

Ekstraksi *Click Stream* Data Web *E-Commerce* Menggunakan *Web Usage Mining*

Kartina Diah¹

^{1,2,1} Jurusan Teknologi Informasi, Program Studi Teknik Informatika, Politeknik Caltex Riau
¹diah@pcr.ac.id

Abstrak

E-Commerce berkembang pesat dalam *world wide web* hingga menghasilkan berbagai jenis data yang dapat dianalisa lebih lanjut untuk berbagai keperluan seperti personifikasi *web*, *profiling customer*, dan sebagainya. Salah satu jenis data yang dihasilkan *e-Commerce* adalah *click stream* data web yang merekam aktivitas *visitor web* dalam bentuk *log* data selama berinteraksi pada laman web. Penelitian ini mengekstraksi *click stream* data *web e-commerce* untuk mendapatkan pola interaksi konsumen terhadap halaman web selama mengunjungi *web e-commerce*. Berdasarkan jenis data yang diekstrak maka *web usage mining* digunakan untuk ekstraksi pola dari *click stream* data yang berbentuk *log* data. Teknik *mining* yang dianalisa terhadap *log* data *e-commerce* pada penelitian ini terdiri dari *frequent itemset*, *asociation rules*, dan *frequence sequence mining*. *Frequent itemset* menghasilkan halaman web yang paling sering diakses oleh *visitor*. *Association rules* menghasilkan pola kemungkinan halaman web yang akan diakses *visitor* jika *visitor* mengakses halaman-halaman tertentu. *Frequence sequence mining* mendapatkan pola urutan halaman web yang paling sering diakses oleh *visitor web e-commerce* saat berinteraksi pada laman web. Pola urutan halaman yang diakses *visitor* menunjukkan urutan kebiasaan *visitor* mengunjungi *e-commerce*. Sedangkan teknik *mining* yang diimplementasikan untuk menghasilkan pola akses *visitor* pada penelitian ini adalah *Frequence sequence mining*. Hasil ekstraksi dari penelitian ini menunjukkan ada enam halaman web yang paling sering diakses oleh konsumen dengan berbagai pola urutan aksesnya.

Kata kunci : *e-commerce*, *web usage mining*, *click stream data*, *frequence sequence*

1. Pendahuluan [*Times New Roman 10, bold*]

Aplikasi *electronic commerce* atau *e-commerce* adalah aplikasi enterprise berbasis web yang dewasa ini semakin berkembang seiring pesatnya perkembangan *world wide web*. Dengan *e-commerce* seorang kostumer dapat memanfaatkan fasilitas web untuk berbelanja berbagai kebutuhan barang maupun jasa dari penjual barang dan jasa yang ada diberbagai tempat diseluruh *world-wide*(Tatsiopoulos et al., 2002).

E-Commerce didefinisikan sebagai kegiatan jual beli produk, layanan, dan informasi melalui jaringan komputer khususnya internet (Wen et al., 2001). Lebih jauh disebutkan bahwa *e-commerce* tidak terbatas hanyapada aktivitas jual-beli saja, tetapi mencakup berbagai proses yang dilakukan dalam organisasi yang mendukung tujuan dari bisnisnya (Diah & Yunanto, 2017).

Data web yang merekam aktivitas *visitor web* disebut sebagai data *click stream*. Data *click stream* pada aplikasi *e-commerce* dapat terdiri dari data *access logs*, *browser logs*, *user profiles*, *registration data*, *user session*, *cookies*, dan data lainnya yang merupakan hasil transaksi user dengan web. Dari data tersebut banyak informasi yang dapat dihasilkan bagi *web site management*. Selain

