

Efektifitas Penggunaan *Association Rules Mining* dalam Personalisasi *Website*

Edi Priyanto ^{(1)*}, Arief Hermawan ⁽²⁾, Rianto ⁽³⁾, Donny Avianto ⁽⁴⁾

Magister Teknologi Informasi, Program Pascasarjana, Universitas Teknologi Yogyakarta,
Yogyakarta

e-mail : edi.priyanto@student.uty.ac.id, {ariefdb,rianto,donny}@staff.uty.ac.id.

* Penulis korespondensi.

Artikel ini diajukan 9 Juli 2020, direvisi 18 Juli 2020, diterima 21 Juli 2020, dan dipublikasikan 20 Januari 2021.

Abstract

As the usage of the internet grows, more and more information is obtained, thus presenting challenges, especially for users and website owners. Website users often have difficulty finding products or services that are relevant to their needs caused by abundant amounts of products and services delivered on a website. Website owners often find it difficult to convey information about the right products and services to certain target users. Based on the problem given above, we can conclude that a recommendation system approach that can improve personalization on their website is needed. The recommendation system approach must be able to provide navigation on the website to make it more adaptive towards the interests and information needed by the user. This study uses Association Rules formed from Microsoft web access log data by finding visitor patterns based on frequently visited web site pages. From the results of the research conducted, the performance of the method used has a precision value of 0.896, 0.058 recall, and F-measure 0.104. Whereas the measurement of the accuracy value resulted in a performance recommendation of exactly 3%, an acceptable rate of 87%, and 10% incorrect. This research shows that the Association Rules method can increase the effectiveness of website personalization to provide relevant information recommendations for visitors. For further research, it can concentrate on improving existing methods thus website personalization becomes more adaptive.

Keywords: *Personalization, Website, Recommendation System, Association Rules Mining, Data Mining*

Abstrak

Seiring berkembangnya pemanfaatan internet maka semakin banyak informasi yang diperoleh, sehingga menghadirkan tantangan terutama bagi pengguna dan pemilik *website*. Pengguna *website* sering mengalami kesulitan mencari produk atau layanan yang relevan terhadap kebutuhannya karena banyaknya produk dan layanan yang disampaikan pada suatu *website*. Bagi pemilik *website* sering mengalami kesulitan untuk menyampaikan informasi tentang produk dan layanan yang tepat pada target pengguna tertentu. Maka dengan permasalahan tersebut, diperlukan pendekatan sistem rekomendasi yang dapat meningkatkan personalisasi pada *website* mereka. Pendekatan sistem rekomendasi harus mampu memberikan navigasi pada *website* agar lebih adaptif terhadap minat dan kebutuhan informasi untuk pengguna. Penelitian ini menggunakan *Association Rules* yang dibentuk dari data log akses web Microsoft dengan menemukan pola pengunjung berdasarkan halaman situs web yang sering dikunjungi. Dari hasil penelitian yang dilakukan, performa metode yang digunakan memiliki nilai *precision* sebesar 0.896, *recall* 0.058, dan *F-measure* 0.104. Sedangkan pengukuran nilai ketepatan menghasilkan performa rekomendasi tepat sebesar 3%, dapat diterima sebesar 87%, dan salah 10%. Penelitian ini menunjukkan bahwa metode *association rules* dapat meningkatkan efektifitas pada personalisasi *website* sehingga memberikan rekomendasi informasi yang relevan bagi pengunjung. Untuk penelitian selanjutnya dapat berkonsentrasi pada peningkatan metode yang ada sehingga personalisasi *website* menjadi lebih adaptif.

Kata Kunci: *Personalisasi, Website, Sistem Rekomendasi, Association Rules Mining, Penggalan Data*



1. PENDAHULUAN

Internet telah memberikan dampak besar dalam kemudahan memasarkan produk bagi pebisnis. Dalam beberapa dekade terakhir, sebagai contoh *e-commerce* telah banyak bermunculan dan menjadi tren bagi pengguna (seperti konsumen) yang menginginkan kemudahan dalam membeli kebutuhan. Dengan adanya interaksi antara konsumen dan *website* toko online (*e-commerce*) memberikan tantangan bagi pebisnis online dalam memberikan informasi produk di internet (Prasetya, 2017). Informasi yang diberikan dapat memberikan berbagai masalah seperti data tersembunyi, skalabilitas, pencarian informasi, kelebihan *item*, dan lain sebagainya. Ini mengharuskan pengguna untuk menemukan, mengekstrak, memfilter dan mengevaluasi informasi yang diinginkan. Maka penambangan web dianggap cara efektif dalam operasional bisnis dan membantu pengguna dalam memberikan informasi yang tepat (Das et al., 2017).

Web mining merupakan teknik penambangan data yang membantu dalam proses ekstraksi pengetahuan berupa informasi yang berguna dari *website*. *Web mining* dapat dikategorikan sebagai *web content mining*, *web usage mining*, dan *web structure mining*. *Web content mining* adalah proses mengekstraksi informasi bermanfaat dari konten dokumen web. *Web structure mining* membantu dalam menganalisis struktur web seperti tautan antar halaman web. *Web usage mining* digunakan untuk mengekstrak dan menganalisis pola penggunaan dari data sehingga membantu dalam pemodelan berbagai aplikasi seperti personalisasi dan sistem rekomendasi (Das et al., 2017).

