

# ANALISIS POLA PERMINTAAN PUBLIKASI DATA BADAN PUSAT STATISTIK MENGGUNAKAN ASSOCIATION RULE APRIORI

Fachruddin Mansyur<sup>1</sup>, Farid Ridho<sup>2</sup>

<sup>1</sup> BPS Kabupaten Majene

Jl. Jend. Ahmad Yani, Banggae Majene Sulawesi Barat

<sup>2</sup>Program Studi Komputasi Statistik

Politeknik Statistika STIS

Jalan Otto Iskandardinata No 64C Jakarta

email: <sup>1</sup>fachruddin.mansyur@bps.go.id, <sup>2</sup>faridr@stis.ac.id

## **Abstract**

*BPS is a data provider body in Indonesia. In publishing, BPS uses a variety of media, one of which is the BPS website. To get data through the BPS website, users can visit the website then download the data they need. The services obtained by data users on the BPS website depend on the quality of the website. The better the quality, the better the service experience gained by data users. The method that can be used to improve the quality of a website is the web usage mining method. Web usage mining is the application of data mining techniques on web repositories to study usage patterns. The purpose of this study is to determine the pattern of data publication requests on the BPS website which can later be used as a reference to improve the quality of BPS website services. Based on the results of the study, it was found that data users tend to access the same data with different years simultaneously. For results by grouping data by title without year, obtained quite diverse rules.*

**Keywords:** web usage mining, association rule, apriori

## **Abstrak**

*BPS merupakan badan penyedia data di Indonesia. Dalam mempublikasikan datanya, BPS menggunakan berbagai media, salah satunya adalah website BPS. Untuk mendapatkan data melalui website BPS, pengguna dapat mengunjungi website kemudian mengunduh data yang mereka butuhkan. Layanan yang didapatkan oleh pengguna data pada website BPS tergantung dari kualitas website tersebut. Semakin baik kualitasnya, semakin baik pula pengalaman pelayanan yang didapatkan oleh pengguna data. Metode yang dapat digunakan untuk meningkatkan kualitas suatu website adalah metode web usage mining. Web usage mining merupakan penerapan teknik data mining pada web repositori untuk mempelajari pola penggunaan. Tujuan dari penelitian ini adalah untuk mengetahui pola permintaan publikasi data pada website BPS yang nantinya dapat digunakan sebagai acuan untuk meningkatkan kualitas layanan website BPS. Berdasarkan hasil penelitian, didapatkan bahwa pengguna data cenderung mengakses data yang sama dengan tahun yang berbeda secara bersamaan. Untuk hasil dengan mengelompokkan data berdasarkan judul tanpa tahun, diperoleh rules yang cukup beragam.*

**Kata kunci:** web usage mining, association rule, apriori

## 1. PENDAHULUAN

Data merupakan suatu yang penting, terutama bagi para pembuat keputusan dan penentu kebijakan. Data merupakan gambaran tentang suatu keadaan ataupun peristiwa dalam bentuk nilai. Agar sesuai dengan kebutuhan dan kemampuan, digunakan data sebagai dasar dari sebuah perencanaan sehingga perencanaan yang kurang ataupun berlebihan dapat dicegah. Selain sebagai dasar perencanaan, data juga digunakan sebagai alat pengendalian dalam suatu pelaksanaan dan sebagai dasar evaluasi dari hasil kerja yang telah ditargetkan.

Untuk memperoleh data, dapat dilakukan pengumpulan data. Pengumpulan data dapat dilakukan dengan berbagai cara dan metode yang telah ada. Misal pada suatu perusahaan yang ingin mendapatkan data tentang produksi yang dilakukan, dapat diperoleh dengan mencatat produksi yang dihasilkan, baik dalam harian, mingguan, maupun bulanan. Selain dengan mengumpulkan sendiri, dapat juga diperoleh data yang telah dikumpulkan oleh badan-badan penyedia data seperti untuk data kependudukan dan perekonomian negara.

Badan Pusat Statistik atau biasa disingkat dengan BPS adalah lembaga penyedia data di Indonesia. BPS merupakan Lembaga Pemerintah non Kementerian yang bertanggung jawab langsung kepada Presiden. Berdasarkan undang-undang nomor 16 tahun 1997 tentang statistik, peranan yang harus dijalankan oleh BPS adalah: Menyediakan kebutuhan data bagi pemerintah dan masyarakat yang didapatkan dari sensus ataupun survei yang dilakukan sendiri dan juga dari departemen atau lembaga pemerintahan lainnya sebagai data sekunder; membantu kegiatan statistik di kementerian, lembaga pemerintah atau institusi lainnya dalam membangun sistem perstatistikan nasional; mengembangkan dan mempromosikan standar teknik dan metodologi statistik, dan menyediakan pelayanan pada bidang pendidikan dan pelatihan statistik; membangun kerjasama dengan institusi internasional dan negara lain untuk kepentingan perkembangan statistik Indonesia [8].

