

Web Usage Mining Guna Analisis Pola Akses Pengunjung Website dengan Association Rule

Hopi Siti Hopipah¹, Jajam Haerul Jaman², Ultach Enri³

¹Universitas Singaperbangsa Karawang, hopi.siti17022@student.unsika.ac.id, Jl. HS Ronggowaluyo Telukjambe Timur, Karawang, Indonesia

²Universitas Singaperbangsa Karawang, jajam.haeruljaman@staff.unsika.ac.id, Jl. HS Ronggowaluyo Telukjambe Timur, Karawang, Indonesia

³Universitas Singaperbangsa Karawang, ultach@staff.unsika.ac.id, Jl. HS Ronggowaluyo Telukjambe Timur, Karawang, Indonesia

Informasi Makalah

Submit : September 2, 2021
Revisi : November 25, 2021
Diterima : Desember 12, 2021

Kata Kunci :

Apriori
Association Rule
Hashing
Modified Apriori
Web Usage Mining

Abstrak

Pengaruh internet erat dengan kehidupan masyarakat, terutama dalam menyediakan kemudahan akses informasi melalui *website*. *Website* digunakan oleh lembaga pendidikan khususnya kampus sebagai media promosi, media informasi, publikasi, dan pengenalan profil kampus. Pemanfaatan *website* secara optimal dapat memberikan pelayanan terbaik bagi pengunjung, sehingga kepercayaan dan citra positif terhadap kampus pun dapat meningkat. Penting bagi pengelola untuk memperhatikan juga meningkatkan kualitas *website*, salah satunya dengan menerapkan *web usage mining*. *Web usage mining* bermanfaat untuk menggali informasi yang didapatkan dari *web*, dengan memahami data aktivitas pengunjung agar dapat mengetahui kelebihan dan kekurangan *website*. Penelitian ini bertujuan untuk mengetahui serta melakukan analisis pola akses pengunjung *website* Unsika dengan *web usage mining* menggunakan *Association Rule*. Algoritma yang digunakan adalah *Modified Apriori* dengan teknik *hashing*. Teknik *hashing* digunakan untuk mengurangi waktu pencarian dengan menyimpan data ke dalam *array* sebagai *key* dan *value* pada saat proses iterasi. Berdasarkan hasil penelitian, dengan nilai minimum *support* 2 dan minimum *confidence* 65%, *rule* yang terbentuk yaitu sebanyak 27 dengan nilai *support* tertinggi 2.20%, nilai *confidence* tertinggi adalah 100%, dan *lift ratio* tertinggi sebesar 91.

Abstract

The influence of the internet is closely related to people's lives, especially in providing easy access to information through the website. The website is used by educational institutions, especially campuses as a promotional media, information media, publications, and introduction of campus profiles. Optimal use of the website can provide the best service for visitors, so that trust and a positive image of the campus can also be increased. It is important for developers to pay attention

to also improve the quality of the website, one of which is by implementing web usage mining. Web usage mining is useful to dig up information obtained from the web, by understanding the data of visitor activity in order to know the advantages and disadvantages of the website. This research aims to find out and analyze the access patterns of unsika website visitors with web usage mining using association rules. The algorithm used is Modified Apriori with hashing technique. Hashing techniques are used to break the search time by storing data into an array as keys and values during the iteration process. Based on the results of the study, with a minimum value of support 2 and a minimum confidence of 65%, the rule formed as much as 27 with the highest support value is 2.20%, the highest confidence value is 100%, and the highest lift ratio is 91.

1. Pendahuluan

Website menjadi salah satu media yang digunakan untuk mengakses informasi. Menurut Setia et al., (2020), *website* telah menjadi tempat penting bagi orang untuk berbagi informasi. Salah satu lembaga yang memanfaatkan *website* adalah kampus. Berdasarkan artikel di www.kompasiana.com ada beberapa keuntungan kampus jika memiliki *website*, diantaranya sebagai sarana promosi, memudahkan kampus untuk berkomunikasi dengan mahasiswanya, media publikasi kegiatan, penelitian, maupun kejuaraan, serta media pengenalan profil resmi kampus (Dewi, 2020).

Universitas Singaperbangsa Karawang (Unsika) merupakan kampus yang telah memanfaatkan *website*. *Website* Unsika (unsika.ac.id) dapat diakses oleh berbagai pihak, baik sivitas akademika maupun masyarakat umum. Pemanfaatan *website* secara optimal dapat meningkatkan kepercayaan masyarakat dan mempengaruhi citra positif kampus (Apriananta & Wijaya, 2018), sehingga penting bagi pengelola agar senantiasa memperhatikan kualitas *website*, salah satunya dengan menggunakan *web server logs* (data kunjungan *web*).

Objek penelitian yang akan digunakan berasal dari *web server logs website unsika.ac.id*. Berdasarkan data kunjungan *web* dari 31 Desember 2020 hingga 14 Februari 2021, rata-rata aktivitas mingguan pada *website* Unsika adalah 3646 *request*, dengan total jumlah aktivitas sebanyak 25.527 *request*. Sejauh ini belum diketahui

bagaimana aturan asosiatif pengaksesan pengunjung pada halaman *website unsika.ac.id*. Teknik *web usage mining* dapat diterapkan untuk mengatasi permasalahan tersebut.

