

# Analisis Keranjang Belanja dengan Algoritma Apriori Klasik pada Data Mining

Tri Dharma Putra <sup>1,\*</sup>

<sup>1</sup> Fakultas Teknik, Universitas Bhayangkara Jakarta Raya; Jl. Raya Perjuangan, Marga Mulya, Bekasi Utara, Jawa Barat 17121. Telp: 021-88955882, 889955883, e-mail: tri.dharma.putra@dsn.ubharajaya.ac.id

\* Korespondensi: e-mail: tri.dharma.putra@dsn.ubharajaya.ac.id

---

## Abstract

*Association Rule Mining is an area of data mining that focus on pruning candidate keys, to find frequent item set. For example, a set of items, such as milk and bread, that appear frequently together in a transaction data set is a frequent itemset. A subsequence, such as buying first PC, then a digital camera, and then a memory card, if it occurs frequently in a shopping history database, is a (frequent) sequential pattern, also known as market basket analysis. This paper describes the step by step classical apriori on market basket analysis.*

**Keywords:** apriori algorithm, frequent item set, market basket analysis, association rule

## Abstrak

Penambangan Aturan Asosiasi adalah area data mining yang fokus pada pemangkasan kunci kandidat, untuk menemukan frequent itemset. Sebagai contoh, satu set item, misalnya susu dan roti, yang muncul sering bersama-sama di set data transaksi adalah frequent itemset. Berikutnya, pelanggan, misalnya membeli PC dahulu, lalu kamera digital, lalu kartu memori, jika ini sering terjadi dalam riwayat basisdata belanja, adalah pola sekuensial berurutan (sering), juga dikenal sebagai analisis keranjang belanja. Tulisan ini menjelaskan langkah demi langkah algoritma apriori klasik pada analisis keranjang belanja.

**Kata kunci:** algoritma apriori, frequent itemset, analisis keranjang belanja, aturan asosiasi

## 1. Pendahuluan

Data mining adalah domain penelitian yang penting sekarang ini yang fokus pada penemuan pengetahuan di basisdata. Di kehidupan sehari-hari, informasi dikumpulkan dari hampir setiap tempat. Sebagai contoh di tempat loket pembayaran pasar swalayan, informasi tentang pembelian pelanggan didata. Ketika pembayaran dan kartu diskon digunakan, informasi tentang perilaku pembelian pelanggan dan informasi pribadi dapat ditautkan (Revathi, 2015).

Evaluasi informasi ini dapat membantu peritel memberlakukan strategi yang lebih efisien dan dimodifikasi. Informasi ini juga penting untuk penentuan letak barang yang dijual di rak belanja.

Pola sering adalah pola (seperti itemset, berurutan, atau substruktur) yang muncul dalam set data dengan sering. Sebagai contoh, satu set item, seperti susu dan roti, yang muncul sering bersama-sama dalam set data transaksi adalah frequent itemset. Sebuah urutan, seperti membeli PC dulu, lalu kamera digital, lalu membeli kartu memori. Ini sering terjadi dalam basisdata riwayat berbelanja, adalah pola berurutan (sering). Sebuah substruktur dapat merujuk pada bentuk struktural berbeda, seperti subgrafik, subtree, atau sublattice, yang mungkin dikombinasikan dengan itemset atau urutan. Jika substruktur terjadi sering, ini disebut pola struktural (sering).

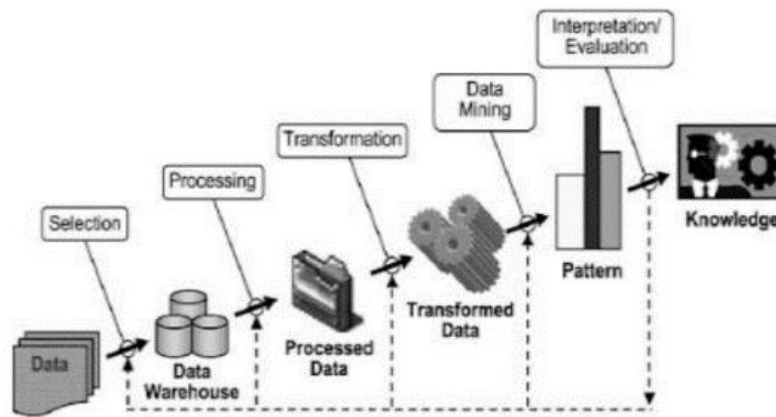
Menemukan pola sering seperti itu memainkan peran penting dalam asosiasi penambangan, korelasi, dan banyak relasi menarik lainnya di antara data. Lebih dari itu, ini membantu dalam klasifikasi data, clustering, dan tugas data mining lain juga. Karena itu, pola penambangan sering telah menjadi tugas data mining penting dan memfokuskannya dalam penelitian data mining.