menemukan *web site* yang paling sering dikunjungi pengguna, dapat juga diidentifikasi *web site* mana yang kemungkinan akan dikunjungi bersamaan oleh user serta kebiasaan pengguna pada saat mengunjungi sebuah web. Salah satu teknologi yang dapat digunakan untuk mengolah dan menggali pengetahuan dari data dalam jumlah besar dengan tipe data yang bervariasi adalah *data mining*. Sedangkan metode *data mining* yang digunakan untuk mengekstrak pengetahuan yang berasal dari data web disebut sebagai *web mining* (Cooley, 2000). Srivastava,dkk dalam “Web Mining – Concept, Application, and Research Direction” menyebutkan, ada dua pendekatan yang dapat digunakan untuk mendefinisikan *web mining*. Pertama, pendekatan berbasis proses (*process-centric-view*) yang mendefinisikan *web mining* sebagai urutan aktivitas (*sequence of task*). Kedua, pendekatan berbasis data (*data-centric view*) yang mendefinisikan *web mining* sebagai terminology tipe data web yang digunakan untuk proses *data mining*.(Abdurrahman et al., 2008;Srivastava & Desikan, Prasanna, Kumar, n.d.;Cooley et al., 1997).

WUM adalah aplikasi teknik data mining untuk menemukan pola penggunaan dari data web dengan tujuan memahami dan memberikan layanan yang lebih baik pada aplikasi berbasis web(Cooley, 2000;

Wang, 2000). Dalam penelitian (Senkul, Pinar, 2012), *web usage mining* juga digunakan untuk memberikan halaman rekomendasi bagi pelanggan *web e-commerce* berdasarkan profil pelanggan yang diperoleh dari data *log*. Penelitian (Mishra, 2012) melakukan ekstraksi terhadap data *log web* untuk mendapatkan pola akses pengunjung *web* yang paling *frequent* menggunakan algoritma *FP-Growth*. *Mining data log web* menggunakan klasifikasi juga digunakan dalam (Azam, & Tabrez, 2014) untuk mengetahui pola kunjungan pelanggan yang berasal dari dalam India dan pola kunjungan pelanggan yang berasal dari luar India. *Association rule* dalam penelitian (Wani, K., & Bakal, 2014) juga digunakan untuk mendapatkan pola kunjungan pelanggan serta mengetahui hubungan antara halaman yang dikunjungi untuk dapat memberikan rekomendasi halaman yang tepat bagi pelanggan.

Dari latar belakang yang diuraikan penulis diatas maka tujuan dari penelitian ini adalah menghasilkan ekstraksi data *click stream* dari aplikasi *e-commerce* menggunakan teknologi *web usage mining*.

2. Tinjauan Pustaka

➤ *Web Usage Mining*

Berdasarkan jenis data yang diekstrak, *web mining* dapat dibagi menjadi tiga kategori, yaitu (Diah & Yunanto, 2017; Hamid Mughal, 2018; Srivastava et al., 2005; Roy & Appa Rao, 2020; Iváncsy & Vajk, 2006):

1. *Web Content Mining* (WCM); teknik *data mining* untuk menghasilkan informasi tentang *content* sebuah dokumen web. Dokumen web tersebut dapat berupa teks, gambar, audio, video, *structured records* seperti *list* dan *table*.
2. *Web Structure Mining* (WSM); teknik *data mining* untuk menghasilkan informasi tentang struktur web. Terbagi menjadi 2, yaitu *hyperlinks*, *document structure*.
3. *Web Usage Mining* (WUM) (sedikit ulasan ttg WCM); teknik *data mining* yang digunakan untuk menemukan pola penggunaan *web* dari *web usage data*, dengan tujuan untuk memahami dan memberikan layanan yang lebih baik pada aplikasi berbasis web (Srivastava et al., 2000). Pola yang dihasilkan berasal dari klasifikasi *usage data* sebagai berikut, *web server data*, *application server data*, *application level data*.

Langkah-langkah yang dilakukan dalam WUM terbagi menjadi 3 tahap, yaitu (Diah & Yunanto, 2017; Rathod, 2012; Cooley, 2000; Vellingiri, J; Pandian, 2011):

1. *Pre-processing*. Data yang tersedia memiliki kecenderungan mengandung *noise*, tidak lengkap dan tidak konsisten. Pada tahap ini, data akan diolah untuk disesuaikan dengan kebutuhan pada fase berikutnya. Tahap ini meliputi pembersihan data, integrasi data, transformasi data dan reduksi data.

