ISSN: 2339-2541
JURNAL GAUSSIAN, Volume 9, Nomor 4, Tahun 2020, Halaman 486 - 494
Online di: https://ejournal3.undip.ac.id/index.php/gaussian/

ANALISIS *WEB USAGE MINING* MENGGUNAKAN METODE *MODIFIED GUSTAFSON – KESSEL CLUSTERING* DAN *ASSOCIATION RULE* PADA *WEBSITE* UNIVERSITAS DIPONEGORO

Galuh Nurvinda Kurniawati[1], Rukun Santoso[2], Sugito[3]
[1,2,3]Departemen Statistika, Fakultas Sains dan Matematika, Universitas Diponegoro
galuhnurvinda@gmail.com

ABSTRACT

The comprehension of web visitors patterns are needed to develop website in an optimal fashion. The visitor pattern contained in the web log file of Diponegoro University's website is clustered by Modified Gustafson-Kessel method. In general, this method produces two until six clusters. Two kinds of results are outlined in this paper. The first is the result contains two clusters, and the second is containing three clusters. In the first result, the visitors are divided into information seekers of student capacity and Engineering Faculty. In the second result, the visitors are divided into information seekers of Medicine Faculty, student admission and Engineering Faculty.
Keywords: website, web usage mining, web log file, Modified Gustafson-Kessel Clustering, Association Rule

1. PENDAHULUAN

Di era modern saat ini, perkembangan teknologi sangat pesat, khususnya teknologi yang berhubungan dengan internet yang semakin mudah diakses.Perkembangan internet berbanding lurus dengan pemanfaatan *website* dalam berbagai bidang.Seiring dengan meningkatnya pemanfaatan *website*, aktivitas *user* dalam penggunaan *website* pun semakin meningkat.Semakin banyak aktivitas *user* dalam laman *website* menghasilkan data *server log-file* yang besar mengenai riwayat interaksi *user* yang tersimpan di dalam *web log server.Web mining* adalah teknik untuk mengolah data *server log-file* dari *website*.*Web mining* dibagi menjadi tiga jenis yaitu, *web structure mining*, *web content mining*, dan *web usage mining*.*Web usage mining* adalah teknik untuk menemukan pola atau *pattern* dari user dalam mengakses web[4].

C*lustering* adalah pengelompokan data berdasarkan karakteristik umum[5]. *Gustafson-Kessel Clustering* merupakan pengembangan dari *Fuzzy C-Means* dengan mengasosiasikan setiap *cluster* dengan pusat *cluster* dan matriks kovariannya untuk memperoleh jumlah cluster optimum menggunakan nilai *partition coefficient* dan *coefficient entropy*. Hasil akhir penelitian ini adalah pengelompokan *user* yang memiliki kemiripan pola navigasi untuk memahami tingkah laku *user* dalam mengakses *website* sehingga hasil dari penelitian ini dapat digunakan sebagai acuan dalam perbaikan kualitas *website* dan mengoptimalkan fitur-fitur yang tersedia di *website*.

2. TINJAUANPUSTAKA

2.1. *Web Mining* dan *Web Usage Mining*

Pengambilan data *website* menggunakan teknik *data mining* disebut dengan *web mining.Web mining* mempermudah pencarian informasi termasuk penemuan dan analisis data, dokumen serta multimedia karena berhubungan dengan *World Wide Web*. *Web mining* dibedakan menjadi tiga jenis yaitu, *Web Structure Mining, Web Content Mining, dan Web Usage Mining*[6].