Web usage mining bermanfaat untuk mengetahui pola akses pengunjung berdasarkan *web server logs* (Kurniawati et al., 2020). *Web usage mining* digunakan dalam menggali informasi yang didapatkan dari *web*, dengan memahami data pengunjung untuk mengetahui kelebihan dan kekurangan *website* (Harika & Sudha, 2020).

Association rule merupakan salah satu metode untuk mencari pola yang sering muncul dari suatu transaksi (Masnur, 2015). *Association rule* dapat digunakan dalam menemukan aturan asosiatif dari halaman *web* yang sering diakses pengguna. Teknik ini dapat membantu *website* dalam mengorganisir kontennya secara efisien (Pramana, 2019). Algoritma yang paling berpengaruh dalam menemukan aturan yang berguna dari *association rule* adalah Apriori (Liu & Wang, 2020). Namun, algoritma Apriori masih memiliki kelemahan dalam melakukan iterasi yang berulang kali pada proses pemindaian *database*, terutama dalam dataset yang berjumlah banyak. Sehingga konsumsi waktu dan memori yang dibutuhkan pun semakin banyak (Anggraeni et al., 2019). Maka dari itu, Apriori banyak dijadikan dasar pengembangan oleh peneliti untuk mengatasi kelemahan dan meningkatkan performanya (Syahrir & Fatimatuzzahra, 2020).

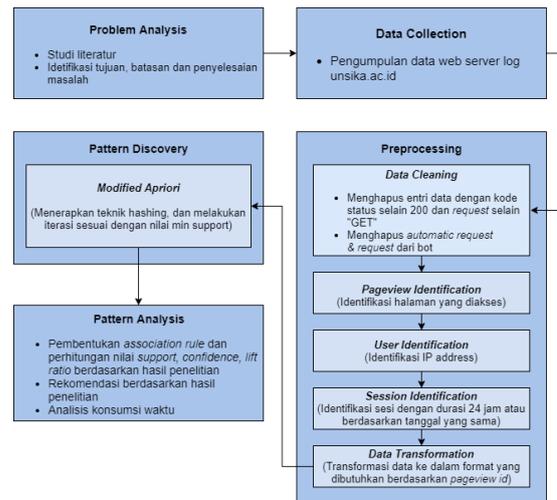
Penelitian sebelumnya dilakukan Anggraeni et al., (2019) dengan menerapkan *web usage mining* menggunakan algoritma Modifikasi Apriori (*Tree-Based Approach*), yang tujuannya untuk menemukan pola navigasi pengunjung *website* UPN “Veteran” Jawa Timur. Algoritma tersebut menggunakan teknik *graph* dan kombinasi *tree* dalam melakukan seleksi *itemset* pada tiap node. Hasilnya terdapat dua rules dengan masing-masing nilai *confidence* 74% dan 93% dan waktu eksekusi yang dibutuhkan yaitu 1.26 detik.

Adapun Kaliappan et al., (2019) menerapkan algoritma *Modified Apriori* dengan teknik *hashing*. Teknik *hashing* ini digunakan dalam menghitung frekuensi setiap *item* pada dataset, dan menghapus data yang tidak relevan untuk mengosongkan ruang sehingga dapat meningkatkan kecepatan eksekusi. *Modified Apriori* ini berhasil mengurangi konsumsi waktu sebesar 7.896 detik, dengan nilai minimum *support* 10% pada jumlah data sebanyak 100.000 record. Bahkan *Modified Apriori* 89.4% lebih efisien dibandingkan dengan algoritma Apriori.

Berdasarkan permasalahan dan pemaparan penelitian sebelumnya, maka penulis melakukan penelitian dengan judul “*Web Usage Mining* Guna Analisis Pola Akses Pengunjung *Website* dengan *Association Rule*”, dengan menerapkan algoritma *Modified Apriori* menggunakan teknik *hashing* pada data *web server logs* unsika.ac.id, sehingga dapat lebih efisien dalam menemukan pola akses pengunjung *website*.

2. Metode Penelitian

Terdapat lima tahapan metode yang digunakan dalam penelitian ini, diantaranya yaitu *Problem Analysis*, *Data Collection*, *Preprocessing*, *Pattern Discovery*, dan *Pattern Analysis*. Adapun tahapan metode terdapat pada Gambar 1.



Gambar 1. Diagram alir penelitian

2.1 Problem Analysis

Website unsika.ac.id merupakan *website* utama dari Universitas Singaperbangsa Karawang yang berfungsi sebagai media informasi, publikasi, pengenalan profil, serta media promosi, yang dikelola oleh Unit Pelaksana Teknis Teknologi Informasi dan Komunikasi (UPT TIK) Unsika. Dalam memaksimalkan pelayanan bagi pengunjung *website*, perlu dilakukan peningkatan dari berbagai aspek, salah satunya dengan memanfaatkan data kunjungan *web* (*web server log*).

Maka pada penelitian ini dilakukan penerapan *web usage mining* menggunakan *association rule* dan algoritma *Modified Apriori* dengan teknik *hashing* untuk mengetahui pola akses pengunjung *website* Unsika. Hasil dari penelitian dapat digunakan sebagai rekomendasi atau saran yang membantu menunjang keputusan *developer* dalam meningkatkan kualitas *website* sehingga fungsionalitas *web* dapat berjalan dengan lebih maksimal.