2. *Pattern discovery*. Beberapa metode dan algoritma seperti *statistic*, *data mining*, *machine learning*, dan *pattern recognition* dapat diterapkan untuk mendapatkan pola pada tahap ini.

Proses “*mining*” terjadi pada tahapan ini, meliputi (Cooley dkk., 2000; Wang, 2000): *statistical analysis*, *association rules*, *clustering*, *classification*, *frequent pattern*, *sequential pattern*, *dependency modeling*.

3. *Pattern Analysis*. Pola yang telah ditemukan selanjutnya dianalisa dan ditampilkan dengan visualisasi dan interpretasi yang lebih difahami oleh pengguna.

➤ *Data Web Log*

Untuk melakukan analisis terhadap *log web*, pertama-tama adalah dengan membagi data *log web* ke sesi, di mana sesi adalah serangkaian referensi halaman satu situs *web* selama satu periode logis. Sesi didefinisikan sebagai periode *browsing web* berkelanjutan atau urutan tampilan halaman web. Sesi juga dapat dilihat sebagai awal pengguna mengunjungi situs web, melakukan pekerjaan, dan kemudian meninggalkan situs web (Sun & Zhang, 2004).

Ratnes Kumar Jain dan kawan-kawan dalam (Jain & Kasana, 2009) menyatakan bahwa *log file* berisi informasi sebagai berikut (Grace & Nagamalai, 2011):

1. *Visited path*: path yang terekam oleh *log* ketika pengguna mengakses halaman-halaman web.
2. *Username*: dibebberapa *website* pengguna diminta untuk mengisikan nama sebelum mengakses *website*. Selain itu, identitas pengguna juga dapat diidentifikasi dari *IP Address* yang tercatat pada *log* setiap pengguna mengakses *website*.
3. *Path traversed*: mengidentifikasi *path-path* yang dilalui oleh pengguna
4. *Timestamp*: durasi pengguna mengakses *website*.
5. *Page last visited*: halaman terakhir yang diakses oleh pengguna
6. *Success rate*: tingkat keberhasilan diukur dari banyaknya jumlah data yang diunduh oleh pengguna
7. *User agent*: *web browser* yang digunakan oleh pengguna untuk mengakses *website*
8. *URL*: alamat sumber yang diakses oleh pengguna
9. *Request type*: *method* yang digunakan untuk transfer data, misalnya *POST* atau *GET*.

3. Analisis Penerapan Web Usage Mining pada Data Click Stream

3.1 Metodologi Penelitian

Penelitian ini terdiri dari beberapa tahap yaitu data collection, data preprocessing, data exploration, data visualization. Tahapan-tahapan tersebut digambarkan pada gambar 1 berikut ini.



Gambar 1 Tahapan Metodologi Penelitian

- **Data Collection**

Data log web yang digunakan dalam penelitian ini adalah data click stream yang diunduh dari ECML/PKDD 2005 Discovery Challenge¹. Setiap file log berisi informasi yang dikumpulkan selama satu jam dari 7 buah aplikasi e-commerce. Data log yang digunakan pada penelitian ini berjumlah 600.000 baris data.

- **Data Preprocessing**

Tidak semua data logs digunakan pada tahap mining data karena beberapa data tidak relevan dengan kebutuhan terhadap mining yang akan dilakukan, untuk itu diperlukan tahap preprocessing data terlebih dahulu. Tahapan preprocessing terdiri dari data cleaning berfungsi untuk membersihkan logs entries dari data yang tidak dibutuhkan pada proses mining. Session identification berfungsi untuk mengidentifikasi semua halaman referensi yang digunakan oleh pengguna untuk disimpan menjadi data dalam user session. Data conversion berfungsi untuk menyesuaikan format data sesuai dengan kebutuhan pada mining proses.

- **Pattern Discovery**

Mengimplementasikan algoritma yang tepat yang sesuai dengan kebutuhan penelitian dilakukan pada tahap ini. Pada penelitian ini algoritma yang diimplementasikan merupakan hasil analisis terhadap 3 buah algoritma yang umum digunakan pada Web Usage Mining yaitu frequent item set, association rule dan frequent sequence item set. Analisis pada tahapan ini dilakukan untuk menentukan algoritma yang paling tepat agar menghasilkan sebuah pola perilaku konsumen pada saat berinteraksi menggunakan web e-commerce.

