

ANALISIS WEB USAGE MINING MENGGUNAKAN TEKNIK K-MEANS CLUSTERING DAN ASSOCIATION RULE (STUDI KASUS : WWW.OWLEXA.COM)

WEB USAGE MINING ANALYSIS USING K-MEANS CLUSTERING AND ASSOCIATION RULE TECHNIQUE (CASE STUDY : WWW. OWLEXA.COM)

Abid Fauzanu¹, Eko Darwiyanto², Gede Agung Ary Wisudiawan³

^{1,2,3}Prodi S1 Teknik Informatika, Fakultas Informatika, Universitas Telkom

¹abidfzn11@gmail.com, ²ekodarwiyanto@telkomuniversity.ac.id, ³ary.wisudiawan@gmail.com

Abstrak

Bagi sebuah perusahaan, mengetahui dan memahami perilaku pengunjung *websitenya* dapat menjadi sebuah acuan dalam melakukan optimasi. Pola akses menjadi salah satu bagian dari perilaku pengunjung ketika mengakses *website*. Dalam Tugas Akhir ini, *web log file* dari *website* resmi Owlexa Healthcare, bisnis unit dari PT. Aplikasi Lintasarta, www.owlexa.com, akan diproses dengan mengimplementasikan teknik *data usage mining* yang dapat menemukan pola perilaku pengunjung. Data log tersebut akan melalui tahap *preprocessing* lalu dilakukan teknik *clustering* dengan algoritma *k-means* dan teknik *association rule*. Dari proses tersebut menghasilkan analisis kecenderungan perilaku pengunjung. Hasil analisis tersebut dapat menjadi acuan bagi Owlexa dalam melakukan optimasi pada *websitenya*

Kata Kunci : *web log file, data usage mining, preprocessing, clustering, algoritma k-means, association rule.*

Abstract

For a company, knowing and understanding the behavior of their website visitors can be a benchmark to optimize their website. Access patterns become part of the visitors' behavior when accessing the website. In this Final Project, web log file from Owlexa Healthcare official website, a PT. Aplikasi Lintasarta business unit, www.owlexa.com, will be processed with data usage mining technique implementation that can be used to find visitor behavior patterns. The log data will go through a preprocessing stage then clustering technique using k-means algorithm and association rule technique. From that process will produce an analysis of trends in visitor behaviour. The result of the analysis can be a reference for Owlexa to optimizing their website.

Keywords : *web log file, data usage mining, preprocessing, clustering, K-Means algorithm, association rule.*

1. Pendahuluan

Dunia bisnis hari ini sangat dipengaruhi oleh kemajuan teknologi. Internet telah mendasari perilaku bisnis, dimulai dari promosi hingga transaksi. Internet telah mengubah bisnis konvensional menjadi e-bisnis. Dan e-bisnis tidak terlepas dari *website*, sebagai sarana bagi publik untuk mengetahui apa dan bagaimana bisnis suatu usaha [1]. Selain sebagai kredibilitas perusahaan dan peningkatan layanan, *website* juga menjadi salah satu media *branding* dan *marketing* perusahaan dalam menyampaikan informasi perusahaan dan produk-produknya. Penting bagi sebuah perusahaan untuk terus meningkatkan kualitas *websitenya* [2].

Owlexa Healthcare adalah sebuah unit bisnis baru dari PT. Aplikasi Lintasarta yang bergerak di bidang layanan kesehatan bagi peserta asuransi atau kelompok perusahaan, dengan *website* resminya, www.owlexa.com. Demi meningkatkan kualitas dan layanan dalam penyampaian informasi pada *websitenya*, Owlexa harus melakukan optimasi untuk mencapai target tersebut. Salah satu bentuk optimasi pada sebuah *website* adalah pada struktur dan navigasinya [3] [4].

Salah satu acuan dalam melakukan optimasi pada struktur dan navigasi sebuah *website* dapat dilihat dari kecenderungan pola akses pengunjung. Untuk memperoleh informasi perilaku pengunjung dilakukan analisis pada pola *click* di dalam *clickstream* pengunjung *website*. Hal tersebut dapat diperoleh dengan mengekstrak informasi dari *web log file* [5].

Penggunaan *web usage mining* memungkinkan untuk menemukan informasi tersembunyi tentang preferensi pengunjung. Dalam tugas akhir ini dilakukan teknik *k-means clustering* dan *association rule* pada *web log file*. Algoritma *k-means* digunakan karena merupakan algoritma yang paling banyak digunakan pada proses *clustering*, baik pada *web mining* maupun *data mining*. Alasannya adalah, karena *K-Means* merupakan salah satu teknik *clustering* yang paling mudah, cepat, tidak memakan banyak memori, serta bisa diaplikasikan kepada *dataset* berukuran kecil atau besar [6].