Untuk mendapatkan data yang telah disediakan BPS, pengguna data dapat datang langsung ke BPS atau mengunjungi *website* BPS. Pengguna data yang datang langsung ke BPS dapat langsung ke pusat layanan untuk memberitahukan kebutuhan mereka dan nantinya akan dipandu oleh pegawai yang bertugas. Berbeda dengan pengguna yang melalui *website* BPS, mereka harus mencari sendiri data yang dibutuhkan. Mudah tidaknya pengguna data menemukan data yang dibutuhkan tergantung dari pengetahuan mereka tentang nama atau istilah data yang akan mereka cari dan menu-menu serta fitur yang tersedia pada *website* yang dapat membantu menemukan data.

Keberadaan *website* BPS adalah untuk memberikan layanan dan kemudahan kepada pengguna data untuk memperoleh data. Namun berdasarkan analisis hasil survei kebutuhan yang dilakukan BPS, kemudahan memperoleh data dari *website* BPS selalu memiliki nilai rendah jika dibandingkan dengan indikator pelayanan lainnya. Pada tahun 2016, kemudahan memperoleh data dari *website* BPS memiliki kepuasan sebesar 73,95% yang merupakan terendah kedua setelah kepuasan ketersediaan fasilitas *wifi*. Pada tahun 2017 mengalami hal yang sama, kemudahan memperoleh data dari *website* BPS sebesar 79,09% yang juga terendah setelah

ketersediaan fasilitas *wifi*. Kemudian pada tahun 2018 kemudahan memperoleh data dari *website* BPS memiliki nilai yang paling rendah yaitu sebesar 87,69% dibandingkan dengan indikator yang lainnya. Hal ini menunjukkan belum terpenuhinya tujuan *website* BPS untuk memudahkan pelayanan data dan perlunya untuk meningkatkan kualitas *website* BPS.

Untuk meningkatkan kualitas suatu *website*, tidak cukup hanya memperhatikan sisi struktur dan kontennya saja, sisi penggunaannya juga harus diperhatikan. Siapa saja yang sering mengunjungi, pada waktu apa saja sering dikunjungi, berapa banyak yang sering mengunjungi, dan hal lainnya yang berkaitan dengan penggunaan *website* perlu diperhatikan. Dari pola penggunaan *website* ini, nantinya dapat dilakukan personalisasi *website* dan beberapa hal lainnya serta pengoptimalan *server* pada waktu-waktu tertentu yang dianggap perlu.

*Web usage mining* merupakan metode *web mining* untuk mencari pola penggunaan *website*. Berdasarkan pengertian yang kemukakan oleh Cooley *web usage mining* adalah penerapan *teknik data mining* pada *web repositori* untuk mempelajari pola penggunaan [1]. Pola penggunaan *website* diporeliah dari hasil analisis data *web log*. *Web usage mining* dimulai dari proses mengambil data hingga proses mencari dan menganalisis pola penggunaan. Untuk mengetahui pola pengguna data *website* dapat dilakukan dengan beberapa teknik, diantaranya: *statistical analysis*, *assosiation rules*, *clustering*, *classification*, *sequential pattern*, dan *dependency modeling* [4].

Pada tahun 2018, BPS melakukan perekaman data transaksi pengunjung *website* BPS. Sebelum mengunduh data yang dibutuhkan, pengguna terlebih dahulu harus membuat akun pada *website* BPS. Setiap kali pengguna data mengunduh data dan informasi pada *website* akan direkam dan disimpan dalam *database*. Beberapa variabel yang direkam dari transaksi pengunjung *website* ini adalah data diri pengunjung, konten yang diakses, dan tanggal pengaksesan.

