

PERBANDINGAN METODE DATA MINING DALAM MENGETAHUI POLA PEMBELIAN MENGGUNAKAN ALGORITMA APRIORI DAN FP-GROWTH

Ferdian Susanto¹, Desiyanna Lasut^{2*}

^{1,2} Program Studi Teknik Informatika, Fakultas Sains dan Teknologi, Universitas Buddhi Dharma

*Corresponding Author, email: desiyanna.lasut@ubd.ac.id

ABSTRAK

Di era digital saat ini, sebuah data transaksi konsumen menjadi sebuah sumber informasi yang penting dalam pengambilan keputusan bisnis dalam sektor ritel. salah satu cara untuk memanfaatkan data transaksi tersebut dengan menggunakan *data mining* untuk menganalisa suatu data, khususnya dengan menggunakan salah satu teknik yaitu *association rule*. Penelitian ini memiliki tujuan untuk membandingkan antara dua algoritma *data mining*, yaitu *Apriori* dan *FP-Growth*, dalam mengidentifikasi pola pembelian konsumen. Data yang digunakan dalam penelitian ini merupakan data sekunder yang di peroleh dari situs *kaggle* jumlah data sebelum di bersihkan 541,910 dan setelah dilakukan pembersihan adalah 396,353 dan memiliki 8 *atribut*. Masalah utama yang diangkat adalah kurangnya pemanfaatan data transaksi untuk memberikan sebuah strategi bisnis yang tepat, serta belum diketahui algoritma mana yang paling efisien dan efektif dalam menemukan pola pembelian. Penelitian ini menggunakan pendekatan CRISP-DM dengan implementasi *algoritma Apriori* dan *FP-Growth* menggunakan perangkat lunak *RStudio*. Hasil analisis menunjukkan bahwa algoritma menghasilkan aturan asosiasi yang serupa, namun berbeda dari sisi efisiensi. *Apriori* menunjukkan performa lebih baik dengan penggunaan memori sebesar 138 MB dan waktu proses 1,91 detik, dibandingkan dengan *FP-Growth* yang menggunakan 140,3 MB memori dan waktu proses 2 detik. Oleh karena itu *algoritma apriori* memiliki performa lebih baik di bandingkan *algoritma fp-growth*.

Kata kunci: Apriori, Asosiasi, Data mining, *Fp-Growth*, Rstudio.

I. PENDAHULUAN

Dalam era digital yang semakin maju, data menjadi sebuah aset yang berharga. Sebuah data dapat diperoleh dari bidang seperti ritel. Setiap terjadinya transaksi akan menghasilkan sekumpulan data transaksi. Data transaksi ini dapat di manfaatkan untuk memberikan sebuah strategi bisnis untuk kedepannya jika diolah dengan baik (Musdalifah & Jananto, 2022). Salah satu cara untuk mengolah data transaksi dengan menggunakan data mining untuk menemukan hubungan yang jelas pada data tersebut (Lienata et al., 2021). *Data mining* adalah sebuah ilmu pada komputer yang dapat digunakan untuk mengumpulkan informasi dari banyak sumber yang terseedia yang nantinya akan di proses untuk menghasilkan sebuah

informasi (Febrian et al., 2023). Dalam mengetahui pola pembelian data mining dapat berkontribusi dalam memberikan informasi strategi bisnis yang berguna seperti mengetahui kebiasaan pelanggan, mengidentifikasi tren pembelian (Rahmi & Mikola, 2021). Informasi ini sangat berharga untuk membantu pengusaha dalam mengambil sebuah keputusan bisnis seperti penyediaan produk dan promosi produk (Styawati et al., 2021). Pada data mining, asosiasi digunakan untuk mengidentifikasi sebuah hubungan diantara berbagai kejadian yang terjadi pada waktu yang sama (Rahmi & Mikola, 2021). Aturan asosiasi akan menghasilkan sebuah frequent itemset dengan ketentuan jika kondisi sudah memenuhi minimum nilai support dan confidence untuk menentukan aturan mana yang cocok di implementasikan dalam mengambil keputusan (Nisa & Khasanah, 2023)