- **Visualization**

Pada tahapan ini akan menampilkan pola yang dihasilkan pada tahap sebelumnya dalam bentuk diagram agar pola interaksi konsumen saat menggunakan web e-commerce dapat lebih mudah digunakan lebih lanjut pada penelitian selanjutnya.

Evaluasi model yang dihasilkan pada penelitian ini baru akan dilakukan pada penelitian lanjutan yaitu mengimplementasikan model yang diperoleh menjadi sebuah web e-commerce, sehingga dapat

ditentukan seberapa akurat model yang dihasilkan dapat memenuhi semua fungsional untuk sebuah aplikasi e-commerce.

3.2 Web Usage Mining

WUM melakukan mining untuk mengekstraksi pola yang berasal dari usage data yaitu log data. Log akses web berisi urutan kejadian atau aktivitas atau events (items) dalam satu session, dengan informasi mencakup session identifier dan informasi akses pengguna. Misalkan, dalam data log web terdiri dari informasi beberapa halaman web yang dikunjungi oleh pengunjung dalam 1 session. Sebagai contoh, format informasi dari sebuah akses log adalah <SessionId,PageId>, dan data log akses web dari beberapa session sesuai format adalah sebagai berikut:

```

    <100,a><100,b><200,e><200,a><300,b><200,e><100,d><200,b><400,a><400,f><100,a><400,b><300,a><100,c><200,c><400,a><200,a><300,b><200,c><300,f><400,c><400,f><400,c><300,a>><300,e><300,c>
    
```

Sebuah session dilihat sebagai awal pengguna mengunjungi situs web, melakukan aktivitas, dan kemudian meninggalkan situs web. Aktivitas yang dimaksud dalam sebuah session tersebut adalah kegiatan mengunjungi halaman-halaman web yang ada pada situs web dalam selang waktu tertentu. Sehingga jika dalam sebuah session didefinisikan sebagai <100,a>, <100,b>, <100,d>, <100,a>, <100,c> maka dalam sebuah session dengan Id 100 melakukan aktivitas mengunjungi halaman web a,b,d,a,c.

Selanjutnya data log tersebut dikelompokkan berdasarkan session-nya seperti yang terlihat pada table 1 Contoh Sequence Database. Urutan log akses dalam database transaksi pada pada masing-masing baris terdiri dari ID Transaksi (TID) yang berasal dari session dan urutan akses (access sequence) yang berasal dari event. Misalkan untuk session 100, event yang dilakukan terdiri dari a,b,d,a,dan c.

Tabel 1 Contoh Sequence Database



TID	Web access sequences
100	abdac
200	eaebcac
300	babfaec
400	afbacfc

Dari tabel 1 diatas dapat diperoleh pola urutan akses yang paling sering dilakukan oleh pengunjung. Sebagai contoh, dari data diatas diperoleh bahwa halaman yang paling sering dikunjungi oleh pengunjung adalah sebagai berikut: halaman a = 4, b = 4, c = 4, d = 1, e = 2, f = 2.

Sehingga diketahui bahwa setiap pengunjung pasti mengunjungi halaman a, b, c sedangkan halaman d, e, f ada yang mengunjungi ada yang tidak. Selanjutnya dapat juga diketahui pola urutan halaman yang paling sering dikunjungi oleh pengunjung yaitu a,b,a,c. Artinya setiap pengunjung akan mengunjungi halaman web dengan urutan halaman a diikuti halaman b kemudian kembali ke halaman a lalu ke halaman c. Urutan halaman yang dikunjungi tersebut jika digambarkan dalam sebuah graph maka akan terlihat pada gambar 1 sebagai berikut:



Gambar 2 Urutan Kunjungan Halaman Web Dengan WUM

Notasi  menggambarkan halaman web, sedangkan  menunjukkan urutan halaman yang diakses.



