

WEB USAGE MINING MENGGUNAKAN METODE ASSOCIATION RULE DENGAN ALGORITMA APRIORI UNTUK MENGETAHUI POLA BROWSING PENGUNJUNG PADA SITUS SUARAMERDEKA.COM

Hendra Wirawan

Jurusan Sistem Informasi, Fakultas Ilmu Komputer, Universitas Dian Nuswantoro
Jl. Nakula I No. 5-11, Semarang, 50131, (024) 351721
E-mail : 112201104239@mhs.dinus.ac.id

Abstrak

E-business merupakan suatu proses bisnis yang dijalankan menggunakan teknologi informasi dan internet untuk meningkatkan produktivitas keuntungan suatu bisnis. E-business sangat dipengaruhi oleh traffic jumlah pengunjung yaitu dengan semakin tingginya traffic jumlah pengunjung suatu situs e-business, maka keuntungan yang diperoleh akan semakin banyak. Namun jika traffic jumlah pengunjung rendah, maka laba yang diperoleh pun hanya sedikit. Hal ini terjadi pada situs berita suaramerdeka.com. Situs ini hanya memiliki pengunjung sebanyak 25.163 orang per hari. Untuk meningkatkan jumlah pengunjung, salah satu strategi yang dapat dilakukan adalah dengan melakukan penemuan terhadap pola pengunjung dari situs tersebut. Dewasa ini perkembangan web usage mining berkembang pesat dalam mengatasi berbagai masalah pengolahan data web dengan jumlah besar. Untuk menemukan pola pengunjung berupa aturan asosiasi adalah menggunakan metode *association rule*. Apriori merupakan salah satu algoritma dari metode tersebut. Dalam penelitian ini algoritma apriori diterapkan untuk mendapatkan pola pengunjung, mengenai berita dengan kategori apa yang sering dibaca dan dilihat oleh pengunjung pada situs www.suaramerdeka.com. Dari hasil pengolahan dataset dengan algoritma apriori, didapati bahwa syarat nilai minimum *support* adalah sebesar 17% dan nilai minimum *confidence* sebesar 42% menghasilkan aturan asosiasi yaitu jika membaca berita dengan kategori bisnis maka membaca berita dengan kategori nasional dan jika membaca berita dengan kategori sepakbola maka membaca berita dengan kategori nasional dengan atribut yang digunakan ialah IP (Internet Protocol) dan Hyperlink (Halaman Kunjungan). Pengolahan data diproses menggunakan RapidMiner dengan menerapkan algoritma apriori Untuk memperoleh pengetahuan atau knowledge, dari pola tersebut akan diterjemahkan ke dalam bentuk yang dapat dimengerti. Pola tersebut digunakan untuk membantu admin dalam melakukan pengembangan pada situs www.suaramerdeka.com dengan memanfaatkan knowledge dari pola asosiasi tersebut, untuk mengatur desain tata letak/ layout penempatan kategori-kategori yang tersedia pada situs tersebut atau biasa disebut *web personalization*.

Kata Kunci: *Web Usage Mining, Association Rule, Apriori, RapidMiner, Desain Tata Letak*

Abstract

E-business is a business process that is executed using information technology and internet to improve productivity advantages of a business. E-business was greatly influenced by the number of visitors traffic where the increasing number of visitors, lead to the increasing of the profits. However, if the number of visitors has low traffic, then the profit obtained was only a little bit. This happens on suaramerdeka.com news sites. This site only has the visitors as much as 25.163 people per day. To increase the number of visitors, one strategy that can be done is to do a discovery against the pattern of the visitors of the site. Nowadays the development of web usage mining was booming in overcoming various problems processing web data with large numbers. To find patterns of visitors in the form of rules of the Association is using the method of association rule. Apriori algorithm is one of those methods. In this study a priori algorithm applied to the visitor pattern, to get news by category what is often read and viewed by visitors on the website www.suaramerdeka.com. From the results of the processing of the dataset with the a priori algorithm, have a minimum value of support is 17% and the minimum value of confidence is 42%

generating the rules of the association are if read news by business category then read the news with a national category and if read news by soccer category then read the news with a national category attributes used is the Internet Protocol (IP) and hyperlinks (page Visits). The processing of such data is processed using RapidMiner by applying a priori algorithm for obtaining knowledge or knowledge, of the pattern will be translated into a form that is understandable. The pattern used to help admin in doing development on the www.suaramerdeka.com site by utilizing the knowledge of patterns of the Association, to set layout design or layout placement of categories that are available on the site or web personalization.

