

Market Basket Analysis Data Groceries Menggunakan Equivalence Class Transformation

Musthofa Galih Pradana ¹, Khoironi ²

¹ Program Studi Sains Data, Fakultas Ilmu Komputer, Universitas Pembangunan Nasional Veteran Jakarta

²Departemen Teknik Informatika dan Komputer, Politeknik Elektronika Negeri Surabaya

INFORMASI ARTIKEL	ABSTRAK
Kata Kunci: Asosiasi, ECLAT, Groceries, Data;	Peningkatan volume data, pengelolaan dan analisis data menjadi elemen penting bagi pengambilan keputusan bisnis. Salah satu metode yang populer adalah dengan Market Basket Analysis untuk menganalisis data transaksi yang ada pada sebuah data. Pendekatan dengan metode ECLAT adalah salah satu yang bisa dilakukan, karena keunggulannya pada teknik depth-first search. Data yang diolah pada penelitian ini sebanyak 38.766 record data transaksi. Hasil penelitian menunjukkan kombinasi 1 itemset dengan minimal support 20 dan 250 mampu menghasilkan data yang relevan, sementara pada kombinasi 2 itemset hanya mampu menghasilkan analisis pada nilai minimal support 10 dikarenakan keterbatasan data pembelian dengan kombinasi 2 item. Waktu eksekusi algoritma ECLAT juga menunjukkan kondisi 1 itemset dengan waktu 0,0402 detik dan eksekusi terlambat pada kondisi 2 item-set 8,773 detik.
Diserahkan 04/02/2025 Direvisi 07/03/2025 Diterima 20/03/2025	© Author's (2025) Creative Commons Attribution-NonCommercial 4.0 International License



1. PENDAHULUAN

Peningkatan volume data, pengelolaan dan analisis data menjadi elemen penting bagi pengambilan keputusan bisnis. Salah satu pendekatan yang populer dalam analisis data adalah Market Basket Analysis (MBA), sebuah metode yang digunakan untuk mengidentifikasi pola pembelian konsumen dengan menganalisis hubungan antar produk dalam keranjang belanja [3]. Market Basket Analysis sering digunakan dalam ritel, e-commerce, dan sektor lainnya untuk mengungkap asosiasi tersembunyi antar produk, yang dapat membantu dalam menyusun strategi pemasaran, mengelola inventaris, dan meningkatkan pengalaman pelanggan.

Market Basket Analysis umumnya memanfaatkan konsep association rule mining untuk menemukan hubungan antar item dalam transaksi. Setiap algoritma tentu memiliki kelebihan dan kekurangannya masing-masing seperti FP-Growth memiliki kecepatan eksekusi dibandingkan apriori dan eclat [14]. Di beberapa kasus apriori juga memiliki keunggulan dibandingkan kedua algoritma lain terkait penggunaan memori [13]. Algoritma seperti Apriori dan FP-Growth telah menjadi pilihan utama dalam analisis ini. Namun, metode-metode ini memiliki beberapa keterbatasan, terutama dalam hal efisiensi ketika berhadapan dengan dataset berukuran besar. Metode Apriori, misalnya, memerlukan iterasi berulang untuk menghasilkan kandidat itemset yang sering, yang dapat meningkatkan kompleksitas komputasi. Sementara itu, FP-Growth memerlukan pembangunan struktur pohon yang kompleks, yang dapat menjadi kurang efektif dalam beberapa kasus. Sebagai alternatif, metode Eclat (Equivalence Class Transformation) menawarkan pendekatan yang lebih sederhana dan efisien dalam menggali pola asosiasi. Eclat menggunakan teknik depth-first search dan representasi berbasis tidset (transaction ID set) untuk menemukan itemset yang sering tanpa perlu melakukan iterasi berulang seperti pada Apriori. Pendekatan ini memungkinkan Eclat untuk menangani dataset besar dengan lebih efisien dan menghasilkan hasil yang akurat dalam waktu yang lebih singkat. Dimana hasil eclat juga menunjukkan hasil ukuran transaksi maksimum dan kerapatan frequent itemset jauh lebih baik dibandingkan kedua algoritma lain [2].