3.2 Pendekatan Teknik WUM untuk Pattern Discovery

Frequent Itemset

Frequent Pattern adalah teknik mining yang dilakukan dengan menganalisa jumlah itemset yang kemunculannya melebihi nilai *support*-nya. Berdasarkan contoh yang diberikan pada lampiran B tentang *frequent itemset*, maka dihasilkan pola akses halaman *web* seperti pada gambar 2 Pola Akses Halaman Web Dengan *Frequent Itemset* dibawah ini.



Gambar 3 Pola Akses Halaman Web Dengan *Frequent Itemset*

Notasi  pada gambar diatas menggambarkan halaman *web*, sedangkan  menggambarkan halaman yang diakses dengan mengabaikan urutan. Dari analisa yang telah dilakukan diketahui bahwa *frequent itemset mining* menghasilkan sebuah model dari *itemset* yang paling sering dikunjungi oleh pengunjung dalam 1 *session*. Dalam *frequent itemset mining* tidak menunjukkan informasi temporal yang menggambarkan sebuah urutan akses halaman web dari waktu ke waktu secara runtun.

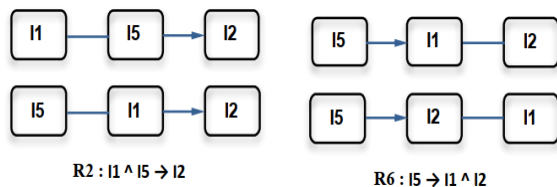
Association rules

Association rules atau aturan asosiasi adalah teknik mining yang digunakan untuk menemukan hubungan antara atribut dalam sebuah *itemset*. Di



dalam data *log web itemset* adalah sekumpulan set halaman-halaman *web* dalam satu *session*. Aturan diterapkan untuk menunjukkan hubungan kemungkinan halaman-halaman *web* yang sering diakses bersama-sama setiap kali pengunjung mengunjungi web. Berdasarkan hasil *confidence* yang berasal dari tabel B-5 maka aturan asosiasi yang memenuhi *minimum support 30%* dan *minimum confidence 70%* adalah R2, R3, R6. Aturan asosiasi yang terbentuk menjadi,

- R2: $I1 \wedge I5 \rightarrow I2$, seorang pengunjung yang mengunjungi halaman halaman I1 dan I5 punya kemungkinan 70% untuk mengunjungi halaman I2 dengan jumlah pola kunjungan tersebut selama ini adalah 30%.
- R3: $I2 \wedge I5 \rightarrow I1$, seorang pengunjung yang mengunjungi halaman halaman I2 dan I5 punya kemungkinan 70% untuk mengunjungi halaman I1 dengan jumlah pola kunjungan tersebut selama ini adalah 30%.
- R6: $I5 \rightarrow I1 \wedge I2$, seorang pengunjung yang mengunjungi halaman halaman I5 punya kemungkinan 70% untuk mengunjungi halaman I2 dan I2 dengan jumlah pola kunjungan tersebut selama ini adalah 30%.

Maka pola yang dihasilkan dari aturan asosiasi diatas adalah seperti gambar 3 Pola Akses Halaman Web Menggunakan *Association Rule* berikut.



Gambar 4 Pola Akses Halaman Web Menggunakan *Association Rule*

Notasi  menggambarkan kemungkinan sebuah halaman diakses jika halaman sebelum notasi diakses oleh pengunjung, sedangkan notasi  menggambarkan halaman yang diakses bersamaan.

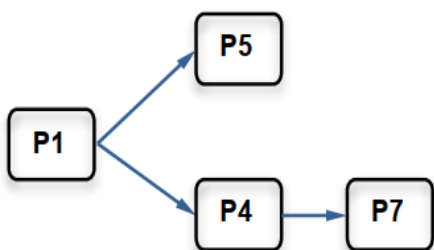
Berdasarkan aturan asosiasi yang dihasilkan diketahui bahwa satu halaman web yang dikunjungi saling mempengaruhi kemungkinan halaman web yang lain dikunjungi juga. Namun urutan kunjungan terhadap halaman web tersebut bukanlah menjadi hal yang dipertimbangkan, seperti misalkan aturan asosiasi dari R6 yang berarti setelah mengunjungi halaman I5 akan diikuti dengan mengunjungi halaman I1 lebih dulu baru halaman I2 atau sebaliknya asal kedua halaman tersebut dikunjungi setelah halaman I5 dikunjungi.