Personalisasi merupakan cara efektif dalam merancang sistem dengan mengamati preferensi pengguna dan melakukan tindakan yang diperlukan untuk memprediksi informasi yang tepat. Sebagai contoh, personalisasi dapat dilakukan dengan menyarankan halaman web yang menarik perhatian pengguna sebagai dasar untuk rekomendasi selanjutnya (Das et al., 2017). Dengan memahami pengguna bersama dengan preferensi, kekhasan dan perilakunya untuk menyesuaikan rekomendasi maka personalisasi dapat memberikan pengalaman yang menarik. Penelitian yang dilakukan oleh Jasberg et al. telah membuktikan bahwa personalisasi dapat beradaptasi lebih dekat dengan pola berpikir pengguna serta menafsirkan perilaku secara lebih memadai, menghasilkan pengalaman web yang lebih baik dan meningkatkan kepuasan pengguna (Sulikowski et al., 2018).

Personalisasi *website* dapat dikembangkan ketika data pengguna tersedia untuk digunakan sebagai *user model* (Hawalah & Fasli, 2015). Proses pengumpulan data dalam aktivitas pengguna disebut pembuatan profil pengguna. Pembuatan profil pengguna didefinisikan sebagai proses identifikasi pengguna melalui minat pengguna ketika mereka berinteraksi dengan *website*. Pembuatan profil pengguna secara umum berkembang dengan pendekatan *data mining & machine learning* (Kanoje et al., 2015). Profil pengguna hampir mirip dengan preferensi pribadi pengguna yaitu minat, hobi, dan favorit. Selain itu, untuk memudahkan proses identifikasi, pengguna dapat diklasifikasikan ke dalam usia, jenis kelamin, dan profesi atau tingkat pendidikan (Garrigós et al., 2010). Sementara itu, personalisasi *website* tidak hanya untuk pengguna individu tetapi juga untuk pengguna grup yaitu personalisasi grup. Perbedaan antara personalisasi individu dan personalisasi grup dapat dilihat pada Tabel 1 (Rianto et al., 2016). Untuk mencapai hasil personalisasi yang tinggi, maka penting untuk membuat profil pengguna yang tepat yang dapat mewakili topik pengguna dan untuk merancang fungsi peringkat yang lebih canggih yang mengukur relevansi antara pengguna dan data produk (Han et al., 2019).

Tabel 1. Perbedaan Personalisasi Individu dan Grup.

Variabel	Individu	Grup
Identifikasi	Bakat, kemampuan, dan preferensi	Kesamaan usia, jenis kelamin, atau profesi
Kebutuhan data	Membutuhkan banyak data sesuai dengan jumlah pengguna	Mengurangi kebutuhan jumlah data berdasarkan kesamaan profil pengguna
Kelengkapan data	Informasi individu yang dikumpulkan tidak lengkap	Data dikumpulkan berdasarkan kesamaan profil
Sasaran	Individu	Umum



Sistem rekomendasi dapat membantu dalam mengidentifikasi produk yang sesuai dengan kebutuhan, kesenangan, dan keinginan pengguna (Prasetya, 2017). Cara kerja sistem rekomendasi yaitu memodelkan selera (preferensi) pengguna untuk menyarankan (merekomendasikan) konten yang tidak terlihat yang menurut pengguna menarik. Preferensi profil pengguna dapat ditangkap secara implisit dan eksplisit (Eke et al., 2019).

Metode eksplisit mengumpulkan informasi pribadi pengguna seperti atribut demografis yang terdiri nama pengguna, alamat, nomor telepon, status pernikahan, minat, hobi. Sedangkan metode implisit dengan mengamati tindakan pengguna atau pola konsumsi (Eke et al., 2019). Selain itu interaksi yang berlangsung seperti preferensi pengguna diperoleh berdasarkan data preferensi yang dikumpulkan. Pengguna memberikan beberapa *item* yang memiliki nilai prediksi tertinggi (Knijnenburg et al., 2012).

Menurut Lee & Koubek (2010) membagi situs web menjadi empat kategori yaitu hiburan, informasi, komunikasi, dan bisnis (*commerce*). YouTube adalah contoh situs web dalam kategori hiburan; Berita BBC adalah salah satu kategori informasi; Facebook adalah kategori komunikasi; sedangkan amazon.com adalah jenis kategori bisnis. Level tertinggi dari situs web yang sedang tumbuh adalah kategori bisnis (*commerce*). *Commerce* atau *e-commerce* mencapai perkembangannya yang luar biasa karena memberikan banyak manfaat bagi pengguna (Purwati, 2011). Untuk vendor, *e-commerce* adalah media pemasaran yang luas, dimana pelanggan dapat dengan mudah menemukan produk (Al-Qaed & Sutcliffe, 2006). Meskipun demikian, keragaman informasi di situs web menghasilkan masalah lain bagi pengguna, yaitu banjir informasi.