Berdasarkan uraian diatas dan dalam rangka pemberian saran untuk meningkatkan kualitas layanan dan bahan evaluasi terhadap *website* BPS, penelitian ini bertujuan menganalisis pola permintaan publikasi data pengunjung *website* BPS. Oleh sebab itu, digunakan salah satu teknik yaitu *association rules* untuk menganalisis pola perilaku pengunjung *website* BPS. *Association rules* merupakan suatu metode untuk menemukan hubungan antara *item* dalam suatu *dataset*. Terdapat beberapa algoritma yang dapat digunakan dalam *association rules*, dalam penelitian ini menggunakan salah satu algoritma asosiasi sederhana dan mudah diterapkan yang dikenal dengan algoritma *apriori*.

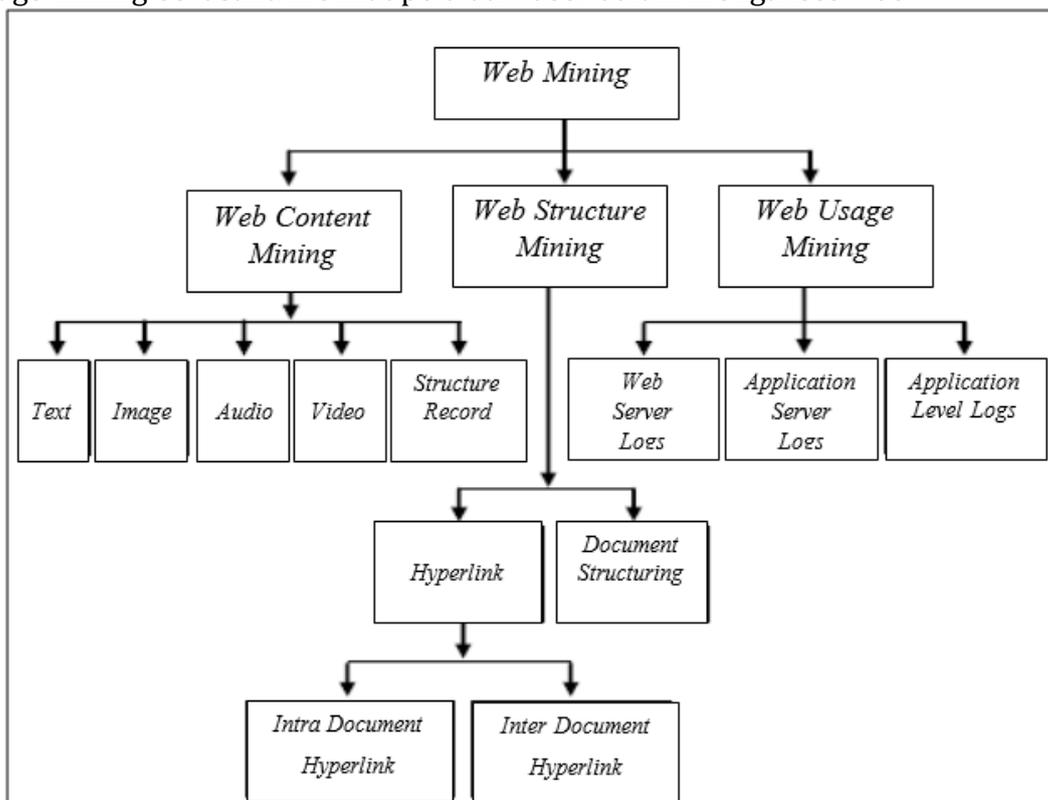
## 2. METODOLOGI PENELITIAN

### 2.1 Studi Literatur

#### 2.2.1 Data Mining dan Web Mining

Data mining dapat didefinisikan sebagai proses yang terlibat dalam penggalian informasi yang menarik, dapat ditafsirkan, berguna dan baru dari data. Beberapa metode data mining diantaranya *classification*, *regression*, *clustering*, *summarization*, *dependency modeling*, dan *Change and deviation detection* [6].

Web mining merupakan pengaplikasian teknik data mining kedalam World Wide Web. Web mining dapat didefinisikan sebagai pencarian dan analisis informasi yang berguna dari World Wide Web [2]. Terdapat 3 (tiga) kategori web mining yaitu web content mining, web structure mining, dan web usage mining [7]. Web content mining berfokus pada usaha menggali informasi dari isi atau content yang disajikan pada web. Web structure mining membahas mengenai penggalian informasi dari struktur halaman web yang saling berhubungan. Sementara web usage mining berusaha melihat pola dari user dalam mengakses web.