Keywords: Web Usage Mining, Association Rule, Apriori, RapidMiner, Layout Design

1. PENDAHULUAN

Perkembangan teknologi internet saat ini sangatlah pesat. Internet merupakan suatu jaringan komputer berskala global [1]. Pengguna internet di Indonesia terdiri dari para pelaku bisnis. Para pelaku bisnis tersebut menjalankan atau mendirikan bisnis mereka dilakukan secara online atau biasanya disebut dengan istilah *e-business* yang merupakan suatu proses bisnis yang dijalankan menggunakan teknologi informasi dan internet untuk meningkatkan produktivitas keuntungan suatu bisnis [4]. *E-business* sangat dipengaruhi oleh *traffic* jumlah pengunjung yaitu dengan semakin tingginya *traffic* jumlah pengunjung suatu situs *e-business*, maka keuntungan yang diperoleh akan semakin banyak. Namun jika *traffic* jumlah pengunjung rendah, maka laba yang diperoleh pun hanya sedikit. Hal ini terjadi pada situs [suaramerdeka.com](http://www.suaramerdeka.com) yang hanya memiliki pengunjung sebanyak 25.163 orang per hari [5]. Pada tabel 1, terlihat bahwa dari ranking situs berita nasional, [suaramerdeka.com](http://www.suaramerdeka.com) menempati ranking 742, jauh tertinggal dari para pesaingnya. Faktor yg mempengaruhi ketertarikan pengunjung untuk mengunjungi suatu situs antara lain adalah faktor mutu yang sangat dipengaruhi oleh isi/konten dan tata letak yg baik [8].

Tabel 1: Ranking Situs Berita Nasional [7]

No	Situs Berita Indonesia	Ranking Nasional
1	detik.com	6
2	kompas.com	9
3	tribunnews.com	15
4	okezone.com	20
5	sindonews.com	43
6	republika.co.id	58
7	jawapos.com	616
8	suaramerdeka.com	742

Untuk meningkatkan jumlah pengunjung, salah satu strategi yang dapat dilakukan adalah dengan melakukan analisis terhadap pola pengunjung tersebut dengan ilmu *web usage mining* menggunakan metode *association rule* dengan algoritma *apriori*. Pola perilaku pengunjung yang dihasilkan berupa aturan asosiasi pola yang sering muncul serta hubungan sebab-akibat diantara himpunan suatu objek [9]. Hasil penelitian ini diharapkan dapat menghasilkan suatu pola pengunjung yang sesuai dengan katagori berita yang dibaca dari situs [suaramerdeka.com](http://www.suaramerdeka.com), sehingga dapat membantu *admin* untuk melakukan pengembangan pada situs www.suaramerdeka.com dengan memanfaatkan knowledge dari pola asosiasi tersebut, untuk mengatur penempatan kategori-kategori yang tersedia pada situs tersebut atau biasa disebut *usage conceptual categories* dalam *web personalization*.

2. ASSOCIATION RULE

Association rule merupakan salah satu peran data mining. Association rule merupakan suatu teknik yang berfungsi untuk menemukan suatu pola asosiasi terhadap suatu kombinasi produk. Terdapat dua parameter untuk mengukur tingkat kepentingan suatu aturan asosiatif, yaitu *support* dan *confidence*. *Support* merupakan persentase kombinasi produk dalam *database*, sedangkan *confidence* merupakan sebuah nilai yang memastikan tingkat kekuatan hubungan antar produk dalam aturan asosiasi. Analisis asosiasi merupakan suatu proses untuk menemukan semua aturan asosiasi yang memenuhi syarat *minimum support* dan *minimum confidence*. *Frequent pattern mining* merupakan analisis asosiasi yang paling sering digunakan oleh para analis [17]. Tahapan dasar analisis asosiasi terbagi menjadi dua yaitu :

1. Analisis pola frekuensi tinggi

Pada tahap ini pencarian terhadap kombinasi item yang memenuhi syarat minimum nilai support. Nilai support sebuah item dapat diperoleh dengan persamaan (1) :