2.2 Data collection

Pada tahap *data collection* ini penulis melakukan permohonan penggunaan *web log file* unsika.ac.id sebagai objek penelitian kepada pihak pengelola. Data dikumpulkan melalui akses direktori *apache access log*. Data awal berekstensi .gz (GZip file) yang

kemudian dibaca menggunakan *tools* http Logs Viewer. Hanya 4 atribut yang digunakan dari data *web log file* tersebut, diantaranya terdapat pada Tabel 1.

Tabel 1. Deskripsi atribut yang digunakan

Atribut	Deskripsi
<i>IP Address</i>	Alamat protokol internet web (IP) yang menjadi ID pengunjung
<i>Date</i>	Tanggal dan waktu saat request terjadi
<i>Request</i>	Baris request dari pengunjung yang terdiri dari jenis <i>request</i> , URL, dan protokol
<i>Status</i>	Kode status HTTP yang dikembalikan ke pengunjung.

2.3 Preprocessing

Tahap *preprocessing* bertujuan untuk mempersiapkan data agar sesuai dengan kebutuhan penerapan algoritma pada tahap selanjutnya. Proses pada tahap *preprocessing* disesuaikan dengan keadaan data sampai memenuhi kebutuhan penelitian. *Preprocessing* terdiri dari beberapa tahap, yaitu:

A. Data Cleaning

Data cleaning bertujuan untuk membersihkan data dari *request* yang tidak relevan dengan penelitian seperti *file* gambar (.png, .jpg, .jpeg), *file text* (.txt), *file javascript* (.js), dan lainnya, serta pembersihan data dari bot atau *nonhuman access*.

Selain itu, pada *data cleaning* juga dilakukan penghapusan entri data dengan kode status selain 200 dan permintaan selain “GET” (Raphaeli et al., 2017). Kode status 200 artinya halaman *website* yang diminta (*request*) *client* berhasil dijalankan oleh *server*. Metode “GET” akan mempermudah dalam pembacaan halaman *website* karena nilai form dapat dilihat langsung dalam URL, metode “GET” ini berfungsi untuk menampilkan atau mengambil data dari *database*.

Pada data *web server log* unsika.ac.id, sebanyak 40.76% data tidak sesuai dengan kriteria seperti kode status selain 200 dan

request selain “GET”. Selain itu, sebanyak 57.91% data tidak relevan dengan kebutuhan penelitian, misal pada *request* “GET / HTTP/1.1” yang tidak memiliki halaman tujuan yang jelas, serta *request* dengan *file* HTML (.css), *file* JavaScript (.js), *file text* (.txt), gambar (.jpg, .jpeg, .png) dan lainnya, termasuk data yang terdeteksi bot.

B. Pageview Identification

Identifikasi *pageview* dilakukan untuk mengetahui setiap halaman yang muncul ketika pengunjung mengklik tautan tertentu. Halaman dikategorikan berdasarkan *request* menu yang diakses pengunjung. Setiap halaman akan diidentifikasi dengan id unik untuk mempermudah proses *data transformation*. Terdapat 11 kategori yang dibuat mulai dari Akademik, Beasiswa, Beranda, Berita, Kegiatan, Kemahasiswaan, Kerjasama, Penelitian dan Inovasi, Pengumuman, Prestasi, dan Tentang.

C. User Identification

Untuk melakukan identifikasi user, setiap aktivitas disegmentasi berdasarkan atribut *IP Address* yang digunakan. *IP Address* diubah menjadi id unik setiap *user*. Hasilnya terdapat sebanyak 100 *user* setelah proses identifikasi.

D. Session Identification

Proses segmentasi aktivitas setiap pengguna dibedakan ke dalam sesi, masing-masing mewakili satu kunjungan ke situs. Dalam satu sesi kunjungan, tentukan halaman yang di-*request*, urutan *request*, dan durasi aksesnya.

Pengunjung mungkin memiliki satu atau beberapa sesi selama periode tertentu. Pada *session identification* ini setiap aktivitas disegmentasikan ke dalam sesi berdasarkan alamat IP dengan durasi waktu sesi 24 jam, atau bisa dikatakan bahwa setiap sesi akan memiliki tanggal akses yang berbeda. Dari sesi tersebut dapat diketahui rangkaian lengkap aktivitas yang dilakukan oleh pengunjung dalam waktu tertentu. Total sesi

setelah proses identifikasi berjumlah 182 transaksi *user*.

E. Data Transformation

Untuk memudahkan proses mining, identifikasi dilakukan berdasarkan *IP address*, ataupun *page item* yang dikunjungi. *Item pageview* dari *user* yang sama dikelompokkan menjadi satu, lalu diubah menjadi format data yang sesuai dengan kebutuhan penelitian.

2.4 Pattern Discovery

Pada penelitian ini, dilakukan penerapan teknik *web usage mining* dengan *association rule* menggunakan algoritma *Modified Apriori*, berupa pola-pola akses pengunjung yang memenuhi nilai minimum *support* dan *confidence*.

