

Performance Evaluation of Agentic Workflow-Driven Trend-Aware Rule Mining for Dynamic Menu Bundling

Andri Triyono^{a*}, Kartika Imam Santoso^b, Rohman Hadi Al Haq^c

^{a,b,c} Universitas An Nuur, Grobogan, Indonesia

* correspondence : andriTriyono1@gmail.com

Abstract— Digital transformation in the culinary industry currently demands moving beyond writing static lines of code, instead acting as an AI orchestrator adaptive to real-world conditions. This research focuses on addressing significant challenges in traditional data mining methods, such as the Apriori and FP-Growth algorithms, which often lack the flexibility to handle dynamic variables like ambient temperature fluctuations. Through the innovative orchestration of the Trend-Aware Rule Mining (TARM) algorithm and a LangGraph-based Agentic Workflow, this study transforms raw association rules into strategic business decisions via an iterative reasoning process and self-correction mechanism. Experimental results on a dataset of 52,494 rows demonstrate TARM's computational superiority, with memory usage of only 8.04 MB, significantly more efficient than Apriori's 127.44 MB. Furthermore, the synergy between the Strategy Agent and Evaluator Agent achieved a logic consistency score of 100%, validated by an independent audit with an average score of 96.25%. These findings confirm that the developed system is in a ready-to-use state to support precise and adaptive decision-making automation in production environments.

Index Terms— Agentic Workflow; Dynamic Menu Bundling; LangGraph; Trend-Aware Rule Mining.

Abstrak— Transformasi digital pada industri kuliner saat ini menuntut kita untuk tidak lagi sekadar menulis baris kode statis, melainkan bertindak sebagai orkestrator AI yang adaptif terhadap realita di lapangan. Penelitian ini berfokus pada penyelesaian PR besar dalam metode penambangan data tradisional, seperti algoritma Apriori dan FP-Growth, yang sering kali kaku dalam menangani variabel dinamis seperti fluktuasi suhu lingkungan. Melalui orkestrasi inovatif antara algoritma Trend-Aware Rule Mining (TARM) dan Agentic Workflow berbasis LangGraph, riset ini mentransformasi aturan asosiasi mentah menjadi keputusan bisnis strategis melalui proses penalaran iteratif dan mekanisme koreksi mandiri. Hasil eksperimen pada dataset sebesar 52.494 baris membuktikan keunggulan komputasi TARM, di mana penggunaan memori hanya sebesar 8,04 MB, jauh lebih efisien dibandingkan Apriori yang mencapai 127,44 MB. Selain itu, sinergi antara Strategy Agent dan Evaluator Agent berhasil mencapai skor konsistensi logika sebesar 100%, yang divalidasi melalui audit independen dengan skor rata-rata 96,25%. Temuan ini menegaskan bahwa sistem yang dikembangkan sudah dalam kondisi ready-to-use untuk mendukung otomatisasi pengambilan keputusan yang presisi dan adaptif di lingkungan produksi.

Kata Kunci— Agentic Workflow; Dynamic Menu Bundling; LangGraph; Trend-Aware Rule Mining.

I. PENDAHULUAN

Transformasi digital pada industri kuliner telah menggeser paradigma operasional secara fundamental, di mana strategi pemasaran statis kini bertransformasi menjadi intervensi strategis berbasis presisi data [1]. Dalam ekosistem yang kompetitif ini, product bundling tetap menjadi salah satu mekanisme promosi paling efektif untuk meningkatkan pendapatan sekaligus mengoptimalkan pengalaman pelanggan melalui pemanfaatan interdependensi produk [2]. Secara konvensional, penyusunan paket promosi tersebut dilakukan melalui analisis keranjang pasar (*market basket analysis*) dengan memanfaatkan *association rule mining* guna mengidentifikasi pola signifikan dalam dataset transaksi yang masif [3]. Namun demikian, perilaku konsumsi saat ini bersifat sangat volatil dan responsif terhadap variabel eksternal, terutama fluktuasi suhu lingkungan dan tren temporal [4]. Meskipun korelasi antara kondisi lingkungan dan pola pembelian telah diakui secara luas, kerangka kerja penambangan data tradisional sering kali bersifat kaku dan belum mampu mengintegrasikan kesadaran kontekstual (*context-aware*) secara *real-time* ke dalam proses pengambilan keputusan otomatis [5].

Algoritma klasik seperti Apriori dan FP-Growth, meskipun efektif dalam mengekstraksi pola, menghadapi tantangan signifikan dalam mengintegrasikan variabel eksternal yang dinamis seperti suhu lingkungan [6]. Pendekatan tradisional tersebut cenderung menghasilkan aturan statis yang kurang

mempertimbangkan konteks, sehingga menghambat fleksibilitas penyesuaian rekomendasi menu terhadap perubahan tren cuaca secara *real-time* [7]. Selain itu, terdapat diskrepansi operasional di mana hasil penambangan data mentah (*raw rules*) sering kali tidak memiliki basis penalaran logis yang memadai untuk implementasi langsung ke dalam strategi bisnis [8]. Sistem konvensional juga belum dilengkapi dengan mekanisme validasi otomatis guna menjamin kepatuhan paket *bundling* terhadap batasan ekonomi serta relevansi praktis di lapangan [9]. Akibatnya, pemanfaatan hasil penambangan data dalam industri ritel masih sangat bergantung pada interpretasi manual, yang pada akhirnya membatasi efektivitas otomatisasi pengambilan keputusan yang cerdas dan adaptif [10].