Data Mining adalah langkah analisis terhadap proses penemuan pengetahuan didalam basisdata atau penemuan pengetahuan dari basisdata yang disingkat KDD (Knowledge Discovery in Database). Pengetahuan bisa berupa pola data atau relasi antar data yang valid (yang tidak diketahui sebelumnya).

Data Mining merupakan gabungan sejumlah disiplin ilmu komputer yang didefinisikan sebagai proses penemuan pola-pola baru dari kumpulan-kumpulan data sangat besar, meliputi metode-metode yang merupakan irisan dari kecerdasan buatan, machine learning, statistik, dan sistem basisdata. Data Mining ditujukan untuk mengekstrak (mengambil intisari) pengetahuan dari sekumpulan data sehingga didapatkan struktur yang dapat dimengerti manusia serta meliputi basisdata dan manajemen data, pemrosesan data, pertimbangan model dan inferensi, ukuran ketertarikan, pertimbangan kompleksitas, pasca pemrosesan terhadap struktur yang ditemukan, visualisasi.

Tahapan yang dilakukan pada proses Data Mining diawali dari seleksi data dari data sumber ke data target, tahap preprocessing untuk memperbaiki kualitas data, transformasi, Data Mining serta tahap interpretasi dan evaluasi yang menghasilkan output berupa pengetahuan baru yang diharapkan memberikan kontribusi yang lebih baik. Rinciannya dijelaskan pada Gambar 1 sebagai berikut: 1) **Data selection (Pemilihan data)**, Pemilihan (seleksi) data dari sekumpulan data operasional perlu dilakukan sebelum tahap penggalian informasi dalam KDD dimulai. Data hasil seleksi yang digunakan untuk proses Data Mining, disimpan dalam suatu berkas, terpisah dari basisdata operasional. 2) **Pre-processing/cleaning (Prapemrosesan/pembersihan)**, Sebelum proses Data Mining dapat dilaksanakan, perlu dilakukan proses cleaning pada data yang menjadi fokus KDD. Proses cleaning mencakup antara lain membuang duplikasi data, memeriksa data yang inkonsisten, dan memperbaiki

kesalahan pada data. 3) **Transformation Coding (Transformasi Pengkodean)**, Adalah proses transformasi pada data yang telah dipilih, sehingga data tersebut sesuai untuk proses Data Mining. Proses pengkodean dalam KDD merupakan proses kreatif dan sangat tergantung pada jenis atau pola informasi yang akan dicari dalam basisdata. 4) **Data Mining**, Adalah proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu. Teknik, metode, atau algoritma dalam Data Mining sangat bervariasi. Pemilihan metode atau algoritma yang tepat sangat bergantung pada tujuan dan proses KDD secara keseluruhan. 5) **Interpretation/evaluation (Interpretasi/evaluasi)**, Pola informasi yang dihasilkan dari proses Data Mining perlu ditampilkan dalam bentuk yang mudah dimengerti oleh pihak yang berkepentingan. Tahap ini merupakan bagian dari proses KDD yang disebut interpretation. Tahap ini mencakup pemeriksaan apakah pola atau informasi yang ditemukan bertentangan dengan fakta atau hipotesis yang ada sebelumnya.