Association rule merupakan salah satu teknik data mining yang berfungsi untuk menemukan asosiasi antar variabel, korelasi atau suatu struktur diantara item atau objek-objek didalam database transaksi, maupun pada penyimpanan informasi lainnya [7]. Pada penelitian Tugas Akhir ini, algoritma asosiasi yang digunakan adalah Apriori. Karena algoritma apriori memiliki performansi yang lebih baik ketimbang algoritma lainnya disegala ukuran data [8].

Dari pola yang dihasilkan, dapat dianalisis dan ditarik kesimpulan menjadi sebuah acuan dan saran dalam melakukan optimasi pada *website* dari segi struktur dan bar navigasinya.

2. Tinjauan Pustaka

2.1. Web Mining

Web mining mengacu pada keseluruhan proses dari penemuan informasi yang berpotensi berguna yang belum diketahui sebelumnya dari data *web* [9]. *Web mining* juga dapat digunakan untuk mengambil informasi tentang konsumen atau pengguna *web*. Analisis dan penemuan untuk informasi yang berguna dari *World Wide Web* (WWW) menjadi tantangan bagi pelaku riset pada bidang ini. Fenomena pengambilan data dengan menggunakan teknik *data mining* disebut juga sebagai *web mining*. Terdapat tiga kategori *web mining*, yaitu, *Web Content Mining* (WCM), *Web Usage Mining* (WUM), dan *Web Structure Mining* (WSM) [5].

2.1.1. Web Usage Mining

Web usage mining merupakan proses untuk mengaplikasikan teknik *web mining* dalam melakukan analisis terhadap pola penggunaan (*usage pattern*) dan analisis aktivitas pengunjung yang tersimpan dalam *web log file*. Tipe data yang akan diolah berupa data *web log file* dalam *web*. Tujuan dari *web usage mining* adalah menemukan dan memprediksi tingkah laku *user* [10]. *Web usage mining* menjadi hal yang sangat penting untuk pengelolaan *website*, menciptakan *website* yang adaptif, layanan pada bisnis dan *support*, personalisasi, serta analisis alur trafik pada jaringan. [11] Di dalam *web usage mining* terdapat tiga proses utama, yaitu *Preprocessing*, *Pattern Discovery*, dan *Pattern Analysis* [12].

2.1.1.1. Preprocessing

Web log file harus melalui tahap *preprocessing* sebelum diolah lebih lanjut. *Preprocessing* dilakukan dengan maksud untuk :

- a) Membersihkan data, yaitu melakukan filterisasi terhadap data yang dihasilkan oleh *web page* berdasarkan *automatic requests* (contohnya *file gambar*), yang secara spesifik tidak di-request oleh *user*.
- b) Menyingkirkan *web log file* dari *nonhuman access behavior*, seperti *spider*, *crawler*, dan *automatic web bot* lainnya. Perilaku dari *bot* secara kualitatif berbeda dengan tingkah laku manusia serta dianggap tidak memiliki informasi yang berguna dari sudut pandang *web usage mining*.
- c) Identifikasi *user* yang berbeda. Dalam mengidentifikasi *user*, digunakan informasi alamat IP sebagai identitas unik dari tiap *user*.
- d) Identifikasi *user session*. Maksudnya, untuk setiap kunjungan, tentukan halaman apa saja yang di-request, urutan dari *request*, dan durasi pengaksesan tiap halaman. Tiap *session* memiliki waktu *time out* antar *request* yang membedakan suatu *session* dengan *session* selanjutnya, yaitu 30 menit.
- e) Identifikasi kategori, yaitu melakukan pengidentifikasian dari setiap kategori request sub halaman yang bisa dikelompokkan menjadi satu kategori halaman. Dengan cara pemberian ID unik ke setiap request halaman yang unik sebelumnya agar memudahkan pengkategorian [13].

2.1.1.2. Pattern Discovery

Pattern discovery adalah tahap penerapan proses yang mengacu pada satu atau beberapa metode dan algoritma. Pada tugas ini, *pattern discovery* yang digunakan adalah *k-means clustering* dan *association rule* [14].

2.1.1.3. Pattern Analysis

Pattern Analysis adalah tahap terakhir pada keseluruhan proses *web usage mining*. Tahap ini berguna untuk dan membuang hal-hal yang tidak diperlukan dari hasil *pattern discovery* lalu mengubah fakta-fakta menjadi informasi [14].