Dalam data mining terdapat 2 algoritma yang populer digunakan dalam menganalisis data untuk memberikan sebuah pola pembelian yaitu apriori dan fp-growth. Pada kedua algoritma ini termasuk ke dalam kategori association rule. Association rule berfungsi untuk menemukan hubungan antara setiap item atau pola kombinasi dalam satu data transaksi yang sama (G. Setiawan et al., 2022). Walaupun ke dua algoritma memiliki tujuan yang sama, algoritma apriori dan fp-growth memiliki pendekatan yang berbeda dalam melakukan proses pencarian pola pembelian. Algoritma apriori merupakan algoritma pengambilan data dengan menggunakan teknik association rule yang bertujuan untuk mengetahui hubungan suatu item dengan item lainnya (Freddie et al., 2021). Algoritma ini akan membuat sebuah iterasi yang nantinya akan menghasilkan sebuah kandidat itemset, proses iterasi akan terus dilakukan sampai tidak dapat memenuhi nilai minimum yang sudah ditentukan (Soepriyono & Triayudi, 2023). Algoritma apriori biasanya digunakan untuk menganalisis pola transaksi untuk menghindari overstock (Chandra & Wijaya, 2024)

Pada algoritma fp-growth juga menggunakan teknik association rule tetapi memiliki perbedaan dalam mengidentifikasi frequent itemset, dengan menghilangkan pembuatan kandidat secara berulang seperti pada algoritma apriori. Algoritma fp-growth akan membangun sebuah struktur pohon yang disebut dengan *fp-tree* (Rachman & Hunaifi, 2020). *Fp-tree* yang bertugas untuk menemukan *frequent itemset* tanpa perlu menggunakan proses pembuatan *generate candidat*

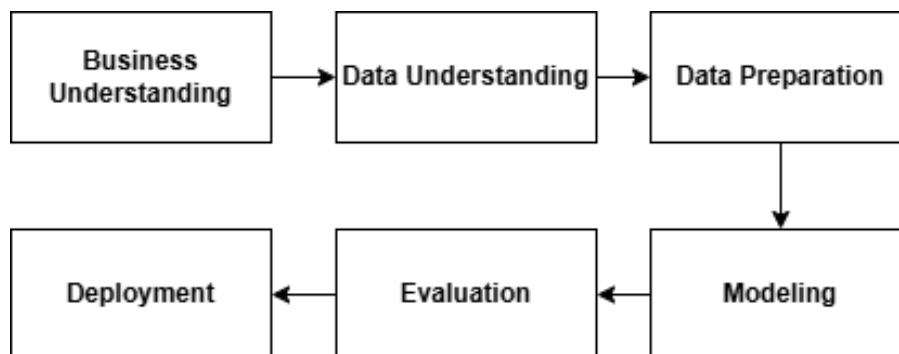
yang memakan waktu seperti pada algoritma apriori, oleh karena itu fp-growth lebih efisien dalam memproses data yang besar untuk menghasilkan pola asosiasi (Prasetyo & Hasugian, 2024). *Software* yang digunakan untuk melakukan perhitungan adalah *Rstudio* yang menggunakan bahasa R untuk menjalankan *software* tersebut (R. R. Setiawan & Jananto, 2023). *Rstudio* digunakan untuk mengimplemmentasikan algoritma apriori dan fp-growth. Selanjutnya akan dilakukannya perbandingan terhadap aturan yang dibuat dari kedua algoritma tersebut serta waktu eksekusi dan penggunaan *memory* (Idris et al., 2022)

Pada penelitian ini bertujuan untuk melakukan perbandingan hasil dari metode asosiasi algoritma apriori dan fp-growth dalam membuat pola pembelian berdasarkan dataset transaksi. Manfaat penelitian ini adalah mengetahui metode apa yang baik digunakan untuk mengetahui pola pembelian. Dan penelitian ini memiliki batasan sebagai berikut:

1. Algoritma yang digunakan adalah apriori dan fp-growth.
2. Pengujian dilakukan dengan membandingkan aturan yang dibuat, waktu proses dan sumber daya yang digunakan.
3. Objek yang diteliti merupakan transaksi yang terjadi pada retail.
4. Data yang digunakan berupa data sekunder yang di dapatkan dari kaggle.
5. *Software* yang digunakan untuk melakukan perhitungan yaitu Rstudio.