Gambar 1 Taksonomi Web Mining [7]

### 2.2.2 Web Usage Mining

*Web usage mining* adalah penerapan teknik *data mining* pada web repositori untuk mempelajari pola penggunaan. Sumber data *web usage mining* dapat berasal dari *web access logs*, *referrer logs*, dan *user registration*. Secara garis besar proses *web usage mining* terbagi menjadi 3 (tiga) fase, yaitu *preprocessing*, *pattern discovery*, dan *pattern analysis* [1][2]. Pada fase *pattern discovery* dapat digunakan beberapa teknik, antara lain: *statistical analysis*, *association rules*, *clustering*, *classification*, *sequential pattern*, dan *dependency modeling* [3].

### 2.2.3 Association Rules Mining

*Association rule mining* adalah teknik data mining untuk menemukan aturan menarik antara suatu kombinasi item. Definisi *association rule mining* yang diperkenalkan oleh Agrawal adalah sebagai berikut:  $I = \{i_1, i_2, \dots, i_m\}$  sebagai atribut

biner, disebut *items* [3].  $T$  sebagai set transaksi dalam *database*. Setiap transaksi  $t$  direpresentasikan sebagai vector biner, dengan  $t[k] = 1$  jika  $t$  membeli item  $i_k$ , dan  $t[k] = 0$  untuk yang lainnya. Terdapat satu *tuple* dalam *database* untuk setiap transaksi.  $X$  merupakan set item di  $I$ . Sebuah *rule* didefinisikan sebagai sebuah *implication* dengan bentuk  $X \Rightarrow i_j$  dimana  $i_j$  adalah *single* item yang tidak terdapat pada set item  $X$ . Dalam Agrawal et al (1994), bentuk umum dari sebuah *rule* didefinisikan dengan  $X \Rightarrow Y$  dan  $X \cap Y = \emptyset$ .  $X$  dan  $Y$  adalah set item yang berbeda, dimana  $X$  disebut *antecedent* (*left-hand-side* atau LHS) dan  $Y$  *consequent* (*right-hand-side* atau RHS) [4]. Untuk menentukan *rules* yang menarik dari sekumpulan *rules*, digunakan ukuran *support*, *confidence*, dan *lift*.

$supp(X)$  dari sebuah itemset  $X$  didefinisikan sebagai proporsi transaksi dalam data set yang mengandung *itemset* [10]. *Support* digunakan untuk melihat berapa sering *itemset* muncul dalam seluruh transaksi.

$$supp(X) = \frac{\text{transaksi itemset } X}{\text{total transaksi}} \quad (1)$$

*Confidence* adalah indikator yang memperlihatkan kemungkinan suatu item dibeli dengan syarat item tertentu juga dibeli. *Confidence* didefinisikan sebagai berikut (Hornik et al, 2005):

$$conf(X \Rightarrow Y) = \frac{supp(X \cup Y)}{supp(X)} \quad (2)$$

*Lift* merupakan indikator yang digunakan untuk menemukan *rules* yang menarik. Nilai *lift* menunjuk an kevalidan proses transaksi dan memberikan informasi apakah suatu item benar dibeli bersama dengan item yang lain. *Lift* didefinisikan sebagai berikut:

$$lift(X \Rightarrow Y) = \frac{supp(X \cup Y)}{supp(X)supp(Y)} \quad (3)$$

$Lift = 1$  mengartikan *antecedent* dan *consequent* independen. *Rules* tidak dapat dibentuk dari keduanya.  $Lift < 1$ , berarti bahwa *items* dalam *rules* saling menggantikan. Kehadiran *item* yang satu menggantikan *item* yang lain.  $Lift > 1$ , berarti *items* dalam rule saling bergantung satu sama lain [11].

### Algoritma Apriori

Algoritma Apriori merupakan algoritma *association rule* yang diperkenalkan oleh Arawal & Srikant pada tahun 1994. Algoritma Apriori adalah algoritma yang melakukan *scanning* pada item dalam *database* kemudian mengeliminasi item yang tidak memenuhi minimum *support* yang telah ditentukan. Selanjutnya dari kombinasi item yang memenuhi dilakukan *scanning* kembali dan dihitung lagi minimum *support*-nya, seterusnya hingga tidak dapat dibuat kombinasi item.