2.1.2. Web Log File

Web log file atau *web server log*, merupakan *records data* yang mencatat informasi aktivitas ketika pengguna *web* mengajukan permintaan ke *web server* [15].

Hasil catatan aktivitas tersebut yang menjadi sumber data utama dalam *web usage mining*. Dari sebuah *web log file*, dapat diketahui beberapa informasi mengenai pola akses dan kelakuan (*behaviour*) pengguna dalam mengakses halaman *web*. *Web log file* memiliki format yang standar, sehingga dapat mempermudah dalam proses *mining*.

2.2. Clustering

Clustering merupakan teknik untuk mengelompokkan objek yang memiliki karakter yang memiliki kemiripan. Dalam ranah *web usage mining*, terdapat dua jenis cluster yang menarik untuk dipelajari, yaitu *usage cluster* dan *page cluster*. *Clustering* yang dilakukan pada user cenderung dilakukan untuk mengelompokkan *user* yang memperlihatkan kemiripan pola. Biasanya *usage cluster* berguna untuk menentukan demografis *user* dalam upaya untuk melakukan segmentasi. Di lain pihak, *page cluster* akan mengelompokkan halaman yang memiliki korelasi. Kedua jenis *cluster* ini berguna untuk menghasilkan rekomendasi acuan optimasi [14].

2.2.1. K-Means Clustering

Algoritma *K-Means* merupakan algoritma yang paling banyak digunakan untuk proses *clustering*, baik pada *web mining* maupun *data mining*. Alasannya adalah, karena *K-Means* merupakan salah satu teknik *clustering* yang paling mudah, cepat, tidak memakan banyak memori, serta bisa diaplikasikan kepada *dataset* berukuran kecil atau besar [6]. *K* adalah jumlah *cluster* yang dibentuk. Cara kerja *k-means* adalah dengan memposisikan titik sebagai titik tengah (*centroid*), satu untuk setiap *cluster*. Selanjutnya adalah mengaitkan setiap titik kumpulan data ke *centroid* terdekat dengan menghitung jarak (*Euclidean*). Lalu dilakukan berulang-ulang hingga tidak ada perubahan pada *centroid*. Nilai *k* biasanya ditentukan secara acak, akan sangat sulit untuk menentukan nilai *k* yang optimum [16].

2.2.1.1. Sum of Squared Error (SSE)

Karena *k-means* menggunakan jarak kuadrat *euclidean* untuk mengalokasikan objek ke *cluster*, dapat diasumsikan bahwa data mempunyai skala jarak yang kira-kira sama untuk mendapatkan alokasi ke suatu *cluster*, maka dapat dihitung nilai *Sum of Squared Error (SSE)*. *SSE* didefinisikan sebagai jumlah jarak kuadrat antara masing-masing anggota *cluster* dan *centroid*-nya. Nilai *SSE* menandakan kemiripan variabelnya. Artinya, semakin kecil nilai *SSE*, semakin mirip variabel suatu *cluster*. Oleh karena itu, *SSE* dapat juga acuan sebagai penentu nilai *k* optimum [17].

2.2.1.2. Elbow Method

Elbow method adalah metode lama yang digunakan untuk menentukan jumlah *cluster* yang tepat pada sebuah *dataset*. Ini merupakan metode visual. Cara kerjanya dengan memulai nilai *k*=2 dan terus bertambah 1 poin setiap step sembari menghitung nilai variabel tertentu pada *cluster*, pada tugas ini yaitu nilai *SSE*. Pada suatu nilai *k*, grafik akan *drop* secara dramatis sebelum akhirnya stabil. Maka, disitulah nilai *k* yang optimum [16].

2.2. Association Rule [18]

Association rule mining adalah teknik *data mining* untuk menemukan aturan asosiatif antara suatu kombinasi item. Analisis asosiasi menjadi terkenal karena aplikasinya untuk menganalisis isi keranjang belanja di pasar swalayan. Analisis asosiasi juga sering disebut dengan istilah *market basket analysis*.

Penting tidaknya suatu aturan asosiatif dapat diketahui dengan dua parameter, *support* (nilai penunjang) yaitu persentase kombinasi item tersebut dalam database dan *confidence* (nilai kepastian) yaitu kuatnya hubungan antar item dalam aturan asosiatif.

Aturan asosiatif biasanya dinyatakan dalam bentuk : {roti, mentega} → {susu} (*support* = 40%, *confidence* = 50%). Yang artinya : "50% dari transaksi di database yang memuat item roti dan mentega juga memuat item susu. Sedangkan 40% dari seluruh transaksi yang ada di database memuat ketiga item itu."