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

Sedangkan nilai support dari dua item dapat diperoleh dengan persamaan (2)

$$\text{Support}(A, B) = \frac{P(A \cap B)}{\text{Total Transaksi}} \dots (2)$$

2. Pembentukan aturan asosiasi

Tahap selanjutnya adalah pencarian terhadap aturan asosiatif yang memenuhi syarat minimum *confidence* dengan menghitung *confidence* aturan asosiasi A B. Nilai *confidence* dari aturan asosiasi

A B dapat diperoleh dengan persamaan (3) [11] :

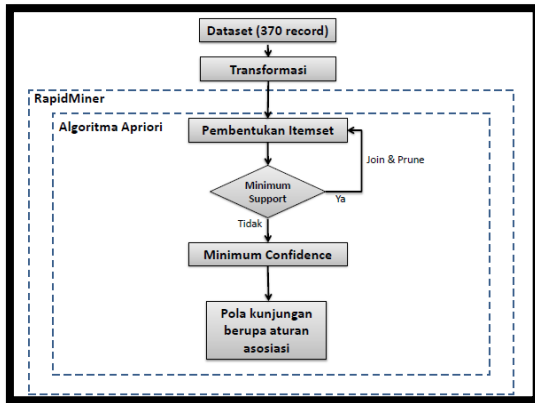
$$\text{Confidence} = P(B|A) = \frac{\text{Jumlah transaksi mengandung } A \text{ dan } B}{\text{Jumlah transaksi mengandung } A} \dots (3)$$

Jika nilai minimum *support*nya terlalu tinggi, maka akan muncul kemungkinan tidak ditemukannya rules yang memenuhi. Sedangkan jika nilai minimum *confidence* terlalu tinggi, maka akan muncul kemungkinan tidak ditemukannya rules yang memenuhi [17].

3. ALGORITMA APIRORI

Algoritma apriori berfungsi untuk menentukan *frequent itemset* pada sebuah kumpulan data yang berisi banyak data, dengan memperhatikan *minimum support*. *Itemset* digunakan untuk menentukan pola keterkaitan barang. Pencarian aturan asosiasi didasarkan pada nilai *minimum support* dan *minimum confidence*. Tahap pencarian aturan asosiasi terdiri dari tahap pencarian *frequent itemset* dan tahap menemukan aturan asosiasi dari *frequent itemset*. Pencarian *frequent itemset* terdiri dari tahap Join (mengkombinasikan setiap item dengan item yang lain sampai kombinasinya habis) dan Prune (penggabungan, lalu setiap kombinasi yang telah digabungkan dipangkas sesuai minimum *support* yang telah ditentukan. Sedangkan pada tahap menemukan aturan asosiasi dari *frequent itemset* yang dilakukan mencari nilai *support* dan *confidence* dari anggota *itemset* yang memenuhi syarat, lalu cari yang melebihi minimum *support* dan minimum *confidence* yang telah ditentukan, pola yang memiliki *confidence* lebih besar dari *minimum confidence* itulah yang menjadi aturan asosiasi. Kelebihan algoritma apriori

adalah memudahkan pencarian itemset karena mampu menyusutkan ruang pencarian data yang sangat besar [18].



4. METODE PENELITIAN

Penelitian ini meliputi terdiri dari beberapa inisiasi, Inisiasi dilakukan dengan identifikasi masalah dan kajian pendukungnya untuk menemukan solusi alternatif. Terlihat pada gambar 1, beberapa inisiasi tersebut yaitu:

1. Data Selection

Sekumpulan data mentah atau raw data tidak semuanya akan digunakan, hanya data yang sesuai dengan kriteria yang akan dipakai untuk dianalisis maka perlu dilakukannya proses seleksi dengan cermat.

2. Preprocessing/Cleaning

Data yang telah diseleksi banyak terdapat kesalahan misalnya seperti data yang tidak konsisten atau data yang tidak relevan. Data yang diperoleh baik dari database suatu perusahaan, memiliki isian-isian yang tidak sempurna seperti data yang hilang, tidak valid, serta atribut data yang tidak relevan. Sehingga perlu dilakukannya preprocessing/cleaning.