Algoritma Apriori secara lengkap sebagai berikut: Tahap pertama menghitung semua kemunculan *itemset* untuk menentukan *large 1-itemset*. Tahapan selanjutnya, katakan tahap  $k$ , terdiri dari dua fase. Pertama *large itemset*  $L_{k-1}$  yang didapat dari tahap  $k-1$  digunakan untuk membangkitkan *candidate*

itemset  $C_k$ , menggunakan fungsi apriori-gen. Selanjutnya, lakukan *scanning* pada database lalu hitung *support* dari candidate  $C_k$ .

```
1)  $L_1 = \{large\ 1\text{-itemset}\}$ 
2) for ( $k = 2; L_{k-1} \neq \emptyset; k++$ ) do begin
3)    $C_k = \text{apriori-gen}(L_{k-1}); // \text{New candidate}$ 
4)   forall transaction  $t \in T$  do begin
5)      $C_t = \text{subset}(C_k, t); // \text{Candidatea contained in } t$ 
6)     forall candidates  $c \in C_t$  do
7)        $c.\text{count}++;$ 
8)   end
9)    $L_k = \{c \in C_k \mid c.\text{count} \geq \text{minsup}\}$ 
10) end
11) Answer =  $U_k L_k;$ 
Apriori-gen
insert into  $C_k$ 
select  $p.\text{item}_1, p.\text{item}_2, \dots, p.\text{item}_{k-1}, q.\text{item}_{k-1}$ 
from  $L_{k-1} p, L_{k-1} q$ 
where  $p.\text{item}_1 = q.\text{item}_1, \dots, p.\text{item}_{k-2} = q.\text{item}_{k-2},$ 
 $p.\text{item}_{k-1} < q.\text{item}_{k-1};$ 

forall itemset  $c \in C_k$  do
  forall  $(k-1)$ -subsets  $s$  of  $c$  do
    if ( $s \notin L_{k-1}$ ) then
      delete  $c$  from  $C_k;$ 
```

#### *Package arules dan arulesviz*

*Package arules* merupakan *library* untuk R yang menyediakan infrastruktur yang diperlukan untuk membuat dan memanipulasi *dataset* untuk algoritma *data mining* dan untuk menganalisis *itemset* dan *rules* yang dihasilkan [10]. *Package* ini menggunakan *sparse matrix* untuk meminimalkan penggunaan memori. Dalam *package* ini tersedia dua algoritma *association rules* yaitu *Apriori* dan *Eclat*.

*Package arulesViz* merupakan *tools* pada R yang mengimplementasikan teknik visualisasi untuk mengeksplorasi *association rules*. Teknik visualisasi yang digunakan seperti *scatter plots*, *matrix visualizations*, *graphs*, *mosaic plots*, *parallel coordinates plots*, dan metode baru yang dikembangkan yaitu *grouped matrix-based visualization* [9].

## 2.2 Cakupan Penelitian

Data yang digunakan dalam penelitian ini adalah data pengunjung website BPS Pusat tahun 2018. Data yang digunakan dalam penelitian ini merupakan data sekunder. Data tersebut didapatkan dari *subject matter* dari Subdirektorat Layanan dan Promosi Statistik BPS Republik Indonesia.

## 2.3 Metode Analisis

### 2.3.1 Pre-processing Data

*Pre-processing* dalam penelitian ini dilakukan dalam beberapa tahapan, yaitu sebagai berikut:

- a. File inti di-load ke dalam R-studio kemudian difilter berdasarkan variabel email. Email @bps dianggap sebagai pegawai BPS sehingga dikeluarkan dari data.
- b. Pada data inti terdapat banyak variabel, sehingga variabel dilakukan *filter* pada variabel dengan menghilangkan variabel yang tidak digunakan. Variabel yang digunakan telah disebutkan sebelumnya.
- c. Pada variabel tanggal akses, bulan dipisahkan dan dibuatkan variabel baru yang disebut variabel bulan.
- d. Data ini terdiri dari data akhir tahun 2017 hingga awal tahun 2019, sehingga dilakukan *filter* dan diambil data pada tahun 2018.
- e. Data dibagi berdasarkan kontennya. Hal ini dilakukan agar penggabungan dengan file judul konten lebih mudah. Penggabungan dilakukan dengan variabel id konten sebagai variabel unik. Sebelum dibagi berdasarkan jenis konten, id konten ini tidak unik.
- f. File judul konten di-load ke dalam R-studio kecuali konten *statictable*. Konten ini memiliki id konten yang tidak unik, sehingga sulit untuk menggabungkannya.
- g. File konten *pressrelease* digabungkan ke data yang telah dibagi tadi sehingga variabel judul konten bergabung. Untuk konten *statictable*, id konten dijadikan sebagai pengganti judul konten. Pada konten *publication*, terdapat id konten yang tidak memiliki judul konten, sehingga beberapa judul konten *publication* menggunakan id konten sebagai judul.
- h. Sebelum digabungkan kembali, pada data konten *publication* dan *pressrelease* variabel judul konten, karakter seperti tanda koma (,) dihilangkan, hal ini dilakukan agar tidak mengganggu ketika pembuatan file *transaction* dimana tanda koma (,) dijadikan sebagai pemisah antar *item*.
- i. Data yang telah dipisah tadi digabungkan kembali sesuai dengan variabelnya.
- j. Menghilangkan transaksi oleh user yang sama dengan *item* yang sama pada hari yang sama. Pada data, terdapat user yang mengunduh data yang sama lebih dari dua kali pada hari yang sama. Dalam transaksi, kejadian tersebut dianggap hanya satu *item* yang diunduh, sehingga *item* yang *double* harus dihilangkan.

