



Natural Language Query Hybrid Architecture for Cross-Sectoral Analytics Dashboards on Public Statistical Data

Yusuf Fadlila Rachman*

Universitas Sebelas Maret,
Indonesia

**Puput Surya
Ningtyas**

Universitas Sebelas Maret,
Indonesia

Masbahah

Universitas Sebelas Maret,
Indonesia

Nur Azizul Haqimi

Universitas Sebelas Maret,
Indonesia

Ahmad Faisal Sani

Universitas Sebelas Maret,
Indonesia

***Corresponding author:**

Yusuf Fadlila Rachman,
Universitas Sebelas Maret, Indonesia.

✉ yusuf_fadil@staff.uns.ac.id

Article Info:

Article history:

Received: August 08, 2026

Revised: September 07, 2026

Accepted: October 05, 2026

Available Online: October 08, 2026

Keywords:

Analytics Dashboard;

Apriori;

Natural Language Query;

Public Statistical Data;

Text-to-SQL



This is an open access article under the
[CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.

Copyright © 2026 by Author. Published by
Politeknik Siber Cerdika International

Abstract

Background: Public statistical data are essential for evidence-based policymaking, yet the complexity of cross-sectoral indicators often limits accessibility for nontechnical users. This challenge is particularly relevant in East Java, where regional development disparities require integrated analytical tools.

Objective: This study aims to develop a cross-sectoral analytics dashboard integrating Natural Language Query (NLQ), Apriori association-rule mining, and a supporting local large language model (LLM).

Methods: The system was developed using the SDLC Waterfall model and utilized more than 50 Statistics Indonesia (BPS) indicators covering 38 districts/cities in East Java from 2016 to 2026. A deterministic, rule- and regular-expression (regex)-based Text-to-SQL mechanism processed natural-language queries, while Apriori identified cross-sectoral association patterns. A local LLM (Ollama, llama3.2:1b) generated supporting narratives. Evaluation included black-box testing, 50-query intent classification, association-rule metrics, and an exploratory LLM assessment.

Results: The NLQ module achieved 96.00% accuracy (48/50 queries). Apriori generated 55,480 high-level rules with an average confidence of 0.909 and lift of 15.44. The exploratory LLM evaluation achieved a composite score of 4.35/5.00, a 2.5% hallucination rate, a 26.07-second P95 latency, and a 100% fallback success rate.

Conclusion: The hybrid architecture effectively supports accessible statistical-data retrieval and cross-sectoral pattern discovery, while the LLM remains a supporting narrative component that requires further formal validation.

To cite this article: Rachman, Y. F., Suryaningtyas, P., Masbahah, Haqimi, N. A., & Sani, A. F. (2026). Natural language query hybrid architecture for cross-sectoral analytics dashboards on public statistical data. *Equivalent: Jurnal Ilmiah Sosial Teknik*, 8(4), 1656–1666. <https://doi.org/10.59261/jequi.v8i4.430>

INTRODUCTION

East Java Province experiences a sharp disparity in development across regions, despite being one of the largest economic centers in Indonesia, with 38 regencies/cities and a population of more than 41 million people. Based on BPS data from 2023, the highest Human Development Index (HDI) was recorded in the city of Surabaya (83.99), while the lowest was recorded in Sampang Regency (64.84). This disparity affects the equitable distribution of development across essential sectors such as health, education, infrastructure, and the economy; therefore, effective

policy formulation requires data-driven cross-sectoral analysis.

BPS routinely releases thousands of macroeconomic and socioeconomic indicators, providing a strong basis for data-driven policy governance. However, two implementation barriers remain. First, identifying specific observations in large BPS tables is time-consuming for analysts without SQL skills. Second, cross-sector indicators are often analyzed separately, even though development outcomes are interrelated; without an integrated computational layer, discovering multivariate patterns requires substantial manual effort (Adhitya et al., 2022; Khoirunnisa et al., 2025; Muis et al., 2025).

Previous studies have developed public-data visualization dashboards, but most emphasize descriptive presentation rather than integrated pattern mining (Asparosa, 2026; Putri & Supriyadi, 2024; Sari & Darmanto, 2025). Association-rule studies on social data typically require manual parameter selection, while Text-to-SQL research demonstrates both the value and the difficulty of enabling cross-domain natural-language access to structured data (Solimun et al., 2023; Wang et al., 2020; Wong et al., 2024; Yu et al., 2018; Zhong et al., 2017). The research gap is therefore the absence, in this context, of a single auditable workflow that combines deterministic natural-language retrieval with cross-sector association-rule discovery for non-technical users.