Dapat juga diartikan : "Seorang konsumen yang membeli roti dan mentega punya kemungkinan 50% untuk juga membeli susu. Aturan ini cukup signifikan karena mewakili 40% dari catatan transaksi selama ini."

Analisis asosiasi didefinisikan suatu proses untuk menemukan semua aturan asosiatif yang memenuhi syarat minimum untuk *support* (*minimum support*) dan syarat minimum untuk *confidence* (*minimum confidence*).

Metodologi dasar analisis asosiasi terbagi menjadi dua tahap :

a) Analisis pola frekuensi tinggi

Tahap ini mencari kombinasi *item* yang memenuhi syarat minimum dari nilai *support* dalam database. Nilai *support* sebuah *item* diperoleh dengan rumus berikut:

$$\text{Support}(A) = \frac{\sum \text{transaksi mengandung } A}{\text{Total Transaksi}} \quad (1)$$

sedangkan nilai *support* dari 2 *item* diperoleh dari rumus berikut :

$$\text{Support}(A \cap B) = \frac{\sum \text{transaksi mengandung } A \text{ dan } B}{\text{Total Transaksi}} \quad (2)$$

b) Pembentukan aturan asosiatif

Setelah semua pola frekuensi tinggi ditemukan, barulah dicari aturan asosiatif yang memenuhi syarat minimum untuk *confidence* dengan menghitung *confidence* aturan asosiatif $A \rightarrow B$.

Nilai *confidence* dari aturan $A \rightarrow B$ diperoleh dari rumus berikut [19] :

$$P(A | B) = \frac{\sum \text{transaksi mengandung } A \text{ dan } B}{\text{Total Transaksi}} \quad (3)$$

Pada Bahasa pemrograman R, terdapat variable *Lift*. *Lift* adalah pengukuran independensi dari A dan B. Nilai *lift* antara 0 dan tak terhingga.

- Nilai *lift* =1 berarti A dan B adalah independen
- Nilai *lift* >1 berarti B lebih cenderung terjadi terhadap A
- Nilai *lift* <1 berarti B tidak cenderung terjadi terhadap A [20]

Sehingga rule yang dapat digunakan adalah yang hanya memiliki nilai *lift* >1.

Nilai *lift* dapat dirumuskan sebagai berikut:

$$\text{Lift} = \frac{P(A | B)}{P(A) \cdot P(B)} \quad (4)$$

Aturan asosiatif yang dihasilkan mengartikan bahwa pola yang dihasilkan menunjukkan presentase *if antecedent then consequent* dengan minimal *support* yang telah ditentukan, bukan *sequence pattern*.

2.3. Owlexa Healthcare

Owlexa Healthcare merubakan unit bisnis e-Health dari PT Aplikasi Lintasarta yang bertujuan untuk memenuhi kebutuhan layanan administrasi asuransi dan jaminan kesehatan yang handal, cepat dan akurat. Owlexa memberikan pelayanan terhadap proses administrasi klaim perusahaan atau asuransi mulai dari proses membership & policy administration, eligibility member, case monitoring, discharge sampai pada proses pembayaran tagihan klaim biaya pengobatan [21].

2.4. R Programming

R adalah salah satu bahasa pemrograman yang *open source* dengan system yang mendukung perhitungan dan statistik. Inti dari bahasa R adalah Bahasa pemrograman yang diinterpretasikan memungkinkan untuk melakukan *branching* dan *looping* serta pemrograman modular menggunakan fungsi.

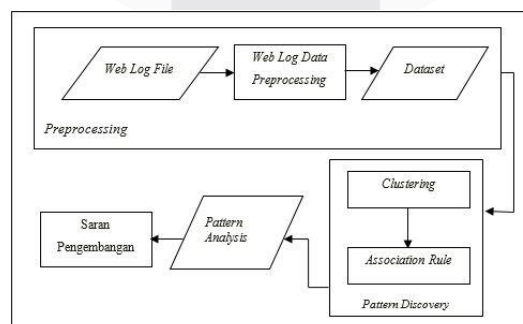
Keunggulan dalam bahasa pemrograman r adalah memiliki fungsionalitas untuk mengolah data statistik yang besar. Terdapat banyak fungsi untuk menciptakan berbagai jenis tampilan data secara visual [22]. Serta kemudahan dalam pengoperasiannya, sehingga Bahasa pemrograman ini banyak digunakan oleh korporasi dan akademisi dalam melakukan *data mining*.

Pada Tugas Akhir ini, *compiler* yang digunakan adalah RStudio.