Data pola penggunaan *(usage pattern)* dan aktivitas *user website* dapat diproses menggunakan metode *web usage mining*. Data tersebut merupakan *server log-file.Server*

log- file mengumpulkan berbagai data tentang permintaan (*request)* informasi ke server *web*. Berikut ini merupakan contoh*server log-file*

37.9.113.49 - - [01/Nov/2019:03:25:04 +0700] "GET /language/id/penerimaan-mahasiswa-baru/daya-tampung?lang=id HTTP/1.1" 200 88736 "-" "Mozilla/5.0 (compatible; YandexBot/3.0; +http://yandex.com/bots)" "-"

2.2. Preprocessing

Tujuan *preprocessing* adalah untuk mendapatkan data pengguna mentah dari serangkaian *request* yang dicatat dalam bentuk *server log-file*[2].Secara singkat *preprocessing* dilakukan dengan maksud untuk:

a. *Cleaning data*, yaitu memfilter data yang tidak di-*request* oleh *user*. Data ini terjadi karena adanya *automatic request*.
b. Menyingkirkan aktivitas yang dilakukan oleh bot yang dianggap tidak memiliki informasi yang berguna berkaitan dengan *web usage mining*.
c. Identifikasi *user* menggunakan alamat IP sebagai identitas unik dari tiap *user*.
d. Indentifikasi *usersession*, yaitu menentukan halaman-halaman yang di-*request* untuk setiap kunjungan *user*.
e. Indentifikasi kategori, yaitu mengelompokkan *request* sub halaman yang bisa digabung.

2.3.*Translation*

Pada tahap ini data*page request*berbahasa Inggris akan diubah menjadi bahasa Indonesia agar *session* yang bermakna sama tidak terhitung dua kali.

2.4.*Text Representation*

Text representation bertujuan untuk secara *numeric* mewakili teks yang tidak terstruktur agar dapat dihitung secara matematis[7]. Setelah *server log-file* melewati proses-proses sebelumnya, maka data telah siap direpresentasikan menjadi bentuk *document term matrix*.Matriks berisi frekuensi *page request* yang muncul (TF) dan pembobotan menggunakan pembobotan TF-IDF.Algoritma pembobotan TF-IDF terdiri dari dua *term* yaitu:

a. *Term Frequency* (TF), yaitu jumlah *page request* yang ingin diboboti dibagi dengan jumlah total *page request* yang diakses oleh *satu user*. Semakin besar jumlah kemunculan suatu *page request* maka semakin besar pula bobot yang diberikan.
b. *nvers Document Frequency* (IDF) merupakan hasil log dari jumlah total *user* yang me-*request page request* dibagi dengan jumlah seluruh *user*

Rumus umum pembobotan TF-IDF sebagai berikut

$$W_{j,i} = \frac{n_{j,i}}{\sum_{j=1}^{p} n_{j,i}} \cdot \log_2 \frac{D}{d_i} \quad (1)$$

dengan:

$W_{j,i}$: pembobotan TF-IDF untuk *page request* ke-j pada dokumen ke-i.
$n_{j,i}$: jumlah kemunculan *page request* ke-j pada dokumen ke-i.
$\sum_{j=1}^{p} n_{j,i}$: jumlah kemunculan seluruh *page request* pada dokumen ke-i.
D : banyaknya dokumen yang dibangkitkan
d_i : banyaknya dokumen yang mengandung *page request* ke-i.
p : banyaknya *page request* yang terbentuk

2.5. *Modified Gustafson – Kessel Clustering*

Algoritma Gustafson-Kessel mengubah fungsi perhitungan jarak menjadi fungsi jarak adaptif (*adaptive distance norm*) yang selalu diperbaharui pada setiap iterasi dengan menggunakan matriks *fuzzy covariance*[3]. Algortima Gustafson-Kessel menggunakan fungsi jarak mahalanobis sehingga lebih dapat menyesuaikan bentuk geometris untuk sebuah himpunan data, tidak seperti *Fuzzy C-Means* yang mengasumsikan bahwa bentuk gemoetris suatu *cluster* adalah bulat sempurna.Meskipun Gustafson-Kessel lebih unggul dari algoritma Fuzzy C-Means, masih terdapat masalah saat matriks *fuzzy covariance* dari data merupakan matriks singular maka perhitungan matriks $\boldsymbol{A_k}$ tidak dapat diterapkan.

