

# Sistem Rekomendasi Content Based Filtering Menggunakan Algoritma Apriori

**Tessy Badriyah, Ronny Fernando, Iwan Syarif**  
Politeknik Elektronika Negeri Surabaya (PENS)  
Jl. Arief Rahman Hakim Kampus PENS, Surabaya 60111  
Tel: (031) 594 7280; Fax: (031) 594 6114  
e-mail: [tessy@pens.ac.id](mailto:tessy@pens.ac.id) (email korespondensi)

## **Abstrak**

*Pada saat ini perkembangan e-commerce di Indonesia telah banyak berkembang dan diterima dengan baik di dalam masyarakat. Masyarakat Indonesia sangat terbuka dengan teknologi baru dan juga jumlah pemakai internet di Indonesia semakin meningkat pesat dari tahun ke tahun. Pada saat ini juga telah berdiri banyak sekali perusahaan e-commerce di Indonesia seperti Lazada, Tokopedia, OLX, Bukalapak, Blibli.com dan masih banyak lagi yang lainnya. Masing-masing perusahaan e-commerce tersebut memiliki berbagai macam strategi pemasaran dan bisnis untuk bersaing di pasar dan untuk menarik lebih banyak lagi pembeli di tokonya. Salah satu strategi yang bisa digunakan antara lain dengan penggunaan Sistem Rekomendasi. Penelitian ini menerapkan penggunaan sistem rekomendasi pada e-commerce dengan metode content-based filtering yaitu algoritma apriori yang dapat menganalisis pola-pola kombinasi item. Dengan adanya fitur sistem rekomendasi di dalam e-commerce diharapkan agar pembeli dapat dengan cepat menemukan barang yang dibutuhkan. Dengan strategi ini e-commerce akan dapat meningkatkan pelayanan dan kepuasan pada pelanggan, serta meningkatkan pendapatan pada e-commerce itu sendiri.*

**Kata kunci:** Sistem Rekomendasi, Algoritma Apriori, Content-based Filtering, E-Commerce

## **1. Pendahuluan**

Pada saat ini perkembangan e-commerce di Indonesia telah banyak berkembang dan diterima dengan baik di dalam masyarakat. Masyarakat Indonesia sangat terbuka dengan teknologi baru dan juga jumlah pemakai internet di Indonesia semakin meningkat pesat dari tahun ke tahun. Hal ini merupakan kesempatan yang bagus sekaligus menjadi tantangan tersendiri untuk menjadikan e-commerce sebagai media pemasaran dan penjualan online bagi para produsen dan penjual di Indonesia.

Masing-masing perusahaan e-commerce tersebut memiliki berbagai macam strategi pemasaran dan bisnis untuk bersaing di pasar dan untuk menarik lebih banyak lagi pembeli di tokonya. Salah satu strategi yang bisa digunakan antara lain dengan penggunaan Sistem Rekomendasi.

Sistem rekomendasi dalam penelitian ini dibuat dengan menggunakan metode content-based filtering dengan algoritma apriori yang dapat menganalisis pola-pola kombinasi item. Dengan adanya fitur sistem rekomendasi di dalam e-commerce diharapkan agar pembeli dapat dengan cepat menemukan barang yang sesuai dengan ketertarikan pembeli. Dengan strategi ini e-commerce akan dapat meningkatkan pelayanan dan kepuasan pada pelanggan, serta meningkatkan pendapatan pada e-commerce itu sendiri.

## **2. Metode Penelitian**

Pada bagian ini akan dijelaskan landasan teori yang digunakan untuk pembentukan sistem rekomendasi dan perancangan sistem rekomendasi content based filtering yang menggunakan algoritma apriori.

### **2.1. Sistem Rekomendasi**

Sistem rekomendasi adalah sebuah aplikasi yang berfungsi untuk memprediksi suatu item yang menarik bagi user, contohnya rekomendasi film, musik, buku, berita dan lain sebagainya.

Ada dua tipe metode yang diterapkan pada sistem rekomendasi, yaitu collaborative filtering dan content based filtering. Collaborative filtering adalah algoritma sistem rekomendasi dimana rekomendasi diberikan berdasarkan pertimbangan data dari user yang lain. Sedangkan Content based filtering,

pemberian rekomendasi diberikan dengan mengeksplorasi isi dari profil user, deskripsi produk atau hal-hal yang berhubungan dengan pembentukan pilihan user atas sebuah item. Penelitian ini menggunakan metode content based filtering pada pembentukan item-item yang muncul pada rekomendasi yang diberikan pada user.

### 2.1.1. Content-Based Filtering

Metode content-based filtering membentuk profil pengguna berdasarkan atribut pembentuk suatu item.