2.5 Pattern Analysis

Pattern analysis merupakan pembahasan hasil analisis pola akses pengunjung berdasarkan hasil penerapan *association rule* dengan algoritma *Modified Apriori*. Pada tahap ini ditentukan mana pola yang paling relevan dengan tujuan penelitian, lalu dilakukan perhitungan nilai *support*, *confidence*, *lift ratio* serta konsumsi waktu yang digunakan.

A. Support

Nilai *support* dari suatu *association rule* merupakan persentase kombinasi *item* tersebut dalam *database* (Desyanti & Sari, 2019). Misalkan *n* adalah jumlah total transaksi. *Support X* diwakili sebagai berapa kali *X* muncul dalam *database* dibagi dengan *n*, terdapat pada Persamaan 1.

$$\text{Support}(X) = \frac{\text{Jumlah transaksi mengandung } X}{\text{Total transaksi}} \quad (1)$$

Sementara *support* untuk *X* dan *Y* bersama-sama diwakili sebagai berapa kali mereka muncul bersama-sama dibagi oleh *n*, seperti pada Persamaan 2.

$$\text{Support}(X, Y) = \frac{\text{Jumlah Transaksi mengandung } X \text{ dan } Y}{\text{Total Transaksi}} \quad (2)$$

Maka, *support X* adalah probabilitas *X* sedangkan *support X,Y* adalah probabilitas $X \cap Y$.

B. Confidence

Support dan *confidence* adalah metrik penting untuk menilai kualitas *association rule mining*. Semakin tinggi tingkat *confidence*, artinya menunjukkan bahwa aturan tersebut benar cukup sering dan baik digunakan untuk keputusan. Nilai *confidence* dapat dihitung menggunakan Persamaan 3.

$$\text{Conf.}(X, Y) = \frac{\text{Jumlah transaksi mengandung } X \text{ dan } Y}{\text{Jumlah transaksi mengandung } X} \quad (3)$$

C. Lift Ratio

Nilai *lift ratio* biasanya digunakan untuk validitas dari aturan asosiasi (Fauzy et al., 2016). Untuk mengetahui nilai *lift* terdapat pada Persamaan 4.

$$\text{Lift Ratio} = \frac{\text{Confidence}(X, Y)}{\text{Benchmark Confidence}(Y)} \quad (4)$$

Sedangkan untuk mencari nilai *benchmark confidence* terdapat pada Persamaan 5.

$$\text{Benchmark Confidence} = \frac{\text{Support Count}(Y)}{\text{Total Transaksi}} \quad (5)$$

3. Hasil dan Pembahasan

Data *web server logs* yang digunakan memiliki rentang waktu 31 Desember 2020 sampai dengan 14 Februari 2021. Data awal berjumlah 25.527 *request* data, namun setelah diolah melalui tahap *data cleaning* dan proses seleksi, jumlah *dataset* menjadi 337 data. Hal tersebut disebabkan 98.67% data tidak relevan dan perlu disingkirkan sehingga tersisa 337 data atau sekitar 1.32% data yang siap untuk diproses pada tahap selanjutnya. Hasil ini sejalan dengan penelitian U. M. Patil (2020), sejumlah 29.605 dari 47.748 data atau sekitar 62% *records* data-nya dihapus setelah *preprocessing*, karena banyaknya data yang tidak relevan dengan kebutuhan penelitian.

A. Penerapan Algoritma

Perhitungan *association rule* dimulai dengan membuat tabel *candidate 1-itemset* (C1) lalu tentukan nilai *support count*. Terdapat 11 jenis halaman sesuai dengan kategori halaman yang sudah dibentuk pada proses *Pageview Identification*. Nilai *support count* menunjukkan jumlah total berapa kali halaman tersebut diakses oleh pengunjung. Berdasarkan hasil pencarian, halaman Beranda paling banyak diakses pengunjung yaitu sebanyak 136 kali kunjungan. Berikut hasil pencarian C1, terdapat pada Tabel 2.

Tabel 2. Tabel C1 (Kandidat 1-itemset)

<i>Itemset</i>	<i>Pageview id</i>	<i>Support count</i>
Akademik	I1	5
Beasiswa	I2	4
Beranda	I3	136
Berita	I4	15
Kegiatan	I5	1
Kemahasiswaan	I6	1
Kerjasama	I7	2
Penelitian dan Inovasi	I8	1
Pengumuman	I9	29
Prestasi	I10	3
Tentang	I11	6

Setelah tabel C1 terbentuk, ditentukan nilai *minimum support* yaitu 2, yang berarti kunjungan yang muncul kurang dari 2 akan dipangkas (*prune*) sedangkan kunjungan yang muncul lebih dari sama dengan 2 akan dimasukkan ke dalam tabel *Large 1-itemset* (L1). Berdasarkan tabel C1, halaman “Kegiatan”, “Kemahasiswaan”, serta halaman “Penelitian dan Inovasi” hanya diakses oleh masing-masing 1 pengunjung. Maka tiga halaman tersebut akan dipangkas karena kurang dari nilai *minimum support*, sedangkan 8 *item* halaman lainnya akan masuk ke dalam tabel L1, seperti yang terdapat pada Tabel 3.