3. Pembahasan

3.1. Gambaran Umum Sistem

Gambar 3.1 merupakan gambaran umum dari rancangan sistem yang akan dibuat :

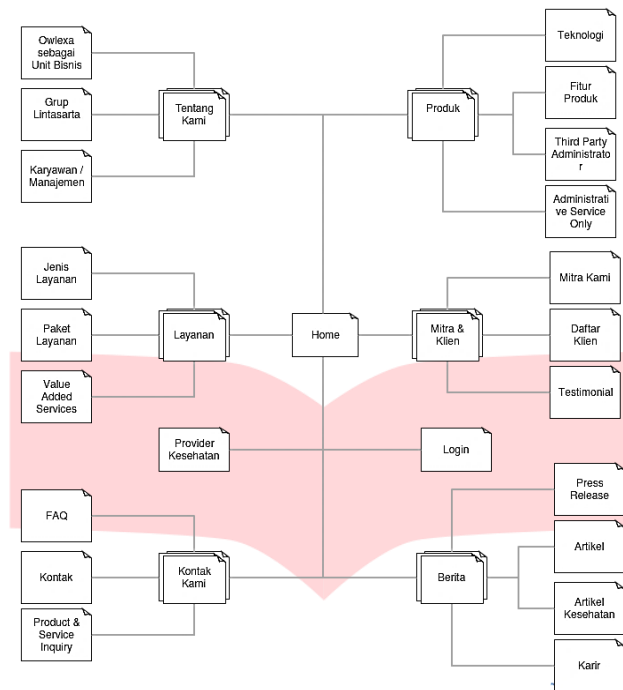


Gambar 3.1 Gambaran Umum Sistem

Sistem yang dibangun pada tugas akhir ini adalah sistem yang dapat menghasilkan sebuah pola pengunjung owlexa.com dengan mengelompokkan ke dalam *cluster* lalu dilakukan *association rule* dari data *web log file* owlexa.com. Pola yang terbentuk nantinya akan dianalisis dan ditarik saran pengembangan bagi *website*

Owlexa. Sistem dibagi menjadi tiga bagian, yaitu: *preprocessing* data, *pattern discovery* dengan k-means *clustering* dan *association rule* dan *pattern analysis*.

3.2. Owlexa Web Structure



Gambar 3.2 Owlexa Web Structure

3.3. Web Log File

Dataset yang digunakan berupa *web log file* yang masih *raw* dari *www.owlexa.com* yang hanya dapat diakses oleh pihak perusahaan. *Web log* yang dianalisis adalah *records* dari tanggal 18 September 2016 – 25 September 2016.

Data tersebut memiliki 759248 *records*. Tabel 3.1 menunjukkan bentuk contoh datanya :

Table 3.1 Sampel Web Log Data

182.23.65.83 - - [18/Sep/2016:03:20:30 +0700] "GET /index.php/id/ " 200 - "-" "Mozilla/5.0 (X11; Linux i686; rv:8.0) Gecko/20100101 Firefox/8.0"
157.55.39.122 - - [18/Sep/2016:03:23:47 +0700] "GET /index.php/id/ HTTP/1.1" 200 54947 "-" "Mozilla/5.0 (X11; Linux i686; rv:8.0) Gecko/20100101 Firefox/8.0"
182.23.65.83 - - [20/Sep/2016:03:30:00 +0700] "GET /index.php/id/news/articles HTTP/1.1" 302 35754 "-" "Mozilla/5.0 (X11; Linux i686; rv:8.0) Gecko/20100101 Firefox/8.0"

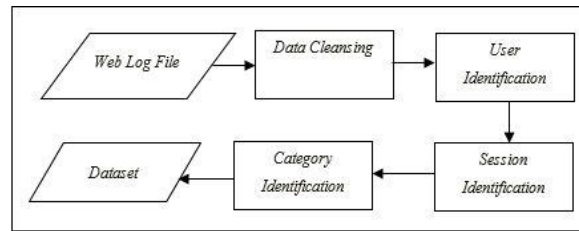
Jika didefinisikan ke dalam table akan terbentuk table informasi seperti pada table 3.2 :

Table 3.2 Tabel Raw Data

IP Address	User Identifier	User Identity	Date & Time	Page Request	Status	Size Request	Reference	Browser
182.23.65.83	-	-	18/Sep/2016:03:20:30	/index.php/id/	200	621	-	Mozilla
157.55.39.122	-	-	18/Sep/2016:03:23:47	/index.php/id/	200	54947	-	Mozilla
182.23.65.83	-	-	18/Sep/2016:03:30:00	/index.php/id/news/articles	302	35754	-	Mozilla

3.4. Preprocessing

Preprocessing adalah sebuah proses untuk mempersiapkan data agar data bisa diolah lebih lanjut oleh proses selanjutnya. Gambar 3.3 ini adalah langkah dalam *preprocessing* :



Gambar 3.3 Proses Preprocessing

Setelah proses *preprocessing* selesai, terbentuk *dataset* yang berisi 56333 records terdiri dari *request page*, *user_id*, *session_id*, *id_kategori*. Seperti pada table 3.3.