Algoritma metode content-based filtering dijelaskan dalam tahap-tahap berikut ini :

- (1) Suatu item barang dipisah-pisah berdasarkan suatu vektor komponen pembentuknya.
- (2) Pengguna akan memberikan nilai suka atau tidak suka pada item tersebut.
- (3) Sistem akan membentuk profil pengguna berdasarkan bobot vektor komponen pembentuk suatu item. Pembuatan profil pengguna dapat menggunakan algoritma TF-IDF (*term frequency-invers document frequency*). TF adalah jumlah term dalam suatu dokumen. Sedangkan nilai IDF dapat dihitung menggunakan rumus:

$$idf_i = \log\left(\frac{n}{df_i}\right)$$

Dimana :

n = jumlah semua

df = jumlah dokumen yang memiliki term *i*.

Sistem akan melakukan penilaian berdasarkan analisis kemiripan profil pengguna dengan vektor komponen pembentuk item. Jika item tersebut akan disukai oleh pengguna maka item tersebut akan direkomendasikan ke pengguna.

Keurangan utama yang pada metode ini yaitu metode ini tidak mampu merekomendasikan jenis item yang baru atau belum pernah dilihat kepada seorang pengguna. Hal ini dikarenakan metode ini dibuat berdasarkan item-item yang pernah dinilai oleh pengguna tersebut.

### 2.1.2. Association Rule

Association rule adalah teknik data mining yang digunakan untuk menemukan aturan asosiatif di dalam suatu kombinasi item (Kusrini dan Taufiq Luthfi, 2009). Terdapat beberapa parameter dalam Association rule yaitu support, confidence dan correlation. Parameter support dan confidence digunakan hanya untuk asosiasi data yang menghasilkan beberapa aturan untuk menentukan metode. Namun, hasilnya dapat ditingkatkan dengan parameter correlation (Han dan Kamber, 2006).

Suatu kelompok item dalam association ruledinamakan itemset. Support adalah nilai penunjang atau suatu persentase yang menunjukkan besarnya tingkat kombinasi itemset dari transaksi keseluruhan. Nilai ini menentukan layak atau tidaknya suatu itemset untuk dicari nilai confidence-nya. Apabila nilai support suatu itemset lebih besar atau sama dengan nilai minimum support, maka itemset dapat disebut sebagai frequent itemset, yang tidak memenuhi disebut infrequent. Confidence adalah nilai kepastian atau nilai yang menunjukkan hubungan antar item. Confidence dapat dicari setelah ditemukannya pola frekuensi munculnya sebuah item. Metode dasar analisis asosiasi dibagi menjadi dua tahap yaitu :

- (1) Analisa *frequent item-sets*

Pada tahap ini dilakukan pencarian kombinasi item yang memenuhi syarat minimum dari nilai support dalam database. Minimum support dan minimum confidence nilainya bebas ditentukan sesuai kebutuhan (Larose, 2009).

Nilai *support* dari suatu item diperoleh dengan rumus sebagai berikut

$$\text{Support (A)} = \frac{\sum \text{Transaksi mengandung A}}{\sum \text{Transaksi}}$$

Sedangkan, nilai *support* dari 2 item diperoleh dari rumus sebagai berikut:

$$\text{Support (A, B)} = \frac{\sum \text{Transaksi mengandung A dan B}}{\sum \text{Transaksi}}$$

- (2) Pembentukan aturan asosiasi  
Apabila semua pola frekuensi tinggi sudah ditemukan, maka dicari *association rule* yang memenuhi syarat minimum *confidence* dengan cara menghitung *confidence* aturan asosiatif  $A \rightarrow B$  dari *support* pola frekuensi tinggi A dan B.

Nilai *confidence* diperoleh dengan rumus sebagai berikut:

$$\text{Confidence } (A \rightarrow B) = \frac{\sum \text{Transaksi mengandung A dan B}}{\sum \text{Transaksi mengandung A}}$$