Guna mengatasi keterbatasan tersebut, penelitian ini mengusulkan sebuah orkestrasi baru: integrasi inovatif antara Agentic Workflow berbasis Large Language Models (LLM) dengan algoritma Trend-Aware Rule Mining (TARM) sebagai solusi adaptif untuk dynamic menu bundling [11]. Melalui pemanfaatan framework LangGraph, sistem ini dirancang bukan hanya untuk mengolah aturan asosiasi mentah, tetapi mentransformasinya menjadi keputusan bisnis strategis melalui proses penalaran iteratif dan evaluasi mandiri oleh agen cerdas [12]. Pendekatan ini menjadi krusial karena metode data mining tradisional sering kali tidak memiliki lapisan kognitif yang diperlukan untuk menginterpretasikan variabel lingkungan—seperti indikator suhu—ke dalam strategi pemasaran yang ready pakai [13]. Oleh karena itu, tujuan utama dari riset ini adalah mengevaluasi secara sistematis efisiensi komputasi dan reliabilitas penalaran dari arsitektur agen tersebut melalui Layered Evaluation Framework [14]. Hipotesis penelitian ini menyatakan bahwa sinergi antara kemampuan ekstraksi pola dari TARM dan Reasoning Node pada alur kerja agen akan secara signifikan meningkatkan integritas pola sekaligus menjaga konsistensi logika pada paket menu yang dihasilkan [15].

II. METODE

Kerangka Evaluasi Berlapis (*Layered Evaluation Framework*)

Penelitian ini mengadopsi pendekatan Experimental Quantitative yang diartikulasikan melalui sebuah Layered Evaluation Framework. Metodologi ini dirancang bukan sekadar untuk mengumpulkan data, melainkan sebagai "jembatan" sistematis untuk memvalidasi integrasi antara algoritma penambangan data tradisional dengan arsitektur *Agentic AI* yang kompleks. Strategi evaluasi ini dibagi menjadi tiga fase yang saling berkesinambungan untuk memastikan sistem memiliki fondasi komputasi yang kuat dan penalaran yang reliabel:

- **Fase 1: Computational Performance & Pattern Integrity** Tahap ini berfokus pada validasi teknis algoritma TARM melalui *benchmarking* komparatif terhadap Apriori dan FP-Growth guna menjamin efisiensi basis data sebelum diproses oleh lapisan *Agentic Reasoning* [16].
- **Fase 2: Agentic Reasoning Reliability (The Core Innovation)** Fase ini merepresentasikan inti inovasi penelitian (*core innovation*), di mana fokus evaluasi diarahkan pada arsitektur kognitif (*reasoning core*) yang dikelola melalui kerangka kerja LangGraph [17]. Kapabilitas agen diuji dalam mentransformasikan aturan asosiasi mentah (*discovered rules*) menjadi keputusan bisnis strategis melalui skenario Ablation Study. Studi ini membandingkan performa sistem secara komprehensif antara konfigurasi yang melibatkan Reasoning Node dengan konfigurasi tanpa modul penalaran, guna membuktikan nilai tambah fungsional dari *Large Language Model* (LLM) dalam proses pengambilan keputusan. Metrik evaluasi yang diterapkan mencakup Inference Latency untuk mengukur kecepatan respon, Logic Consistency Score untuk memitigasi risiko halusinasi logika, serta Token Efficiency untuk menganalisis efisiensi operasional sistem di lingkungan produksi.
- **Fase 3: Multi-Perspective Validation (The "Proof")** Sebagai tahap pembuktian akhir, validasi yang bersifat subjektif digantikan dengan metodologi audit objektif yang diakui secara ilmiah [18]. Proses ini melibatkan teknik LLM-as-a-Judge menggunakan model dengan kapabilitas lebih tinggi (seperti Gemini 1.5 Pro) sebagai juri independen untuk menilai keselarasan strategis (*strategic alignment*). Selain itu, dilakukan *Expert Semantic Review* untuk mengukur reliabilitas hasil bundling berdasarkan relevansi dan kesiapan implementasi (*actionability*) di lapangan.

Fase 1: Computational Performance & Pattern Integrity

Fase pertama merupakan tahapan validasi fundamental untuk menguji keandalan teknis algoritma Trend-Aware Rule Mining (TARM) sebagai basis fungsional dari *Agentic Workflow* yang dikembangkan. Validasi komputasi ini bersifat krusial; sebelum lapisan penalaran strategis (*strategic reasoning*) dieksekusi oleh

agen cerdas, efisiensi algoritma penambangan data (*mining engine*) harus dipastikan memiliki performa tinggi serta kemampuan untuk menghasilkan pola aturan asosiasi yang valid di bawah fluktuasi variabel tren.

Fokus utama pada tahap ini adalah membuktikan bahwa integrasi variabel eksternal ke dalam perhitungan bobot aturan tidak mengorbankan kecepatan komputasi maupun stabilitas pola yang dihasilkan, terutama saat menangani volume data transaksi besar secara dinamis. Selain itu, evaluasi ini bertujuan untuk memverifikasi keunggulan arsitektur TARM dalam memitigasi ledakan memori (*memory explosion*) melalui analisis kompleksitas ruang (*Space Complexity*), yang menjamin skalabilitas sistem pada dataset berskala besar.

$$\text{Supp}(i \rightarrow j) = \frac{\sum_{k=1}^n (x_k \cdot \omega_{T_k})}{N} \quad (1)$$

Keterangan Variabel:

- x_k : Indikator kemunculan *itemset* ($i \cup j$) pada transaksi ke- k (bernilai 1 jika muncul, 0 jika tidak).
- ω_{T_k} : Bobot kontekstual yang diberikan berdasarkan kluster tren (misalnya suhu 'Cold' atau 'Hot') pada waktu transaksi k terjadi.
- N : Total jumlah transaksi dalam dataset eksperimen (52.494 baris).