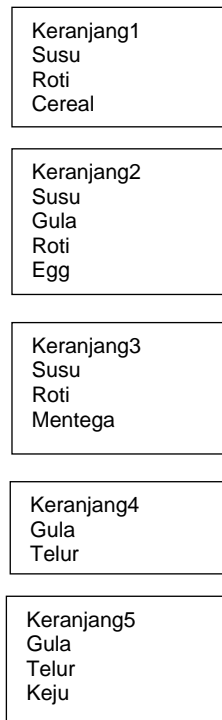


Gambar 1. Tahapan Data Mining

Aturan Asosiasi, dapat dijelaskan sebagai berikut: Penambangan Frequent Itemset menuju penemuan asosiasi dan korelasi di antara item dalam transaksi sangat besar dan set data relasional. Dengan jumlah masif data terus dikumpulkan dan disimpan, banyak industri menjadi tertarik dalam penambangan pola seperti ini dari basisdata mereka. Penemuan relasi korelasi menarik di antara record transaksi bisnis yang sangat banyak dapat membantu dalam banyak proses pengambilan keputusan bisnis, seperti desain katalog, pemasaran silang, dan analisis perilaku belanja.

Analisis Keranjang Belanja, dapat dijelaskan sebagai berikut: Sebuah contoh umum penambangan frequent itemset adalah analisis keranjang belanja di pasar swalayan. Proses ini menganalisis kebiasaan membeli pelanggan dengan menemukan asosiasi antara item yang berbeda yang ditempatkan pelanggan dalam 'keranjang belanja' (Gambar 2). Penemuan asosiasi seperti ini dapat membantu peritel mengembangkan strategi pemasaran dengan memperoleh wawasan pada item apa saja yang sering dibeli bersamaan oleh pelanggan.

Sebagai contoh, jika pelanggan membeli susu, seberapa mungkin mereka juga membeli roti (dan roti jenis apa) pada pembelian yang sama di pasar swalayan? Informasi ini dapat menunjang peningkatan penjual dengan membantu peritel melakukan pemasaran selektif dan merencanakan tata letak ruang rak penjualan item mereka.



Sumber: Hasil Penelitian (2020)

Gambar 2. Keranjang Belanja

Menambang frequent itemset adalah bagian fundamental dari penambangan aturan asosiasi. Sejumlah algoritma diajukan untuk menghasilkan frequent item set. Algoritma Apriori adalah algoritma terkemuka yang diperkenalkan oleh Agrawal & Srikant (1994).

## 2. Metode Penelitian

Metode penelitian pada jurnal ini adalah dengan studi pustaka mendalam. Literatur studi dilakukan dari berbagai sumber literatur dan jurnal.

Metode yang digunakan dalam sebuah penelitian sangat mempengaruhi kinerja sistem penelitian untuk dapat bekerja secara optimal. Dalam sebuah penelitian, metode yang sesuai dengan kebutuhan diharapkan dapat berjalan dengan baik, sehingga bisa mengikuti metode atau prosedur yang diberikan.

Metode penelitian pada tulisan ini adalah menggunakan studi literatur yang mendalam. Beberapa penelitian eksperimen relevan dibandingkan berdasarkan literatur.

Ada berbagai macam varian apriori algoritma yang ditemukan di literatur. Dari algoritma di literatur kemudian dilakukan perbandingan tahapan proses penerapan algoritma termodifikasi dengan algoritma klasik apriori Akhirnya pada tahap akhir studi literatur, dipilih algoritma apriori klasik dari Sheila A. Abaya.

Jurnal ini diatur dalam langkah-langkah berikut: Dibagi menjadi lima bab. Bab pertama adalah Pendahuluan, bab kedua adalah tentang Landasan Teori Frequent Itemset. Bab tiga adalah tentang Metode Penelitian dan Struktur Organisasi, Bab empat adalah Pembahasan tentang Algoritma Apriori Klasik dan memerinci bahasan tentang analisis keranjang belanja. Bab ini menjelaskan langkah demi langkah algoritma apriori klasik pada analisis keranjang belanja. Bab terakhir, bab lima adalah Kesimpulan dan Saran. Saran untuk penelitian lanjutan diberikan.