Tabel 3. Hasil tabel large 1-itemset (L1)

<i>Itemset</i>	<i>Pageview id</i>	<i>Support count</i>
Akademik	I1	5

Beasiswa	I2	4
Beranda	I3	136
Berita	I4	15
Kerjasama	I7	2
Pengumuman	I9	29
Prestasi	I10	3
Tentang	I11	6

Selanjutnya yaitu mencari kandidat *frequent 2-itemset*, yang merupakan hasil penggabungan (*selfjoin*) antara L1 dengan L1. Berdasarkan hasil penggabungan, terdapat 28 *itemset* halaman yang terbentuk, dengan *support count* masing-masing ditentukan untuk mempermudah proses pemangkasan (*prune*). *Itemset* dengan *support count* terbanyak adalah “Berita, Pengumuman” dan “Pengumuman, Tentang” yang masing-masing muncul sebanyak 4 kali. 2 *itemset* muncul sebanyak 3 kali, yaitu “Akademik, Beranda” dan “Beranda, Berita”, sebanyak 6 *itemset* muncul 2 kali, dan sisanya sebanyak 18 *itemset* hanya muncul 1 kali atau bahkan tidak ada sama sekali. Hasil dari tabel C2 ditunjukkan pada Tabel 4.

Tabel 4. Hasil tabel C2 dengan *frequent 2-itemset*

<i>Itemset</i>	<i>Pageview id</i>	<i>Support count</i>
Akademik, Beasiswa	I1,I2	1
Akademik, Beranda	I1,I3	3
Akademik, Berita	I1,I4	1
Akademik, Kerjasama	I1,I7	0
Akademik, Pengumuman	I1,I9	2
Akademik, Prestasi	I1,I10	1
Akademik, Tentang	I1,I11	1
Beasiswa, Beranda	I2,I3	1
Beasiswa, Berita	I2,I4	1
Beasiswa, Kerjasama	I2,I7	0
Beasiswa, Pengumuman	I2,I9	1
Beasiswa, Prestasi	I2,I10	1
Beasiswa, Tentang	I2,I11	1
Beranda, Berita	I3,I4	3
Beranda, Kerjasama	I3,I7	0
Beranda, Pengumuman	I3,I9	2
Beranda, Prestasi	I3,I10	1
Beranda, Tentang	I3,I11	1
Berita, Kerjasama	I4,I7	0

Berita, Pengumuman	I4,I9	4
Berita, Prestasi	I4,I10	2
Berita, Tentang	I4,I11	2
Kerjasama, Pengumuman	I7,I9	0
Kerjasama, Prestasi	I7,I10	0
Kerjasama, Tentang	I7,I11	0
Pengumuman, Prestasi	I9,I10	2
Pengumuman, Tentang	I9,I11	4
Prestasi, Tentang	I10,I11	2

Setelah pembentukan tabel C2, dilakukan kembali pemangkasan *item* yang tidak memenuhi nilai *minimum support*. Hasilnya terdapat 10 kunjungan yang lebih dari sama dengan nilai *minimum support* (≥ 2). Hasil tersebut akan dimasukkan ke tabel *Large 2-itemset* atau L2 yang ditunjukkan pada Tabel 5.

Tabel 5. Hasil dari tabel *large 2-itemset* (L2)

<i>Itemset</i>	<i>Pageview id</i>	<i>Support count</i>
Akademik, Beranda	I1,I3	3
Akademik, Pengumuman	I1,I9	2
Beranda, Berita	I3,I4	3
Beranda, Pengumuman	I3,I9	2
Berita, Pengumuman	I4,I9	4
Berita, Prestasi	I4,I10	2
Berita, Tentang	I4,I11	2
Pengumuman, Prestasi	I9,I10	2
Pengumuman, Tentang	I9,I11	4
Prestasi, Tentang	I10,I11	2

Setelah itu tabel C3 yang merupakan kandidat *frequent 3-itemset* dibentuk dari hasil penggabungan tabel L2 dengan L2, dengan mencari semua kemungkinan *3-itemset* yang dapat terbentuk. Tabel 6 berikut merupakan tabel C3 beserta *support count* nya.

Tabel 6. Hasil dari tabel *frequent 3-itemset* (C3)

<i>Itemset</i>	<i>Pageview id</i>	<i>Support Count</i>
Akademik, Beranda, Pengumuman	I1,I3,I9	2

Akademik, Pengumuman, Prestasi	I1,I9,I10	1
Akademik, Pengumuman, Tentang	I1,I9,I11	1
Beranda, Berita, Pengumuman	I3,I4,I9	1
Beranda, Berita, Prestasi	I3,I4,I10	1
Beranda, Berita, Tentang	I3,I4,I11	1
Beranda, Pengumuman, Prestasi	I3,I9,I10	1
Beranda, Pengumuman, Tentang	I3,I9,I11	1
Berita, Pengumuman, Prestasi	I4,I9,I10	2
Berita, Pengumuman, Tentang	I4,I9,I11	2
Berita, Prestasi, Tentang	I4,I10,I11	2
Pengumuman, Prestasi, Tentang	I9,I10,I11	2

Untuk mendapatkan tabel L3, dilakukan kembali pemangkasan (*pruning*) pada data yang tidak memenuhi nilai *minimum support*. Tabel 7 berikut merupakan hasil eliminasi data yang memenuhi *support count* ≥ 2 .