Table 3.3 Contoh Hasil Preprocessing

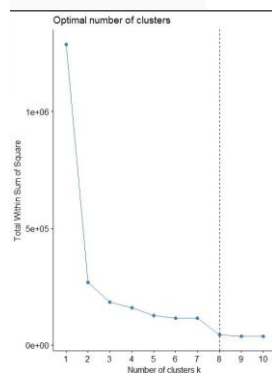
Page Request	UserID	SessionID	ID Kategori
/index.php/id	1	1	1
/index.php/id/news/artikel-kesehatan	1	2	7

3.5. Pattern Discovery

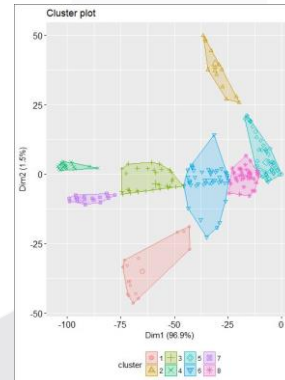
Data hasil *preprocessing* kemudian di proses dalam *pattern discovery*. Berikut adalah proses yang akan dilakukan di dalam *pattern discovery*.

3.5.1. K-Means Clustering

Untuk memulai *clustering* terlebih dahulu mencari nilai *k* optimum dengan menggunakan metode Elbow berdasarkan grafik nilai SSE terhadap nilai *k*. didapatkan hasil grafik seperti pada gambar 3.4:



Gambar 3.5 Grafik SSE Terhadap K



Gambar 3.4 Cluster Plot

Bedasarkan metode Elbow, dari hasil grafik pada gambar 3.4 dapat diambil kesimpulan bahwa nilai *k* optimum terdapat pada *k*=8. Dari hasil penentuan nilai *k* optimum dengan metode Elbow, dilakukan *clustering* dengan *k*=8.

Gambar 3.5 menunjukkan visualisasi *cluster* yang terbentuk. Terdapat 8 buah *cluster*.

Table 3.4 Cluster Size k=8

	C1	C2	C3	C4	C5	C6	C7	C8
Size	18	18	62	345	95	77	71	76

Tabel 3.4 menunjukkan jumlah anggota *cluster* pada *cluster* 1 sebanyak 18, *cluster* 2 sebanyak 18, *cluster* 3 sebanyak 62, *cluster* 4 sebanyak 345, *cluster* 5 sebanyak 95, *cluster* 6 sebanyak 77, *cluster* 7 sebanyak 71, dan *cluster* 8 sebanyak 76 anggota.

3.5.2. Association Rule

Setelah melakukan *clustering*, dilakukan asosiasi di setiap *cluster* yang terbentuk.

Table 3.5 Hasil asosiasi

Cluster ke-	Rules				
1	lhs	rhs	support	confidence	lift
	[1] {Layanan}	=> {Tentang.Kami}	0.05555556	0.5	2.25
2	[2] {Tentang.Kami}	=> {Kontak.Kami}	0.11111111	0.5	1.50
	[1] {Produk, Tentang.Kami}	=> {Layanan}	0.05555556	1.0000000	9.00
3	[2] {Tentang.Kami}	=> {Layanan}	0.05555556	0.5000000	4.50
	[3] {Layanan}	=> {Produk}	0.11111111	1.0000000	4.50
	[4] {Layanan}	=> {Kontak.Kami}	0.05555556	0.5000000	3.00
	[5] {Kontak.Kami}	=> {Produk}	0.11111111	0.6666667	3.00
	[6] {Tentang.Kami}	=> {Produk}	0.05555556	0.5000000	2.25
	[1] {Produk}	=> {Tentang.Kami}	0.08064516	0.5555556	3.131313
4	lhs	rhs	support	confidence	lift
	[1] {Produk, Tentang.Kami}	=> {Layanan}	0.02608696	0.6923077	14.049774
	[2] {Layanan}	=> {Tentang.Kami}	0.04057971	0.8235294	9.165085
5	[3] {Layanan}	=> {Produk}	0.02608696	0.5294118	6.764706
	[1] {Layanan}	=> {Mitra...Klien}	0.13684211	1.0000000	1.032609
	[2] {Layanan}	=> {Berita}	0.13684211	1.0000000	1.010638
6	[3] {Berita}	=> {Home}	0.98947368	1.0000000	1.010638
	[4] {Mitra...Klien}	=> {Berita}	0.96842105	1.0000000	1.010638
	[1] {Produk, Layanan}	=> {Tentang.Kami}	0.06493506	0.8333333	4.583333
7	[2] {Produk}	=> {Tentang.Kami}	0.06493506	0.5000000	2.750000
	[3] {Tentang.Kami}	=> {Layanan}	0.11688312	0.6428571	2.605263
	[4] {Produk}	=> {Layanan}	0.07792208	0.6000000	2.431579
8	lhs	rhs	support	confidence	lift
	[1] {Layanan}	=> {Tentang.Kami}	0.04225352	0.6	8.52
8	lhs	rhs	support	confidence	lift
	[1] {Tentang.Kami, Kontak.Kami}	=> {Layanan}	0.02631579	1	8.444444