Penelitian yang dilakukan oleh Payal Das, Jisha R C dan G P Sajeew yaitu personalisasi *website* sebagai sistem rekomendasi menggunakan metode *Splay Tree*. Hasil penelitian menunjukkan bahwa personalisasi *website* yang adaptif dilakukan dengan menganalisis minat pengguna yang berubah-ubah serta informasi halaman situs web sehingga memberikan hasil rekomendasi yang tepat dan komputasi lebih rendah. Performansi yang diperoleh yaitu sebesar 87.4% ketepatan rekomendasi yang relevan (Das et al., 2017). Selain itu, Prasetya juga melakukan penelitian tentang sistem rekomendasi pada *e-commerce* menggunakan metode *K-Nearest Neighbor*. Penelitian yang dilakukan menggunakan tiga pendekatan metode rekomendasi untuk perbandingan metode yang paling baik digunakan dalam sistem rekomendasi. Hasil penelitian menunjukkan bahwa pendekatan *Content Based* menghasilkan nilai rata-rata *precision* dan *F-measure* paling tinggi dibandingkan *Collaborative Filtering* dan *Hybrid* pada $k=10$ yaitu sebesar 0.080 dan 0.148. Sedangkan *recall* tertinggi dihasilkan dengan metode *Collaborative Filtering* pada $k=100$ yaitu 0.404. Maka apabila nilai k semakin tinggi, maka nilai *recall* yang dihasilkan juga semakin tinggi. Hal ini disebabkan karena semakin banyak rekomendasi produk yang diberikan kepada pembeli, maka nilai *true positive* yaitu barang yang direkomendasikan dan dibeli akan meningkat.

Data mining didefinisikan sebagai hubungan dan pola yang tersembunyi dalam data. Secara umum, model *data mining* dibagi menjadi dua sebagai prediktif dan deskriptif. Salah satu model deskriptif adalah aturan asosiasi (*association rules*). *Association rules* adalah metode yang membantu mengungkap hubungan antara data yang tampaknya tidak relevan. Ini menggunakan kombinasi analisis statistik, penambangan data, dan manajemen basis data untuk mengungkap hubungan tersembunyi yang ada (Percin et al., 2019). Akan ada 2 nilai yang dipertimbangkan pada *association rules* yaitu nilai *support* dan nilai *confidence*. *Association rules* hasil dari proses *mining* selanjutnya dapat digunakan untuk membentuk sistem rekomendasi dalam navigasi situs web.

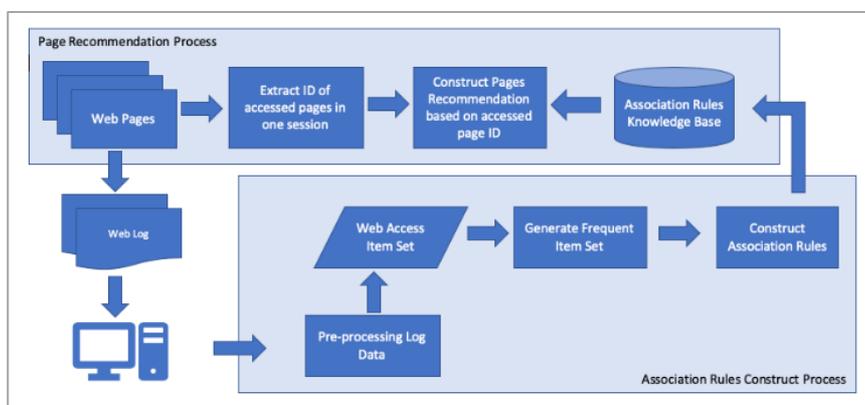
Penelitian ini menggunakan *Association Rules Mining (ARM)* sebagai salah satu proses *data mining* dengan mengidentifikasi hubungan antar *item* pada dataset dan mengetahui efektivitas metode tersebut dalam personalisasi *website*. Teknik tersebut digunakan dalam menentukan rekomendasi halaman web yang relevan sesuai dengan preferensi pengguna. Makalah disusun menjadi beberapa bagian sebagai berikut. Bagian II menjelaskan metodologi, termasuk



penjelasan teori *association rules mining*. Bagian III membahas tentang pokok penelitian, termasuk hasil Analisa performansi. Bagian IV mengekstrak kesimpulannya.

2. METODE PENELITIAN

Pada bagian ini model personalisasi *website* dengan menerapkan metode *Association Rules Mining* dipaparkan. Tujuan utama dari sistem adalah untuk menyarankan halaman web kepada pengguna sesuai dengan preferensi mereka. Untuk mencapai tujuan tersebut maka dilakukan dengan 2 fase yaitu fase pembentukan *association rules* dan fase pembentukan rekomendasi pada saat pengguna mengakses *website*. Fase pertama terdiri dari pembuatan *knowledge base* dengan menggunakan metode *association rules* yang dibentuk dari data log akses server web pada periode waktu tertentu. *Tools* yang digunakan dalam pembentukan *association rules* pada penelitian ini adalah RapidMiner Studio versi 9.6. Pada fase kedua, pembentukan rekomendasi dilakukan berdasarkan ID halaman-halaman yang diakses pengguna dalam satu *session* sebagai data *premises* untuk mendapatkan *conclusion* dari *knowledge base*. Usulan arsitektur sistem disajikan pada Gambar 1.



Gambar 1. Arsitektur Sistem.