### 2.1.3. Algoritma Apriori

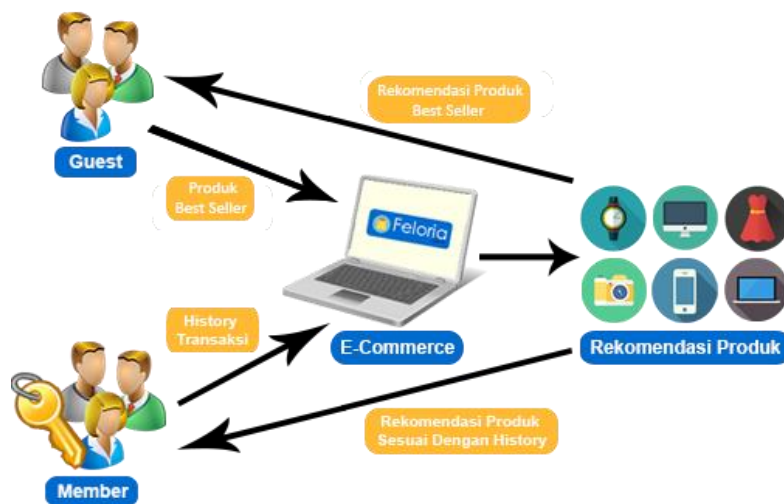
Algoritma apriori adalah salah satu jenis metode dalam asosiasi pada *data mining*. Berikut ini penjelasan cara kerja algoritma apriori secara umum:

- (1) Pembentukan kandidat itemset, Kandidat k-itemset dibentuk dari kombinasi (k-1)k-itemset yang diperoleh dari iterasi sebelumnya. Satu ciri dari algoritma Apriori yaitu adanya pemangkasan kandidat k-itemset yang *subset*-nya yang berisi k-1 *item* tidak termasuk dalam pola frekuensi tinggi dengan panjang k-1.
- (2) Penghitungan *support* dari tiap kandidat k-itemset. *Support* dari tiap kandidat k-itemset diperoleh dengan men-*scan database* untuk menghitung jumlah transaksi semua item di dalam kandidat k-itemset tersebut. Ini juga merupakan ciri dari algoritma Apriori yang dimana memerlukan penghitungan dengan *scan* seluruh *database* sebanyak k-itemset terpanjang.
- (3) Tetapkan pola frekuensi tinggi. Pola frekuensi tinggi yang memiliki k *item* atau k-itemset ditentukan dari kandidat k-itemset yang *support*-nya lebih besar dari minimum *support*.
- (4) Jika tidak ditemukan pola frekuensi tinggi baru maka seluruh proses akan dihentikan. Jika tidak, maka k akan ditambah satu dan kembali ke bagian 1.

## 2.2. Perancangan Sistem Rekomendasi

Penelitian ini membangun sebuah sistem rekomendasi pada *e-commerce* yang menggunakan algoritma apriori untuk menampilkan rekomendasi produk berdasarkan data transaksi pelanggan. Algoritma Apriori telah dijelaskan pada bab 2.1.3 sebelumnya.

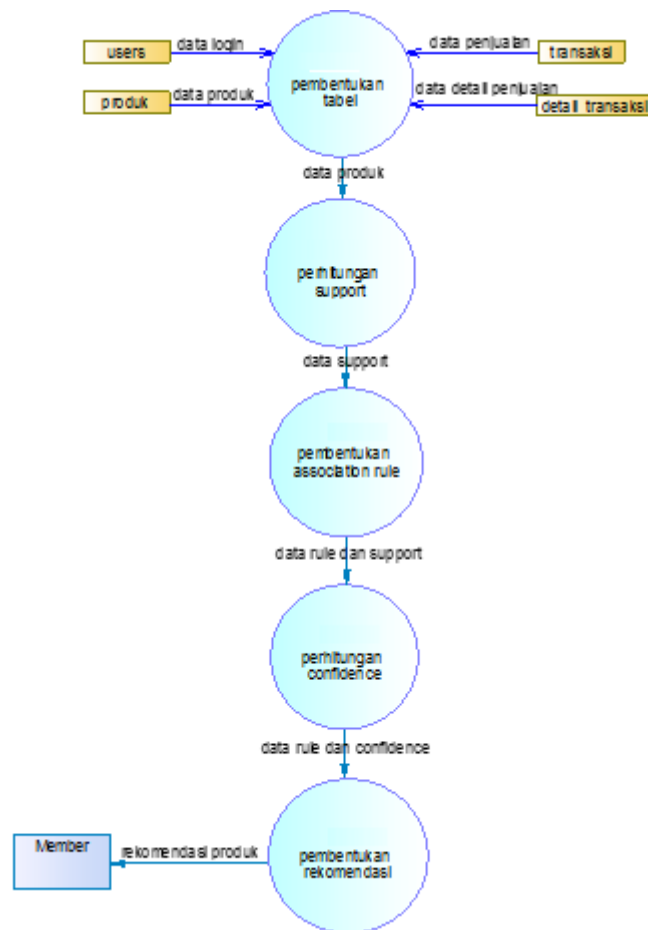
Dan berikut akan dibahas perancangan sistem yang merupakan gambaran secara umum serta pembahasan detail dari alur sistem yang dimulai dengan perancangan diagram umum sistem seperti yang ada pada Gambar 1 berikut ini.