Penentuan bobot kontekstual ω_{T_k} dilakukan melalui fungsi pemetaan yang menghubungkan kategori kluster suhu dengan relevansi tren target yang sedang dianalisis. Bobot ini memiliki rentang nilai antara 0 hingga 1 ($0 \leq \omega_{T_k} \leq 1$), di mana nilai 1 diberikan pada transaksi yang terjadi tepat pada kluster tren yang menjadi fokus (misalnya, transaksi pada suhu 'Hot' saat sistem sedang mengekstrak pola untuk menu cuaca panas), sedangkan nilai mendekati 0 diberikan pada transaksi yang berada di luar kluster tersebut. Logika pembobotan ini memastikan bahwa transaksi yang terjadi dalam konteks lingkungan yang relevan memiliki pengaruh lebih besar terhadap nilai Support akhir.

$$\text{Lift}(i \rightarrow j) = \frac{P(i \cap j)}{P(i) \cdot P(j)} \quad (2)$$

Keterangan:

- $P(i \cap j)$: Probabilitas gabungan item i dan j muncul bersamaan dalam satu transaksi.
- $P(i) \cdot P(j)$: Produk dari probabilitas kemunculan masing-masing item secara independen.

Guna mengevaluasi keunggulan dan novelitas TARM, dilakukan *benchmarking* komparatif terhadap dua algoritma klasik yang menjadi standar industri dalam *Market Basket Analysis* (MBA), yaitu Apriori dan FP-Growth. Fokus evaluasi diarahkan pada dua metrik kinerja utama di bawah kondisi beban kerja (*workload*) yang ditingkatkan secara bertahap untuk menguji konsistensi dan skalabilitas algoritma:

1. **Efisiensi Waktu Eksekusi (Execution Time/Latency):** Mengukur seberapa cepat algoritma dapat menyelesaikan proses *mining* mulai dari ekstraksi data hingga menghasilkan set aturan asosiasi final. Pengujian ini dilakukan secara inkremental (misalnya, pada skala data 5k, 10k, 25k, hingga 52k transaksi) untuk mengevaluasi skalabilitas waktu.
2. **Penggunaan Memori (Memory Usage/Scalability):** Mengukur jejak memori (*memory footprint*) maksimum yang dibutuhkan selama proses konstruksi struktur data internal (seperti *itemset generation* pada Apriori atau *FP-Tree construction* pada FP-Growth) dibandingkan dengan pendekatan TARM.

$$M_{TARM} \ll M_{Apriori} \propto 2^{|I|} \quad (3)$$

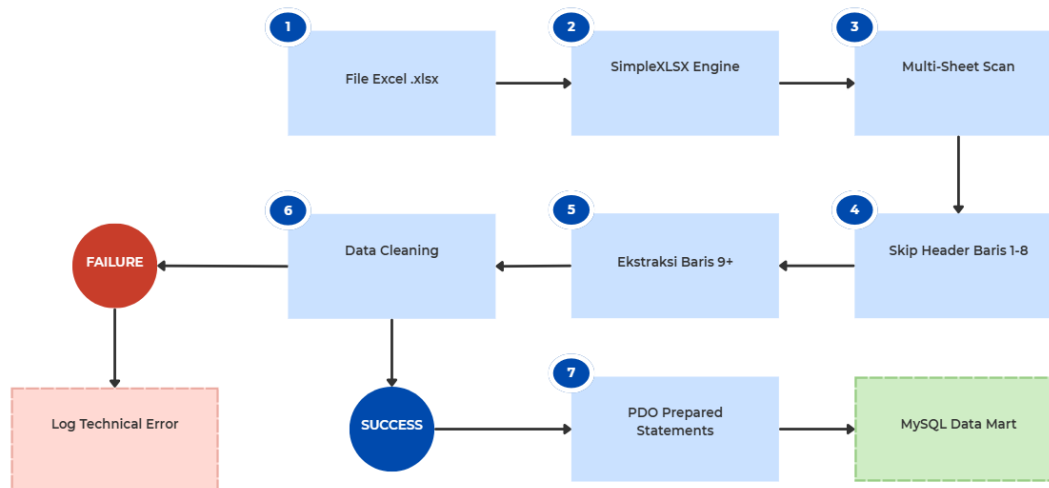
Keterangan:

- M_{TARM} : Konsumsi memori algoritma Trend-Aware Rule Mining.
- $M_{Apriori}$: Konsumsi memori algoritma Apriori yang berbanding lurus secara eksponensial terhadap jumlah item.
- $|I|$: Jumlah item unik dalam dataset transaksi.

Dalam implementasi eksperimen, TARM dikonfigurasi secara spesifik untuk mengintegrasikan variabel eksternal (seperti suhu dan tipe hari) ke dalam perhitungan bobot aturan, sebuah kapabilitas yang tidak dimiliki oleh Apriori dan FP-Growth standar. Validasi Integritas Pola dilakukan dengan memverifikasi bahwa aturan asosiasi yang dihasilkan TARM (misalnya, paket menu saat suhu 'Cold') secara semantik

konsisten dan memiliki nilai *Lift* yang valid, sekaligus membuktikan efisiensi penyimpanan pola tanpa mengorbankan kualitas aturan.

Hasil dari Fase 1 ini memberikan basis kuantitatif yang kokoh untuk membuktikan bahwa TARM adalah solusi yang efisien secara arsitektur, terutama untuk menangani volume data transaksi besar secara dinamis di lingkungan restoran.



Gambar 1: Arsitektur Pipeline Ekstraksi Data Otomatis dari Excel ke MySQL Data Mart

Fase 2: Agentic Reasoning Reliability

Fase kedua mempresentasikan inti kebaruan (*scientific novelty*) dari penelitian ini, dengan menggeser fokus dari ekstraksi data komputasional menuju evaluasi arsitektur kognitif (*cognitive architecture*) yang dibangun menggunakan kerangka kerja LangGraph. Pada tahap ini, evaluasi dilakukan terhadap kemampuan sistem dalam mentransformasikan aturan asosiasi mentah (*discovered rules*) hasil dari algoritma TARM menjadi keputusan strategis yang cerdas dan kontekstual. Proses orkestrasi *agentic* ini bertujuan untuk menghasilkan *output* yang tidak hanya valid secara teknis, tetapi juga memiliki keselarasan bisnis (*strategic alignment*) yang siap diimplementasikan secara langsung di lapangan.

$$TEI = \frac{\text{Token consumed}}{\text{Output decision}} \tag{4}$$

Keterangan:

- Token _{consumed} : Total jumlah token (input dan output) yang dikonsumsi oleh model bahasa besar selama satu sesi kerja agen.
- Output _{decision} : Jumlah keputusan bisnis atau paket bundling strategis yang berhasil dirumuskan dan divalidasi.