Algoritma *modified* Gustafson-Kessel *Clustering* secara lengkap adalah sebagai berikut[1]: input data yang akan dikelompokkan sebagai X (matriks a x b), menentukan jumlah *cluster* yang akan dibentuk (c $\geq$ 2), *weighting exponent* (m $>$ 1), maksimum iterasi (t_{max}), *error* terkecil yang diharapkan (ε), nilai *threshold* (β), dan parameter pembobot $\gamma \in$ [0,1]. Membangkitkan bilangan random u_{ik} ,$1 \leq i \leq$ n; $1 \leq k \leq$ c sebagai elemen – elemen matriks partisi awal $\boldsymbol{U_0}$dan hitung matriks kovarian $\boldsymbol{F_0}$dari keseluruhan data.

$$\sum_{k=1}^{c} u_{ik}^{m} = 1,\ 1 \leq i \leq \text{n};\ 1 \leq k \leq \text{c} \qquad (2)$$

Lalu dilanjutkan untuk $t = 1, 2, \ldots, t_{max}$

Step 1.menghitung pusat *cluster* ke-k ($\boldsymbol{v_k}$) dengan rumus:

$$v_k^{(t)} = \frac{\sum_{i=1}^{n} (u_{ik}^{(t-1)})^m \boldsymbol{x_i}}{\sum_{i=1}^{n} (u_{ik}^{(t-1)})^m},\ 1 \leq k \leq \text{c} \qquad (3)$$

dengan:

u_{ik} = derajat keanggotaan dari data ke-i pada *c luster* ke-k
m = pangkat pembobot untuk fungsi keanggotaan *fuzzy*
t = banyaknya iterasi
n = banyaknya data
c = banyaknya *cluster*

Step 2. Menghitung matriks kovarian *cluster* dengan rumus:

$$\boldsymbol{F_k} = \frac{\sum_{i=1}^{n} (u_{ik}^{(t-1)})^m (x_i - v_k^{(t)})\ (x_i - v_k^{(t)})^T}{\sum_{i=1}^{n} (u_{ik}^{(t-1)})^m} \qquad (4)$$

$$1 \leq k \leq c$$

dengan:

$\boldsymbol{x_i}$ = vektor data ke-i
$\boldsymbol{v_k}$ = pusat *cluster* ke-k
u_{ik} = derajat keanggotaan data ke-i pada *cluster* ke-k
m = pangkat pembobot untuk fungsi keanggotaan *fuzzy*
t = banyaknya iterasi
n = banyaknya data
c = banyaknya *cluster*

Ekstraksi nilai *eigenvactors* (ϕ) dan *eigenvalues* (λ) dan menghitung nilai

$$\boldsymbol{F_k} = (1 - \gamma)\boldsymbol{F_k} + \gamma \det(\boldsymbol{F_0})^{1/p}\mathbf{I},\ k = 1,2,\ldots..,c \qquad (5)$$

dengan:

γ = parameter untuk mengatur bentuk matriks *fuzzycovariance*, $\gamma \in [0,1]$
$\boldsymbol{F_0}$ = matriks kovarian dari seluruh data
$\boldsymbol{F_k}$ = matriks *fuzzy covariance cluster* ke-k (pada persamaan 4)
p = banyaknya variabel
I = matriks indentitas

Jika rasio antara nilai eigen maksimal dan minimal melewati nilai threshold yang ditentukan, maka rekonstruksi $\boldsymbol{F_k}$ dengan penjabaran sebagai berikut:

$$\boldsymbol{F_k} = \boldsymbol{\phi \Lambda \phi^{-1}}, k = 1, 2, \ldots, c \quad (6)$$

dengan:

$\boldsymbol{\phi}$ = vektor eigen dari matriks *fuzzy covariance cluster* ke-k

$\boldsymbol{\Lambda}$ = matriks diagonal dari nilai-nilai eigen (*eigen values*) matriks *fuzzy covariance cluster* ke-k