Penerapan metode Eclat dalam MBA menjadi semakin relevan seiring dengan meningkatnya kebutuhan analisis data yang cepat dan akurat. Misalnya, dalam industri e-commerce, di mana jumlah transaksi meningkat setiap hari, Eclat dapat digunakan untuk mengidentifikasi pasangan produk yang sering dibeli bersama sehingga dapat digunakan untuk rekomendasi produk secara real-time. Selain itu, metode ini juga dapat diterapkan dalam manajemen inventaris untuk mengoptimalkan stok berdasarkan pola permintaan yang terdeteksi. Meskipun memiliki keunggulan dalam efisiensi, penggunaan metode Eclat masih menghadapi beberapa tantangan, seperti kebutuhan akan optimalisasi dalam menangani dataset yang sangat besar dan kompleks, serta kemampuan untuk menghasilkan hasil yang dapat diinterpretasikan dengan mudah oleh pengguna bisnis. Oleh karena itu, penelitian ini bertujuan untuk mengeksplorasi potensi penerapan metode Eclat dalam MBA, mengidentifikasi kekuatan dan kelemahannya, serta mengusulkan pendekatan yang dapat meningkatkan kinerja metode ini dalam skenario dunia nyata. Penelitian ini diharapkan dapat memberikan kontribusi dalam mengembangkan solusi analitik berbasis data yang lebih efisien dan aplikatif untuk kebutuhan industri modern. Dengan memanfaatkan keunggulan metode Eclat, diharapkan analisis pola pembelian konsumen dapat dilakukan dengan lebih cepat dan tepat, sehingga mendukung pengambilan keputusan strategis yang lebih baik..

2. TEORI DAN PENELITIAN RELEVAN

2.1 Data Mining

Data mining adalah proses mengekstraksi informasi berharga dari dataset yang besar dan kompleks dengan menggunakan metode statistik, algoritma pembelajaran mesin, atau teknik berbasis aturan. Dalam konteks bisnis, data mining digunakan untuk menemukan pola tersembunyi, tren, dan hubungan yang dapat mendukung pengambilan keputusan strategis [5].

Proses data mining biasanya melibatkan beberapa langkah utama [15], seperti

- a. Preprocessing: Membersihkan dan mengorganisir data agar siap dianalisis.
- b. Pattern Discovery: Menggunakan algoritma untuk menemukan pola, seperti klasifikasi, klusterisasi, atau asosiasi.
- c. Evaluation: Mengukur relevansi dan validitas pola yang ditemukan untuk memastikan aplikabilitasnya.

Data mining menjadi inti dari Market Basket Analysis (MBA), karena memungkinkan identifikasi hubungan antara item dalam data transaksi besar dengan cepat dan efisien.

2.2 Market Basket Analysis

Market Basket Analysis (MBA) adalah metode analisis data yang bertujuan untuk mengidentifikasi pola pembelian pelanggan berdasarkan hubungan antar produk dalam keranjang belanja. MBA sering digunakan dalam bidang ritel untuk memahami kebiasaan pelanggan, seperti produk apa saja yang sering dibeli bersama, guna mengembangkan strategi pemasaran, rekomendasi produk, dan pengelolaan inventaris. Analisis ini bekerja dengan mengidentifikasi itemset yang sering muncul dalam transaksi menggunakan algoritma tertentu, seperti Apriori, FP-Growth, atau Eclat [4].

2.3 Association Rules

Association rules adalah aturan yang menggambarkan hubungan atau asosiasi antara item dalam dataset. Aturan ini biasanya dinyatakan dalam bentuk implikasi $A \rightarrow B$. $A \rightarrow B$, yang berarti jika item A dibeli, maka item B cenderung juga dibeli. Association rules menjadi dasar untuk menemukan pola-pola yang dapat diterapkan dalam MBA, seperti penempatan produk di toko atau rekomendasi produk di platform e-commerce. Dengan memahami ketiga konsep ini, MBA menggunakan analisis data dan association rules untuk memberikan nilai strategis dalam bisnis, memungkinkan organisasi untuk memahami perilaku pelanggan secara mendalam dan meresponsnya dengan strategi yang lebih efektif [6].