2.1. Association Rules Mining

Association Rules Mining (ARM) adalah salah satu teknik *data mining* yang digunakan untuk menemukan pola-pola dataset. Dataset ARM biasanya dalam bentuk rangkaian waktu yang diperoleh dari transaksi berkala yang berulang. Pola yang diperoleh disebut *frequent pattern*, terdiri dari kombinasi *item* dengan hubungan kuat atau lemah. Biasanya hubungan kuat dari *frequent pattern* akan digunakan untuk membuat keputusan (Siswanto & Thariqa, 2018).

Ada 2 parameter yang sangat dipertimbangkan dalam *association rules mining* yaitu nilai penunjang (*support*) dan nilai keyakinan (*confidence*). Nilai *support* akan menunjukkan seberapa sering suatu *item* muncul terhadap semua transaksi catatan dataset dan nilai *confidence* menunjukkan seberapa kuat hubungan antara satu *item* dengan *item* lainnya (Siswanto & Thariqa, 2018).

Pada *item* X kemungkinan terdapat *item* Y dinotasikan $X \rightarrow Y$, sedangkan X dan Y adalah disjoint *itemset* dinotasikan $X \cap Y$. Kumpulan dari transaksi-transaksi ini disebut dengan *itemset*, yang dinotasikan dengan I_k ($k=1, 2, \dots, m$). Jika terdapat *itemset* yang mempunyai *item* sebanyak k, maka disebut dengan *k-itemset*.

Association rule ini nantinya akan menghasilkan *rules* yang menentukan seberapa besar hubungan antar X dan Y tadi, dan diperlukan dua ukuran untuk *rules* ini, yakni *support* dan *confidence*. Formula untuk mendapatkan nilai *support* dilihat pada Pers. (1) dari satu *item* dan Pers. (2) untuk 2 *item*, sedangkan formula untuk mendapatkan nilai *confidence* dilihat pada Pers. (3) (Percin et al., 2019).



$$\text{Support}(X) = \frac{\Sigma X}{\Sigma T} \quad (1)$$

$$\text{Support}(X \cap Y) = \frac{\Sigma(X \cap Y)}{\Sigma T} \quad (2)$$

$$\text{Confidence}(X \rightarrow Y) = \frac{\Sigma(X \cap Y)}{\Sigma X} \quad (3)$$

Dimana ΣX merupakan jumlah transaksi mengandung X , $\Sigma(X \cap Y)$ merupakan jumlah transaksi yang mengandung X dan Y , dan ΣT merupakan total transaksi.

Pada umumnya, *association rule* dapat dilihat sebagai dua proses, yaitu (Mulya, 2019):

- 1) Temukan semua *frequent itemsets*, dengan pengertian setiap *itemset* ini akan muncul sesering sekurangnya *minimum support* (min-sup) yang telah ditentukan.
- 2) Hasilkan *association rules* yang kuat dari *frequent itemsets*: dengan pengertian *rule* ini haruslah memenuhi nilai *minimum support* dan *minimum confidence*.

Adapun langkah-langkah proses pembentukan *association rule* pada sistem rekomendasi yaitu terdapat beberapa tahap sebagai berikut (Mulya, 2019):

- 1) Sistem memindai *database* untuk mendapatkan kandidat 1-*itemset* (himpunan *item* yang terdiri dari 1 *item*) dan menghitung nilai *support*. Lalu nilai *support* tersebut dibandingkan dengan *minimum support* yang telah ditentukan, jika nilainya lebih besar atau sama dengan *minimum support* maka *itemset* tersebut termasuk dalam *large itemset*.
- 2) *Itemset* yang tidak termasuk dalam *large itemset* tidak diikuti dalam iterasi selanjutnya.
- 3) Pada iterasi kedua sistem akan menggunakan hasil *large itemset* pada iterasi pertama (L1) untuk membentuk kandidat *itemset* kedua (L2). Pada iterasi selanjutnya sistem akan menggunakan hasil *large itemset* pada iterasi sebelumnya (Lk-1) untuk membentuk kandidat *itemset* berikut (Lk). Sistem akan menggabungkan (*join*) Lk-1 dengan Lk-1 untuk mendapatkan Lk, seperti pada iterasi sebelumnya sistem akan menghapus (*prune*) kombinasi *itemset* yang tidak termasuk dalam *large itemset*.
- 4) Setelah dilakukan operasi *join*, maka pasangan *itemset* baru hasil proses *join* tersebut dihitung *support*-nya.
- 5) Proses pembentuk kandidat yang terdiri dari proses *join* dan *prune* akan terus dilakukan hingga himpunan kandidat *itemset*-nya null, atau sudah tidak ada lagi kandidat yang akan dibentuk.
- 6) Setelah itu, dari hasil *frequent itemset* tersebut dibentuk *association rule* yang memenuhi nilai *support* dan *confidence* yang telah ditentukan.
- 7) Pada pembentukan *association rule*, nilai yang sama dianggap sebagai satu nilai.
- 8) *Association rule* yang terbentuk harus memenuhi nilai minimum yang telah ditentukan.
- 9) Untuk setiap *large itemset* L, kita cari himpunan bagian L yang tidak kosong. Untuk setiap himpunan bagian tersebut, dihasilkan *rule* dengan bentuk aB(L-a) jika *support*-nya (L) dan *support*-nya (a) lebih besar dari *minimum support*.