Data yang diperoleh dari beberapa proses di atas disimpan dalam dua format, yaitu format *xlsx* untuk disiapkan dalam proses analisis deskriptif dan format *transaction* untuk proses *association rule mining*. Format *transaction* diperoleh dengan menggabungkan semua *item* dalam satu *row* yang dipisahkan oleh tanda koma (,) berdasarkan *userID* dan *date* kemudian disimpan dalam format *csv*. Format ini dikenal dengan format *basket*.

### 2.3.2 Knowledge Discovery

*Knowledge discovery* dalam *web usage mining* dapat dilakukan dengan beberapa *teknik*. Dalam penelitian ini dilakukan dalam dua *teknik*, yaitu analisis deskriptif dan *association rule mining*.

Pada tahap analisis deskriptif, digunakan data yang telah telah disimpan dalam format *xlsx*. Banyak ukuran yang dapat diunakan dalam analisis deskriptif, seperti frekuensi, rata-rata, median dan sebagainya. Dalam penelitian ini hanya menggunakan satu ukuran, yaitu ukuran frekuensi. Analisis dilakuakn terhadap beberapa variabel, diantaranya Konten, Judul Konten, Bulan, jenis kelamin, Pendidikan terakhir, dan Pekerjaan.

Berikutnya, data yang telah disimpan dalam format *basket* diolah dengan teknik *association rule mining* dengan algoritma *Apriori*. Penentuan *minimum support* dilakukan dengan membandingkan jumlah transaksi dan jumlah *item*. Hasil yang didapatkan dari perbandingan tersebut hanya sebagai batas penentuan *minimum support*, sehingga dilakukan percobaan dengan *min-support* yang berbeda-beda, yaitu pada 0,0007 hingga 0,0013. *Min-confiden* yang digunakan tetep, yaitu 0,8 dengan asumsi *item* diakses bersama telah memiliki kemungkinan yang besar. Dari beberapakali percobaan, diambil hasil pertimbangan *rules* yang dihasilkan sesuai dengan tujuan penelitian.

Selain dilakukan percobaan dengan *minimum support* yang berbeda, juga dilakukan percobaann dengan membagi data berdasarkan jenis konten dan juga bulan. Pembagian ini dilakukan dengan asumsi didapatkan *rules* yang lebih beragam.

### 2.3.3 Pattern Analysis

Tahapan ini dilakukan dengan memvisualisasikan hasil yang telah didapatkan pada tahap sebelumnya. Visualisasi dilakukan menggunakan tabel ataupun grafik yang memudahkan memahami pola yang dihasilkan.