Step 3. Menghitung jarak dengan persamaan (3) dari *norm inducing matrix* di persamaan (5) dengan $i = 1, 2, \ldots, n$ dan $k = 1, 2, \ldots, c$. sehingga didapat:

$$D^2_{ikA\text{k}} = (x_i - v_k^{(t)})^T \left[\rho_k \det(\boldsymbol{F_k})^{\frac{1}{p}} \boldsymbol{F_k^{-1}}\right](x_i - v_k^{(t)}) \quad (7)$$

dengan:

$D^2_{ikA\text{k}}$: jarak data ke-i terhadap pusat *cluster* dengan *normincluding matrix* $\boldsymbol{A}_\text{k}$

$\boldsymbol{x_i}$: vektor data ke-i

$\boldsymbol{v_k}$: pusat *cluster* ke-k

$\boldsymbol{F_k}$: matriks *fuzzy covarian cluster* ke-k

ρ_k : volume *cluster* ke-k

p : banyaknya variabel

n : banyaknya data

c : banyaknya *cluster*

Step 4.Memperbarui matriks fungsi keangotaan

Untuk $1 \leq i \leq n$

Jika $D^2_{ikA\text{k}} > 0$ untuk $1 \leq k \leq c$

$$u_{ik}^{(t)} = \left[\sum_{l=1}^{c} \left(\frac{D_{ikA_k}}{D_{ilA_k}}\right)^{\frac{2}{(m-1)}}\right]^{-1} \quad (8)$$

Jika tidak, maka:

$$u_{ik}^{(t)} = 0 \text{jika} D^2_{ikA\text{k}} > 0 \text{ dan } u_{ik}^{(t)} \in [0,1] \quad (9)$$

$$\text{dengan} \sum_{k=1}^{c} u_{ik}^{(t)} = 1$$

Iterasi dihentikan jika$\|\boldsymbol{U}^{(t)} - \boldsymbol{U}^{(t-1)}\| < \varepsilon$ atau jika $t >$ iterasi maksimum.

Nilai *threshold* (β) yang digunakan biasanya ditentukan dalam angka yang besar, seperti 10^{15}.

2.6. *Association Rule*

Association rule mining adalah teknik *data mining* untuk menemukan aturan asosiatif antara suatu kombinasi item. Ada dua parameter untuk mengetahui penting atau tidaknya suatu aturan asosiasi yaitu *support* (nilai penunjang) dan *confidence* (nilai kepastian). Analisis asosiasi terbagi menjadi dua tahap:

a. Analisis pola frekuensi tinggi

Nilai *support* sebuah *item* diperoleh dengan rumus sebagai berikut:

$$Support\,(\text{A}) = \frac{\sum transaksi\ yang\ mengandung\ item\ A}{total\ transaksi} \quad (12)$$

Sedangkan nilai *support* dari 2 *item* adalah sebagai berikut:

$$Support\,(\text{A} \cap B) = \frac{\sum transaksi\ yang\ mengandung\ item\ A\ dan\ B}{total\ transaksi} \quad (13)$$

b. Pembentukan aturan asosiatif

Nilai *confidence* dari aturan A → B diperoleh dari rumus berikut:

$$\text{P(B|A)} = \frac{\sum transaksi\ yang\ mengandung\ item\ A\ dan\ B}{total\ transaksi\ A} \quad (14)$$

Pada bahasa pemrograman R, terdapat variable *lift*, yaitu pengukuran independensi dari A dan B. Nilai *lift* antara 0 hingga tak hingga.Sehingga *rule* yang dapat digunakan adalah yang hanya memiliki nilai *lift* > 1.