Pada tabel 3.5, setiap *cluster* memiliki jumlah *rules* yang berbeda-beda. Nilai *support* dan *confidence* dikalikan 100% untuk mendapatkan hasil dalam bentuk presentase. Lhs menandakan *antecedent* 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.

4. Kesimpulan

Berdasarkan hasil dari pengujian yang telah dilakukan maka dapat disimpulkan hal-hal sebagai berikut:

- Berdasarkan perhitungan nilai SSE terhadap nilai k dengan dilihat melalui metode Elbow, didapatkan nilai k optimum yang didapat adalah 8.
- Dari pengujian yang telah dilakukan, diambil kesimpulan bahwa terdapat satu pola user dalam melakukan akses pada website yang memiliki frekuensi kemunculan paling tinggi :

- Layanan ⇒ Tentang Kami

Pada cluster 1 dengan *confidence* 60%, cluster 4 dengan *confidence* 82%, dan cluster 7 dengan *confidence* 60%.

Hal tersebut dapat menandakan bahwa mayoritas pengunjung cenderung melakukan akses dari kategori halaman layanan menuju kategori halaman tentang kami. Artinya, pengunjung *website* cenderung membutuhkan informasi mengenai Owlexa beserta layanannya. Sedangkan kategori halaman yang tidak pernah diakses adalah provider kesehatan.

- Dari pola yang didapat, dapat diberikan beberapa saran:

- Karena kecenderungan pengunjung *website* membutuhkan informasi mengenai layanan dan informasi perusahaan, ada baiknya Owlexa menampilkan informasi tambahan guna meningkatkan daya market dan nilai jual kepada pengunjung, seperti kelebihan Owlexa dari kompetitornya, kemudahan-kemudahan yang diberikan, atau sertifikasi yang dimiliki perusahaan
- Melihat masih adanya kategori halaman yang tidak diakses, ada baiknya Owlexa menggunakan URL halaman lain sebagai media promosi dan iklan (*web ads*) agar penyampaian informasi spesifik dapat ditargetkan secara langsung. Misalkan pada halaman provider kesehatan. Contoh: "Owlexa telah tersedia di kotamu! Klik disini untuk info lebih lanjut."
- Penggabungan kategori provider kesehatan pada kategori mitra & klien, pada sub-menu mitra kami. Dikarenakan informasi yang disampaikan sama. Atau,
- Karena halaman provider kesehatan berisi daftar klinik dan rumah sakit di beberapa kota yang telah bekerjasama dengan Owlexa, sebaiknya halaman tersebut dipindahkan dan diberi *section* sendiri pada halaman utama atau pada seluruh akhir halaman dengan judul "Cek Layanan Owlexa di Kotamu" berupa peta wilayah Indonesia dengan titik-titik dimana Owlexa dapat dilayani. Jika diklik, akan merujuk ke halaman provider kesehatan yang menampilkan daftar klinik dan rumah sakit yang telah *filter* oleh kota yang dipilih.

5. Saran

Berdasarkan apa yang telah dilakukan pada Tugas Akhir ini, maka saran yang dapat disampaikan adalah:

1. Untuk penelitian serupa selanjutnya agar menggunakan subjek penelitian pada *website e-commerce* yang terdapat fungsionalitas transaksi pembelian dan pembayaran di dalamnya.
2. Tahap *preprocessing* merupakan tahap yang paling penting dan riskan. Pastikan telah dilakukan dengan baik dan seksama, seperti penghapusan data yang benar-benar tidak dibutuhkan. Karena akan sangat berpengaruh pada tahap-tahap berikutnya dan mempengaruhi hasil.
3. Untuk penelitian lebih lanjut, ada baiknya menggunakan metode *Improved k-means* dan *K-Apriori*.

