

WEB USAGE MINING MENGGUNAKAN K-MEANS UNTUK MENGETAHUI KECENDERUNGAN AKSES PENGGUNA (STUDI KASUS: GANTO.CO)

Rizky Maulana¹⁾, Nurindah Dwiyani²⁾

¹⁾Program Studi Pendidikan Teknik Informatika, Fakultas Teknik, Universitas Negeri Padang

²⁾Dosen Jurusan Teknik Elektronika, Fakultas Teknik, Universitas Negeri Padang

Jalan Prof. Hamka, Kampus UNP Air Tawar, Padang

e-mail: ¹⁾r14076073@gmail.com, ²⁾nurindahkamar@gmail.com

ABSTRAK

Situs web Ganto, *ganto.co*, merupakan salah satu sumber informasi bagi mahasiswa, dosen, dan juga khalayak ramai. Setiap kunjungan ke situs web Ganto dicatat di dalam *server log*. Jika dibiarkan begitu saja, *server log* hanya akan memenuhi penyimpanan server. Oleh karena itu, perlu dilakukan tindakan seperti menerapkan *Web usage mining*. *Web usage mining*, menggunakan *k-means clustering*, merupakan salah satu cara untuk menjamin kepuasan *user* yang mengakses situs web Ganto, yaitu dengan mengetahui kecenderungan akses *user* melalui analisis *server log*. *Cluster* pertama, September 2017. *Cluster* kedua, Oktober 2017, November 2017, Desember 2017, Februari 2018, Maret 2018, dan April 2018. Sedangkan *cluster* ketiga, Juli 2017, Agustus 2017, Januari 2018, Mei 2018, dan Juni 2018. Hasil analisis, Berita merupakan rubrik yang paling sering di-view, kedua Artikel, dan ketiga E-Paper. Sedangkan rubrik lainnya (Info Kampus, Sastra Budaya, Ganto TV, dan Ganto Foto) tidak terlalu sering di-view oleh *user* atau pengunjung situs web Ganto.

Kata kunci: *Web usage mining, clustering, k-means, Ganto*

ABSTRACT

The Ganto website, ganto.co, is one source of information for students, lecturers, and the public. Every visit to the Ganto website is recorded on the server log. If left unchecked, the server log will only fulfill server storage. Therefore, actions need to be taken such as implementing Web usage mining. Web usage mining, using k-means clustering, is one way to ensure the satisfaction of users who access the Ganto website, by knowing the tendency of user access through server log analysis. First Cluster, September 2017. Second Cluster, October 2017, November 2017, December 2017, February 2018, March 2018, and April 2018. While the third cluster, July 2017, August 2017, January 2018, May 2018, and June 2018. Results of analysis, Berita is the most frequently viewed rubric, the second is Artikel, and the third is E-Paper. While the other rubrics (Info Kampus, Sastra Budaya, Ganto TV, and Ganto Foto) are not too often viewed by users or visitors of the Ganto website.

Keywords: *Web usage mining, clustering, k-means, Ganto*

PENDAHULUAN

Di era digital seperti sekarang ini, orang lebih memilih mencari informasi di internet melalui telepon genggam, komputer desktop, atau perangkat komputer yang sejenis. Informasi tersebut biasanya dimuat dalam situs-situs web sehingga mudah diakses kapan saja dan di mana saja—selama terhubung ke internet.

Situs web adalah keseluruhan halaman web yang terdapat dalam sebuah domain yang mengandung informasi [1]. Situs web merupakan sebuah jaringan komputer *client-server* yang menggunakan *software* khusus [2].

Surat Kabar Kampus Ganto (SKK Ganto atau *Ganto* saja) merupakan sebuah koran kampus yang berbasis di Universitas Negeri Padang (UNP). Menurut Rektor UNP [3], Ganto tidak hanya sebagai

media penerbitan tetapi juga berfungsi sebagai sebuah wadah yang edukatif dan komunikatif.

Dalam perkembangannya, Ganto tidak hanya memuat berita atau informasi dalam bentuk media cetak (koran). Pada periode 2005–2006, Ganto merilis situs web dengan nama domain *ganto.unp.ac.id* di bawah pimpinan Adriyanto. Kemudian pada periode-periode selanjutnya, Ganto mengalami beberapa perubahan nama domain, dari *ganto.unp.ac.id* menjadi *ganto.web.id*, *ganto.or.id*, dan sekarang menjadi *ganto.co* [3].

Ganto berperan penting dalam menyediakan informasi yang edukatif, komunikatif, dan relevan bagi mahasiswa, dosen, maupun khalayak ramai. Setiap kunjungan ke situs web Ganto dicatat di dalam suatu *file* yang tersimpan—biasanya—di sisi server, yaitu *server log*. *File* tersebut berisi catatan kunjungan, seperti alamat IP, waktu akses, halaman yang diakses, dan lain-lain.