II. METODOLOGI

Pada penelitian proses data mining ini menggunakan beberapa siklus pada CRISP-DM siklus proses data mining terdapat enam fase yaitu, *Business Understanding*, *Data Understanding*, *Data Preparation*, *Modeling*, *Evaluation* dan *Deployment* (Schröer et al., 2021).



Gambar 1. Tahapan CRISP-DM

III. HASIL DAN PEMBAHASAN

3.1 Business Understanding

Pada tahap business understanding adalah untuk memahami tujuan serta kebutuhan pada bisnis, tujuan dari penelitian ini untuk membandingkan algoritma apriori dan fp-growth dalam memproses aturan asosiasi dalam segi aturan yang terbentuk, waktu yang digunakan dan penggunaan *memory* untuk mendapatkan sebuah informasi yang berguna untuk mengembangkan bisnis

3.2 Data Understanding

Pada tahap data understanding, data yang di gunakan pada penelitian ini berasal dari website kaggle. Pada *dataset* ini memiliki 541.910 baris transaksi, dan dataset ini memiliki 8 atribut yang berguna dalam melakukan asosiasi pola pembelian, berikut penjelasan terkait masing-masing atribut.

Tabel 1. Daftar Seluruh Atribut

Atribut	Keterangan
Invoice	Nomor unik setiap transaksi
StockCode	Nomor unik setiap produk
Description	Nama produk
Quantity	Jumlah produk yang terjual
InvoiceDate	Waktu terjadinya transaksi
Price	Harga produk
Customer ID	Id pelanggan
Country	Negara tempat terjadinya transaksi

3.3 Data Preparation

Pada tahap data preparation, data akan di olah yang bertujuan untuk memudahkan dalam tahapan data mining selanjutnya, dengan dilakukannya tahap ini diharapkan dataset yang digunakan dapat optimal dalam proses perhitungan. Dataset dilakukan pembersihan berdasarkan data yang memiliki nilai *null*, data tidak normal. Sebelum dilakukan pembersihan total data berjumlah 541.910 setelah dilakukan pembersihan total data menjadi 396.353.

3.4 Modeling

Pada tahap modeling, di lakukan menggunakan *software Rstudio* terhadap dataset transaksi:

Tabel 2. Kode Perhitungan Apriori

Kode	Penjelasan
profvis((Digunakan untuk memonitoring dan menyimpan penggunaan memory dan waktu esksekusi
library(arules)	Digunakan untuk menggunakan library arules
data <- read.csv("supermarket_data.csv", stringAsFactors = FALSE)	read.csv digunakan untuk memuat file bertipe csv dan di simpan pada variable data
transaksi <- as(split(data\$Description, data\$Invoice), "transactions")	Variable transaksi digunakan untuk merubah data ke format transaksi
rules <- apriori(transaksi, parameter = list(support = 0.02, confidence = 0.7))	Variable rules digunakan untuk menyimpan hasil perhitungan apriori() dengan menggunakan nilai minimum support 0.002 dan nilai confidence 0.7
inspect(rules))	inspect(rules) digunakan untuk menampilkan hasil aturan asosiasi yang sudah dibuat sebelumnya

Tabel 3. Kode Perhitungan Fp-Growth

Kode	Penjelasan
support_threshold <- 0.02 confidence_threshold <- 0.7	Variable yang digunakan untuk menyimpan nilai min support dan min confidence
association_rules <- fim4r(transactions, method = "fpgrowth", support=support_threshold, confidence = confidence_threshold, target = "rules")	Variable association_rules digunakan untuk menyimpan hasil perhitungan fim4r() dengan menggunakan metode fp-growth dan menggunakan variable support_threshold dan confidence_threshold
inspect(association_rules))	inspect(association_rules) digunakan untuk menampilkan hasil aturan asosiasi yang sudah dibuat sebelumnya