2.2. Persiapan Data

Dataset yang digunakan dalam penelitian ini merupakan *sparse DST file* yang perlu dilakukan proses pengolahan untuk menjadi data yang terstruktur. *Dataset* diambil dari <http://mlr.cs.umass.edu/ml/machine-learning-databases/anonymous/>.

Dataset ini dipersiapkan oleh Jack S. Breese, David Heckerman dan Carl M. Kadie pada bulan November 1998 memuat data log akses web secara anonim dari situs www.microsoft.com. Data ini mencatat log akses *website* dengan 38.000 pengguna anonim yang dipilih secara acak. Untuk setiap pengguna, data mencantumkan semua area situs web (Vroots) sebagai atribut yang dikunjungi dalam jangka waktu satu minggu. Atribut Vroots terdiri dari 294 yang diidentifikasi berdasarkan judul (mis. "NetShow for PowerPoint") dan URL (mis. "/Stream"). Untuk setiap baris data terdapat beberapa informasi yaitu ID area situs web atau halaman *website* (ID Vroots), Judul Vroots, URL. Sampel *dataset* dapat dilihat pada Gambar 2.



```
A,1287,1,"International
AutoRoute","/autoroute"
A,1288,1,"library","/library"
A,1289,1,"Master Chef Product
Information","/masterchef"
```

Gambar 2. Sampel Data Vroots.

Sedangkan data log pengguna terdiri dari data *training* sejumlah 32.711 dan data *testing* sejumlah 5.000 dengan memilih secara acak dari pengguna *website*. Pada data log pengguna terdiri dari beberapa atribut yaitu ID *User* dan ID Vroots atau ID halaman yang pengguna kunjungi. Sampel dataset dapat dilihat pada Gambar 3.

```
C,"10001",10001
V,1000,1
V,1001,1
V,1002,1
C,"10002",10002
V,1001,1
V,1003,1
C,"10003",10003
V,1001,1
V,1003,1
V,1004,1
C,"10004",10004
V,1005,1
```

Gambar 3. Sampel Data Log Akses Pengguna.

3. HASIL DAN PEMBAHASAN

3.1. *Pre-Processing*

Merujuk pada penelitian yang dilakukan oleh Das et al., (2017) Teknik data *pre-processing* digunakan untuk mengubah data mentah dari web menjadi format terstruktur. Data mentah dari web memiliki berbagai entri yang tidak diinginkan sehingga mempengaruhi kinerja sistem. Pada penelitian menggunakan dataset file log server *website* anonim www.microsoft.com. Dengan menggunakan metode pengikisan web informasi dari halaman web diekstraksi untuk membuat dataset konten web. Ini terdiri dari konten teks dan informasi tautan seperti ID halaman, URL halaman, dan judul halaman. *Pre-processing* pada data log dilakukan dalam dua langkah yaitu pembersihan dan identifikasi pengguna.

Tahap pembersihan meliputi penghapusan data yang salah dan tidak diinginkan dari file log. Identifikasi pengguna dilakukan berdasarkan *Session ID*. *Session ID* unik diambil untuk mengidentifikasi pengguna yang berbeda. Identifikasi pengguna digunakan untuk mengolah data log server dengan mengelompokkan data halaman yang mereka kunjungi dalam satu *session* menjadi data *itemset* yang menunjukkan halaman apa saja yang dikunjungi pengguna dalam satu *session*. Dari proses ini menghasilkan data *itemset* dari *training* dataset seperti pada Gambar 4.

Row No.	caseid	vrootset
1	10001	1000 1001 1002
2	10002	1001 1003
3	10003	1001 1003 1004
4	10004	1005
5	10005	1006
6	10006	1003 1004
7	10007	1007
8	10008	1004

Gambar 4. Sampel Training Itemset.



3.2. Generate Frequent Itemset

Dalam proses membentuk *association rules*, hal pertama yang harus diketahui adalah halaman mana yang paling sering muncul dalam *itemset* akses pengunjung *website*. Untuk itu algoritma FP-Growth kami gunakan dalam proses pembentukan *frequent itemset*. Dalam penelitian ini nilai *support* minimal yang diperhitungkan adalah 0.3%, sehingga *frequent itemset* yang diperhitungkan dalam penelitian ini adalah *itemset* yang minimal memiliki frekuensi 100 kali dari 32.711 *itemset*.

Dari proses ini menghasilkan 1.014 *frequent itemset* dengan *maximal size itemset* 6. Contoh dari *frequent itemset* yang dihasilkan dapat dilihat pada Gambar 5.

Size	Sup... ↑	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6
1	0.003	1063					
1	0.003	1019					
1	0.003	1090					
1	0.003	1050					
1	0.003	1083					
1	0.003	1095					
2	0.003	1008	1075				
2	0.003	1008	1021				
2	0.003	1008	1049				
2	0.003	1008	1064				

Gambar 5. Sampel *Frequent Itemset*.