Frequent Sequence Mining



Sequential Pattern pada dasarnya tersusun dari himpunan *frequent itemset* yang merepresentasikan

pola berdasarkan urutan waktu terjadinya peristiwa, disebut juga dengan istilah *frequent sequence*.

Pola yang dihasilkan adalah pola dengan nilai maksimal diatas nilai *minimum support*-nya. Pola {(P1)(P5)} tersebut berarti, secara umum pengunjung web mengakses halaman P1 kemudian mengunjungi halaman P5 diwaktu yang lain dalam rentang waktu yang sama. Begitu juga dengan pola {(P1), (P4P7)} yang artinya, pengunjung mengunjungi halaman P1 kemudian mengunjungi halaman P4 lalu halaman P7 pada kunjungan selanjutnya dalam rentang waktu yang sama pula. Dengan demikian jika pola tersebut digambarkan dalam urutan proses maka akan terlihat seperti gambar 4 Pola Akses Halaman Web Dengan Frequence Sequence dibawah ini.



Gambar 5 Pola Akses Halaman Web Dengan Frequent Sequence

Notasi  menggambarkan halaman web, sedangkan notasi  menunjukkan urutan halaman web yang diakses.

Berdasarkan analisis terhadap 3 algoritma diatas, hanya algoritma *frequency sequence item set* yang dapat menunjukkan pola urutan akses konsumen terhadap halaman web pada saat berinteraksi dengan web e-commerce. Pola urutan akses halaman web dapat dijadikan salah satu acuan dalam pengembangan aplikasi web e-commerce sehingga algoritma *frequency sequence item set* akan diimplementasikan pada penelitian ini.

4. Implementasi Web Usage Mining pada Data Click Stream

4.1 Click Stream Data

Data *log web* yang digunakan dalam penelitian ini adalah data *click stream* yang diunduh dari *ECML/PKDD 2005 Discovery Challenge1*. Setiap file log berisi informasi yang dikumpulkan selama satu jam dari 7 buah aplikasi e-commerce. Data log yang digunakan pada penelitian ini berjumlah 600.000 baris data. Tabel 2 Data Log Web berikut ini adalah contoh data *log* dari *click stream* data e-commerce yang digunakan pada penelitian ini.

Tabel 2 Data Log Web

Shop ID	Unix ime()	IP Address	Session	Visited Page
11	1074628802	212.20.106.210	fadb9b666eab173cfc0900148ebd21cb	/
14	1074628802	217.11.233.2	8a8ad32c814ea45c41f1709df7967a19	/ct/?c=164
11	1074628802	82.34.184.151	be5d5d6dc24bf3e54530d7dc8945ff25	/znacka/?&pozice=100&c=102
12	1074628802	80.188.4.42	7a7f25ac8ed0f4e64a1c524803c526e8	/dt/?c=9674&komentare=1
14	1074628803	195.250.148.121	ed03ab09cc0ac7d67a5735599bc6b88e	/dt/?c=7308
11	1074628805	212.20.99.79	2cbb61aefadb5fd25f6d7c9e7a596ce6	/dt/?c=927
10	1074628805	212.111.9.109	344ce202b8094d3aecd2fff84cf5da31	/znacka/?&pozice=20&c=2
15	1074628806	194.108.242.78	2feff86721f48bea10510e1594e5ebca	/dt/?c=5231
14	1074628808	213.220.194.175	680dd7669205aed163dbb06bfd8ecabd	/findp/reber3.php
14	1074628808	194.228.166.69	a955d38c204e572ab635418178806168	/ls/preber.php
10	1074628808	62.177.77.138	4e38b17a019a399d018319e337c395ad	/dt/prisl.php?id=7578

Selanjutnya dilakukan *pre-processing data* terhadap data log tersebut. bagian pertama pada tahap *pre-processing* dilakukan untuk menghasilkan *sequence database* seperti terlihat pada tabel 3 *Sequence Database*.