Tabel 7. Hasil dari tabel *large 3-itemset* (L3)

<i>Itemset</i>	<i>Pageview id</i>	<i>Support Count</i>
Akademik, Beranda, Pengumuman	I1,I3,I9	2
Berita, Pengumuman, Prestasi	I4,I9,I10	2
Berita, Pengumuman, Tentang	I4,I9,I11	2
Berita, Prestasi, Tentang	I4,I10,I11	2
Pengumuman, Prestasi, Tentang	I9,I10,I11	2

Setelah proses 3 kali iterasi, tersisa 5 data dengan *3-itemset* yang memenuhi nilai *minimum support* ≥ 2 seperti pada Tabel 7. Namun dari tabel tersebut masih terdapat kemungkinan terbentuknya tabel C4 dengan *frequent 4-itemset*. Hasil pembentukan *rule* dengan *frequent 4-itemset* tertera pada Tabel 8.

Tabel 8. Hasil dari tabel C4 sekaligus L4

Itemset	Pageview Id	Support Count
Berita, Pengumuman, Prestasi, Tentang	I4,I9,I10, I11	2

Berdasarkan Tabel 8, hanya terdapat 1 *itemset* yang memenuhi nilai *support count* sehingga *itemset* tersebut sekaligus menjadi tabel L4.

B. Pembentukan *association rule* dan perhitungan nilai *support*, *confidence*, dan *lift ratio*

Berdasarkan *itemset* yang terbentuk, dapat ditentukan kandidat *rule* dengan menggunakan *antecedent* atau *consequent* dari tabel L2, L3 dan L4. Setelah itu dilakukan perhitungan nilai *support* dan nilai *confidence* dari setiap *rule* yang mungkin terbentuk. Kandidat *rule* yang memenuhi syarat nilai minimum *support* 2, yaitu $\geq 0,01$ atau 1,10% dan nilai minimum *confidence* 65% akan menjadi sebuah *rule* asosiasi. Berikut contoh cara menentukan nilai *support* dan *confidence* dari masing-masing *rule*.

Support {Prestasi → Berita}

$$= \frac{\text{Support Count Prestasi dan Berita}}{\text{Total Transaksi}} \times 100\%$$

$$= \frac{2}{182} \times 100\% = 1.10\%$$

Confidence {Prestasi → Berita}

$$= \frac{\text{Support count Prestasi dan Berita}}{\text{Support Count Prestasi}} \times 100\%$$

$$= \frac{2}{3} \times 100\% = 66.67\%$$

Tabel 9. Hasil seluruh *rule* beserta *support*, *confidence* dan *lift ratio*

Pageview	Support	Confidence	Benchmark	Lift Ratio
If access Prestasi Then access Berita	1,10%	66,67%	0,08	8,09
If access Prestasi Then access Pengumuman	1,10%	66,67%	0,16	4,18
If access Tentang Then access Pengumuman	2,20%	66,67%	0,16	4,18
If access Prestasi Then access Tentang	1,10%	66,67%	0,03	20,22
If access Akademik, Beranda Then access Pengumuman	1,10%	66,67%	0,16	4,18
If access Akademik, Pengumuman Then access Beranda	1,10%	100,00%	0,75	1,34
If access Beranda, Pengumuman Then access Akademik	1,10%	100,00%	0,03	36,40
If access Prestasi Then access Berita, Pengumuman	1,10%	66,67%	0,03	24,27

Setelah proses eliminasi *rule* yang memenuhi nilai minimum *confidence* 65%, maka tentukan nilai *lift ratio* untuk mengetahui seberapa kuat ikatan dari *rule* yang sudah terbentuk. Untuk penentuan nilai *lift ratio*, terdapat contoh sebagai berikut.

Lift Ratio {Prestasi → Berita}

$$= \frac{\text{Confidence \{Prestasi \rightarrow Berita\}}}{\text{Benchmark Confidence \{Prestasi \rightarrow Berita\}}}$$

Confidence {Prestasi → Berita}

$$= \frac{\text{Support count Prestasi dan Berita}}{\text{Support Count Prestasi}} = \frac{2}{3} = 0.67$$

Benchmark Confidence {Prestasi → Berita}

$$= \frac{\text{Support count Berita}}{\text{Total Transaksi}} = \frac{15}{182} = 0.08$$

Lift Ratio {Prestasi → Berita}

$$= \frac{0.67}{0.08} = 8.09$$

Rule yang digunakan adalah if X then Y, dimana X disini sebagai *antecedent* dan Y sebagai *consequent*. Misalnya, untuk *rule* “Prestasi → Berita”, maka pola yang didapatkan yaitu “Jika *user* mengakses Prestasi, maka akan mengakses Berita dengan *support* 1.10% dan *confidence* 66.67%”. Berdasarkan hasil perhitungan nilai *lift ratio* pun menunjukkan bahwa semua *rule* yang terbentuk menghasilkan nilai lebih dari 1, yang artinya *rule* yang dihasilkan memiliki kekuatan yang bagus. *Rule* pola lainnya yang didapatkan tertera pada Tabel 9.