Server log situs web Ganto hanya akan bertambah banyak dan memenuhi penyimpanan *server* jika dibiarkan begitu saja. Oleh karena itu, perlu dilakukan tindakan seperti menerapkan *Web usage mining*.

Web Usage mining, which is the process of applying data mining techniques to the discovery of usage patterns from Web data, targeted towards various applications [4].

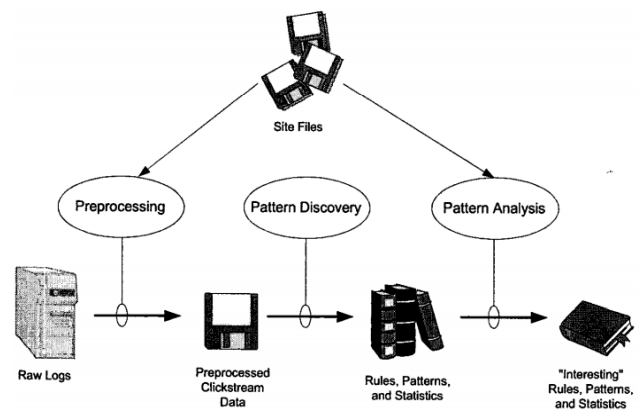
Web usage mining merupakan penerapan teknik *data mining* untuk menemukan pola penggunaan ataupun kecenderungan akses *user* terhadap sebuah situs web. *Data mining is the process of discovering useful patterns and trends in large data sets* [5]. *Data mining* merupakan analisis terhadap data (biasanya data yang berukuran besar) menggunakan metode tertentu untuk menemukan hubungan yang jelas serta menyimpulkan yang belum diketahui sebelumnya agar berguna bagi si pemilik data tersebut [6]. Fungsi-fungsi dalam *Data Mining* antara lain fungsi deksripsi, estimasi, prediksi, klasifikasi, klusterisasi dan asosiasi [7].

Teknik *data mining* yang paling sering digunakan [5], yaitu (1) *Description*, (2) *Estimation*, (3) *Prediction*, (4) *Classification*, (5) *Clustering*, dan (6) *Association*.

Clustering (penggugusan) berkenaan dengan pengelompokan *record*, observasi, atau kasus menjadi kelas dengan objek-objek yang mirip. Sebuah *cluster* (gugus) adalah kumpulan *record* yang serupa satu sama lain, tetapi takserupa dengan *record* dalam gugus lain. Algoritme penggugusan mencari dan menyegmentasikan seluruh data ke dalam subgolongan atau gugus, memaksimalkan kemiripan *record* dalam satu gugus, dan meminimalkan kemiripan *record* antargugus.

METODE

Alur penelitian ini menyesuaikan dengan proses *Web usage mining* [4].



Gambar 1. Proses *Web usage mining*

1. Pengumpulan Data

Data set (*server log*) yang digunakan, yaitu kunjungan ke situs web Ganto selama satu tahun (Juli 2017–Juni 2018). Setelah *data set* terkumpul, maka lanjut ke *preprocessing*.

2. *Preprocessing*

Hal-hal yang dilakukan dalam *preprocessing* adalah (1) *data cleaning*, (2) *user identification*, (3) *session identification*.

a. *Data cleaning*

Tahap ini bertujuan untuk membersihkan *server log* dari data yang tidak relevan dengan proses mining, seperti data multimedia (misalnya *.png, *.gif, *.pdf, dan lain-lain), *http request method* selain GET, *http status code* atau *request status* selain 200, kolom *rfc1413*, *authuser*, dan juga *bytes*.

b. *User identification*

Proses untuk mengidentifikasi *user* yang mengakses suatu situs web. Hal tersebut ditentukan dengan menyeleksi IP dan UA (*User Agent*) yang sama. Jika IP dan UA-nya sama, maka dihitung sebagai satu pengguna.

c. *Session identification*

Proses untuk mengidentifikasi *session* dari tiap-tiap *user* yang mengakses situs web. Jika seorang *user* mengakses suatu halaman web (termasuk *page loading* dan membaca konten) dalam waktu tidak lebih dari tiga puluh menit, maka itu dihitung sebagai satu *session*.

3. *Pattern Discovery*

Tahap pencarian pola akses yang dilakukan oleh *user* (pengguna) situs web. Tahap ini sangat penting dan menentukan keluaran dari

proses *Web usage mining*. Beberapa teknik yang sering digunakan dalam proses *pattern discovery* ini adalah *statistical analysis, association rules, clustering, classification, sequential pattern, dan dependency modeling*. Pada penelitian ini digunakan *k-means clustering*.

Adapun algoritme *k-means* adalah sebagai berikut [5] [6].

- a. Tentukan banyak *cluster k* yang akan dibentuk dari *data set*.