### **3. Hasil dan Pembahasan**

#### **a) Analisis awal**

Agrawal R. et. al, mengajukan Apriori algoritma yang diajukan menggunakan pendekatan iteratif yang dikenal sebagai pencarian levelwise. Di mana k-itemset digunakan untuk menelusuri (k+1)-itemset. Pertama kali, set frekuensi 1-itemset ditemukan. Set ini didefinisikan sebagai L1. L1 digunakan untuk menemukan L2, set frekuensi 2-itemset, yang digunakan untuk menemukan L3, dan seterusnya, hingga tidak ada lagi frekuensi k-itemset yang dapat ditemukan. Penemuan masing-masing Lk mensyaratkan pemindaian basisdata penuh. Untuk menemukan semua frekuensi itemset, algoritma ini mengadopsi metode rekursif (Abaya, 2012).

Ini adalah pseudo-code algoritma apriori klasik (Han J., 2006):

*C<sub>k</sub>*: Candidate itemset of size *k*

*L<sub>k</sub>* : frequent itemset of size *k*

*L*1 = {frequent items};

for (*k* = 1; *L<sub>k</sub>* != ∅; *k*++) do begin

*C<sub>k+1</sub>* = candidates generated from *L<sub>k</sub>*;

        for each transaction *t* in database do

            increment the count of all candidates in *C<sub>k+1</sub>*

that are contained in *t*

*L<sub>k+1</sub>* = candidates in *C<sub>k+1</sub>* with min\_support

end

return  $\bigcup_k L_k$ ; (1)

#### **b) Analisis Keranjang Belanja**

Algoritma berbeda telah diajukan untuk menemukan frequent itemset. Algoritma Apriori adalah adalah pendekatan yang terkemuka yang di perkenalkan oleh Agrawal & Srikant (1994).

Algoritma ini adalah menggunakan pendekatan iteratif dan ada dua langkah di setiap masing-masing iterasi. Langkah pertama menghasilkan satu set set item kandidat. Lalu di langkah kedua, kita menghitung kejadian teknik pemangkasan. Pertama pada basis hitungan support (harus lebih besar dari ambang batas support yang ditentukan pengguna) dan kedua untuk set item yang sering, semua subsetnya harus berada di set frequent item terakhir (Pragaladan, 2015).

Mari kita tentukan keranjang belanja terlebih dahulu. Ada 4 keranjang, yang berisi Roti, Keju, Bir, Gula, Telur, dan Susu. Setiap keranjang adalah satu transaksi (Putra, 2018).

Diberikan: Support Minimum = 75%

Support Minimum: sebuah nilai yang ditentukan oleh peneliti untuk memotong kombinasi set item menjadi lebih sedikit (Suprianto Panjaitan, 2019).

Tabel 1. Set Data Awal

| ID Transaksi | Item                         |
|--------------|------------------------------|
| T1           | Susu, Roti, Gula, Keju       |
| T2           | Gula, Roti, Bir, Telur, Keju |
| T3           | Bir, Roti, Keju, Telur       |
| T4           | Keju, Roti, Gula             |

Sumber: Hasil Penelitian (2020)

Langkah pertama, Hitung jumlah transaksi yang terjadi di setiap item.

Tabel 2. Jumlah transaksi yang terjadi di setiap item

| Item  | Jumlah Transaksi |
|-------|------------------|
| Roti  | 4                |
| Keju  | 4                |
| Bir   | 2                |
| Gula  | 3                |
| Telur | 2                |
| Susu  | 1                |

Sumber: Hasil Penelitian (2020)

Dari tabel di langkah pertama kemudian lakukan pemangkasan data item yang tidak memenuhi support minimum 75%.

Tabel 3. Pemangkasan data item

| Item yang dibeli | Jumlah transaksi | Perhitungan support | support |
|------------------|------------------|---------------------|---------|
| Roti             | 4                | 4/4                 | 100%    |
| Keju             | 4                | 4/4                 | 100%    |
| Bir              | 2                | 2/4                 | 50%     |
| Gula             | 3                | 3/4                 | 75 %    |
| Telur            | 2                | 2/4                 | 50 %    |
| Susu             | 1                | 1/4                 | 25 %    |

Sumber: Hasil Penelitian (2020)

Karena data item Bir, Telur, dan Susu tidak memenuhi support minimum 75 %, item ini dipangkas. Langkah ketiga, pasang data dalam langkah 2 setelah dipangkas.