2.4 Eclat

Eclat (Equivalence Class Transformation) adalah salah satu algoritma data mining yang digunakan untuk menemukan frequent itemsets dalam dataset transaksi. Berbeda dengan algoritma Apriori yang berbasis breadth-first search, Eclat menggunakan pendekatan depth-first search untuk mengeksplorasi dataset. Metode ini bekerja dengan merepresentasikan hubungan antar item dalam bentuk tidset (Transaction ID Set), yaitu kumpulan ID transaksi yang mengandung item tertentu. Prosesnya meliputi: Membuat daftar item dan tidset awal dari dataset transaksi; Menyusun frequent itemsets berdasarkan irisan tidset antar item; Melanjutkan eksplorasi hingga tidak ada lagi kombinasi yang memenuhi batas support [4].

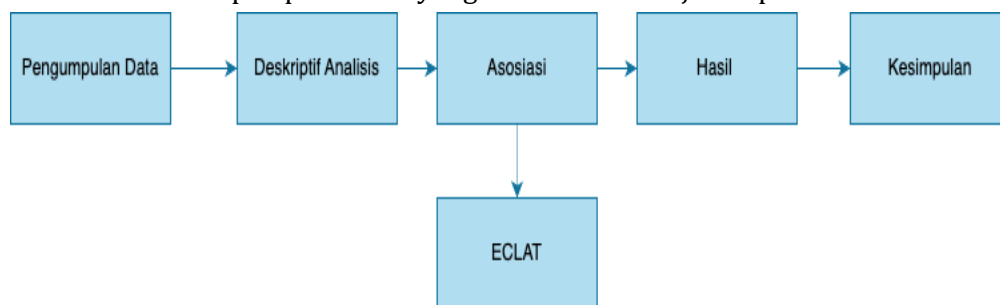
2.5 State of The Art

Teknik ini kemudian diperluas dengan menggunakan algoritma Eclat Inkremental (i-Eclat). Hasil percobaan menunjukkan peningkatan 32% dalam eksekusi waktu untuk algoritma yang diusulkan [9], meskipun ECLAT memiliki waktu eksekusi untuk kumpulan data besar [8]. Namun ECLAT memiliki kelebihan pada pembacaan data secara vertical [11]. Perbandingan metode ECLAT, Apriori dan FP-Growth hasil menunjukkan keunggulan Eclat daripada Apriori dan Fpgrowth [13]. Pada penelitian yang lain menunjukkan Apriori memberikan opsi aturan yang lebih banyak dibandingkan ECLAT dengan perbandingan 9 aturan dan 11 aturan, namun eksekusi program menunjukkan keunggulan ECLAT selama 0,2 second [7]. Perbandingan ECLAT dan apriori dalam data groceries menunjukkan nilai konstan pada nilai confidence dan apriori cenderung mengalami penurunan [10]. Kombinasi Filter Bloom pada Eclat mampu meningkatkan kemampuan asosiasi Eclat secara signifikan [1].

3. METODE PENELITIAN

3.1 Tahapan Penelitian

Berikut ini adalah tahapan penelitian yang dilakukan ditunjukkan pada Gambar 1.



Gambar 1. Alur Penelitian

3.2 Data

Data yang digunakan adalah data public dimana data ini adalah data transaksi yang digunakan pada perusahaan retail, yang memiliki karakteristik data pada **Tabel 1**

Tabel 1. Dataset Penelitian

Member Number	Date	ItemDescription
1808	21-07-2024	tropical fruit
2552	05-01-2024	whole milk
2300	19-09-2024	pip fruit
1187	12-12-2024	other vegetables
3037	01-02-2024	whole milk
4941	14-02-2024	rolls/buns
4501	08-05-2024	other vegetables
3803	23-12-2024	pot plants
2762	20-03-2024	whole milk

4119	12-02-2024	tropical fruit
1340	24-02-2024	citrus fruit
2193	14-04-2024	beef
1997	21-07-2024	frankfurter
4546	03-09-2024	chicken
4736	21-07-2024	butter
1959	30-03-2024	fruit/vegetable juice
1974	03-05-2024	packaged fruit/vegetables
2421	02-09-2024	chocolate

Data yang didapatkan untuk penelitian ini adalah data public pada sebuah toko grosir yang digunakan untuk pengujian model ECLAT pada data groceries..