$$Lift = \frac{P(B|A)}{P(A)P(B)} \quad (15)$$

Nilai *lift* = 1, artinya A dan B adalah independen
Nilai *lift* < 1, artinya B tidak cenderung terjadi terhadap A
Nilai *lift*> 1, artinya B lebih cenderung terjadi terhadap A

3. METODE PENELITIAN

Data yang digunakan dalam penelitian ini merupakan data primer yang didapat dari data *web log serverwebsite* Universitas Diponegoro (*www.undip.ac.id*) yang hanya bisa diakses oleh admin pada tanggal 1 November 2019 hingga 7 November 2019. Analisis pengelompokkan menggunakan metode Gustafson-Kessel *Clustering* menggunakan *software* Matlab 2016a dan dilanjutkan dengan *Association rule* untuk mengetahui pola *user* menggunakan *software* Rstudio. Proses *text representation* untuk membentuk matriks TF-IDF dan *tokenizing* menggunakan *software* Rstudio.

4. HASIL DAN PEMBAHASAN

Web log yang dianalisis adalah *records* dari tanggal 1 November 2019 hingga 7 November 2019. Data tersebut terdiri dari 6500 *page request* dengan jumlah *IP Address* sebanyak 2496.

4.1. *Preprocessing*

Preprocessing merupakan proses untuk mempersiapkan data agar data bisa diolah lebih lanjut. Langkah – langkah *preprocessing* adalah *data cleansing*, menyingkirkan aktivitas bot, Mengidentifikasi *user* menggunakan alamat IP sebagai identitas unik, mengidentifikasi *user session*, dan mengidentifikasi kategori. Contoh *preprocessing* terlihat pada Tabel 1.

Tabel 1. Contoh *Preprocessing*

No.	Alamat IP	Kode Alamat IP	*Page Request*
1.	125.163.220.242	10	fakultas/fakultas-teknik/ fakultas/
2.	114.125.52.228	24	beasiswa-bidikmisi/ beasiswa-bidikmisi/

4.2. *Translation*

Pada proses *translationpage request* berbahasa Inggris diterjemahkan mejadi bahasa Indonesia agar *session* yang bermakna sama tidak terhitung dua kali. Pada Tabel 2 terlihat kata '*leaders*' berubah menjadi 'pimpinan' dan '*academic-regulations*' berubah menjadi 'peraturan-akademik'.

Tabel 2. Contoh *Translation*

No.	kode	Sebelum *Translation*	Setelah *Translation*
1.	28	profil/leaders/	profil/pimpinan/
2.	60	academic-regulations/	peraturan-akademik/

4.3. *Text Representation*

Proses ini bertujuan untuk mengetahui *page request*yang di-*request* oleh seorang *user* secara keseluruhan dari tanggal 1 November 2019 hingga 7 November 2019 dan dibentuk menjadi *Documen Term Matrix*menggunakan pembobotan TF-IDF dengan fungsi '`as.matrix(weightTfIdf(m=DocumentTermMatrix(data),normalize=TRUE))`'.

Berdasarkan hasil *text representation*, dari total 6500 *page request* ternyata hanya disusun oleh 94 *page request.*Artinya, satu *page request* dapat di-*request* lebih dari satu kalioleh *user*.Contoh matriks TF dan matriks IDFterlihat pada Tabel 3dan Tabel 4.

Tabel 3. Contoh Matriks TF

Kode *user*	*Gabungan Page request*	Page request		
		fakultas/fakultas-ilmu-budaya/	fakultas/	fakultas/fakultas-psikologi/
13	fakultas/fakultas-ilmu-budaya/ fakultas/	1	2	0
527	fakultas/fakultas-psikologi/	0	0	1

Tabel 4. Contoh Matriks TF-IDF

Kode *user*	*Gabungan Page request*	Page request		
		fakultas/fakultas-ilmu-budaya/	fakultas/	fakultas/fakultas-psikologi/
13	fakultas/fakultas-ilmu-budaya/ fakultas/	1,6485	0,9747	0
527	fakultas/fakultas-psikologi/	0	0	4,9279