3.5 Evaluation

Berdasarkan model yang sudah dibuat maka memberikan hasil sebagai berikut:

Tabel 4. Aturan Asosiasi Apriori Dan Fp-Growth

No	Aturan
1	Jika membeli "PINK REGENCY TEACUP AND SAUCER" maka akan membeli "GREEN REGENCY TEACUP AND SAUCER" dengan <i>support</i> 0.02, <i>confidence</i> 0.82 dan <i>lift ratio</i> 22

No	Aturan
2	Jika membeli "PINK REGENCY TEACUP AND SAUCER" maka akan membeli "ROSES REGENCY TEACUP AND SAUCER" dengan <i>support</i> 0.02, <i>confidence</i> 0,78 dan <i>lift ratio</i> 18
3	Jika membeli "GREEN REGENCY TEACUP AND SAUCER" maka akan membeli "ROSES REGENCY TEACUP AND SAUCER" dengan <i>support</i> 0,02, <i>confidence</i> 0,78 dan <i>lift ratio</i> 18
4	Jika membeli "GARDENERS KEELING PAD CUP OF TEA" maka akan membeli "GARDENERS KNEELING PAD KEEP CALM" dengan <i>support</i> 0,02, <i>confidence</i> 0,72 dan <i>lift ratio</i> 17
5	Jika membeli "GREEN REGENCY TEACUP AND SAUCER" dan "PINK REGENCY TEACUP AND SAUCER" maka akan membeli "ROSES REGENCY TEACUP AND SAUCER" dengan <i>support</i> 0.02, <i>confidence</i> 0,84 dan <i>lift ratio</i> 19
6	Jika membeli "PINK REGENCY TEACUP AND SAUCER" dan "ROSES REGENCY TEACUP AND SAUCER" maka akan membeli "GREEN REGENCY TEACUP AND SAUCER" dengan <i>support</i> 0,02, <i>confidence</i> 0,89 dan <i>lift ratio</i> 23
7	Jika membeli "GREEN REGENCY TEACUP AND SAUCER" dan "ROSES REGENCY TEACUP AND SAUCER" maka akan membeli "PINK REGENCY TEACUP AND SAUCER" dengan <i>support</i> 0.02, <i>confidence</i> 0.72 dan <i>lift ratio</i> 23

Tabel 5. Hasil perbandingan

Algoritma	Waktu proses	Memory	Jumlah aturan
Apriori	1,91 detik	138 MB	7
Fp-growth	2 detik	140,3 MB	7

3.6 Deployment

Penelitian ini bertujuan untuk membandingkan antara algoritma apriori dan fp-growth dalam melakukan proses asosiasi data transaksi. Setelah di dapatkan hasil diharapkan dapat mengetahui metode yang terbaik dalam pembuatan pola pembelian pada data transaksi.

IV. SIMPULAN

Berdasarkan hasil penelitian ini menggunakan *software Rstudio* pada *dataset* transaksi di dapatkan bahwa algoritma apriori dan fp-growth menghasilkan 7 aturan asosiasi yang sama, dan untuk waktu proses apriori hanya membutuhkan 1.91 detik sedangkan fp-growth 2 detik dan penggunaan memory apriori 138 MB dan fp-growth 140,3 MB, oleh karena itu algoritma apriori memiliki performa lebih baik di bandingkan algoritma fp-growth.