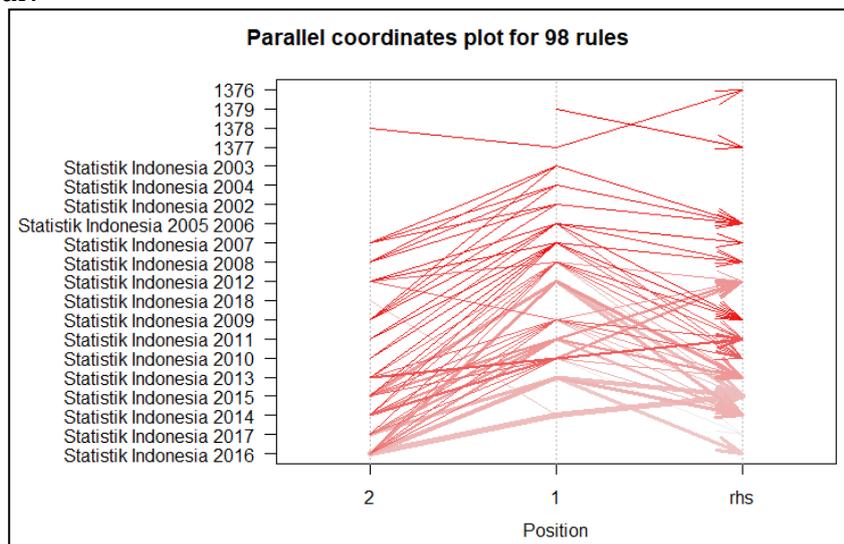
## 3. HASIL DAN PEMBAHASAN

### 3.1. Data dan publikasi yang sering diakses bersamaan secara keseluruhan pada *website* BPS

Publikasi data yang sering diakses bersamaan merupakan suatu pola yang menarik yang dapat dijadikan sebagai bahan untuk meningkatkan kualitas layanan dan evaluasi *website* BPS RI. Seringnya suatu publikasi data diakses bersamaan dapat diidentifikasi bahwa publikasi data tersebut kegunaannya saling terkait bagi pengguna data. Dalam penelitian ini, keterkaitan kegunaan publikasi data dianggap menarik. Hal tersebut dikarekanakan latar belakang penelitian ini sendiri.

Pada data yang digunakan, terdapat sebanyak 312.274 transaksi publikasi data dengan 5.578 *item*. Dengan menggunakan *minimum support* 0,0007 dan *minimum confidence* 0,8 seperti yang telah dijelaskan sebelumnya, diperoleh 98 *rules* yang diperlihatkan pada gambar 1. Panah pada gambar mengidentifikasi *support* dan *lift*. Ukuran panah untuk *support*, makin besar ukuran panah maka

*support* semakin besar. Kemudian warna merah untuk *lift*, makin merah maka *lift* makin besar.



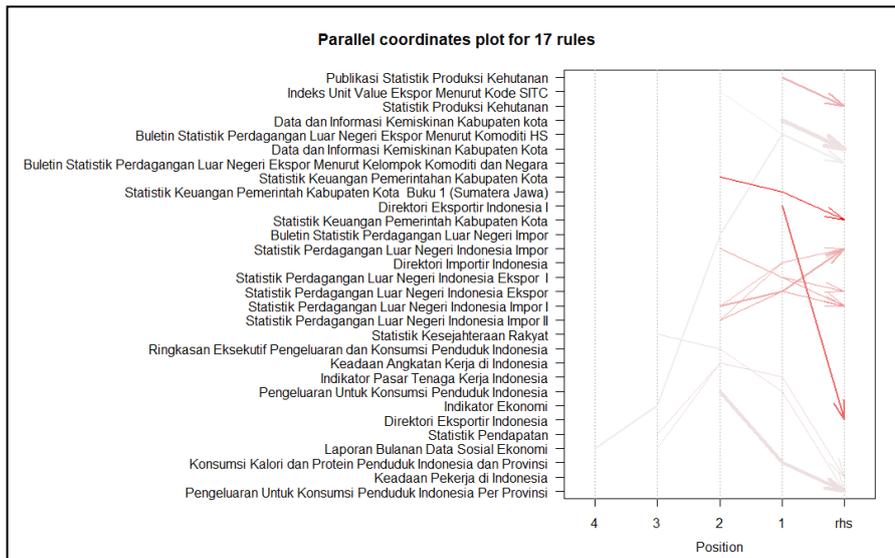
Gambar 2. Publikasi data yang sering diakses bersamaan pada *website* BPS

### 3.2 Pengelompokan data berdasarkan judul tanpa tahun

Pada bagian ini, judul konten dikategorisasi dengan menghilangkan tahun, bulan, triwulan, jilid, dan kontruksi. Data yang digunakan hanya data dengan konten *publication*. Hal ini dikarenakan hanya konten ini jika bagian bagian tertentu yang disebutkan tadi dihilangkan, kemungkinan adanya kata pengganggu lainnya lebih sedikit.

Dengan transaksi sebanyak 205.985 dan 1.449 *item*, dan dengan menggunakan *minimum support* 0,0007 didapatkan 7 *rules* seperti pada gambar 2. Dapat dilihat bahwa *rules* yang dihasilkan mulai beragam, namun karena pada data terdapat judul konten yang sama dengan penggunaan huruf yang berbeda sehingga dalam proses dianggap berbeda yang menyebabkan *rules* yang dihasilkan masih kurang baik. Selain itu, banyaknya *item* yang masih belum dihilangkan perbedaannya seperti {Statistik Perdagangan Luar Negeri Indonesia Impor I} yang memiliki *item* lain yaitu Impor II.