Banyaknya *cluster* yang akan dibentuk ditentukan oleh peneliti [5]. “*How does the k-means algorithm work?*” ... *First, it randomly selects k* (banyaknya *cluster* yang akan dibentuk) *of the objects in D* (*data set* yang akan diolah), [8].

- b. Pilih *k record* untuk dijadikan sebagai lokasi pusat awal *cluster*.

Penentuan *centroid* pada tahap awal, umumnya dilakukan secara acak [9]. Pilih secara acak *k record* untuk dijadikan *cluster center (m_i)* awal [5]. Sedangkan untuk iterasi selanjutnya menggunakan rumus berikut [6].

$$C(i) = \frac{X_1 + X_2 + \dots + X_n}{\sum x} \quad (1)$$

Keterangan:

C(i) = *centroid* dari *cluster* ke-*i*

X₁ = nilai record ke-1

X₂ = nilai record ke-2

X_n = nilai record ke-*n*

$\sum x$ = banyaknya anggota *cluster*

- c. Cari jarak terdekat ke pusat *cluster* pada setiap *record*.

Jarak terdekat dengan *centroid* yang digunakan adalah *Euclidean Distance*, dihitung menggunakan persamaan berikut [9].

$$D(i, j) = \sqrt{(X_{1i} - X_{1j})^2 + (X_{2i} - X_{2j})^2 + \dots + (X_{ki} - X_{kj})^2} \quad (2)$$

Keterangan:

D(i, j) = jarak data ke *i* ke pusat *cluster j*

X_{ki} = data ke *i* pada atribut data ke *k*

X_{kj} = titik pusat ke *j* pada atribut ke *k*

- d. Hitung rasio antara *Between-Cluster Variation (BCV)* dan *Within-Cluster Variation (WCV)*.

$$BCV = \frac{d(m_1 - m_2) + d(m_1 - m_3) + d(m_2 - m_3)}{WCV} \quad (3)$$

$$\frac{BCV}{WCV} \quad (4)$$

- e. Untuk setiap *k cluster*, temukan *cluster center*-nya dan perbarui lokasi data dalam *cluster* berdasarkan nilai *centroid* baru.
- f. Ulangi langkah *c-f* hingga *convergence* (*centroid* tidak berubah).

Perhitungan pada tahap *pattern discovery* dilakukan secara manual dan menggunakan RapidMiner Studio. RapidMiner Studio adalah *visual workflow designer* yang membantu peneliti data agar lebih produktif, yaitu dari membuat prototipe gagasan dengan cepat hingga perancangan model prediksi tugas kritis [10].

4. Pattern Analysis

Proses ini merupakan akhir dari *Web usage mining*. Pada tahap ini dilakukan proses visualisasi hasil *pattern discovery*. Hasilnya disajikan sesuai dengan kebutuhan, misalnya dalam bentuk grafik, diagram, dan lain-lain.

HASIL DAN PEMBAHASAN

Berikut adalah *server log* situs web Ganto (Juli 2017–Juni 2018). *Data set (server log)* di bawah ini terdiri dari 12 *record* (baris) dan 7 *attribute* (kolom).

Tabel 1. *Server log* situs web Ganto Juli 2017–2018

bulan	berita	info_kampus	sastra_budaya
Juli	771	19	15
Agustus	1251	14	74
September	25256	71	39
Oktober	9867	55	19
November	10584	38	33
Desember	15303	30	21
Januari	3190	21	7
Februari	11031	42	18
Maret	13135	78	280
April	10476	55	30
Mei	6524	35	13
Juni	4039	27	18

ganto_tv	ganto_foto	artikel	e_paper
24	14	109	18
19	20	877	64
44	61	467	40
27	42	323	29
23	35	352	31
24	101	206	45
14	13	211	51
28	21	345	100
154	256	6399	439
32	49	914	219
25	21	1014	145
19	22	780	93

Terdapat tujuh rubrik atau menu utama yang ada di dalam situs web Ganto, yaitu (1) Berita, (2) Info Kampus, (3) Sastra Budaya, (4) Ganto TV, (5) Ganto Foto, (6) Artikel, dan (7) E-Paper. Gambar 2 merupakan *server log* situs web Ganto yang telah melalui tahap *preprocessing*.

Selanjutnya adalah *pattern discovery*. Berikut adalah perhitungan yang dilakukan pada tahap tersebut.

1. Perhitungan Manual

- a) Tentukan banyaknya *cluster* yang akan dibentuk: *k* = 3.

b) Tentukan *centroid* awal secara acak.