4. HASIL DAN PEMBAHASAN

4.1 Hasil

Hasil dari penelitian ini akan dibagi menjadi beberapa bagian untuk memudahkan interpretasi dan analisis hasil.

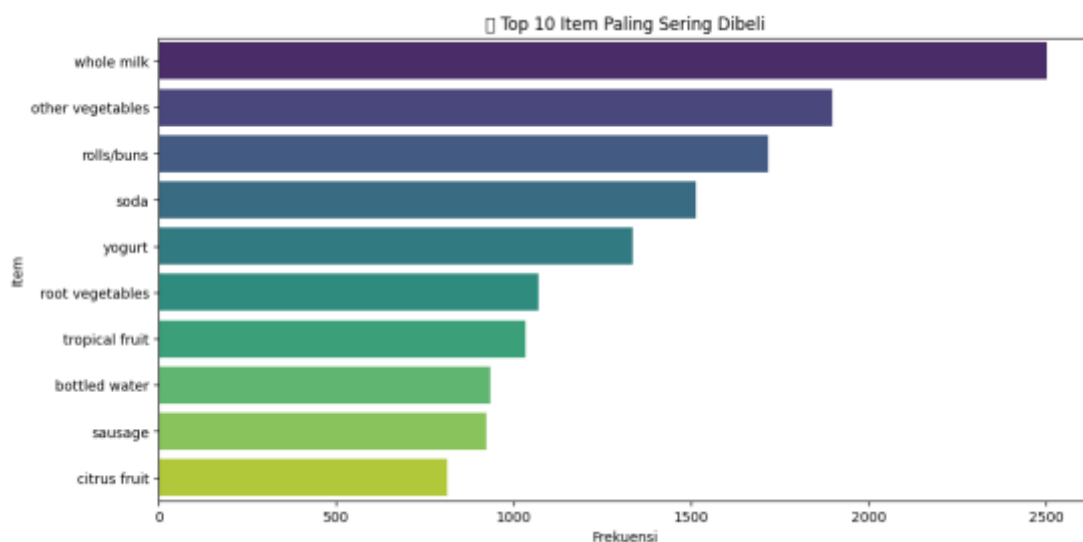
1) Analisis Deskriptif

Hasil dari deskriptif data yang ada pada dataset yang digunakan dimana jumlah data sebanyak 38.766 transaksi. Proses analisis secara deskriptif ini memberikan gambaran secara umum dari beberapa identifikasi terhadap data yang digunakan dalam penelitian. Data yang digunakan dengan 3 variable/column dengan tipe data pada **Tabel 2**.

Tabel 2. Tipe Data Column

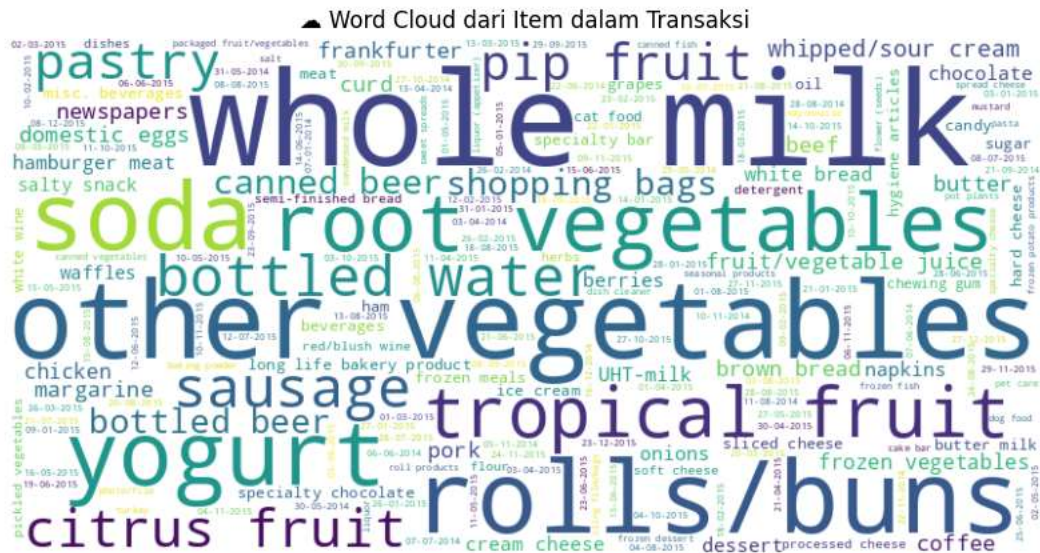
Data	Column	Dtype
0	Member Number	int64
1	Date	object
2	Item Description	object