3. Transformation

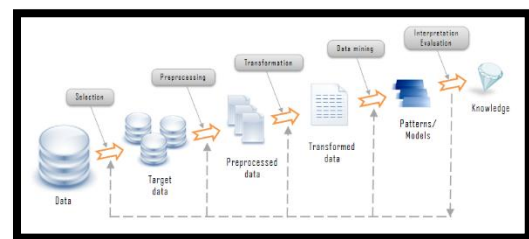
Data yang telah melalui proses selection dan preprocessing tidak bisa langsung digunakan, tahap ini merupakan proses kreatif untuk membentuk data kedalam bentuk yang bias diterapkan untuk proses selanjutnya dan sangat tergantung pada jenis data atau pola informasi yang akan dicari.

4. Data mining

Proses mencari informasi menarik atau pola dalam data terpilih dengan menggunakan metode, algoritma atau teknik tertentu yang sangat bervariasi. Pemilihan algoritma atau metode yang akan digunakan sangat bergantung pada tujuan dan prosesnya.

5. Evaluation/Interpretation

Pola-pola yang telah ditemukan kemudian diterjemahkan ke dalam bentuk yang dapat dimengerti untuk membantu dalam perencanaan strategi bisnis



Gambar 1. Tahapan Proses Knowledge Discovery in Database [16]

3.1 Dataset

Setelah melalui tahap preprocessing/cleaning, dataset akan diproses dalam RapidMiner dengan jumlah 370 record. Atribut yang digunakan untuk mengimplemetasi algoritma apriori adalah IP (Internet Protocol), *Hyperlink* (URL) dan kategori

berita dari data pengunjung situs suaramerdeka.com pada bulan Oktober, November, dan Desember 2014. Serta untuk dilakukannya proses pengujian digunakan record sebanyak 368 record dari data pengunjung pada bulan Agustus 2015.

3.2 Hasil Implementaasi pada RapidMiner

Pada tabel 2, membuktikan bahwa algoritma apriori mendapatkan pola kunjungan berupa aturan asosiasi yang sama antara data pada 3 bulan terakhir tahun 2014 dan pada bulan Agustus 2015. Aturan asosiasi tersebut yaitu: jika pengunjung membaca berita dengan kategori bisnis maka mereka akan membaca berita dengan kategori nasional, dan jika pengunjung membaca berita dengan kategori sepakbola maka mereka akan membaca berita dengan kategori nasional Hasil tersebut sesuai dengan batasan minimum *support* dan minimum *confidence* yang telah ditentukan yaitu 17% dan 42%

Tabel 2: Hasil pola pengunjung berupa aturan asosiasi

<pre> W-Apriori Apriori ===== Minimum support: 0.17 (63 instances) Minimum metric <confidence>: 0.42 Number of cycles performed: 17 Generated sets of large itemsets: Size of set of large itemsets L(1): 7 Size of set of large itemsets L(2): 2 Best rules found: 1. Bisnis=1 142 ==> Nasional=1 66 conf:(0.46) 2. Sepakbola=1 149 ==> Nasional=1 64 conf:(0.43) </pre>	<p>Data transaksi pengunjung pada bulan Oktober, November, dan Desember 2014</p>
--	--

<pre> W-Apriori Apriori ===== Minimum support: 0.17 (63 instances) Minimum metric <confidence>: 0.42 Number of cycles performed: 17 Generated sets of large itemsets: Size of set of large itemsets L(1): 7 Size of set of large itemsets L(2): 2 Best rules found: 1. Bisnis=1 127 ==> Nasional=1 64 conf:(0.5) 2. Sepakbola=1 135 ==> Nasional=1 63 conf:(0.47) </pre>	<p>Data transaksi pengunjung pada bulan Agustus 2015</p>
---	--

5. KESIMPULAN

Berdasarkan penelitian yang telah dilakukan, rendahnya *traffic* pengunjung situs suaramerdeka.com dapat disimpulkan bahwa dengan menggunakan metode association rule dengan algoritma apriori menghasilkan pola pengunjung situs berita suaramerdeka.com berupa aturan asosiasi dengan nilai minimum *support* sebesar 17% dan nilai minimum *confidence* sebesar 42% yaitu: jika pengunjung membaca berita dengan kategori bisnis maka mereka akan membaca berita dengan kategori nasional, dan jika pengunjung membaca berita dengan kategori sepakbola maka mereka akan membaca berita dengan kategori nasional. Sehingga dapat membantu *admin* pada situs suaramerdeka.com dalam pengembangan situs tersebut dalam hal *web personalization* yang berfokus pada *usage conceptual categories*.