Tabel 2. *Centroid* awal

m_i	1	2	3	4	5	6	7
1	25256	71	39	44	61	467	40
2	10584	38	33	23	35	352	31
3	6524	35	13	25	21	1014	145

c) Iterasi 1: Hitung jarak terdekat (persamaan 2)

(1) Juli 2017

$$D(1,1) = \sqrt{\begin{matrix} (771-25256)^2 + (19-71)^2 + \\ (15-39)^2 + (24-44)^2 + \\ (14-61)^2 + (109-467)^2 + \\ (18-40)^2 \end{matrix}}$$

$$= 24487,75$$

$D(1,1)$: perhitungan jarak *record* Juli 2017 ke titik pusat *cluster* 1 (m_1)

$$D(1,2) = \sqrt{\begin{matrix} (771-10584)^2 + (19-38)^2 + \\ (15-33)^2 + (24-23)^2 + \\ (14-35)^2 + (109-352)^2 + \\ (18-31)^2 \end{matrix}}$$

$$= 9816,07$$

$D(1,2)$: perhitungan jarak *record* Juli 2017 ke titik pusat *cluster* 2 (m_2)

$$D(1,3) = \sqrt{\begin{matrix} (771-6524)^2 + (19-35)^2 + \\ (15-13)^2 + (24-25)^2 + \\ (14-21)^2 + (109-1014)^2 + \\ (18-145)^2 \end{matrix}}$$

$$= 5825,16$$

$D(1,3)$: perhitungan jarak *record* Juli 2017 ke titik pusat *cluster* 3 (m_3)

d) Kelompokkan *record* ke dalam *cluster* tertentu berdasarkan jarak terdekat atau nilai terkecil (hasil langkah c ditabulasi).

Tabel 3. Hasil iterasi 1

Bulan	Jarak ke			Jarak terdekat ke <i>cluster</i>
	C1	C2	C3	
Juli	24487.75	9816.07	5825.16	C3
Agustus	24008.65	9347.95	5275.80	C3
September	0.00	14672.53	18740.38	C1
Oktober	15389.72	717.97	3415.77	C2
November	14672.53	0.00	4115.27	C2
Desember	9956.62	4721.76	8817.04	C2
Januari	22067.64	7395.47	3430.69	C3
Februari	14225.76	452.86	4556.61	C2
Maret	13504.62	6585.51	8540.20	C2
April	14787.86	602.85	3954.15	C2
Mei	18740.38	4115.27	0.00	C3
Juni	21219.48	6559.31	2496.56	C3
September	Oktober	Juli		
	November	Agustus		
	Desember	Januari		
	Februari	Mei		
	Maret	Juni		
	April			
	1	6	5	

Berdasarkan Tabel 3, anggota tiap-tiap *cluster*, yaitu *cluster* 1 (September 2017), *cluster* 2 (Oktober 2017, November 2017, Desember 2017, Februari 2018, Maret 2018, dan April 2018), *cluster* 3 (Juli 2017, Agustus 2017, Januari 2018, Mei 2018, Juni 2018).

e) Hitung rasio BCV dan WCV (persamaan 3 dan 4)

Berdasarkan Tabel 2, rata-rata *centroid* awal adalah $m_1 = 3711,14$; $m_2 = 1585,14$; dan $m_3 = 1111$.

$$BCV = (3711,14 - 1585,14) + (3711,14 - 1111) + (1585,14 - 1111) = 5200,29$$

$$WCV = (5825,16)^2 + (5275,8)^2 + (0)^2 + (717,97)^2 + (0)^2 + (4721,76)^2 + (3430,69)^2 + (452,86)^2 + (6585,51)^2 + (602,85)^2 + (0)^2 + (2496,56)^2 = 146516875,00$$

Setelah diketahui nilai BCV dan WCV, maka hitung rasionya.

$$\frac{BCV}{WCV} = \frac{5200,29}{146516875,00} = 0,00003549$$

Belum ada rasio perbandingan (BCV dan WCV) dengan rasio sebelumnya, maka hitung kembali jarak setiap data dengan *centroid* baru.

f) Tentukan *centroid* baru, yaitu rata-rata nilai dari tiap anggota *cluster* yang terbentuk (persamaan 1).

(1) *Centroid* dari *cluster* pertama

Item pada *cluster* pertama hanya satu, September 2017, maka *record* tersebut dapat langsung digunakan sebagai *centroid* baru *cluster* pertama: $m_1 = (25256, 71, 39, 44, 61, 467, 40)$.