Dari tipe data tersebut dipetakan data yang digunakan untuk melihat data dengan jumlah transaksi terbesar, diambil dari 10 transaksi teratas. Data pada 10 transaksi teratas ditunjukkan pada Gambar 2.



Gambar 2. Top 10 Item

Visualisasi dari word cloud kata atau item yang paling sering muncul divisualisasikan dalam Gambar 3.



Gambar 3. Wordcloud Item

2) Kombinasi Itemset

Sebelum dicari kombinasi itemset yang mungkin akan muncul data aturan asosiasi menggunakan ECLAT, berikut ini adalah hasil yang didapatkan dari frequent itemset yang ada. Data yang muncul adalah secara random dan belum merepresentasikan urutan dari kemunculan data item penjualan. Detail pada data ditunjukkan pada **Tabel 3**.

Tabel 3. Frequent Itemset

Frequent Itemset		
No	Itemset	Support
0	{margarine}	491
1	{margarine}	38765
2	{salty snack}	283
3	{salty snack}	38765
4	{canned beer}	717
88	{other vegetables}	1898
89	{other vegetables}	38765
90	{brown bread}	571
91	{brown bread}	38765
92	{hamburger meat}	327

Dari frequent itemset yang terbentuk ada beberapa item yang paling sering dibeli seperti margarine, salty snack dimana nilai support yang cukup tinggi untuk beberapa item tersebut. Tahapan selanjutnya akan dicari untuk kombinasi itemset yang mungkin muncul dalam transaksi. Untuk mencari nilai kombinasi Itemset, digunakan aturan dan skenario dengan perbedaan nilai support yang ditentukan disesuaikan dengan kondisi data.

3) 1-Itemset

Minimal Support : 20

Hasil yang didapatkan dari minimal support 20 adalah data pada Tabel 4.