3.3. Konstruksi *Association Rules*

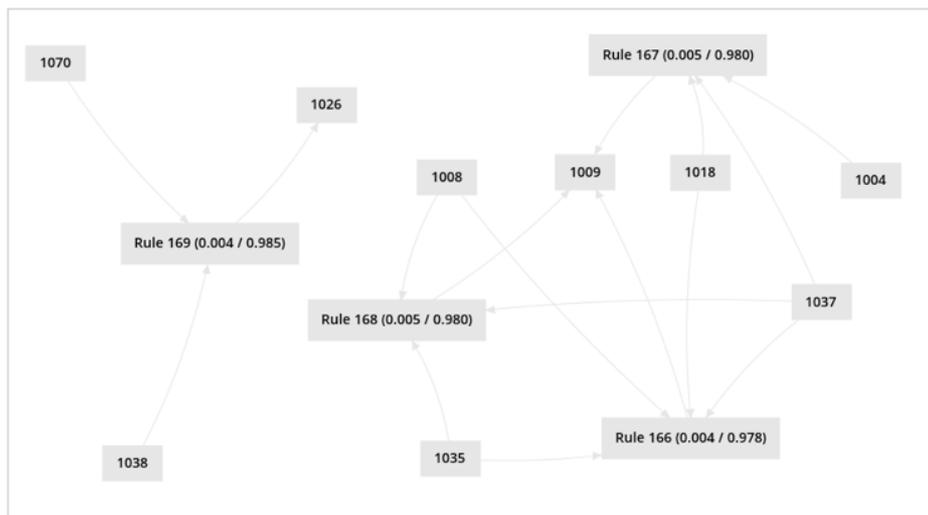
Hasil dari proses pembentukan *frequent itemset* digunakan untuk merekonstruksi model *association rules*, sehingga terbentuk basis pengetahuan sistem berupa pola dari preferensi pengguna. Pola yang terbentuk disebut *frequent pattern*, terdiri dari kombinasi *item* dengan hubungan kuat atau lemah. Hubungan kuat dari *frequent pattern* akan digunakan untuk membuat rekomendasi halaman *website*. Pola akan dihitung berdasarkan nilai *confidence* dan nilai *support* yang mengindikasikan kualitas dari prediksi menggunakan model tersebut.

Proses rekomendasi pada *website* didapat dengan cara menggunakan daftar halaman (*vroot_id*) yang diakses pengguna dalam satu *session* sebagai *premises* dengan memperhatikan nilai *confidence* dan *support* sehingga diperoleh *conclusion* dari *association rules* berupa daftar halaman yang dapat direkomendasikan kepada pengguna. Nilai *confidence* yang diperhitungkan dalam penelitian ini adalah minimal 80%. Proses ini menghasilkan 150 *rules*, ilustrasi dari *association rules* ini dapat dilihat pada Gambar 6 dan Gambar 7.

Premises	Conclusion	Support	Confidence ↓
1038, 1070	1026	0.004	0.985
1008, 1035, 1037	1009	0.005	0.980
1004, 1018, 1037	1009	0.005	0.980
1008, 1018, 1035, 1037	1009	0.004	0.978

Gambar 6. *Association Rules*.





Gambar 7. Ilustrasi Association Rules.

3.4. Analisis Kinerja

3.4.1. Efektivitas Sistem

Pengujian dilakukan dengan mengukur performa metode yang digunakan yaitu menghitung *precision* dan *recall*. *Precision* proporsi *item* yang direkomendasikan yang relevan dari jumlah total *item* yang direkomendasikan (Bobadilla et al., 2013). Untuk menghitung *precision* ditunjukkan pada Pers. (4).

$$\text{precision} = \frac{TP}{TP+FP} \quad (4)$$

Sedangkan *recall* merupakan proporsi *item* yang direkomendasikan yang relevan dari jumlah *item* yang relevan (Bobadilla et al., 2013). Untuk menghitung *recall* ditunjukkan pada Pers. (5).

$$\text{recall} = \frac{TP}{TP+FN} \quad (5)$$

F-Measure merupakan nilai rata-rata yang mewakili kinerja sistem dari nilai *precision* dan *recall* (Garrigós et al., 2010). Pers. (6) untuk menghitung nilai *F-Measure*.

$$F - \text{Measure} = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \quad (6)$$

Penelitian menggunakan *confusion matrix* yang umum digunakan seperti *True Positive* (TP), *True Negative* (TN), *False Positive* (FP) dan *False Negative* (FN). Jika kesamaan antara prediksi dan referensi sama persis maka disebut sebagai TP, atau disebut juga FP. Adapun penjelasan matriks model yang digunakan dapat dilihat pada Tabel 2.

Tabel 2. Pemetaan Matriks.

Metriks	Umum	Model Penelitian
<i>True Positive</i>	Diidentifikasi dengan benar	Minat, Rekomendasi
<i>True Negative</i>	Diidentifikasi dengan salah	Tidak Minat, Tidak Rekomendasi
<i>False Positive</i>	Ditolak dengan benar	Tidak Minat, Rekomendasi
<i>False Negative</i>	Ditolak dengan salah	Minat, Tidak Rekomendasi

Pengukuran performansi dengan menggunakan *test dataset* sejumlah 5.000 *record* terhadap 150 *association rules* yang terbentuk dari *training dataset* didapatkan hasil performansi seperti pada Tabel 3.



Tabel 3. Perfomansi.