If access Berita, Prestasi Then access Pengumuman	1,10%	100,00%	0,16	6,28
If access Pengumuman, Prestasi Then access Berita	1,10%	100,00%	0,08	12,13
If access Berita, Tentang Then access Pengumuman	1,10%	100,00%	0,16	6,28
If access Prestasi Then access Berita, Tentang	1,10%	66,67%	0,01	60,67
If access Berita, Prestasi Then access Tentang	1,10%	100,00%	0,03	30,33
If access Berita, Tentang Then access Prestasi	1,10%	100,00%	0,02	60,67
If access Prestasi, Tentang Then access Berita	1,10%	100,00%	0,08	12,13
If access Prestasi Then access Pengumuman, Tentang	1,10%	66,67%	0,02	30,33
If access Pengumuman, Prestasi Then access Tentang	1,10%	100,00%	0,03	30,33
If access Prestasi, Tentang Then access Pengumuman	1,10%	100,00%	0,16	6,28
If access Berita, Pengumuman, Prestasi Then access Tentang	1,10%	100,00%	0,03	30,33
If access Berita, Pengumuman, Tentang Then access Prestasi	1,10%	100,00%	0,02	60,67
If access Berita, Prestasi, Tentang Then access Pengumuman	1,10%	100,00%	0,16	6,28
If access Pengumuman, Prestasi, Tentang Then access Berita	1,10%	100,00%	0,08	12,13
If access Prestasi Then access Berita, Pengumuman, Tentang	1,10%	66,67%	0,01	60,67
If access Berita, Prestasi Then access Pengumuman, Tentang	1,10%	100,00%	0,02	45,50
If access Berita, Tentang Then access Pengumuman, Prestasi	1,10%	100,00%	0,01	91,00
If access Pengumuman, Prestasi Then access Berita, Tentang	1,10%	100,00%	0,01	91,00
If access Prestasi, Tentang Then access Berita, Pengumuman	1,10%	100,00%	0,03	36,40

Selain perhitungan nilai *support*, *confidence*, dan *lift ratio*, dilakukan juga pengujian konsumsi waktu pada *Modified Apriori*. Pengujian ini dilakukan dengan menginput 3 nilai minimum *support* yang berbeda yaitu 2, 3, dan 4. Nilai minimum *confidence* yang digunakan secara acak adalah 65%. Hasil analisis pengujian terdapat pada Tabel 10.

Tabel 10. Hasil pengujian konsumsi waktu

Minimum <i>Support</i>	Lama eksekusi (dalam detik)
2	0,08
3	0,06
4	0,05

Berdasarkan Tabel 10, rata-rata konsumsi waktu adalah 0.06 detik. Pengujian tersebut menunjukkan bahwa nilai minimum *support* mempengaruhi konsumsi waktu, semakin besar nilai minimum *support*, maka semakin sedikit konsumsi waktu yang diperlukan.

C. Rekomendasi berdasarkan Hasil Penelitian

Berdasarkan hasil penelitian, perlu diperhatikan bagaimana cara agar pengunjung tertarik untuk mengakses menu lain dan bertahan lebih lama di *website*. Sehingga menu lainnya lebih tereksplorasi dan informasi dapat tersampaikan kepada pengunjung. Berikut beberapa rekomendasi yang dapat diterapkan pada *website* Unsika, diantaranya yaitu:

- Menambah konten pada setiap menu dengan lebih *up to date*, runtut dan mudah dibaca.
- Menyediakan fitur personalisasi. Misalnya pada Berita dan Pengumuman, dengan adanya personalisasi berdasarkan tanggal atau kata kunci, pengunjung bisa lebih mudah mendapatkan informasi sesuai dengan kebutuhannya.

- c. Menambah performa atau kecepatan akses *website*. Alternatif yang dapat dilakukan yaitu dengan menentukan gambar atau animasi yang resolusinya tidak terlalu besar, atau dengan melakukan kompresi pada gambar yang akan digunakan. Selain itu, pada pola yang mungkin sering diakses, dapat diterapkan mekanisme *caching web server* atau *proxy*.
- d. Menambahkan menu navigasi lain di suatu menu yang banyak diakses. Misalnya, pada Tabel 9 pola akses dengan nilai *support* 2.20% adalah “Jika mengakses Tentang, maka akan mengakses Pengumuman”. Maka pada halaman Tentang bisa ditambahkan menu navigasi Pengumuman, sehingga ketika pengunjung mengakses Tentang, mereka juga bisa langsung menemukan Pengumuman tanpa perlu kembali ke Beranda terlebih dahulu. Begitu pun pada pola atau *rule* lainnya.

4. Simpulan

Berdasarkan hasil penelitian yang sudah dilakukan, maka dapat diambil kesimpulan sebagai berikut :

1. Penerapan *Association rule* dengan algoritma *Modified Apriori* untuk menentukan pola akses pengunjung *website* Unsika pada nilai minimum *support* = 2 dan *confidence* = 65% dapat membentuk *rule* sebanyak 27 dengan nilai *support* tertinggi 2.20%, pada *rule* “Jika mengakses Tentang, maka mengakses Pengumuman”. Sedangkan *rule* dengan nilai *confidence* 100% salah satunya adalah “Jika mengakses Akademik, Pengumuman, maka mengakses Beranda. Lalu, nilai *lift ratio* tertinggi sebesar 91 terdapat pada *rule* “Jika mengakses Berita dan Tentang, maka mengakses Pengumuman dan Prestasi” dan *rule* “Jika mengakses

