sep-14	357	655	119	-	-	1131
09-Sep-14	220	334	208	-	-	762
10-Sep-14	726	262	-	-	-	988
11-Sep-14	372	387	-	-	-	759
12-Sep-14	555	823	151	-	-	974
13-Sep-14	1912	1	4	-	-	1917
15-Sep-14	3	1282	568	1	-	1854
16-Sep-14	236	339	-	-	-	575
17-Sep-14	0	1403	440	-	-	1843
18-Sep-14	400	362	-	-	-	762
19-Sep-14	1222	563	-	-	-	1785
20-Sep-14	396	365	-	-	-	761

Tabel 4.2 menunjukkan jumlah segmentasi pengunjung web pada setiap *cluster*nya yang dihasilkan dengan *K-Means*

Analisis 1 menghasilkan bagaimana menu-menu yang diakses setiap harinya. Berdasarkan hasil pada lampiran 5, dalam rentang tanggal 24 Agustus 2014-20 September 2014 terdapat beberapa hari dimana memiliki faktor yang banyak, sehingga kita pada hari tersebut terdapat banyak menu-menu yang dikunjungi oleh *user*. Pada *cluster* yang dihasilkan dan memiliki jumlah anggota yang paling banyak pada 25 Agustus 2014, 9 September 2014, 13 September 2014, 17 September 2014, dan 20 September 2014. *User* pada *cluster* tersebut tidak dominan pada satu menu saja, melainkan lebih dari 5 menu yang diakses pada tanggal tersebut.

Analisis 2 menghasilkan bagaimana menu-menu yang diakses setiap harinya dan dianalisis per-minggu. Pada minggu 1, *user* lebih cenderung menu yang diakses adalah tentang menu Registrasi, baik itu dalam tagihan pebayaran, keterlambatan registrasi dan proses registrasi. Pada minggu 2, *user* cenderung dominan pada menu silabus dan akademik mahasiswa, baik kehadiran maupun jadwal mahasiswa. Sementara pada minggu 3 *user* tidak cenderung pada beberapa menu saja melainkan banyak menu yang diakses, namun pada minggu 3 hal yang paling diperhatikan adalah pada menu tentang Tugas Akhir atau Proyek Akhir. Pada minggu 4 yang paling diperhatikan adalah pada menu tentang akademik mahasiswa baik keharidan, presensi maupun jadwal mahasiswa, dimana menu tersebut yang paling dominan diakses.

V. KESIMPULAN DAN SARAN

V.1 Kesimpulan

Berdasarkan analisis hasil pengujian yang dilakukan, maka pada penelitian Tugas Akhir ini didapatkan kesimpulan sebagai berikut :

1. Pengimplementasian metode *clustering* dalam segmentasi pada data log pengunjung website akademik sebagai berikut :
    a) Metode *clustering* yang diimplementasikan adalah Metode Hirarki dan Metode Non-Hirarki.
    b) Algoritma *cluster* yang di implementasikan padametode Hirarki yaitu : *Hierarchical Agglomerative Clustering (HAC)* dan Metode Non- Hirarki yaitu : *K-Means.*
2. Hasil segmentasi pengelompokan *user* untuk informasi evaluasi website sebagai berikut :

ISSN : 2355-9365 e-Proceeding of Engineering : Vol.2, No.1 April 2015 | Page 1767

a) Jumlah *cluster* yang dihasilkan adalah 2-5 *cluster*. Segmentasi pengelompokan *user* yang dihasilkan, banyak menu-menu yang diakses pada 25 Agustus 2014, 9 September 2014, 13 September 2014, 17 September 2014, dan 20 September 2014 *user*. Pada *cluster* tersebut tidak dominan pada satu menu saja, melainkan lebih dari 5 menu yang diakses pada tanggal tersebut.

b) Pada minggu 1, user lebih cenderung menu yang diakses adalah tentang menu Registrasi, baik itu dalam tagihan pembayaran, keterlambatan registrasi dan proses registrasi. Pada minggu 2, user cenderung dominan pada menu silabus dan akademik mahasiswa, baik kehadiran maupun jadwal mahasiswa. Sementara pada minggu 3 *user* tidak cenderung pada beberapa menu saja, melainkan banyak menu-menu yang diakses, namun pada minggu 3 hal yang paling diperhatikan adalah pada menu tentang Tugas Akhir atau Proyek Akhir. Pada minggu 4 yang paling diperhatikan adalah pada menu tentang akademik mahasiswa baik kehadiran, presensi maupun jadwal, dimana menu tersebut yang paling dominan diakses oleh *user*.

V.2 *Saran*

Pengembangan yang dapat dilakukan pada tugas akhir ini antara lain :

1. Perlu di lakukan proses *preprocessing* data yang lebih simple, karena pada penelitian ini menggunakan *Microsoft excel*, sehingga *preprocessing* dilakukan secara manual pada *Microsoft excel* dan membutuhkan waktu yang cukup lama.
2. Komputer dengan kinerja tinggi diperlukan untuk mendapatkan hasil *cluster* dengan data yang sangat besar.

DAFTAR PUSTAKA

[1] B. Santosa, Data Mining. Teknik Pemanfaatan Data untuk Keperluan Bisnis, First Edition ed. Yogyakarta: Graha Ilmu, (2007).

[2] Cooley R. [et al.]. WebSIFT: The Web Site Information Filter System [Conference] // Department of Computer Science, University of Minnesota. - 1999.

[3] Han Jiawei and Kamber Micheline Data Mining: Conceptsand Techniques [Book]. - [s.l.] : Morgan Kaufmann Publisher, 2006.

[4] K. Arai and A. R. Barakbah, "Hierarchical K-means: an algorithm for centroids initialization for K- means," (2007).

[5] Laboratorium Data Mining Jurusan Teknik Industri Fakultas Teknologi Industri Universitas Islam Indonesia, Modul II Clustering. [Online]. Tersedia: http://www.ss354.com/wpcontent/uploads/2014/03/Data-Mining-Modul-Clustering-Modul-Clustering.pdf [03 Mei 2014].

[6] Santoso, S. 2010. Statistik Multivariat. Jakarta: Elex Media Komputindo.

[7] Satriyanto, Edi (2010). Clustering. [Online]. Tersedia : http://student.eepisits.edu/~spydeeyk/download/sem4/statistik/clustering.doc, [14 Mei 2014].

[8] Srivastava J. [et al.] Web Usage Mining: Discovery and Applications of Usage Patterns from Web Data [Conference]. - Minneapolis : Department of Computer Science and Engineering, Universityof Minnesota, 2000.

[9] Yuhefizar, 2008. “10 Jam Menguasai Internet dan Aplikasinya”. Jakarta. PT. Elexmedia Komputindo, ISBN : 978-979-27-3470-6.