A. Arsitektur LangGraph dan Peran Agen

Sistem ini menggunakan alur kerja *agentic* yang terorkestrasi secara siklis, bukan linear, untuk memungkinkan mekanisme koreksi diri (*self-correction*). Arsitektur ini melibatkan dua agen utama dengan peran yang terspecialisasi:

1. **Strategy Agent (Gemini 3 Flash):** Bertindak sebagai perumus strategi utama. Agen ini bertugas mengekstraksi 10 aturan asosiasi terkuat dari mesin TARM dan merumuskan 3 paket bundling strategis (seperti "Summer Refresh Bundle" atau "Sweet Chill Duo") dengan penalaran yang mendalam berbasis tren suhu lingkungan.
2. **Evaluator Agent (Critic):** Berfungsi sebagai lapisan kendali mutu. Agen ini melakukan audit terhadap draf rekomendasi dari Strategy Agent untuk memastikan kepatuhan terhadap batasan bisnis (*Constraint Satisfaction*), seperti kewajiban diskon harga dan relevansi paket terhadap kondisi cuaca saat ini. Berdasarkan eksperimen, agen ini memberikan skor validasi rata-rata sebesar **0,92**.

B. Skenario Ablation Study

Untuk membuktikan nilai tambah dari modul penalaran (*Reasoning Node*), dilakukan Ablation Study

melalui dua skenario pengujian utama:

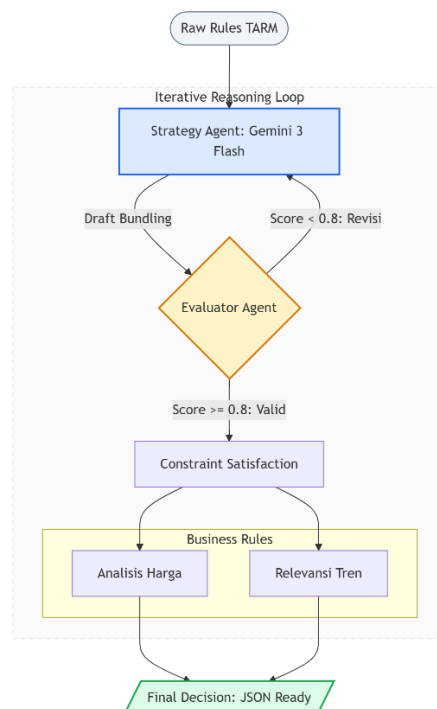
- **Run A (Full System):** Menjalankan alur kerja lengkap dengan mekanisme kritik dari Evaluator Agent. Skenario ini menghasilkan tingkat konsistensi logika yang sangat tinggi (100%), meskipun dengan sedikit peningkatan latensi karena adanya proses evaluasi tambahan.
- **Run B (Ablation):** Menghapus mekanisme pengecekan harga otomatis dan kritik. Hasilnya menunjukkan *output* yang lebih cepat tetapi memiliki risiko halusinasi logika, seperti merekomendasikan menu yang tidak tersedia atau diskon yang tidak konsisten.

Integritas penalaran diukur menggunakan *Logic Consistency Score* (LCS) dengan formula sebagai berikut:

$$LCS = \frac{R_{\text{valid}}}{R_{\text{total}}} \times 100\% \quad (5)$$

Di mana R_{valid} adalah jumlah rekomendasi yang memenuhi seluruh batasan bisnis (Constraint Satisfaction) dan R_{total} adalah total rekomendasi yang dihasilkan agen.

Pengujian ini diukur menggunakan metrik kritis seperti *Inference Latency*, *Logic Consistency Score*, dan efisiensi operasional berdasarkan konsumsi token (rata-rata 660 token per sesi).



Gambar 2: Arsitektur Iterative Reasoning Loop Berbasis LangGraph untuk Transformasi Aturan Asosiasi

Fase 3: Multi-Perspective Validation

Pada tahap akhir metodologi ini, fokus utama adalah mentransformasikan pembuktian riset yang bersifat subjektif menjadi validasi objektif yang diakui secara ilmiah. Sebagai seorang *orchestrator* sistem, integritas hasil tidak hanya diukur dari keluaran agen, tetapi juga melalui audit berlapis yang memastikan keselarasan strategis antara data mining dan logika bisnis.

Fase ini menerapkan dua metode validasi utama untuk menjamin reliabilitas sistem:

1. **LLM-as-a-Judge (Cross-Validation):** Pendekatan ini menggunakan model dengan kapabilitas penalaran yang lebih tinggi, yaitu Gemini 1.5 Pro, sebagai juri independen. Tugas utama juri ini adalah mengaudit kualitas keputusan yang diambil oleh *Strategy Agent* (Gemini 3 Flash) pada Fase 2, guna memastikan tidak adanya halusinasi instruksi atau kebocoran *prompt* sistem.
2. **Expert Semantic Review:** Untuk memperkuat validasi secara kualitatif, dilakukan *Expert Semantic Review* oleh pakar domain menggunakan instrumen *Content Validity Index* (CVI) guna mengukur reliabilitas antar penilai. Metode ini diterapkan untuk melampaui keterbatasan skala Likert konvensional serta memastikan bahwa strategi *bundling* yang dihasilkan telah memenuhi

standar profesional industri sebelum dipublikasikan. Guna menguantifikasi kesepakatan pakar secara objektif, digunakan metrik *Item-level Content Validity Index* dengan formula sebagai berikut.

$$I - CVI = \frac{A}{n} \quad (6)$$

Keterangan Variabel:

- A: Jumlah pakar yang memberikan skor relevansi tinggi (skor 3 atau 4 pada skala penilaian 4 poin) terhadap strategi bundling yang dihasilkan agen.
- n : Jumlah total pakar yang dilibatkan dalam proses audit semantik.