- Pengumuman dan Prestasi, maka mengakses Berita dan Tentang”.
2. Berdasarkan hasil penelitian, terdapat beberapa rekomendasi yang dapat diterapkan pada *website* Unsika, diantaranya, menambah konten pada setiap menu, menambahkan fitur personalisasi berdasarkan kebutuhan pengunjung, meningkatkan performa atau kecepatan akses *website*, serta menambahkan menu navigasi terkait pada pola yang mungkin sering diakses. Dalam menentukan pola akses pengunjung *website* Unsika dengan *Association Rule* menggunakan algoritma *Modified Apriori* dapat menghasilkan *rule* yang sama dengan algoritma *Apriori*. Perbedaan proses algoritma *Modified Apriori* dengan algoritma *Apriori* terdapat pada saat pembentukan *itemset* dan penghapusan *rule* yang tidak memenuhi nilai minimum *support* pada *database* di setiap iterasi.

Adapun saran untuk penelitian selanjutnya adalah menambah rentang waktu akses data *web log file* yang akan digunakan, sehingga *itemset* pada setiap transaksi bisa lebih bervariasi dan dapat menggunakan nilai minimum *support* yang lebih tinggi agar nilai *support* yang dihasilkan semakin bagus. Selain itu, penulis juga menyarankan pengujian proses *session identification* dengan metode lain untuk mengetahui pengaruhnya terhadap hasil yang didapatkan.

5. Referensi

- Anggraeni, D., Anggraeny, F. T., & Purbasari, I. Y. (2019). Penerapan Algoritma Modifikasi Apriori (Tree-Based Approach) untuk Menemukan Pola Navigasi Pengguna Website. *SEMINAR SANTIKA, September*, 159–160.
- Apriananta, Y. J., & Wijaya, L. S. (2018).

- Penggunaan Website Dan Media Sosial Dalam Membangun Citra Positif Perguruan Tinggi. *Jurnal Komunikatif*, 7(2), 187–209. <https://doi.org/10.33508/jk.v7i2.1750>
- Desyanti, & Sari, F. (2019). Penerapan Algoritma Apriori Terhadap Perawatan Tubuh di Kakiku. *SATIN - Sains Dan Teknologi Informasi*, 5(1), 51–59. <https://doi.org/10.33372/stn.v5i1.457>
- Dewi, A. R. (2020). Pentingnya Website Kampus. *Kompasiana*. <https://www.kompasiana.com/adityardewi26/5e3cc136d541df567d28b832/pentingnya-website-kampus>
- Fauzy, M., Saleh W, K. R., & Asror, I. (2016). Penerapan Metode Association Rule Menggunakan Algoritma Apriori Pada Simulasi Prediksi Hujan Wilayah Kota Bandung. *Jurnal Ilmiah Teknologi Informasi Terapan*, II(2), 221–227.
- Harika, B., & Sudha, P. T. (2020). Data Extraction from Web Server Logs using Web Usage Mining. *International Journal Of Innovative Research In Technology*, 6(11), 469–474.
- Kaliappan, J., Sai, S. M., & Preetham, K. S. (2019). Weblog and Retail Industries Analysis using a robust modified Apriori algorithm. *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*, 8(6), 1727–1733.
- Kurniawati, G. N., Santoso, R., & Sugito. (2020). Analisis Web Usage Mining Menggunakan Metode Modified Gustafson-Kessel Clustering dan Association Rule pada Website Universitas Diponegoro. *Jurnal Gaussian*, 9(4), 486–494.
- Liu, Z., & Wang, Y. (2020). CMS System Identification Based on Improved Algorithm of Apriori. *Journal of Physics: Conference Series*, 1631(1). <https://doi.org/10.1088/1742-6596/1631/1/012062>
- Masnur, A. (2015). Analisa Data Mining Menggunakan Market Basket Analysis untuk Mengetahui Pola Beli Konsumen. *SATIN-Sains Dan Teknologi Informasi*, I(2), 32–40.
- Patil, U. M. (2020). User and Session Identification From Web Data Preprocessing. *JAC : A Journal Of Composition Theory*, XIII(3), 376–385.
- Pramana, A. E. (2019). Penerapan Web Usage Mining Menggunakan Teknik Association Rule (Studi Kasus: <http://portal.stikom-db.ac.id>). In *Stikom Dinamika Bangsa*.
- Raphaeli, O., Goldstein, A., & Fink, L. (2017). Analyzing online consumer behavior in mobile and PC devices: A novel web usage mining approach. *Electronic Commerce Research and Applications*, 26, 1–12. <https://doi.org/10.1016/j.elerap.2017.09.003>
- Setia, S., Jyoti, V., & Duhan, N. (2020). HPM : A Hybrid Model for User ' s Behavior Prediction Based on N -Gram Parsing and Access Logs. *Scientific Programming*, 2020(ii).
- Syahrir, M., & Fatimatuzzahra. (2020). Integrasi Pendekatan Metode Custom Hashing dan Data Partitioning untuk Mempercepat Proses Pencarian Frekuensi Item-set pada Algoritma Apriori. *Matrik : Jurnal Manajemen, Teknik Informatika Dan Rekayasa Komputer*, 20(1), 149–158. <https://doi.org/10.30812/matrik.v20i1.853>