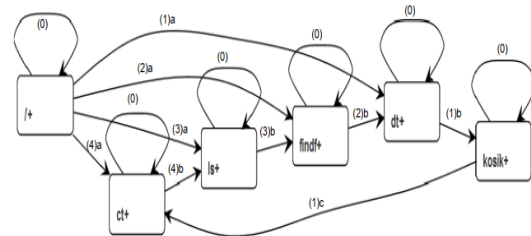
Tabel 3 *Sequence Database*

Session	Visited Page
6cf200741ca492c1666b42c0aa879377	/ ls
24675844746d8353d00b2240db2cd1ae	ls
69fa610df40c7118aa7912e35155f762	/ ct ls dt ls dt ls dt ls dt ls dt ls dt ls dt ls dt ls
ed316ac9b35b4d703920eb43f92790b0	df dt
1e852e124d4aea2cd6c7150a21a51e94	ls
baef21dc9b744e8d77fb1ce1d0e977f0	findp
156bd9732087fcdcaec72376d4061ae4	dt
e2fec82d4715a294e9540bec87023fdf	/ ct ls dt ls dt ls dt znacka dt ls dt ls dt
f22093add2b9dc88d83c1167a0fe037b	/
3b3d093e3bd62c1f737fc3a99106fb66	ls
8f16c7a4d649f341296d44716f2dfe1c	/ ct
387cb0078d37aa354700065f2140dcae	dt poradna ct / findf dt findf dt findf dt findf dt klient
f3cc04154ca0cbd73248c0b07aa8e508	/ znacka / obchod y- elektro / /
b55b3a50fc5abcefc339fc1ac0533ed9	/ ct ls dt ls dt ls dt ls dt
642b5bc5f6f20b43e4091f8f8f62bf	poradna

4e	
e6332b88864cfe866b319e28f5fe9fcf	znacka
1561188c58329d5824eec0b9f85fd967	/ ct / ls dt ls dt ls dt znacka

4.2 *Pattern Discovery WUM*

Pada penelitian ini Teknik mining yang diimplementasi pada tahapan pattern discovery adalah *Frequence Sequence Mining*. Gambar 5 Hasil Ekstraksi WUM Menggunakan *Frequence Sequence Mining* dibawah ini adalah hasil pattern discovery WUM menggunakan *frequence sequence minig* dari data hasil *pre-processing* pada tahap sebelumnya.



Gambar 6 Hasil Ekstraksi WUM Menggunakan *Frequence Sequence Mining*

Keterangan :

- /+: Halaman Home
- ct+: Halaman Kategori Produk
- ls+: Halaman List Produk
- dt+: Halaman Detil Produk
- findf+: Halaman Pencarian Produk dan Aksesories
- kosik+: Halaman Keranjang Belanja, Proses *Check Out*.

Sehingga dari urutan-urutan aktivitas pelanggan tersebut pola urutan halaman web e-commerce yang biasa diakses pelanggan adalah Home → Home, atau Home → Detail Product, atau Home → Detail Product → Shopping Cart, atau Home → Detail Product → Shopping Cart → Product Category.

5. Kesimpulan

5.1 Kesimpulan

Hasil ekstraksi click stream data e-commerce pada penelitian ini memanfaatkan *Web Usage Mining* dengan Teknik mining *frequent sequence mining* menghasilkan 6 halaman web yang paling sering diakses oleh *visitor* dengan pola urutan akses yang berbeda-beda. Nilai minimum *support* yang digunakan sangat mempengaruhi pola halaman akses *visitor web e-commerce* yang dihasilkan.

5.2 Future Works

Future works dari penelitian ini dengan mengeksplorasi lebih lanjut *variable independent* yang dianggap sangat mempengaruhi pola yang dihasilkan seperti nilai *minimum support* agar pola akses *visitor e-commerce* lebih optimal. Urutan pola akses *visitor web e-commerce* dapat dimanfaatkan untuk mendukung dalam tahap desain pada pengembangan *web e-commerce*.