4.4. Pengaplikasian Algortima Gustafson-Kessel *Clustering*

Clustering dimulai dari jumlah *cluster* sebanyak 2 sampai dengan 6 menggunakan fungsi '`result=GKC(X, U0, m, e, beta, gamma)`' pada *command window* Matlab 2016a. Data yang digunakan adalah *Document Term Matrix* dengan pembobotan TF-IDF.Pangkat *fuzzyfier* yang digunakan adalah m = 3,85 yang merupakan hasil yang paling optimal dari proses *trial* dan *error* karena menghasilkan fungsi keanggotaan $u_{ik} \neq 1/c$. untuk batas *error* terkecil dan nilai *threshold* yang digunakan diambil dari jurnal Babuska (2002) sebesar $\varepsilon = 0{,}001$ dan $\beta = 10^{15}$. Sedangkan nilai parameter pembobot yang digunakan sebesar $\gamma = 0{,}8$.

Hasil *clustering* yang didapat yaitu untuk jumlah cluster sebanyak 4, 5, dan 6 cluster, hanya 3 cluster dari total seluruh cluster yang memiliki anggota kelompok. Jadi, dapat disimpulkan bahwa alamat IP yang mengakses *website* Universitas Diponegoro pada tanggal 1 November 2019 hingga 7 November 2019 hanya mampu dikelompokkan dengan jumlah *cluster* maksimal sebanyak 3 *cluster*. Dalam penelitian ini akan dilakukan analisis terhadap pengelompokkan dengan jumlah *cluster* 2 dan 3 saja. Hasil clustering dengan jumlah 3 *cluster* dapat dilihat pada Tabel 5 dan 6.

Tabel 5. Hasil *Clustering* dengan *Algoritma Gustafson-Kessel* untuk jumlah *cluster* 2

Cluster ke-	Kode alamat IP	Jumlah anggota
1	2, 3, 6, 7, 9, 10, 13, 14, 15, 16, 17, 20, 21, 22, 23, 25, …, 2496	1142
2	1, 4, 5, 8, 11, 12, 18, 19, 24, 29, 30, 31, 32, 33, 37, 38, …, 2493	1354

Tabel 6. Hasil *Clustering* dengan *Algoritma Gustafson-Kessel* untuk jumlah *cluster* 3

Cluster ke-	Kode alamat IP	Jumlah anggota
1	15, 45, 65, 67, 132, 152, 155, 250, 275, 328, 347, 352, 356, 396, 467, 470, 504, 506, 507, 537, 560, 587, 646,…,2488	96
2	344, 753, 924, 1327, 1777, 1796, 1926, 1935	8
3	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34,…, 2496	2392

4.5. *Tokenizing*

Tokenizing merupakan proses pemotongan kalimat menjadi kata-kata penyusunnya dengan fungsi yang digunakan pada Rstudio '`unnest_tokens(word,text)`'. *Tokenizing* yang dirancang pada penelitian ini adalah memecah *page request* dengan *white space* atau spasi sebagai pembagi.Pada Tabel 7 terlihat contoh *tokenizing*.

Tabel 7. Contoh *Tokenizing*

Kode alamat IP	Sebelum *Tokenizing*	Setelah *Tokenizing*
15	senat_akademik fakultas fakultas fakultas_ilmu_budaya	senat_akademik fakultas fakultas fakultas_ilmu_budaya

4.6. *Association Rule*

Setelah melakukan clustering, dilakukan asosiasi di setiap cluster yang terbentuk.Proses pembentukan *association rule* menggunakan fungsi '`apriori(trans,parameter = list(supp=0.04,conf=0.04))`' pada Rstudio. *Support* minimum yang digunakan adalah 4% dan *confidence* minimum yang digunakan adalah 4%. Nilai ini didapat dari proses *trial* dan *error* untuk mendapatkan asosiasi terbaik di setiap *cluster*. Tabel 8 dan 9 menunjukkan hasil *association rule* untuk jumlah *cluster* 3 *cluster*.