Parameter	Performance
<i>Precision</i>	0.896
<i>Recall</i>	0.058
<i>F-Measure</i>	0.104

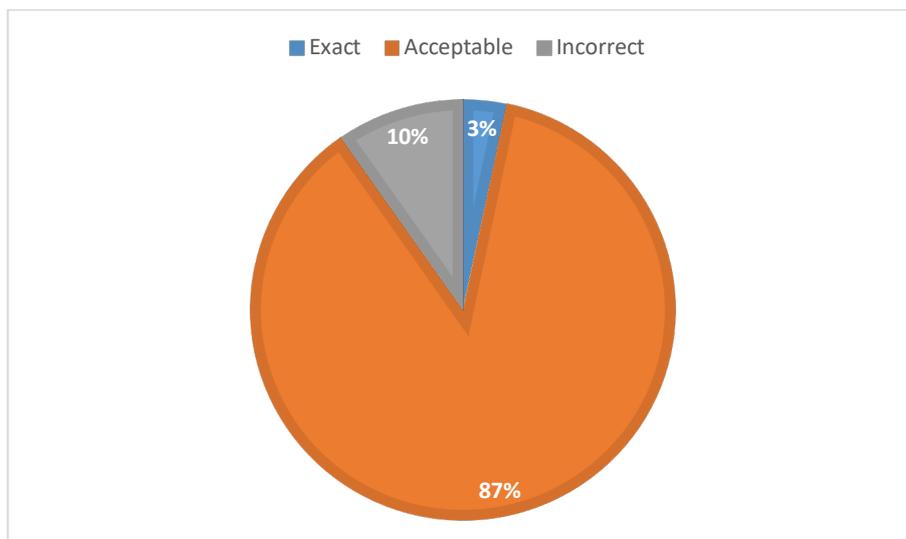
Pada Tabel 3 di atas dijelaskan bahwa nilai *precision* yang diperoleh sebesar 0.896 merupakan proporsi *item* yang direkomendasikan benar dan relevan dengan jumlah total *item* yang direkomendasikan. Nilai *recall* sebesar 0.058 merupakan proporsi *item* yang direkomendasikan benar dan relevan dari jumlah *item* yang relevan. Nilai *F-measure* sebesar 0.104 merupakan nilai rata-rata mewakili kinerja sistem dengan menjumlahkan nilai *precision* dan *recall*.

3.4.2. Ketepatan Rekomendasi

Mempertimbangkan sistem akses web dengan N pengguna. Lalu P didefinisikan sebagai set halaman yang menarik dan R halaman yang direkomendasikan. Kita mengklasifikasikan kedalam 3 kategori yang diberi nilai RC yaitu:

- *exact*, if $|P - R| == 0$
- *acceptable*, if $0 < |P - R| < length(P)$
- *incorrect*, if $|P - R| == P$

Pengukuran performansi ketepatan rekomendasi dengan menggunakan *test dataset* sejumlah 5.000 *record* terhadap 150 *association rules* yang terbentuk dari *training dataset* didapatkan hasil performansi seperti pada Gambar 8.



Gambar 8. Hasil Ketepatan Rekomendasi.

Pada Gambar 8 dijelaskan bahwa hasil rekomendasi kategori *Exact* (tepat) sebesar 3% yang berarti jumlah halaman yang menarik sama dengan jumlah halaman yang direkomendasikan. Kategori *Acceptable* (dapat diterima) sebesar 87% artinya jumlah halaman yang direkomendasikan relevan dengan set halaman yang menarik. Sedangkan kategori *Incorrect* (salah) sebesar 10% artinya jumlah halaman yang direkomendasikan tidak relevan dengan set jumlah halaman yang menarik.

4. KESIMPULAN

Pendekatan yang dilakukan dengan menggunakan sistem rekomendasi terbukti dapat menangani sebagian informasi yang berlebihan pada halaman *website* sehingga menyajikan model personalisasi *website* yang adaptif berdasarkan kebutuhan dinamis pengguna. *Association*



rules merupakan metode *data mining* yang digunakan untuk menemukan pola yang terdapat dalam kumpulan data. Pola yang terbentuk dari *association rules* dapat digunakan dalam merekomendasikan sesuai dengan preferensi pengguna yang berbeda-beda. Model yang diusulkan divalidasi melalui eksperimen.

Association rules dapat mencari *top-n* rekomendasi halaman *website* berdasarkan model basis pengetahuan yang terbentuk dari data server log. Pengujian dilakukan pada 5000 *record data* dan dihasilkan 150 *rules* dari pola yang terbentuk menggunakan metode *association rules mining*. Dari pengujian yang telah dilakukan, efektivitas metode *association rules mining* dalam personalisasi *website* menghasilkan performansi nilai *precision* sebesar 0.896 artinya proporsi *item* yang direkomendasikan benar dan relevan dengan jumlah total *item* yang direkomendasikan. Nilai *recall* sebesar 0.058 merupakan proporsi *item* yang direkomendasikan benar dan relevan dari jumlah *item* yang relevan. Nilai *F-measure* sebesar 0.104 merupakan nilai rata-rata mewakili kinerja sistem dengan menjumlahkan nilai *precision* dan *recall*. Selain itu performansi ketepatan dalam rekomendasi menghasilkan performansi 87% artinya halaman yang direkomendasikan relevan dengan halaman yang menarik dan diminati oleh pengunjung. Sedangkan rekomendasi tepat sebesar 3% dan rekomendasi salah sebesar 10%.