(2) *Centroid* dari *cluster* kedua

Item pada *cluster* kedua berjumlah enam: Oktober 2017, November 2017, Desember 2017, Februari 2018, Maret 2018, dan April 2018. Oleh karena itu, *centroid* baru *cluster* kedua ditentukan dengan mencari rata-rata dari *record* anggota *cluster* kedua. Contoh perhitungan:

$$\text{Berita} = \frac{9867 + 10584 + 15303 + 11031 + 13135 + 10476}{6} = 11732,67$$

$$m_2 = (11732,67; 49,67; 66,83; 48; 84; 1423,17; 143,83)$$

(3) *Centroid* dari *cluster* ketiga: $m_3 = (3155; 23,20; 25,40; 20,20; 18; 598,20; 74,20)$

Tabel 4. *Centroid* baru berdasarkan hasil iterasi 1

m_i	1	2	3	4	5	6	7
1	25256.0	71.0	39.0	44.0	61.0	467.00	40.00
2	11732.67	49.67	66.83	48	84	1423.17	143.83
3	3155.00	23.2	25.4	20.2	18.0	598.20	74.20

Berdasarkan Tabel 4, didapatkan rata-rata tiap-tiap *centroid*, yaitu $m_1 = 3711,14$; $m_2 = 1935,45$; dan $m_3 = 559,17$.

g) Iterasi 2: Hitung kembali jarak terdekat.

h) Kelompokkan kembali *record* ke dalam *cluster* tertentu berdasarkan jarak terdekat atau nilai terkecil (hasil langkah *g* ditabulasi).

Tabel 5. Hasil iterasi 2

Bulan	Jarak ke			Jarak terdekat ke cluster
	C1	C2	C3	
Juli	24487.75	11041.29	2434.36	C3
Agustus	24008.65	10496.49	1924.97	C3
September	0.00	13557.56	22101.53	C1
Oktober	15389.72	2169.97	6717.92	C2
November	14672.53	1576.03	7433.24	C2
Desember	9956.62	3773.84	12154.65	C2
Januari	22067.64	8629.35	389.99	C3
Februari	14225.76	1289.77	7880.14	C2
Maret	13504.62	5186.49	11555.31	C2
April	14787.86	1359.04	7329.38	C2
Mei	18740.38	5225.44	3395.35	C3
Juni	21219.48	7721.16	902.74	C3
September	Oktober	Juli		
	November	Agustus		
	Desember	Januari		
	Februari	Mei		
	Maret	Juni		
	April			
	1	6	5	

Berdasarkan Tabel 5, perincian tiap-tiap *cluster* adalah sebagai berikut.

C1 = September 2017

C2 = (Oktober 2017, November 2017, Desember 2017, Februari 2018, Maret 2018, dan April 2018), dan

C3 = (Juli 2017, Agustus 2017, Januari 2018, Mei 2018, Juni 2018).

i) Hitung kembali rasio BCV dan WCV.

$$BCV = 6303,94$$

$$WCV = 73971705.57$$

$$\frac{BCV}{WCV} = \frac{6303,94}{73971705.57} = 0.00008522$$

Sampai langkah ini, keanggotaan *cluster* tidak berubah. Namun, jika rasio (BCV dan WCV) di atas dibandingkan dengan rasio sebelumnya, $0.00008522 > 0.00003549$. Oleh sebab itu, iterasi dilanjutkan.

j) Tentukan kembali *centroid* baru

Tabel 6. *Centroid* baru berdasarkan hasil iterasi 2

<i>m</i>	1	2	3	4	5	6	7
1	25256.0	71.0	39.0	44.0	61.0	467.00	40.00
2	11732.6	49.6	66.8	48	84	1423.1	143.8
3	3155.00	23.2	25.4	20.2	18.0	598.20	74.20

Berdasarkan langkah *h* dan *i*, keanggotaan tiap-tiap *cluster* tidak berubah. Oleh sebab itu, *centroid* baru (Tabel 6) yang akan digunakan pada iterasi ketiga, sama

dengan *centroid* (Tabel 4) pada iterasi kedua. Begitu juga dengan rata-rata *centroid* dari tiap-tiap *cluster*, yaitu $m_1 = 3711,14$; $m_2 = 1935,45$; dan $m_3 = 559,17$.

k) Iterasi 3: Hitung kembali jarak terdekat.

l) Kelompokkan kembali *record* ke dalam *cluster* tertentu berdasarkan jarak terdekat atau nilai terkecil (hasil langkah *k* ditabulasi).

Tabel 7. Hasil iterasi 3

Bulan	Jarak ke			Jarak terdekat ke cluster
	C1	C2	C3	
Juli	24487.75	11041.29	2434.36	C3
Agustus	24008.65	10496.49	1924.97	C3
September	0.00	13557.56	22101.53	C1
Oktober	15389.72	2169.97	6717.92	C2
November	14672.53	1576.03	7433.24	C2
Desember	9956.62	3773.84	12154.65	C2
Januari	22067.64	8629.35	389.99	C3
Februari	14225.76	1289.77	7880.14	C2
Maret	13504.62	5186.49	11555.31	C2
April	14787.86	1359.04	7329.38	C2
Mei	18740.38	5225.44	3395.35	C3
Juni	21219.48	7721.16	902.74	C3
September	Oktober	Juli		
	November	Agustus		
	Desember	Januari		
	Februari	Mei		
	Maret	Juni		
	April			
	1	6	5	

Berdasarkan Tabel 7, perincian tiap-tiap *cluster* adalah sebagai berikut.