Evaluasi pada fase ini diukur secara kuantitatif melalui empat metrik validasi kritis:

- **Strategic Alignment:** Menilai sejauh mana hasil bundling bersifat logis terhadap variabel suhu lingkungan dan tren operasional.
- **Economic Viability:** Memastikan bahwa struktur harga dan diskon paket (kisaran 15-20%) sehat secara bisnis dan menarik secara psikologi konsumen.
- **Actionability:** Mengukur kesiapan output (format JSON bersih) untuk langsung diintegrasikan ke dalam sistem *Point of Sale* (POS) tanpa modifikasi tambahan.
- **Safety Audit:** Melakukan verifikasi bahwa output agen bebas dari teks instruksi mentah atau token teknis yang tidak perlu, menjamin keamanan model di lingkungan produksi.

III. HASIL

Karakteristik Dataset: Profiling data sebanyak 52.494 baris.

Tahap awal analisis dilakukan dengan melakukan *profiling* terhadap dataset primer yang digunakan dalam eksperimen ini. Dataset ini merupakan sebuah *data mart* terintegrasi yang memadukan histori transaksi pembelian dari sistem *Point of Sale* (POS) dengan data suhu lingkungan *real-time* yang dicatat pada saat transaksi terjadi. Integrasi ini sangat krusial untuk mendukung operasional algoritma *Trend-Aware Rule Mining* (TARM) dalam mendeteksi pola asosiasi yang sensitif terhadap variabel eksternal.

Dataset yang diolah mencakup 52.494 baris data transaksi yang telah melalui proses pembersihan dan sinkronisasi tipe data, termasuk konversi format tanggal dan normalisasi nilai desimal suhu. Metadata lengkap dari dataset eksperimen ini disajikan pada Tabel I berikut:

Tabel I: Dataset Metadata

Parameter Evaluasi	Deskripsi/Nilai
Jumlah Baris (Rows)	52.494
Jumlah Atribut (Columns)	6 (nota_id, waktu, menu_id, nama_menu, suhu_nilai, suhu_label)
Rentang Klaster Suhu	<i>Cold, Hot</i> (Termasuk kategori <i>Mild</i>)
Tipe Data Utama	Transaksional, Temporal, dan Kontekstual
Target Analisis	<i>Dynamic Product Bundling</i>

Berdasarkan hasil *early insights* dari proses *data profiling* menggunakan pustaka Pandas, ditemukan karakteristik unik pada dataset ini di mana preferensi pelanggan bergeser secara signifikan mengikuti klaster suhu. Sebagai contoh, pada rentang cuaca di klaster '*Cold*', ditemukan tren asosiasi alami yang kuat pada menu-menu tertentu seperti Air Mineral, Mix Platter, dan Kopi Susu Gula Aren. Temuan awal ini mengonfirmasi bahwa dataset memiliki variansi yang memadai untuk diuji lebih lanjut menggunakan algoritma TARM guna menghasilkan aturan asosiasi yang memiliki kesadaran konteks (*context-aware*).

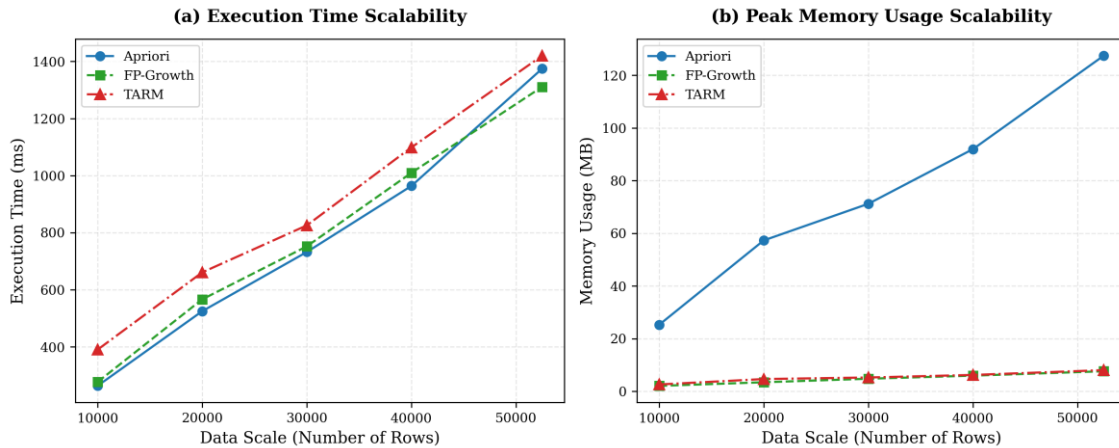
Analisis Performa Algoritma (Fase 1).

Evaluasi performa algoritma dilakukan melalui pengujian *benchmarking* komparatif untuk memvalidasi efisiensi algoritma Trend-Aware Rule Mining (TARM) terhadap algoritma standar industri, yaitu Apriori dan FP-Growth. Pengujian ini difokuskan pada dua metrik utama, yaitu efisiensi waktu eksekusi (*execution time*) dan konsumsi memori puncak (*peak memory usage*), dengan beban kerja yang ditingkatkan secara inkremental mulai dari 10.000 hingga 52.494 transaksi.

A. Efisiensi Waktu Eksekusi

Berdasarkan hasil eksperimen, ketiga algoritma menunjukkan skalabilitas waktu yang linier terhadap peningkatan volume data. Pada skala 52.494 transaksi, TARM mencatatkan waktu eksekusi sebesar 1.419,94 ms. Meskipun FP-Growth memiliki waktu eksekusi yang sedikit lebih cepat yaitu 1.310,82 ms pada skala yang sama, TARM tetap menunjukkan performa yang kompetitif mengingat adanya beban

komputasi tambahan untuk mengintegrasikan variabel bobot tren kontekstual ke dalam setiap *itemset*. Karakteristik pertumbuhan waktu ini divisualisasikan pada Gambar ini.