Tabel 4. 1-Itemset 20 Data

Item	Support
Whole Milk	2502
Other Vegetables	1898

Rolls/Buns	1716
Soda	1514
Yogurt	1334
Beverages	20
Citrus Fruit	20

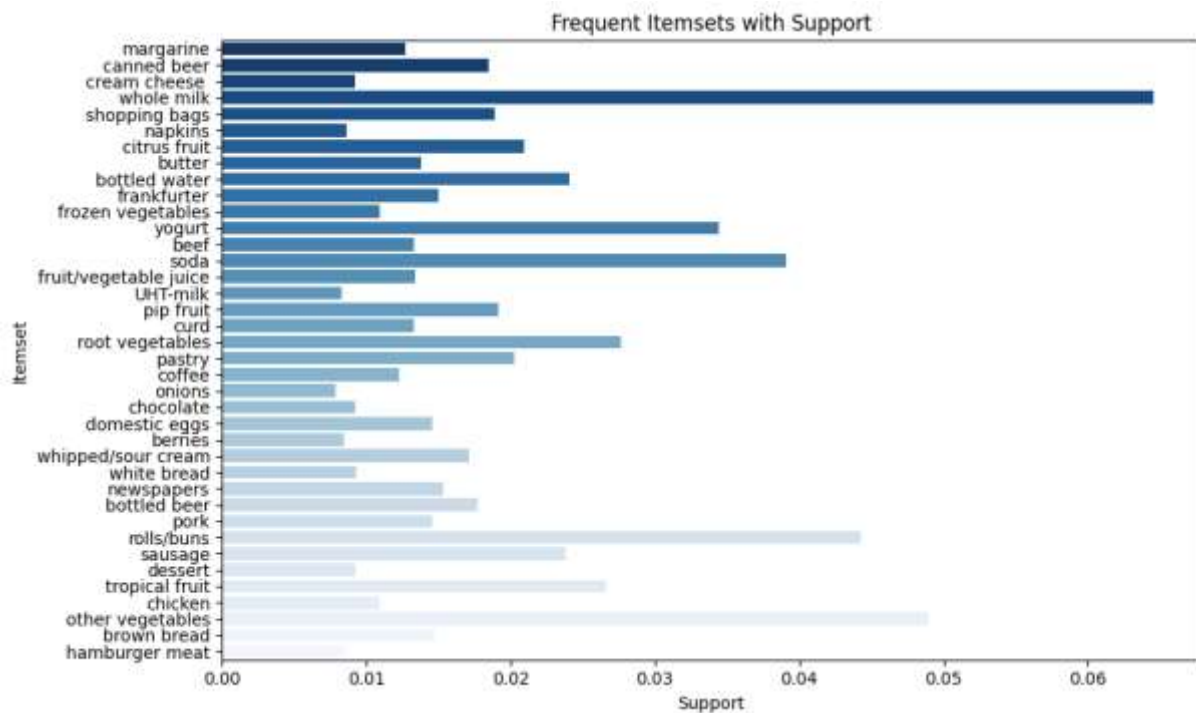
Minimal Support : 250

Hasil yang didapatkan dari minimal support 250 adalah data pada **Tabel 5**.

Tabel 5. 1-Itemset 250 Data

Item	Support
Whole Milk	2502
Other Vegetables	1898
Rolls/Buns	1716
Soda	1514
Yogurt	1334
Meat	252
Frozen Meal	251

Hasil visualisasi pada 250 data support ditunjukan pada Gambar 4.



Gambar 4. Visualisasi Itemset dengan Support

4) 2-Itemset

Minimal Support : 20

Hasil yang didapatkan dari minimal support 20 untuk kombinasi 2 itemset tidak ditemukan, artinya bahwa data ini tidak ditemukan karena tidak ada transaksi melebihi 20 item secara bersamaan untuk 2 item. Hasil ditujukan pada Gambar 5.

```
Frequent 2-Itemsets:  
Empty DataFrame  
Columns: [Itemset, Support]  
Index: []
```

🕒 Waktu eksekusi: 8.5382 detik

Gambar 5. Hasil 2 Itemset 20 Data

Minimal Support : 10

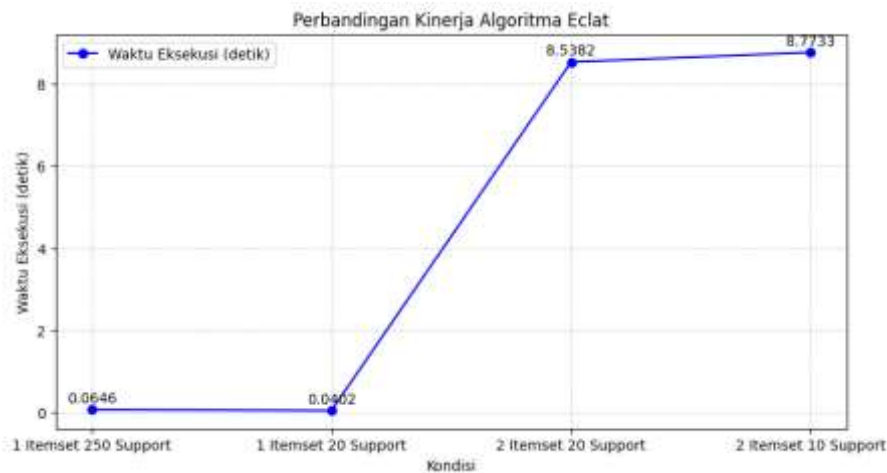
Karena hasil untuk 20 data support tidak berhasil didapatkan, maka secara logis nilai support harus diturunkan, dan digunakan nilai 10 sebagai standar support. Hasil yang didapatkan dari minimal support 10 untuk kombinasi 2 itemset ditunjukkan pada Tabel 6.