Tabel 8. Hasil *Association Rule* 2 *Cluster*

cluster	LHS	RHS	*Support*	*Confidence*	Lift
1	{ fakultas }	{ fakultas_ilmu_sosial_dan_ilmu_politik}	6%	13%	2,3
	{fakultas}	{ fakultas_kesehatan_masyarakat }	5%	100%	2,3
	{fakultas}	{ fakultas_hukum }	5%	100%	2,3
	{fakultas}	{ fakultas_peternakan_dan_pertanian }	6%	100%	2,3
	{fakultas}	{ fakultas_teknik }	8%	17%	2,3
	{fakultas}	{ fakultas_psikologi }	5%	13%	2,3
	{fakultas}	{fakultas_perikanan_dan_ilmu_kelautan}	5%	14%	2,3
	{fakultas}	{fakultas_ilmu_budaya}	6%	10%	2,3
	{fakultas}	{sekolah_vokasi}	5%	10%	2,3
2	{penerimaan_mahasiswa_baru}	{daya_tampung}	5%	100%	2,3
	{fakultas}	{fakultas_kedokteran}	6%	13%	2,3

Tabel 9. Hasil *Association Rule* 3 *Cluster*

cluster	LHS	RHS	*Support*	*Confindence*	Lift
1	{Akademik}	{Pindah_studi}	10,4%	100%	7,4
	{fakultas}	{fakultas_ekonomi_dan_bisnis}	13,5%	24,07%	1,8
	{fakultas}	{fakultas_kedokteran}	26,04%	46,30%	1,8
	{fakultas}	{sekolah_vokasi}	8%	15%	1,8
	{fakultas}	{fakultas_ilmu_budaya}	8%	15%	1,8
2	{Profil}	{Pimpinan_fakultas_sains_dan_matematika}	12,5%	100%	8,0
	{Penerimaan_mahasiswa_baru}	{Kalender_penerimaan_mahasiswa_baru}	62,5%	100%	1,6
3	{penerimaan_mahasiswa_baru}	{daya_tampung}	5%	3,2%	6,6
	{fakultas}	{fakultas_teknik}	5%	10,5%	2,3

Pada Tabel 9, setiap *cluster* memiliki jumlah *rules* yang berbeda-beda. Nilai *support* dan *confidence* dikalikan 100% untuk mendapatkan hasil dalam bentuk presentase. Lhs menandakan *antecendent* dan rhs menandakan *consequent*. Setiap *rule* yang terbentuk mengartikan bahwa *user* melakukan akses dari kategori halaman LHS memiliki presentase sejumlah *confidence* untuk mengakses kategori halaman RHS. Pola tersebut terdapat sejumlah presentase *support* dari total data dengan nilai *lift*>1 yang menandakan pola tersebut dependent.

5. KESIMPULAN

Berdasarkan Metode *Gustafson-Kessel Clustering* dapat disimpulkan bahwa jumlah *cluster* yang terbentuk adalah 2 dan 3 *cluster*.

Dari pengujian yang telah dilakukan pada jumlah *cluster* 2 terdapat 2 pola *user* dalam melakukan akses pada *website* Universitas Diponegoro yang memiliki frekuensi kemunculan tinggi:

Fakultas → Fakultas Teknik pada *cluster* pertama dengan *confidence* 17% dan *support* 6%. Hal tersebut dapat menandakan bahwa sebagian besar *user* di *cluster* pertama cenderung melakukan akses dari *sub session* ‘fakultas’ menuju *sub session* ‘fakultas teknik’. Artinya pengunjung *website* di *cluster* pertama lebih banyak membutuhkan informasi mengenai fakultas teknik.