Tabel 4. Pasangan item set

| Item yang dibeli |
|------------------|
| Roti, Keju       |
| Roti, Gula       |
| Keju, Gula       |

Sumber: Hasil Penelitian (2020)

Langkah keempat, Hitung berapa kali item yang dipasangkan dibeli bersamaan.

Tabel 5. Jumlah pasangan yang terjadi

| Item set   | Total Frekuensi |
|------------|-----------------|
| Roti, Keju | 4               |
| Roti, Gula | 3               |
| Keju, Gula | 3               |

Sumber: Hasil Penelitian (2020)

Langkah kelima, Pangkas pasangan dengan support minimum di bawah 75%.

Tabel 6. Memangkas pasangan

| Item set   | Total frekuensi | Penghitungan support |
|------------|-----------------|----------------------|
| Roti, Keju | 4               | 100%                 |
| Roti, Gula | 3               | 75%                  |
| Keju, Gula | 3               | 75%                  |

Sumber: Hasil Penelitian (2020)

Langkah Keenam, Pasangkan item dengan ukuran 3.

Tabel 7. Hasil akhir

| Item set         | Total frekuensi |
|------------------|-----------------|
| Roti, Keju, Gula | 3               |

Sumber: Hasil Penelitian (2020)

Karena (Roti, Keju, Gula) memenuhi support minimum, ini adalah item set terakhir. Kesimpulan, item dengan frequent itemset adalah Roti, Keju, Gula. Ketika orang membeli roti, mereka juga membeli keju dan gula.

#### 4. Kesimpulan

Dalam tulisan ini algoritma apriori diperkenalkan untuk mencari frequent itemset dari keranjang belanja di pasar swalayan. Dengan data yang diberikan maka disimpulkan item dengan frequent item set adalah Roti, Keju, dan Gula, berarti ketika pelanggan pasar swalayan membeli Roti, mereka juga membeli Keju dan Gula. Prinsip aturan asosiasi dilakukan dengan algoritma apriori pada analisis keranjang belanja ini. Untuk penelitian selanjutnya dimungkinkan melakukan modifikasi algoritma apriori, bisa dilakukan untuk mempercepat proses scanning pada basisdata dan membuat iterasi yang lebih efisien. Ada beberapa cara lain untuk memodifikasi algoritma apriori. Saat ini penelitian tingkat lanjut untuk menemukan algoritma yang lebih efisien dilakukan oleh para ahli untuk menemukan ukuran awal ideal dan ukuran kombinasi. Akan lebih baik jika dilakukan implementasi pada kedua algoritma ini untuk membandingkan efisiensinya secara waktu nyata.

## Daftar Pustaka

- Abaya, S. A. (2012). Association Rule Mining based on Apriori Algorithm in Minimizing Candidate Generation. *International Journal of Scientific & Engineering Research*, Volume 3, Issue 7.
- Han J., K. M. (2006). *Data Mining Concepts and Techniques*, Ed. 2. San Francisco: Elsevier Inc.
- Pragaladan, A. R. (2015). Re-Adapted Apriori Algorithm in E-commerce. *International Journal of Innovative Research in Computer and Communication Engineering*, Volume 3, Issue 8.
- Putra, T. D. (2018). Perbandingan Tahapan Algoritma Apriori Klasik dengan Apriori Termodifikasi. *Jurnal Kajian Ilmiah, Universitas Bhayangkara Jaya*. Volume 18, No.1, January 2018.
- R. Revathi, M. G. (2015). Re-Modified Apriori Algorithm in E-Commerce Recommendation System. *International Journal of Innovative Research in Computer dan Communication Engineering*.
- Panjaitan, S. M. (2019). Implementation of Apriori Algorithm for Analysis of Consumer Purchase. *The International Conference on Computer Science and Applied Mathematic. Journal of Physics: Conference Series*.