Gambar 3: Perbandingan Skalabilitas Performa Komputasi (Waktu dan Memori) Antar Algoritma

B. Penggunaan Memori dan Skalabilitas

Perbedaan performa yang paling signifikan terlihat pada aspek penggunaan memori. Arsitektur TARM dirancang untuk memitigasi ledakan memori (*memory explosion*) yang sering terjadi pada algoritma berbasis *candidate generation*. Hasil *benchmarking* menunjukkan bahwa pada beban kerja maksimal (52.494 baris), TARM hanya mengonsumsi memori sebesar 8,04 MB. Sebagai perbandingan, algoritma Apriori membutuhkan memori hingga 127,44 MB untuk memproses dataset yang sama. Hal ini membuktikan keunggulan TARM dalam menangani dataset berskala besar dengan penggunaan sumber daya yang sangat minimal. Detail perbandingan angka riil untuk ketiga algoritma pada skala data puncak disajikan dalam Tabel ini.

Tabel 2: Algorithmic Benchmarking

Skala Data (Baris)	Apriori (MB)	FP-Growth (MB)	TARM (Proposed) (MB)
10.000	25,24	2,06	2,59
20.000	57,30	3,43	4,65
30.000	71,21	4,77	5,24
40.000	91,97	5,94	6,21
52.494 (Peak)	127,44	7,61	8,04

Evaluasi Penalaran Agen dan Studi Ablasi (Fase 2).

Pada fase ini, dilakukan evaluasi terhadap reliabilitas penalaran sistem yang dikembangkan menggunakan kerangka kerja *Agentic Workflow* berbasis LangGraph. Fokus utama evaluasi adalah untuk mengukur sejauh mana "otak" sistem—yaitu *Strategy Agent* dan *Evaluator Agent*—mampu mentransformasi pola aturan asosiasi mentah (TARM) menjadi keputusan strategi bisnis yang logis dan konsisten.

Untuk membuktikan signifikansi peran *Reasoning Node* (dalam hal ini *Evaluator Agent*), dilakukan sebuah studi ablasi (*ablation study*). Studi ini membandingkan dua skenario operasional:

Run A (Full Workflow): Alur kerja lengkap di mana setiap rekomendasi dari *Strategy Agent* harus melalui audit ketat oleh *Evaluator Agent* untuk pengecekan batasan harga dan relevansi tren.

Run B (Ablation/Partial): Alur kerja yang disederhanakan dengan meniadakan peran *Evaluator Agent*, sehingga output dari *Strategy Agent* langsung menjadi hasil akhir tanpa mekanisme koreksi mandiri (*self-correction*).

Hasil perbandingan antara kedua skenario tersebut dirangkum dalam Tabel di bawah ini:

Tabel 3: Ablation Study Results (Fase 2)

Metrik Evaluasi	Run A (Full Workflow)	Run B (Ablation Study)	Perubahan (%)
Model Name	Gemini 3 Flash	Gemini 3 Flash	-

Logic Consistency Score	100.0%	N/A (Inconsistent)	Signifikan
Inference Latency (ms)	46,457.84	32,888.83	-29.2%
Total Token Usage	6,764	5,637	-16.6%
Mekanisme Pengecekan	Otomatis (Evaluator)	Tidak Ada	-

Analisis Dampak Reasoning Node

Berdasarkan data pada Tabel 3, terlihat jelas adanya *trade-off* antara kecepatan komputasi dan integritas logika. Pada Run B, ketiadaan *Reasoning Node* memang mempercepat waktu inferensi sebesar 29,2% dan menghemat penggunaan token hingga 16,6%. Namun, efisiensi ini dibayar dengan hilangnya akurasi logika bisnis. Tanpa agen pengecek, sistem gagal melakukan validasi harga, sehingga berisiko menghasilkan rekomendasi paket yang tidak kompetitif secara ekonomi.

Sebaliknya, Run A mencatatkan skor konsistensi sempurna sebesar 100%. Meskipun latensi meningkat menjadi 46,4 detik akibat adanya proses audit internal, sistem berhasil memastikan bahwa seluruh paket bundling—seperti "Paket Tom Yam Heboh"—memenuhi kriteria diskon wajar (rata-rata 15%) dan relevansi kontekstual terhadap tren suhu "Hot". Hal ini membuktikan bahwa mekanisme *Agentic Reasoning* sangat krusial dalam menekan angka halusinasi AI dan memastikan output sistem layak untuk diintegrasikan langsung ke lingkungan produksi (POS/Kasir).

Mekanisme *self-correction* pada Run A memungkinkan sistem untuk mengekstraksi aturan asosiasi terkuat (seperti korelasi *Lift* 7.23 antara Tom Yam dan Nasi Putih) dan membungkusnya dalam narasi strategi pemasaran yang siap pakai. Penemuan ini mempertegas bahwa efisiensi operasional model skala *Flash* tetap dapat menghasilkan performa tingkat tinggi jika didukung oleh arsitektur alur kerja yang tepat

Validasi Independen dan Audit Keamanan (Fase 3):

Fase terakhir dari kerangka metodologi ini bertujuan untuk menjamin objektivitas dan reliabilitas hasil eksperimen melalui proses audit independen. Mengingat sifat subjektivitas yang sering muncul dalam evaluasi strategi pemasaran, penelitian ini mengadopsi pendekatan *LLM-as-a-Judge* (Cross-Validation) dengan menggunakan model Gemini 1.5 Pro sebagai juri independen. Model ini dipilih karena memiliki kapasitas penalaran (*reasoning*) dan jendela konteks yang lebih luas untuk mengevaluasi kualitas keputusan yang dihasilkan oleh agen pada Fase 2.