6. Daftar Pustaka

- [1] R. P. Vlosky and T. J. Westbrook, "The State of Forest Products Industry e-Business," Louisiana, 2001.
- [2] N. Leinbach-Reyhle, "3 Reasons Websites Are Vital for Small Businesses," *Forbes*, 29 September 2014. [Online]. Available: <https://www.forbes.com/sites/nicoleleinbachreyhle/2014/09/29/websites-for-small-businesses/>.
- [3] S. Patel, "SEO 101: How Your Website's Structure Affects its SEO," *Search Engine Journal*, 17 December 2012. [Online]. Available: <https://www.searchenginejournal.com/website-structure-and-seo/54156/>.
- [4] C. Cumbrowski, "50 Questions to Evaluate the Quality of Your Website," *Search Engine Journal*, 20 February 2008. [Online]. Available: <https://www.searchenginejournal.com/50-questions-to-evaluate-the-quality-of-your-website/>.
- [5] A. K. Sharma and P. C. Gupta, "Study & Analysis of Web Content Mining Tools to Improve Techniques of Web Data Mining," *IJAR CET*, vol. 1, no. 8, 2012.
- [6] J. B. MacQueen, "Some Methods for classification and Analysis of Multivariate Observations," in *Proceedings of 5-th Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley, 1967.
- [7] M. S. C, "Data Mining," *Paper Individu Topik-Topik lanjutan Sistem Informasi*, 2014.
- [8] R. Agrawal and R. Srikant, "Fast Algorithm for Mining Association Rules," in *Proceeding of the 20th VLDB Conference*, Santiago, 1994.
- [9] F. Johnson and S. K. Gupta, "Web Content Mining Techniques: A Survey," *International Journal of Computer Applications*, 2012.
- [10] F. Aini, "Web Usage Mining Menggunakan Algoritma Adaptive Web Access Pattern Tree (AWAPT)," 2011.
- [11] K. Poongothai, M. Parimala and S. Sathiyabama, "Efficient Web Usage Mining with Clustering," *IJCSI International Journal of Computer Science Issues*, vol. 8, no. 6, pp. 203-209, 2011.
- [12] H. Rana and M. Patel, "A Study of Web Log Analysis Using Clustering Techniques," *International Journal of Innovative Research in Computer and Communication Engineering*, vol. 1, no. 4, pp. 925-929, 2013.
- [13] T. R. Aziz, "Analisis dan Implementasi Algoritma K-Means dan LVQ (Learning Vector Quantization) pada Web Usage Mining (Studi Kasus : Tuneeca Online Store)," 2014.
- [14] J. Srivastava, R. Cooley, M. Deshpande and P.-N. Tan, "Web Usage Mining: Discovery and Applications of Usage Patterns from Web Data," 2000.
- [15] K. R. Suneetha and D. R. Krishnamoorthi, "Identifying User Behavior by Analyzing Web Server Access Log File," *IJCSNS*, 2009.
- [16] P. R. Makwana and T. M. Kodinariya, "Review on determining number of Cluster in," *International Journal of Advance Research in*, vol. 1, no. 6, pp. 90-95, 2013.
- [17] P. Fränti, "Cluster Validation," *Speech and Image Processing Unit School of Computing University of Eastern Finland, Eastern Finland*, 2014.
- [18] Oracle, "Data Mining Concepts," May 2008. [Online]. Available: https://docs.oracle.com/cd/B28359_01/datamine.111/b28129/market_basket.htm#DMCON009.
- [19] M. Fauzy, K. R. Saleh and I. Asror, "Penerapan Metode Association Rule Menggunakan Algoritma Apriori Pada Simulasi Prediksi Hujan Wilayah Kota Bandung," *Jurnal Ilmiah Teknologi Informasi Terapan*, vol. 2, no. 2, pp. 221-227, 2016.
- [20] G. Collins, "How can I interpret the formula of lift ratio in association rule?," Quora, 24 Juny 2016. [Online]. Available: <https://www.quora.com/How-can-I-interpret-the-formula-of-lift-ratio-in-association-rule>.
- [21] O. Healthcare, "Owlexa sebagai Unit Bisnis," Owlexa Healthcare, 2015. [Online]. Available: <https://www.owlexa.com/about-us/owlexa-as-a-business-unit.html>.
- [22] K. Homik, "R FAQ - 2.1 What is R?," CRAN-R Project, 6 June 2016. [Online]. Available: https://cran.r-project.org/doc/FAQ/R-FAQ.html#What-is-R_003f.