Secara keseluruhan, *rules* yang diperoleh dapat dikatakan menarik, baik itu secara objektif maupun subjektif. Terdapat juga *rules* yang memiliki *antecedent* atau bagian yang mendahului lebih dari 2 *item*. Hal ini memperlihatkan telah beragamnya *rules* yang dihasilkan.



Gambar 3. Publikasi data yang sering diakses bersamaan pada website BPS yang telah dikelompokan

### 3.3 Rekomendasi peningkatan layanan website BPS

Berdasarkan hasil analisis publikasi data yang diakses secara bersama, pengguna yang mengakses lebih dari sekali cenderung akan mengakses publikasi data yang sama namun berbeda tahun. Selain itu juga terdapat publikasi data yang berbeda diakses secara bersama seperti yang didapatkan pada hasil sebelumnya. Saran yang dapat diberikan berdasarkan hasil tersebut adalah dengan menyediakan menu konten yang sama atau yang kemungkinan diakses bersama berdasarkan hasil analisis sebelumnya pada halaman saat pengguna mengakses suatu publikasi data, sehingga pengguna tidak perlu lagi ke menu sebelumnya untuk membuka konten yang lain.



Gambar 4. Saran halaman ketika membuka konten publication pada website BPS RI

#### 4. SIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan, dapat didapatkan kesimpulan sebagai berikut:

- a. Publikasi data yang sering diakses secara bersamaan pada *website* BPS kebanyakan merupakan data yang sama namun berbeda tahun.
- b. Publikasi data yang dikelompokkan berdasarkan judul dengan menghilangkan tahun didapatkan hasil yang beragam, konten yang diakses memiliki judul yang berbeda meskipun memiliki hubungan seperti {Konsumsi Kalori dan Protein Penduduk Indonesia dan Provinsi, Pengeluaran Untuk Konsumsi Penduduk Indonesia}.
- c. Berdasarkan hasil yang diperoleh, peneliti memberikan rekomendasi yaitu menyediakan menu konten yang sama atau yang kemungkinan diakses bersama pada halaman saat pengguna mengakses suatu publikasi data.

#### DAFTAR PUSTAKA

- [1] Cooley, R., Mobasher, B., & Srivastava, J. (1997, November). Web Mining: Information and Pattern Discovery on the World Wide Web. In *ictai* (Vol. 97, pp. 558-567).
- [2] Cooley, R., Tan, P. N., & Srivastava, J. (1999, August). Discovery of interesting usage patterns from web data. In *International Workshop on Web Usage Analysis and User Profiling* (pp. 163-182). Springer, Berlin, Heidelberg.
- [3] Agrawal, R., Imieliński, T., & Swami, A. (1993, June). Mining association rules between sets of items in large databases. In *Acm sigmod record* (Vol. 22, No. 2, pp. 207-216). ACM.
- [4] Agrawal, R., & Srikant, R. (1994, September). Fast algorithms for mining association rules. In *Proc. 20th int. conf. very large data bases, VLDB* (Vol. 1215, pp. 487-499).
- [5] Srivastava, J., Cooley, R., Deshpande, M., & Tan, P. N. (2000). Web usage mining: Discovery and applications of usage patterns from web data. *Acm Sigkdd Explorations Newsletter*, 1(2), 12-23.
- [6] Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From data mining to knowledge discovery in databases. *AI magazine*, 17(3), 37-37.
- [7] Patel, K. B., Chauhan, J. A., & Patel, J. D. (2011). Web mining in e-commerce: Pattern discovery, issues and applications. *International Journal of P2P Network Trends and Technology*, 1(3), 40-45.
- [8] BPS. Tentang BPS. Diakses pada Mei 17, 2019, dari <https://www.bps.go.id/menu/1/informasi-umum>.
- [9] Hahsler, M., & Chelluboina, S. (2011). Visualizing association rules: Introduction to the R-extension package *arulesViz*. R project module, 223-238.
- [10] Hornik, K., Grün, B., & Hahsler, M. (2005). *arules*-A computational environment for mining association rules and frequent item sets. *Journal of Statistical Software*, 14(15), 1-25.

- [11] Kumbhare, T. A., & Chobe, S. V. (2014). An overview of association rule mining algorithms. *International Journal of Computer Science and Information Technologies*, 5(1), 927-930.