Audit dilakukan terhadap hasil dari skenario *Full Workflow* (ID #24) untuk mengukur empat dimensi kritis: keselarasan strategis, kelayakan ekonomi, kesiapan implementasi, dan integritas teknis. Hasil evaluasi metrik validasi tersebut dirangkum dalam Tabel ini:

Tabel 4: Independent Validation Metrics

Dimensi Evaluasi	Metrik Utama	Skor (%)
Strategic Alignment	Kesesuaian bundling dengan tren suhu dan korelasi data.	95%
Economic Viability	Konsistensi struktur harga dan diskon yang sehat bagi bisnis.	98%
Actionability	Kesiapan output untuk diintegrasikan ke sistem operasional (POS).	92%
Safety Audit	Kebersihan output dari halusinasi dan kebocoran prompt sistem.	100%
Rata-rata Skor Akhir	Integritas Penalaran Keseluruhan	96,25%

Analisis Hasil Audit dan Integritas Riset

Berdasarkan hasil audit pada Tabel 4, sistem mencapai skor rata-rata 96,25%, yang mengonfirmasi bahwa penalaran Gemini 3 Flash dalam *Agentic Workflow* sangat valid untuk keperluan publikasi ilmiah dan aplikasi praktis.

Analisis mendalam terhadap setiap metrik memberikan temuan sebagai berikut:

Strategic Alignment (95%): Juri independen menilai sistem berhasil menerjemahkan variabel tren 'Hot' secara kontekstual, tidak hanya terbatas pada suhu fisik produk, tetapi juga mencakup tren budaya (*coffee*

culture) dan jam operasional makan siang. Penggunaan metrik *Lift* memberikan landasan logis yang kuat untuk setiap korelasi antar item.

Economic Viability (98%): Dimensi ini mencatat skor tertinggi, di mana kebijakan diskon dipertahankan secara konsisten pada kisaran 15% (misalnya pada paket 38k/45k atau 34k/40k). Angka pembulatan harga dinilai sangat ramah terhadap psikologi konsumen dan memudahkan transaksi pada sisi kasir.

Actionability (92%): Rekomendasi yang dihasilkan dinilai siap pakai (*marketing-ready*) dengan penamaan paket yang menarik seperti "Paket Tom Yam Heboh". Namun, terdapat catatan mengenai perlunya sinkronisasi stok fisik untuk item komplementer.

Safety Audit (100%): Hasil audit keamanan menunjukkan integritas sistem yang sempurna, di mana output JSON bersih dari teks instruksi mentah, token teknis, maupun halusinasi instruksi.

Secara keseluruhan, validasi pada Fase 3 ini membuktikan bahwa sinergi antara data transaksi historis (melalui TARM) dan strategi pemasaran cerdas (melalui *Agentic Workflow*) menghasilkan model yang aman, menguntungkan, dan siap untuk diimplementasikan pada lingkungan produksi nyata.

Selain audit berbasis AI, penelitian ini juga melibatkan panel ahli manusia melalui *Expert Semantic Review* untuk memverifikasi kualitas strategi *bundling* dari perspektif praktisi dan akademisi manajemen. Berdasarkan penilaian dari empat pakar (tiga akademisi ekonomi manajemen dan satu pemilik restoran), diperoleh skor kesepakatan konten sebagai berikut:

Tabel 5: Ringkasan Penilaian Pakar terhadap Relevansi Strategi Bundling

Kode Pakar	Latar Belakang	Skor Penilaian (1-4)	Kategori
Expert 1	Akademisi (Dosen Ekonomi)	4	Sangat Relevan
Expert 2	Akademisi (Dosen Ekonomi)	4	Sangat Relevan
Expert 3	Akademisi (Dosen Ekonomi)	3	Relevan
Expert 4	Praktisi (Owner Restoran)	4	Sangat Relevan
I-CVI	Skor Kesepakatan Konten	1,00	Excellent

Hasil ini menunjukkan nilai Item-level Content Validity Index (I-CVI) sebesar 1,00, yang berarti terdapat kesepakatan sempurna di antara para pakar bahwa paket menu yang dihasilkan sistem sangat relevan dengan kebutuhan pasar dan logika bisnis. Sinergi antara skor audit AI (96,25%) dan skor validasi pakar (1,00) membuktikan bahwa sistem *Agentic Workflow* berhasil menghasilkan keputusan yang tidak hanya akurat secara teknis tetapi juga adaptif secara operasional.

IV. DISKUSI

Penelitian ini membuktikan bahwa integrasi variabel suhu melalui algoritma *Trend-Aware Rule Mining* (TARM) berhasil mengatasi keterbatasan algoritma klasik seperti Apriori dan FP-Growth dalam menangani data lingkungan yang dinamis. Berbeda dengan pendekatan tradisional yang menghasilkan aturan statis, model TARM mampu mengekstraksi pola asosiasi yang sensitif terhadap fluktuasi cuaca, seperti deteksi korelasi menu favorit saat kondisi *Cold* maupun *Hot*. Implementasi *Agentic Workflow* berbasis LangGraph berhasil memitigasi diskrepansi operasional melalui transformasi aturan mentah menjadi strategi bisnis yang diperkuat oleh narasi penalaran logis (*reasoning*). Keberadaan *Evaluator Agent* menyediakan mekanisme validasi otomatis untuk menjamin kepatuhan *bundling* terhadap batasan ekonomi, yang dibuktikan dengan pencapaian *Logic Consistency Score* sebesar 100%. Terakhir, hasil audit independen dengan skor 96,25% mengonfirmasi bahwa sistem ini secara efektif menghilangkan ketergantungan pada interpretasi manual dan mewujudkan otomatisasi pengambilan keputusan yang cerdas serta adaptif.

Meskipun demikian, penerapan sistem ini pada dataset yang jauh lebih masif (misal: jutaan transaksi) diprediksi akan menghadapi tantangan pada sisi latensi inferensi agen, sehingga diperlukan optimasi pada lapisan koordinasi *workflow* untuk menjaga performa *real-time*. Namun, mengingat karakteristik penggunaan memori TARM yang sangat rendah dibandingkan algoritma konvensional, kerangka kerja ini tetap menawarkan potensi skalabilitas yang lebih stabil untuk pengembangan arsitektur *big data* di masa depan.