Penelitian selanjutnya dapat berkonsentrasi pada meningkatkan metode yang ada sehingga personalisasi *website* menjadi lebih adaptif melalui pendekatan sistem rekomendasi. Hal yang dapat dilakukan seperti mendapatkan nilai parameter yang paling optimal dengan waktu komputasi yang rendah, kombinasi yang tepat dari metode rekomendasi yang ada dengan menggunakan berbagai jenis informasi yang tersedia. Dari sisi pengguna, dapat mengikuti tren yang berkaitan dengan kebiasaan, konsumsi dan selera pengguna dalam proses rekomendasi.

UCAPAN TERIMA KASIH

Kami berterimakasih kepada Universitas Teknologi Yogyakarta atas kesempatan yang diberikan dalam penelitian ini. Kami sangat mengapresiasi atas kontribusi yang diberikan.

DAFTAR PUSTAKA

- Al-Qaed, F., & Sutcliffe, A. (2006). Adaptive decision support system (ADSS) for B2C e-commerce. *Proceedings of the 8th International Conference on Electronic Commerce The New E-Commerce: Innovations for Conquering Current Barriers, Obstacles and Limitations to Conducting Successful Business on the Internet - ICEC '06*, 492. <https://doi.org/10.1145/1151454.1151528>
- Bobadilla, J., Ortega, F., Hernando, A., & Gutiérrez, A. (2013). Recommender systems survey. *Knowledge-Based Systems*, 46, 109–132. <https://doi.org/10.1016/j.knosys.2013.03.012>
- Das, P., Jisha, R. C., & Sajeev, G. P. (2017). Adaptive web personalization system using splay tree. *2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, 1582–1587. <https://doi.org/10.1109/ICACCI.2017.8126067>
- Eke, C. I., Norman, A. A., Shuib, L., & Nweke, H. F. (2019). A Survey of User Profiling: State-of-the-Art, Challenges, and Solutions. *IEEE Access*, 7, 144907–144924. <https://doi.org/10.1109/ACCESS.2019.2944243>
- Garrigós, I., Gomez, J., & Houben, G.-J. (2010). Specification of personalization in web application design. *Information and Software Technology*, 52(9), 991–1010. <https://doi.org/10.1016/j.infsof.2010.04.001>
- Han, K., Yi, M. Y., & Kim, J. (2019). Search Personalization in Folksonomy by Exploiting Multiple and Temporal Aspects of User Profiles. *IEEE Access*, 7, 95610–95619. <https://doi.org/10.1109/ACCESS.2019.2927026>
- Hawalah, A., & Fasli, M. (2015). Dynamic user profiles for web personalisation. *Expert Systems with Applications*, 42(5), 2547–2569. <https://doi.org/10.1016/j.eswa.2014.10.032>
- Kanoje, S., Girase, S., & Mukhopadhyay, D. (2015). User Profiling Trends, Techniques and Applications. *International Journal of Advance Foundation and Research in Computer (IJAFRC)*, 1(1).
- Knijnenburg, B. P., Willemsen, M. C., Gantner, Z., Soncu, H., & Newell, C. (2012). Explaining the user experience of recommender systems. *User Modeling and User-Adapted Interaction*,



- 22(4–5), 441–504. <https://doi.org/10.1007/s11257-011-9118-4>
- Lee, S., & Koubek, R. J. (2010). The effects of usability and web design attributes on user preference for e-commerce web sites. *Computers in Industry*, 61(4), 329–341. <https://doi.org/10.1016/j.compind.2009.12.004>
- Mulya, D. P. (2019). Analisa Dan Implementasi Association Rule Dengan Algoritma Fp-Growth Dalam Seleksi Pembelian Tanah Liat. *Teknologi Dan Sistem Informasi Bisnis*, 1(1), 47–57.
- Percin, I., Yagin, F. H., Guldogan, E., & Yologlu, S. (2019). ARM: An Interactive Web Software for Association Rules Mining and an Application in Medicine. *2019 International Artificial Intelligence and Data Processing Symposium (IDAP)*, 1–5. <https://doi.org/10.1109/IDAP.2019.8875885>
- Prasetya, C. S. D. (2017). Sistem Rekomendasi Pada E-Commerce Menggunakan K-Nearest Neighbor. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, 4(3), 194. <https://doi.org/10.25126/jtiik.201743392>
- Purwati, Y. (2011). STANDARD FEATURES OF E-COMMERCE USER INTERFACE FOR THE WEB. *Journal of Arts, Science & Commerce*, 2(3), 77–87.
- Rianto, Nugroho, L. E., & Santosa, P. I. (2016). Pattern discovery of Indonesian customers in an online shop: A case of fashion online shop. *2016 3rd International Conference on Information Technology, Computer, and Electrical Engineering (ICITACEE)*, 313–316. <https://doi.org/10.1109/ICITACEE.2016.7892462>
- Siswanto, B., & Thariqa, P. (2018). Association Rules Mining for Identifying Popular Ingredients on YouTube Cooking Recipes Videos. *2018 Indonesian Association for Pattern Recognition International Conference (INAPR)*, 95–98. <https://doi.org/10.1109/INAPR.2018.8627002>
- Sulikowski, P., Zdziebko, T., Turzyński, D., & Kańtoch, E. (2018). Human-website interaction monitoring in recommender systems. *Procedia Computer Science*, 126, 1587–1596. <https://doi.org/10.1016/j.procs.2018.08.132>