C1 = September 2017

C2 = (Oktober 2017, November 2017, Desember 2017, Februari 2018, Maret 2018, dan April 2018), dan

C3 = (Juli 2017, Agustus 2017, Januari 2018, Mei 2018, Juni 2018).

m) Hitung kembali rasio BCV dan WCV.

$$BCV = 6303,94$$

$$WCV = 73971705.57$$

$$\frac{BCV}{WCV} = \frac{6303,94}{73971705.57} = 0.00008522$$

Sampai langkah ini, baik *centroid* maupun keanggotaan *cluster* tidak berubah. Rasio (BCV dan WCV) dengan iterasi sebelumnya juga tidak berubah: $0.00008522 = 0.00008522$. Oleh karena itu, iterasi dihentikan.

2. Perhitungan Menggunakan RapidMiner Studio

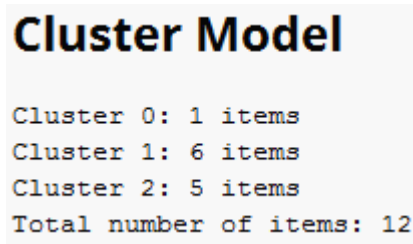
Tabel berikut adalah *centroid* akhir pada RapidMiner Studio.

Tabel 8. *Centroid* akhir di RapidMiner Studio

Attribute	Cluster_0	Cluster_1	Cluster_2
berita	25256	11732.667	3155
info_kampus	71	49.667	23.200
sastra_budaya	39	66.833	25.400

ganto_tv	44	48	20.200
ganto_foto	61	84	18
Artikel	467	1423.167	598.200
e_paper	40	143.833	74.200

Gambar berikut adalah perincian *cluster* dan jumlah anggotanya.



Gambar 2. Cluster model

Sedangkan Gambar berikut adalah perincian *cluster* dengan keterangan yang lebih lengkap, yaitu bulan-bulan apa saja yang berada pada tiap-tiap *cluster* yang terbentuk.

bulan	clust...	berita	info_kampus	sastra_budaya	ganto_tv	ganto_foto	artikel	e_paper
Septem...	cluster_0	25256	71	39	44	61	467	40
Oktober	cluster_1	9867	55	19	27	42	323	29
Novem...	cluster_1	10584	38	33	23	35	352	31
Desem...	cluster_1	15303	30	21	24	101	206	45
Februari	cluster_1	11031	42	18	28	21	345	100
Maret	cluster_1	13135	78	280	154	256	6399	439
April	cluster_1	10476	55	30	32	49	914	219
Juli	cluster_2	771	19	15	24	14	109	18
Agustus	cluster_2	1251	14	74	19	20	877	64
Januari	cluster_2	3190	21	7	14	13	211	51
Mei	cluster_2	6524	35	13	25	21	1014	145
Juni	cluster_2	4039	27	18	19	22	780	93

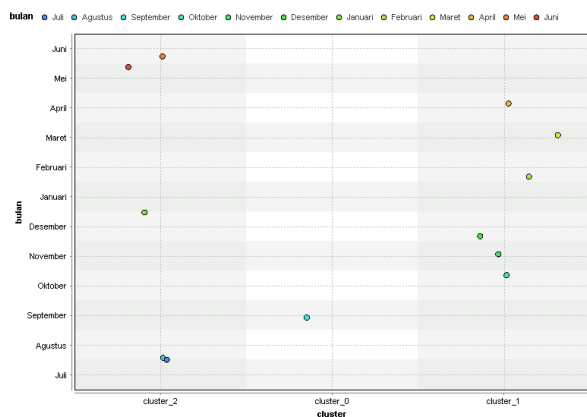
Gambar 3. Perincian *cluster* dan anggotanya

Berdasarkan Gambar 4, perincian *cluster* yang terbentuk adalah sebagai berikut.

C1 = September 2017

C2 = (Oktober 2017, November 2017, Desember 2017, Februari 2018, Maret 2018, dan April 2018), dan

C3 = (Juli 2017, Agustus 2017, Januari 2018, Mei 2018, Juni 2018).



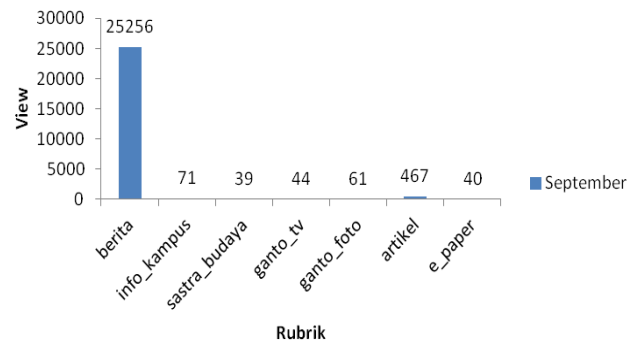
Gambar 4. Sebaran anggota *cluster*

Tahap terakhir, yaitu *pattern analysis*. Berdasarkan perhitungan manual dan perhitungan menggunakan aplikasi RapidMiner Studio, hasilnya sama. Berikut adalah hasil *k-means clustering* tersebut.