Tabel 6. 2 Itemset 10 Data

Item	Support
Whole Milk, Canned Beer	15
Whole Milk, Frankfurter	13
Whole Milk, Sugar	11

5) Kinerja Algoritma

Perbandingan kinerja algoritma ECLAT dalam mengeksekusi skenario yang diberikan menunjukkan waktu eksekusi paling cepat adalah pada kondisi 1 itemset dengan nilai support sebanyak 250, dan yang paling lambat pada kondisi 2 itemset dengan nilai support 10. Hal ini tentu selaras, karena semakin banyak kombinasi yang dimunculkan proses komputasi yang dibutuhkan juga semakin besar. Detail grafis hasil kinerja algoritma ditunjukkan pada Gambar 6.



Gambar 6. Grafik Perbandingan Kinerja ECLAT

4.2 Pembahasan

Hasil yang didapatkan dari eksperimen percobaan analisis yang sudah dilakukan ada beberapa poin berikut ini :

- Kombinasi itemset 1 memang menunjukkan hasil yang bervariasi mengikuti jumlah minimum support yang digunakan

- b. Daftar frequent itemset yang muncul sangat sering dalam transaksi seperti : Whole Milk dengan nilai support sebesar 2.502 Other Vegetables dengan nilai support 1.898.
- c. Rolls/Buns, Soda, Yogurt, Beverages, Citrus Fruit masuk kategori yang cukup sering dibeli.
- d. Kombinasi itemset 2 menunjukkan hasil yang lebih sedikit, dikarenakan tidak ada penjualan produk dengan nilai tertentu yang memenuhi ambang batas support yang ditentukan.
- e. Kombinasi itemset 2 dapat dimunculkan pada nilai support minimal 10 dengan menampilkan data transaksi Whole Milk dan Canned Beer sebanyak 15 transaksi sebagai transaksi yang paling banyak.

Hal ini dapat disimpulkan juga bahwa masih cukup rendah kombinasi yang terjadi untuk pembelian 2 item sekaligus secara bersamaan

5. KESIMPULAN

Kesimpulan yang dapat ditarik dalam penelitian ini adalah sebagai berikut :

- a. Kombinasi itemset 1 menghasilkan rekomendasi yang bervariasi dengan item tertinggi seperti Whole Milk, Other Vegetables.
- b. Kombinasi itemset 2 menunjukkan hasil yang lebih sedikit, dikarenakan tidak ada penjualan produk dengan nilai tertentu yang memenuhi ambang batas support sebanyak 20 support minimum yang ditentukan.
- c. Kombinasi itemset 2 dapat dimunculkan pada nilai support minimal 10 dengan menampilkan data transaksi Whole Milk dan Canned Beer sebanyak 15 transaksi sebagai transaksi yang paling banyak.
- d. Waktu eksekusi algoritma juga menunjukkan semakin banyak kombinasi yang dimunculkan proses komputasi yang dibutuhkan juga semakin besar dengan waktu eksekusi terkecil pada kondisi 1 itemset dengan waktu 0,0402 detik dan eksekusi terlambat pada kondisi 2 itemset 8,773 detik.