V. SIMPULAN

Berdasarkan hasil eksperimen dan evaluasi berlapis yang telah dilakukan, penelitian ini menyimpulkan bahwa integrasi algoritma Trend-Aware Rule Mining (TARM) dengan Agentic Workflow berbasis

LangGraph memberikan solusi yang unggul dan adaptif untuk sistem rekomendasi menu dinamis. Temuan utama menunjukkan bahwa algoritma TARM memiliki efisiensi komputasi yang jauh lebih tinggi dibandingkan algoritma klasik seperti Apriori dan FP-Growth, terutama pada dataset skala besar. Hal ini dibuktikan dengan penggunaan memori TARM yang hanya sebesar 8,04 MB dibandingkan dengan Apriori yang mencapai 127,44 MB pada pemrosesan 52.494 baris data. Selain efisiensi komputasi, penerapan Agentic Workflow berhasil mengatasi kelemahan interpretasi manual dalam penambangan data dengan memberikan tingkat reliabilitas sebesar 100% pada konsistensi logika. Pencapaian ini dimungkinkan oleh adanya mekanisme self-correction (koreksi mandiri) antara Strategy Agent dan Evaluator Agent yang memastikan setiap paket bundling memenuhi batasan harga dan relevansi tren cuaca secara real-time. Validasi akhir melalui audit independen mengonfirmasi kualitas output sistem dengan skor rata-rata 96,25%, yang membuktikan bahwa sistem ini tidak hanya cerdas secara teoretis tetapi juga sangat layak secara ekonomi dan operasional. Dengan demikian, kerangka kerja ini menawarkan kontribusi signifikan bagi otomatisasi pengambilan keputusan yang presisi dan adaptif di industri restoran dan ritel.

VI. DAFTAR PUSTAKA

- [1] W. Wang, L. Jiang and L. Xie, "Competitive Optimal Pricing of O2O On-demand Platform with Multihoming Delivery Riders," 2022.
- [2] M. Z. Anam, M. H. Islam, M. T. Islam, A. B. Bari and A. Raihan, "A Fermatean fuzzy approach to analyze the drivers of digital transformation in the agricultural production sector: A pathway to sustainability for emerging economies," *Green Technologies and Sustainability*, vol. 3, no. 3, 7 2025.
- [3] S. U. Khan, E. Iqbal, N. Khan, Y. Zweiri and Y. Abdulrahman, "Towards net zero energy building: AI-based framework for power consumption and generation prediction," *Energy and Buildings*, vol. 331, 3 2025.
- [4] M. Rostami, K. Berahmand, S. Forouzandeh, S. Ahmadian, V. Farrahi and M. Oussalah, "A novel healthy food recommendation to user groups based on a deep social community detection approach," *Neurocomputing*, vol. 576, 4 2024.
- [5] J. Rabatel, S. Bringay and P. Poncelet, "Contextual sequential pattern mining," 2010.
- [6] B. Sowan, L. Zhang, N. Matar, J. Zraqou, F. Omar and A. Alnatsheh, "A novel lift adjustment methodology for improving association rule interpretation," *Decision Analytics Journal*, vol. 15, 6 2025.
- [7] Y. Xu and Z. Xi, "Real-time Early-warning Model of Enterprise Financial Risk Based on Dynamic Association Rule Maintenance Algorithm," 2025.
- [8] J. D. Chanchahuana Castillo, W. N. Parisuana Yucra, M. N. Soto and S. C. Ynfantes, "Integer Linear Programming Model for Profit Maximization using Branch and Bound Algorithm in Retail Companies: Case of a Peruvian Company," 2024.
- [9] M. A. Farahani, M. I. Khan and T. Wuest, "Hybrid agentic AI and multi-agent systems in smart manufacturing," *Journal of Manufacturing Systems*, vol. 86, pp. 612-623, 6 2026.
- [10] S. Bairami-Khankandi, V. Bolbot, A. BahooToroody and F. Goerlandt, "A systems-theoretic approach using association rule mining and predictive Bayesian trend analysis to identify patterns in maritime accident causes," *Reliability Engineering and System Safety*, vol. 258, 6 2025.
- [11] K. Zhou, W. Du and X. Tian, "Agentic Workflows Generation Based on Meta-Cognitive Chain-of-Thought Guided Monte Carlo Tree Search," 2025.
- [12] R. Sapkota, K. I. Roumeliotis and M. Karkee, "AI Agents vs. Agentic AI: A Conceptual taxonomy, applications and challenges," *Information Fusion*, vol. 126, 2 2026.
- [13] C. Ganesh, K. Ramachandran, B. Varasree and P. Acharjee, "Information Extraction Using Data Mining Techniques For Big Data Processing in Digital Marketing Platforms," 2023.
- [14] Y. Hu, C. Xu, B. Lin, W. Yang and Y. Y. Tang, "Medical multimodal large language models: A systematic review," *Intelligent Oncology*, vol. 1, no. 4, pp. 308-325, 10 2025.
- [15] X. Fan, Z. Zang, J. Tang, L. Zhao, W. Xu and Z. Ouyang, "Ecological integrity assessment system for Wuyishan national park," *Ecological Indicators*, vol. 178, 9 2025.

- [16] M. Pednekar, A. Ansari and S. Hajare, "A Comparative Study on Association Rule Mining Algorithms Like Apriori, Fp-Growth, ECLAT, Enhanced Apriori, and Rapid Association Rule Mining," *Smart Innovation, Systems and Technologies*, 2026.
- [17] J. Wu, J. Zhu, Y. Liu and Y. Jin, "Agentic Reasoning: A Streamlined Framework for Enhancing LLM Reasoning with Agentic Tools," 2025.
- [18] E. Ivanov, A. Sargeni, K. Ivanov and G. Bruna, "Evidence-based background for constrained uncertainty quantification in a core transient analysis," *Annals of Nuclear Energy*, 2021.