Tabel 9. Hasil WUM *k-means clustering*

Cluster 1	Cluster 2	Cluster 3
September 2017	Oktober 2017	Juli 2017
	November 2017	Agustus 2017
	Desember 2017	Januari 2018
	Februari 2018	Mei 2018
	Maret 2018	Juni 2018
	April 2018	
1	6	5

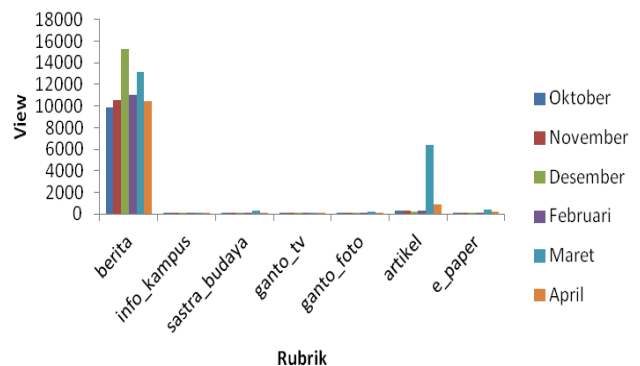
Cluster 1



Gambar 5. Cluster pertama

Cluster pertama merupakan bulan dengan *view* rubrik Berita terbanyak, yaitu 25256 (dua puluh lima ribu dua ratus lima puluh enam). Namun, pada rubrik lainnya, *view* September 2017 tidak sebanyak Maret 2018 (*cluster* kedua).

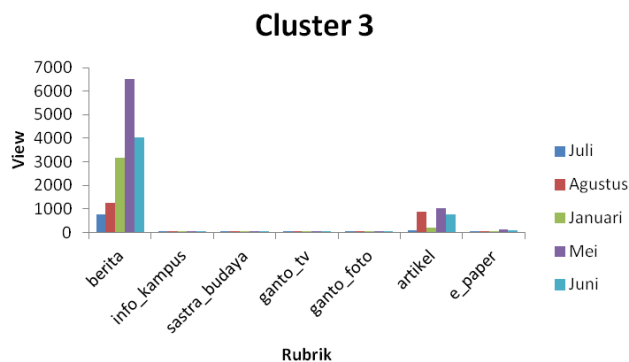
Cluster 2



Gambar 6. Cluster kedua

Sedangkan pada *cluster* kedua, selisih *view* rubrik berita tidak begitu besar, 9867, 10584, 15303, 11031, 13135, dan 10476. Namun, *view* tersebut lebih sedikit dibandingkan dengan *view* rubrik Berita pada

cluster pertama. Maret 2018 merupakan bulan dengan view rubrik Info Kampus, Sastra Budaya, Ganto TV, Ganto Foto, Artikel, dan E-Paper terbanyak, berturut-turut, yaitu 78, 280, 154, 256, 6399, dan 439.



Gambar 7. Cluster ketiga

Lalu, cluster ketiga merupakan bulan-bulan dengan view rubrik berita paling sedikit jika dibandingkan dengan bulan-bulan lainnya, yaitu 771, 1251, 3190, 6524, dan 4039. Begitu pun dengan view rubrik lainnya, seperti Info Kampus, Sastra Budaya, Ganto TV, dan Ganto Foto. Namun, view rubrik Artikel dan rubrik E-Paper lebih banyak jika dibandingkan dengan bulan lainnya (kecuali jika dibandingkan dengan Maret 2018).

Ketiga gambar (Gambar 5, Gambar 6, dan Gambar 7), menunjukkan bahwa yang paling sering di-view oleh pengunjung adalah rubrik Berita. Setelah itu, rubrik yang paling sering di-view adalah Artikel dan E-Paper. Sedangkan rubrik lainnya (Info Kampus, Sastra Budaya, Ganto TV, dan Ganto Foto) tidak terlalu sering di-view oleh pengunjung.

SIMPULAN

Berikut adalah simpulan dari penelitian yang telah dilakukan.

- Berita merupakan rubrik yang paling sering dikunjungi.
- Banyaknya cluster yang dibentuk berjumlah tiga ($k = 3$) dan iterasi dilakukan sebanyak tiga kali.
- Perincian cluster yang terbentuk, yaitu
 - cluster pertama, September 2017;
 - cluster kedua, Oktober 2017, November 2017, Desember 2017, Februari 2018, Maret 2018, dan April 2018; dan
 - cluster ketiga, Juli 2017, Agustus 2017, [7]

SARAN

Berdasarkan pembahasan dan kesimpulan, diajukan saran sebagai berikut.