DAFTAR PUSTAKA

- [1] N. Zaenudin, "Internet," Skipnesia.com, 2014.[Online]. Available:<http://www.skipnesia.com/2014/06/10-manfaat-internet-secara-umum.html>. [Accessed 24 04 2015].
- [2] "Internet Live Stats," Internet Live Stats, 01 07 2014. [Online]. Available: <http://www.internetlivestats.com/internet-users-by-country/>. [Accessed 24 04 2015].
- [3] "Biro Teknologi Informasi," Universitas Katolik Parahyangan, 2008. [Online]. Available: <http://bti.unpar.ac.id/undang-undang-ite/>. [Accessed 20 04 2015].
- [4] B. Wibisono, "Perbedaan E-commerce dan E-bisnis," STMIK AMIKOM Yogyakarta, Yogyakarta, 2011.
- [5] "StatShow," StatShow, [Online]. Available: <http://www.statshow.com/www/suara-merdeka.com>. [Accessed 24 04 2015].
- [6] "SuaraMerdeka," Suara Merdeka Cybernews, 1996. [Online]. Available: <http://www.suaramerdeka.com/>. [Accessed 24 04 2015].
- [7] I. Alexa Internet, "Alexa," Alexa Internet, Inc., 1996. [Online]. Available: www.alexa.com/topsites/countries;0/D. [Accessed 26 Maret 2015].
- [8] H. W. Demartha and D. Ismi, "Perilaku Konsumen Untuk Mengakses Suaramerdeka.com," *Jurnal Studi Manajemen & Organisasi*, vol. 10, no. 2, pp. 121-131, 2013.
- [9] A. Solichin, Ferdiansyah and W. Pramusinto, "Web Usage Mining Dengan Google Analytics: Studi Kasus Situs achmatim.net," *Seminar Nasional Multidisiplin Ilmu (SENMI)*, pp. 295-304, 2010.
- [10] M. Dahria and M. Syahril, "Analisis Web Server Log Dalam Pencarian Pola Pengunjung Web Dengan Teknik Association Rule," *Jurnal Ilmiah SAINTIKOM Sains dan Komputer*, vol. 13, no. 03, pp. 197-208, 2014.
- [11] T. Suratno, T. Prahasto and A. F. Rochim, "Web Usage Mining, Pattern Discovery dan Log File," *Sistem Informasi Bisnis*, vol. II, no. 02, pp. 94-99, 2011.
- [12] B. Kumar and K. Rukmani, "Implementation of Web Usage Mining Using Apriori and FP Growth Algorithms," *Advanced Networking and Applications*, vol. 01, no. 06, pp. 400-404, 2010.
- [13] J. Srivastava, R. Cooley and B. Mobasher, "Data Preperation for Mining World Wide Web Browsing Patterns," *Knowledge and Information Systems*, vol. 01, no. 102, pp. 5-32, 1999.
- [14] E. Magdalini and V. Michalis, "Web Mining for Web Personalization," *ACM Transaction on Internet Tecgnology*, vol. 03, no. 01, pp. 1-27, 2003.
- [15] J. Z. Muhammad and M. Wegner, *Data Mining and Analysis Fundamentals Concepts and Algorithms*, New York: Cambridge University Press, 2014.
- [16] "RITHME - Knowledge Discovery," Rithme Business Intellingence Solutions, 2015. [Online]. Available: <http://www.rithme.eu/?m=resources&p=kdprocess&lang=en>. [Accessed 25 September 2015].
- [17] A. C. Denis, A. B. Donny and A. Lia, *Belajar Data Mining dengan Rapid Miner*, Jakarta, 2013.
- [18] T. Daniel, *Discovering Knowledge In Data*, New Jersey: John Wiley & Sons, Inc., 2005.

- [19] M. Hofmann and R. Klinkenberg, *Rapidminer: Data Mining Use Cases and Business Analytics Applications*, CRC Press, 2013.
- [20] K. Robert V and M. Daryle W, "Determining Sample Size for Research Activities," *Educational and Psychological Measurement*, no. 30, pp. 607-610, 1970.