Gambar 1. Desain Sistem

Sistem rekomendasi yang ada pada *e-commerce* ini dapat memberikan 2 macam rekomendasi produk pada pembeli sesuai dengan hak akses yang dimiliki oleh pembeli tersebut yaitu hak akses sebagai *Guest* dan *member* seperti yang ditampilkan pada gambar 1. *Guest* adalah user yang belum terdaftar di dalam *e-commerce* sedangkan *member* adalah user yang sudah terdaftar di dalam *e-commerce* dan dapat mengakses fitur-fitur *e-commerce* yang tidak bisa diakses *Guest*. Rekomendasi untuk produk yang paling populer adalah rekomendasi untuk *Guest* dan rekomendasi produk berdasarkan history transaksi adalah untuk *member*. Rekomendasi produk *best seller* adalah produk-produk yang direkomendasikan berdasarkan

jumlah pembelian produk terbanyak sedangkan rekomendasi untuk *member* adalah produk-produk yang direkomendasikan berdasarkan history transaksi pembelian yang pernah dilakukan oleh seorang *member*, kemudian sistem akan mengolah data transaksi tersebut dan mencari data transaksi *member* lain yang memiliki kesamaan produk-produk dengan *member* tersebut menggunakan algoritma apriori.



Gambar 2. Data Flow Diagram untuk Pembentukan Rekomendasi

Data flow diagram dari pembentukan rekomendasi dijelaskan pada Gambar 2, dimana pada tiap tahapannya menerapkan landasan teori dari pembentukan sistem rekomendasi yang telah dijelaskan pada sub bab sebelumnya.

### 3. Hasil dan Pembahasan

Bagian berikut ini akan membahas pengujian terhadap sistem rekomendasi yang dibangun pada e-commerce.

#### 3.1. Spesifikasi Perangkat Lunak

Berikut ini merupakan spesifikasi perangkat yang digunakan dalam tahap pengujian sistem:

- Processor : Intel Core i3-2350M, 2.30 GHz
- Memory : 2 GB
- Dengan perangkat lunak: Windows 10, dengan framework Laravel, Composer 1.2. dan editor Sublime Text 3

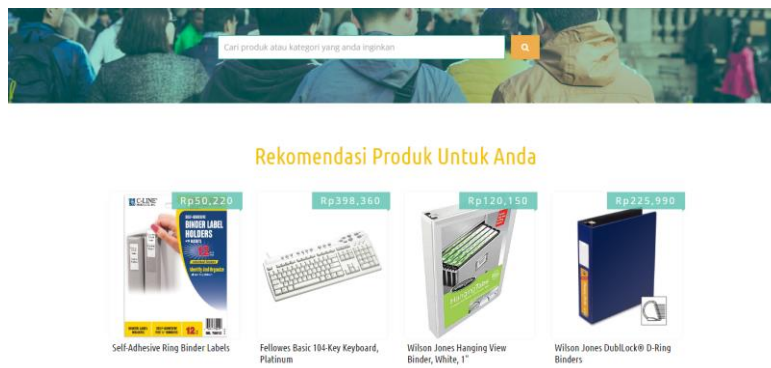
#### 3.2. Pengujian Sistem Rekomendasi pada e-Commerce

Tampilan hasil rekomendasi diberikan dalam dua hal yaitu (1) rekomendasi produk dari member yang sudah login ke dalam sistem e-commerce, dan (2) tampilan rekomendasi untuk setiap detail dari produk yang ditampilkan.

Masing-masing akan dijelaskan berikut ini.

- Rekomendasi Produk Member

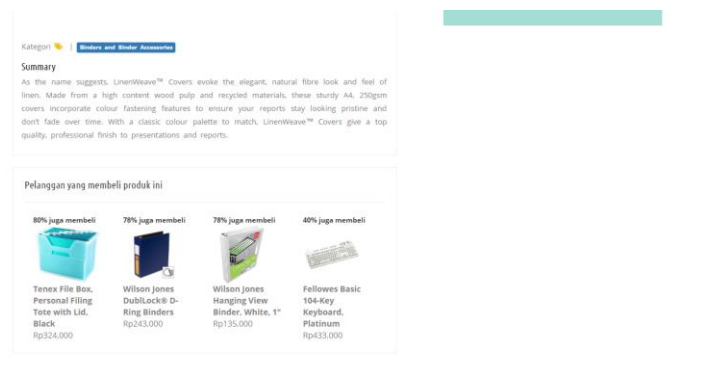
Rekomendasi produk ini akan tampil pada saat member *e-commerce* login ke dalam sistem. Rekomendasi produk ditampilkan berdasarkan transaksi pembelian member di *e-commerce*, sehingga sistem akan menampilkan rekomendasi produk yang bisa jadi berbeda-beda tergantung pada user member yang sedang login saat itu.