Saran pada penelitian ini adalah untuk bisa memperkaya model skenario uji yang dilakukan dan membandingkan kinerja algoritma ECLAT dengan algoritma asosiasi lainnya untuk mendapatkan hasil yang lebih komprehensif

6. DAFTAR PUSTAKA

- [1] Abbasi, S., & Moeini, A. (2021). BloomEclat: Efficient Eclat Algorithm based on Bloom filter. In *Journal of Algorithms and Computation* (Vol. 53, Issue 1). <http://jac.ut.ac.ir>
- [2] Akhmad Khabrun Siregar, B. A. K. A. P. K. S. (2018). Perbandingan Algoritme FP-Growth dan Eclat untuk Analisis Pola Pembelian Konsumen pada Toko "X." CITISEE.
- [3] Business Analytics Institute. (2024). *Market Basket Analysis for Business Growth*.
- [4] Galit Shmueli, Peter C. Bruce, Peter Gedeck, Inbal Yahav, & Nitin R. Patel. (2023). *Machine Learning for Business Analytics: Concepts, Techniques, and Applications* in R. Wiley.
- [5] George Dimitoglou, Leonidas Deligiannidis, & Hamid Arabnia. (2024). *Big Data, Data Mining and Data Science: Algorithms, Infrastructures, Management and Security* (Intelligent Computing Book 2) (George Dimitoglou, Leonidas Deligiannidis, & Hamid Arabnia, Eds.). De Gruyter.
- [6] Iqbal Sarker, Alan Colman, Jun Han, & Paul Watters. (2022). *Context-Aware Machine Learning and Mobile Data Analytics: Automated Rule-based Services with Intelligent Decision-Making*. Springer.
- [7] Krishnan, M. S., Nair, A. S., & Sebastian, J. (2022). Comparative Analysis of Apriori and ECLAT Algorithm for Frequent Itemset Data Mining (pp. 489–497). https://doi.org/10.1007/978-981-16-3675-2_37
- [8] Lin, N., & Win, T. (2024). Optimizing the Performance of the ECLAT Algorithm using K-means Clustering for Frequently Itemset Mining for Large Datasets. 2024 5th International Conference on Advanced Information Technologies (ICAIT), 1–6. <https://doi.org/10.1109/ICAIT65209.2024.10754926>

- [9] Man, M., Jalil, M. A., & Bakar, W. A. W. A. (2023). Fi-Eclat: An enhancement of incremental Eclat algorithm. 060006. <https://doi.org/10.1063/5.0110230>
- [10] Mohapatra, D., Tripathy, J., Mohanty, K. K., & Nayak, D. S. K. (2021). Interpretation of Optimized Hyper Parameters in Associative Rule Learning using Eclat and Apriori. 2021 5th International Conference on Computing Methodologies and Communication (ICCMC), 879–882. <https://doi.org/10.1109/ICCMC51019.2021.9418049>
- [11] Omran, O. M. Ben. (2023). Using ECLAT Algorithm for Privacy Preserving on Publishing Data. 2023 IEEE 3rd International Maghreb Meeting of the Conference on Sciences and Techniques of Automatic Control and Computer Engineering (MI-STA), 407–412. <https://doi.org/10.1109/MI-STA57575.2023.10169272>
- [12] Srinadh, V. (2022a). Evaluation of Apriori, FP growth and Eclat association rule mining algorithms. International Journal of Health Sciences, 7475–7485. <https://doi.org/10.53730/ijhs.v6nS2.6729>
- [13] Srinadh, V. (2022b). Evaluation of Apriori, FP growth and Eclat association rule mining algorithms. International Journal of Health Sciences, 7475–7485. <https://doi.org/10.53730/ijhs.v6nS2.6729>
- [14] Yusuf Husain, Enny Dwi Oktaviyani, & Sherly Christina. (2023). Analisis Perbandingan Algoritma Apriori, FP-Growth, Dan Eclat dalam Menemukan Pola Pembelian Konsumen. KONSTELASI: Konvergensi Teknologi Dan Sistem Informasi, 3(2), 231–243. <https://doi.org/10.24002/konstelasi.v3i2.7007>
- [15] Zunan Setiawan, Muhammad Fajar , Arif Mudi Priyatno, Anggi Yhurinda Perdana Putri, Mediana Aryuni , Siti Yuliyanti, Harya Widiputra , Budanis Dwi MeilaZunan Setiawan, Muhammad Fajar , Arif Mudi Priyatno, Anggi Yhurinda Perdana Putri, Mediana Aryuni , Siti, A. W. (2023). Buku Ajar Data Mining. PT. Sonpedia Publishing Indonesia