DAFTAR PUSTAKA

- Chandra, S., & Wijaya, H. (2024). Implementasi Data Mining Untuk Menentukan Pola Pemilihan Kombinasi Menu Menggunakan Algoritma Apriori (Studi Kasus: The Coffee Theory). *JURNAL ALGOR*, 6(1), 10–22.
- Febrian, T., Rino, & Wijaya, H. (2023). Penerapan Data Mining Untuk Menganalisa Data Penjualan Barang Di Swalayan Dengan Menggunakan Algoritma Fp-Growth. *JURNAL ALGOR*, 4(2), 98–106.
<https://doi.org/https://doi.org/10.31253/algor.v4i2.1546>
- Freddie, Kurnia, Y., Arijanto, R., & Yakub. (2021). Perancangan Aplikasi Web Dengan Menggunakan Algoritma Apriori Pada Data Mining Untuk Mengetahui Pola Pembelian Konsumen PT Cipta Tunggal Elektronik. *JURNAL ALGOR*, 3(1), 71–82.
<https://doi.org/https://doi.org/10.31253/algor.v3i1.640>
- Idris, A. I., Sampetoding, E. A. M., Ardhana, V. Y. P., Maritsa, I., Sakri, A., Ruslan, H., & Manapa, E. S. (2022). Comparison of Apriori, Apriori-TID and FP-Growth Algorithms in Market Basket Analysis at Grocery Stores. *The IJICS (International Journal of Informatics and Computer Science)*, 6(2), 107–112.
- Lienata, B., Fenriana, I., Andre, & Safitri, R. D. (2021). Penerapan Data Mining Pada Penjualan Pakaian Brand Expand Dengan Algoritma Apriori Menggunakan Metode Association Rules PT. Vidiaelok Lestari Garmino. *JURNAL ALGOR*, 3(1), 83–95.
<https://doi.org/https://doi.org/10.31253/algor.v3i1.647>
- Musdalifah, I., & Jananto, A. (2022). Analisis Perbandingan Algoritma Apriori Dan FP-Growth Dalam Pembentukan Pola Asosiasi Keranjang Belanja Pelanggan. *Progresif: Jurnal Ilmiah Komputer*, 18(2), 175–184.
- Nisa, V. C., & Khasanah, F. N. (2023). Algoritma Apriori Dalam Identifikasi Pola Pembelian Konsumen Pada Produk Minuman. *INFORMAL: Informatics Journal*, 8(2), 156–164.
- Prasetyo, F., & Hasugian, H. (2024). Analisis Pola Pembelian Produk Makanan Menggunakan Algoritma FP-Growth Untuk Strategi Penjualan. *IDEALIS: InDonEsiA Journal Information System*, 7(1), 11–20.
- Rachman, R., & Hunaifi, N. (2020). Penerapan Metode Algoritma Apriori dan FP-

- Tree Pada Penentuan Pola Pembelian Obat. *Jurnal Informatika Dan Komputer*, 22(2), 175–182.
- Rahmi, A. N., & Mikola, Y. A. (2021). Implementasi Algoritma Apriori Untuk Menentukan Pola Pembelian Pada Customer (Studi Kasus: Toko Bakoel Sembako). *Information System Journal*, 4(1), 14–19.
- Schröer, C., Kruse, F., & Gómez, J. M. (2021). A systematic literature review on applying CRISP-DM process model. *Procedia Computer Science*, 181, 526–534.
- Setiawan, G., Putra, D. S. D., & Wijaya, H. (2022). Aplikasi Data Mining Berbasis Web Menggunakan Algoritma Apriori Untuk Analisa Pola Pembelian Barang Pada PT Menara Bahagia Bersama. *JURNAL ALGOR*, 3(2), 1–11. <https://doi.org/https://doi.org/10.31253/algov3i2.1020>
- Setiawan, R. R., & Jananto, A. (2023). Implementasi Data Mining Untuk Rekomendasi Penyedia Pupuk Non Subsidi Dengan Menggunakan Metode Algoritma Apriori. *Jurnal Tekno Kompak*, 17(1), 13–24.
- Soepriyono, G., & Triayudi, A. (2023). Implementasi Data Mining dengan Algoritma Apriori dalam Menentukan Pola Pembelian Aksesoris Laptop. *JURNAL MEDIA INFORMATIKA BUDIDARMA*, 7(4), 2087–2096.
- Styawati, S., Nurkholis, A., & Anjumi, K. N. (2021). Analisis Pola Transaksi Pelanggan Menggunakan Algoritme Apriori. *J-SAKTI (Jurnal Sains Komputer Dan Informatika)*, 5(2), 619–626.