Gambar 4. Rekomendasi Produk Member

b. Rekomendasi di Setiap Detail Produk

Halaman rekomendasi setiap detail produk akan ditampilkan rekomendasi produk seperti yang ada pada Gambar 5. Pada tampilan rekomendasi akan diberikan keterangan, bahwa user yang telah membeli produk tersebut, juga akan membeli produk lain berdasarkan prosentasi kemunculan dari produk lain dari user yang sedang menampilkan detail produk pada saat itu.



Gambar 5. Rekomendasi di setiap Detail Produk

#### 4. Simpulan

Penelitian ini menggunakan metode Content based Filtering dengan Algoritma Apriori untuk memberikan rekomendasi produk pada e-commerce. Berdasarkan hasil pengujian sistem yang telah dilakukan maka dapat diambil kesimpulan bahwa Sistem dapat memberikan rekomendasi produk berdasarkan transaksi belanja yang pernah dilakukan oleh pembeli di e-commerce.

Berdasarkan hasil pengujian sistem rekomendasi pada e-commerce, maka dapat ditampilkan produk rekomendasi untuk setiap user member yang aktif. Dan pada setiap detail produk yang ditampilkan, dapat direkomendasikan produk lain berdasarkan prosentasi kemunculan dari produk lain dari user yang sedang menampilkan detail produk pada saat itu.

Penentuan nilai support dan confidence merupakan nilai yang mempengaruhi banyak sedikitnya rekomendasi produk yang ditampilkan. Semakin besar nilai support dan confidence yang diberikan maka akan semakin sedikit rekomendasi yang ditampilkan dan semakin kecil nilai support dan confidence yang diberikan maka akan semakin banyak rekomendasi yang ditampilkan.

**Daftar Pustaka**

- [1] Tessa Badriyah, Erry Tri Wijayanto, Iwan Syarif, Prima Kristalina. 2017. *A Hybrid Recommendation System for e-Commerce based on Product Description and User Profile*, Seventh International Conference on Innovative Computing Technology (INTECH), 2017, 16-18 Aug. 2017. p.95-100.
- [2] Fayyad, U., Piatetsky-Shapiro, G. & Smyth, P. 1996. *Data Mining to Knowledge Discovery in Databases*. AI Magazine, 17(3), 37-54.
- [3] Gorunescu, F. 2011. *Data Mining: Concepts, Models, and Techniques*. India: Springer.
- [4] Jiawei Han, Micheline Kamber, and Jian Pei. 2011. *Data Mining: Concepts and Techniques* (3rd ed.). Morgan Kaufmann Publishers Inc., San Francisco, CA, USA.
- [5] Jonathan Lee Herlocker. 2000. *Understanding and Improving Automated Collaborative Filtering Systems*. Ph.D. Dissertation. University of Minnesota. Advisor(s) Joseph A. Konstan. AAI9983577.
- [6] Pasquale Lops, Marco De Gemmis, Giovanni Semeraro. 2011. *Content-based recommender systems: State of the art and trends*. Recommender Systems Handbook, pp. 73–105.
- [7] Michael J. Pazzani. 1999. *A Framework for Collaborative, Content-Based and Demographic Filtering*. Artif. Intell. Rev. 13, 5-6 (December 1999), 393-408.
- [8] Badrul Sarwar, George Karypis, Joseph Konstan, and John Riedl. 2001. *Item-based collaborative filtering recommendation algorithms*. In Proceedings of the 10th international conference on World Wide Web (WWW '01). ACM, New York, NY, USA, 285-295.
- [9] V. M. Robin, V. S. Maarten. 2000. *Using Content-Based Filtering for Recommendation*. University of Amsterdam, Roeterstraat.
- [10] W. Paik, S. Yilmazel, E. Brown, M. Poulin, S. Dubon, and C. Amice. 2001. *Applying natural language processing (nlp) based metadata extraction to automatically acquire user preferences*. Proceedings of the 1st international conference on Knowledge capture, pp. 116–122.
- [11] Y. AlMurtadha, M. N. Sulaiman, N. Mustapha, and N. I. Udzir. 2011. *Improved web page recommender system based on web usage mining*. Proceedings of the 3rd International Conference on Computing and Informatics (ICOCI), pp. 8–9.
- [12] Y. Seroussi. 2010. *Utilising user texts to improve recommendations*. User Modeling, Adaptation, and Personalization, pp. 403–406.