Daftar Pustaka:

- Abdurrahman, T Riyanto, B., Mandala, R., & Govindaraju, R. (2008). PEMANFAATAN ALGORITMA ANT COLONY UNTUK WEB USAGE MINING. *Konferensi Dan Temu Nasional Teknologi Informasi Dan Komunikasi Untuk Indonesia*.
- Azam, & Tabrez, N. (2014). Data Mining of Web Access Logs Using Classification Techniques. *International Journal For Research in Applied Science Engineering Technology (I J R A S E T)*, 2(Viii), 1–5.
- Cooley, R. (2000). *Web Usage Mining: Discovery and Application of Interesting Patterns from Web Data*.
- Cooley, R., Mobasher, B., & Srivastava, J. (1997). *Web Mining: Information and Pattern Discovery on the World Wide Web*.
- Diah, K., & Yunanto, W. (2017). Heuristics miner for e-commerce visitor access pattern representation. *Communications in Science and Technology*, 2(1), 1–5. <https://doi.org/10.21924/cst.2.1.2017.21>
- Hamid Mughal, M. J. (2018). Data mining: Web data mining techniques, tools and algorithms: An overview. *International Journal of Advanced Computer Science and Applications*, 9(6), 208–215. <https://doi.org/10.14569/IJACSA.2018.090630>
- Ivancsy, R., & Vajk, I. (2006). Frequent pattern mining in web log data. *Acta Polytechnica Hungarica*, 3(1), 77–90. <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.101.4559&rep=rep1&type=pdf>
- Mishra, R. (2012). Discovery of Frequent Patterns from Web Log Data by using FP-Growth algorithm for Web Usage Mining. *International Journal of Advanced Research in Computer Science and Software Engineering*, 2(9), 311–318.
- Rathod, D. M. (2012). A Review On Web Mining. *International Journal of Engineering Research and Technology (IJERT)*, Vol. 1(2), April 2012).
- Roy, R., & Appa Rao, G. (2020). Survey on pre-processing web log files in web usage mining. *International Journal of Advanced Science and Technology*, 29(3 Special Issue), 682–691.
- Senkul, Pinar, and S. S. (2012). Improving pattern quality in web usage mining by using semantic information. *Knowledge and Information Systems*, 30.3, 527– 541.
- Srivastava, J., Cooley, R., Deshpande, M., & Tan, P. (2000). Web usage mining: discovery and applications of usage patterns from Web data. *ACM SIGKDD Explorations Newsletter*, 12–23.
- Srivastava, J., & Desikan, Prasanna, Kumar, V. (n.d.). *Web Mining - Concepts, Applications, and Research Directions*.
- Srivastava, J., Desikan, P., & Kumar, V. (2005). Web Mining - Concepts , Directions. *Foundations and Advances in Data Mining Studies in Fuzziness and Soft Computing*, 180, 2005, 51–71.
- Tatsiopoulou, I. P., Panayiotou, N. A., & Ponis, S. T. (2002). A modelling and evaluation methodology for E-Commerce enabled BPR. *Computers in Industry*, 49(1), 107–121. [https://doi.org/10.1016/S0166-3615\(02\)00062-3](https://doi.org/10.1016/S0166-3615(02)00062-3)
- Vellingiri, J;Pandian, S. . (2011). A Survey on Web Usage Mining. *Global Journal of Computer Science and Technology*, 11(4).
- Wang, Y. (2000). Web mining and knowledge discovery of usage patterns. *CS 748T Project, Part I*, 1–25. <https://cs.uwaterloo.ca/~tozsu/courses/cs748t/surveys/wang.pdf>
- Wani, K., & Bakal, J. W. (2014). Association Rule Mining From Server Log File. *International Journal on Advanced Computer Theory and Engineering (IJACTE)*, 6–11.
- Wen, H. J., Chen, H. G., & Hwang, H. G. (2001). E-commerce Web site design: Strategies and models. *Information Management and Computer Security*, 9(1), 5–12. <https://doi.org/10.1108/09685220110366713>