- Ganto dapat menjadikan hasil penelitian ini sebagai bahan evaluasi untuk memperbaiki atau meningkatkan layanan situs web mereka agar

penyediaan berita atau informasi lebih informatif dan edukatif serta menarik.

- Web usage mining dapat diterapkan pada situs web selain Ganto, asalkan pihak atau pemilik situs web terkait berkenan untuk dilakukan penelitian.
- Gunakan catatan server log dengan rentang waktu yang lebih lama, misalnya dua tahun atau lebih dari lima tahun—jika tersedia—agar hasil lebih akurat.

DAFTAR PUSTAKA

- I. T. Syamnugroho and R. Efendi, "Pengembangan Website SMK Negeri 4 Semarang," *Jurnal Teknologi Informasi dan Komunikasi*, vol. V, no. 2, Agustus 2014.
- S. Hapsari, "Pembuatan Website pada Google Original Movie Rental Pacitan," *Journal Speed: Sentra Penelitian Engineering dan Edukasi*, vol. II, no. 2, pp. 48-54, 2010.
- "Tentang Kami," [Online]. Available: <http://www.ganto.co/profil/1/tentang-kami.html>. [Accessed 7 May 2018].
- J. Srivastava, R. Cooley, M. Deshpande and P.-N. Tan, "Web Usage Mining: Discovery and Applications of Usage Patterns from Web Data," *SIGKDD Explorations*, vol. I, no. 2, pp. 12-23, Januari 2000.
- D. T. Larose and C. D. Larose, *Discovering Knowledge in Data: An Introduction to Data Mining*, 2nd ed., New Jersey: John Wiley & Sons, Inc., 2014.
- H. Hamimi, "Analisis Data Anggaran Pendapatan Belanja Daerah Menggunakan Clustering K-Means dan Forecasting (Studi Kasus pada DPKA Kota Padang)," *Jurnal Vokasional Teknik Elektronika dan Informatika (VOTEKNIKA)*, vol. II, no. 1, 2014.
- H. K. Saputra, "Analisis Data Mining Untuk Pemetaan Mahasiswa Yang Membutuhkan Bimbingan dan Konseling Menggunakan Algoritma Naïve Bayes Classifier," *Jurnal Teknologi Informasi dan Pendidikan*, vol. 11, no. 1, pp. 14-26, 2018.
- J. Han, M. Kamber and J. Pei, *Data Mining: Concepts and Techniques*, 3rd ed., Massachusetts: Morgan Kaufmann, 2012.
- J. O. Ong, "Implementasi Algoritma K-means Clustering untuk Menentukan Strategi Marketing President University," *Jurnal Ilmiah Teknik Industri*, vol. XII, no. 1, pp. 10-20, June 2013.
- "RapidMiner Studio," [Online]. Available: <http://www.rapidminer.com/products/studio/>. [Accessed 9 August 2018].
- I. Lunden, "German Predictive Analytics Startup Rapid-I Rebrands as RapidMiner, Takes \$5M from Open Ocean, Earlybird to Tackle the U.S. Market," [Online]. Available: <https://techcrunch.com/2013/11/04/german-predictive-analytics-startup-rapid-i-rebrands-as-rapidminer-takes-5m-from-open-ocean-earlybird-to-tackle-the-u-s-market/>. [Accessed 12 August 2018].

- [12] G. Deutsch, "RapidMiner from Rapid-I at CeBIT 2010," [Online]. Available: <http://www.data-mining-blog.com/cloud-mining/rapidminer-cebit-2010/>. [Accessed 12 August 2018].
- [13] "Redaksi," [Online]. Available: <http://www.ganto.co/profil/4/redaksi.html>. [Accessed 7 May 2018].
- [14] "Interview with RapidMiner's Ingo Mierswa, Ralf Klinkenberg, part 1," [Online]. Available: <https://kdnuggets.com/2010/02/f-interview-rapid-i-founders.html>. [Accessed 12 August 2018].
- [15] A. and R. Adrian, "Penerapan Metode K-means untuk Clustering Mahasiswa berdasarkan Nilai Akademik dengan Weka Interface Studi Kasus pada Jurusan Teknik Informatika UMM Magelang," *Jurnal Ilmiah Semesta Teknika*, vol. XVIII, no. 1, pp. 76-82, Mei 2015.
- [16] T. Khotimah, "Pengelompokan Surat dalam Alquran Menggunakan Algoritma K-means," *Jurnal SIMETRIS*, vol. V, no. 1, pp. 83-88, April 2014.