Penerimaan mahasiswa baru → daya tampung pada *cluster* kedua dengan *confidence* 13% dan *support* 6%. Hal tersebut dapat menandakan bahwa sebagian besar *user* di *cluster* pertama cenderung melakukan akses dari *sub session* ‘penerimaan mahasiswa baru’ menuju *sub session* ‘daya tampung’. Artinya pengunjung *website* di *cluster* pertama lebih banyak membutuhkan informasi mengenai daya tampung.

Dari pengujian yang telah dilakukan pada jumlah *cluster* 3 terdapat 3 pola *user* dalam melakukan akses pada *website* Universitas Diponegoro yang memiliki frekuensi kemunculan tinggi:

Fakultas → Fakultas Kedokteran pada *cluster* pertama dengan *confidence* 46,30% dan *support* 26,04%. Hal tersebut dapat menandakan bahwa sebagian besar *user* di *cluster* pertama cenderung melakukan akses dari *sub session* 'fakultas' menuju *sub session* 'fakultas kedokteran'. Artinya pengunjung *website* di *cluster* pertama lebih banyak membutuhkan informasi mengenai fakultas kedokteran.

Penerimaan mahasiswa baru → kalender penerimaan mahasiswa baru pada *cluster* kedua dengan *confidence* 100% dan *support* 62,5%. Hal tersebut menandakan bahwa sebagian besar pengunjung cenderung melakukan akses dari *session* 'Penerimaan mahasiswa baru' menuju *sub session* 'kalender penerimaan mahasiswa baru'. Artinya pengunjung *website* di *cluster* kedua membutuhkan informasi mengenai *timeline* penerimaan mahasiswa baru dari mulai tanggal pelaksanaan hingga tahap seleksi.

Fakultas → fakultas teknik pada *cluster* ketiga dengan *confidence* 10,5% dan *support* 5%. Hal tersebut menandakan sebagian besar *user* di *cluster* ketiga cenderung melakukan akses dari *session* 'fakultas' menuju 'fakultas teknik'. Artinya pengunjung *website* di *cluster* ketiga membutuhkan informasi mengenai fakultas teknik.

Beberapa perbaikan yang dapat dilakukan untuk penelitian selanjutnya adalah menggunakan metode *clustering* yang lebih efisien serta memiliki kecocokan dengan data yang relatif homogen, tahap *preprocessing* merupakan tahap yang paling penting dan riskan.Pastikan telah dilakukan dengan baik dan seksama, seperti penghapusan *page request* yang benar-benar tidak dibutuhkan.

DAFTAR PUSTAKA

Babuska, R., Veen, P.v.d. & Kaymak, U., 2002.*Improved Covariance Estimation for Gustafson-Kessel Clustering*.Netherland: Delft University of Technology.

Fauzanu, A., Eko, D. & Gede, A, A, W., 2017. *Analisis Web Usage Mining Menggunakan Teknik K-Means Clustering dan Association Rule (Studi Kasus: www.owlexa.com)*. e-Proceeding of Engineering, Vol. 4(2), Hal. 3284-3291.

Gustafson, D. & Kessel, W., 1979, *Fuzzy Clustering with a Fuzzy Covariance Matrix*. San Diego, Hal. 761-766.

Hikmawan, W. A. 2017. *Analisis Web Usage Mining untuk Pembentukan Profil User Menggunakan Algoritma Clustering K-Means (Studi Kasus: Website etd.ugm.ac.id)*. Universitas Gajah Mada.

Rahmatika, L., Suparti. &Diah, S., 2015.*Analisis Kelompok dengan Algoritma Fuzzy C-Means dan Gustafson Kessel Clustering pada Indeks LQ45*.Jurnal Gaussian, Vol. 4(3), Hal. 543-552.

Scime, A., 2004. *Web Mining: Applications and Tecniques*.USA*:* State University of New York College at Brockport.

Yan J. (2009) *Text Representation*. In: LIU L., ÖZSU M.T. (eds) Encyclopedia of Database Systems. Springer, Boston, MA