Based on these problems, this study proposes a hybrid architecture integrating deterministic natural-language querying (NLQ), Apriori association-rule mining, and a supporting local LLM-based narrative layer within a single cross-sectoral analytics dashboard. The contributions are: (1) First, the architecture combines deterministic rule-based/regex Text-to-SQL, dynamic Apriori association-rule mining, and local LLM-based narrative generation with graceful degradation, allowing non-technical users to search, mine, and interpret BPS data without writing SQL or manually configuring each analytical step; (2) Second, Apriori is applied to multivariate public statistical data comprising more than 50 BPS indicators across 38 regencies/cities from 2016–2026, following the discretization of numerical indicators into categorical items. Association rules are evaluated using support, confidence, and lift, extending the classic association-rule approach (Agrawal & Srikant, 1994; Fernandez-Basso et al., 2024; Han et al., 2012) to cross-sector public-policy exploration.

Third, the system is evaluated through a 50-query NLP intent-classification test and a metric-based assessment of Apriori outputs, while the LLM layer is treated separately as an exploratory post hoc narrative component that requires further formal validation.

METHOD

The system was developed using the System Development Life Cycle (SDLC) Waterfall model, comprising analysis, design, implementation, integration and testing, and maintenance. The dataset contained more than 50 macro indicators from BPS East Java Province for 2016–2026 across 38 regencies/cities. The data were obtained from the BPS SIMDASI API through an ETL (Extract, Transform, Load) process and stored in PostgreSQL as a cross-sectoral aggregate table (DATA_MERGE_FLOAT).

System Architecture

The system architecture consisted of six components: (1) a Web Dashboard for visualization, NLQ search, and Apriori analysis; (2) a Backend Server that processed requests and coordinated the database and AI services; (3) PostgreSQL; (4) an AI Engine containing deterministic query understanding, Apriori mining, and local LLM-based narrative generation; (5) an ETL Pipeline; and (6) the BPS SIMDASI API as the data source. The workflow proceeded from ETL-based data acquisition and PostgreSQL storage to AI processing and dashboard presentation.

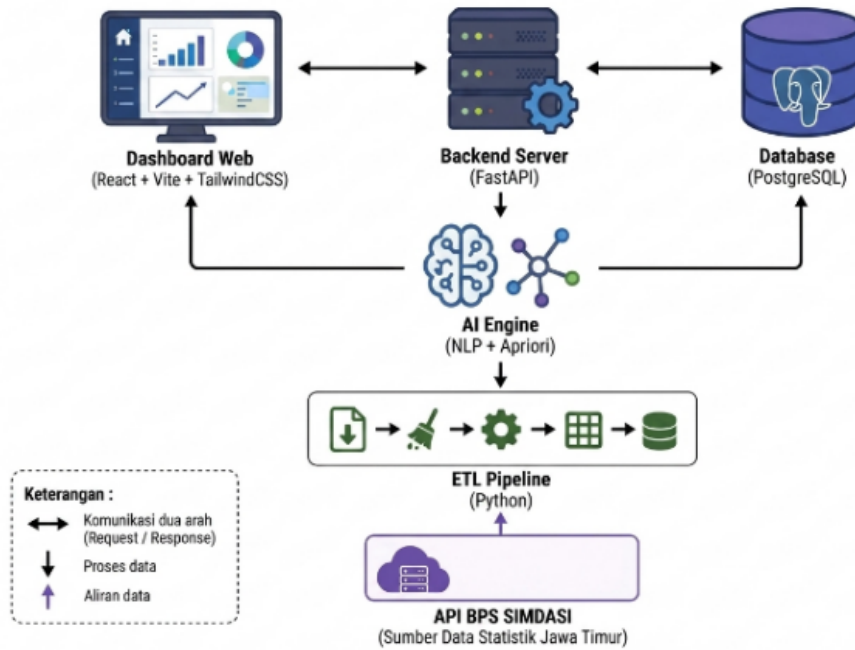


Figure 1. System Architecture

Natural Language Query (Text-to-SQL) Module

The query-understanding module translated natural-language queries into database query commands. The module was designed to operate deterministically and was entirely rule-based and regex-based, without relying on large language models or external services, so that parameter extraction was fast, consistent, and reproducible. The process was carried out in four stages: (1) Named Entity Recognition used a combination of regex patterns for standard formats, such as year ranges, and mapping dictionaries such as WILAYAH_JATIM, DIMENSION_MAP, and METRIC_MAP to capture regions, indicator metrics, and demographic category filters (e.g., gender and age group); (2) Intent Classification used keyword matching against intent dictionaries to detect user objectives, including descriptive, comparative, and analytical categories such as correlation and pattern discovery; (3) query structuring generated structured database queries (Text-to-SQL) based on the detected entities and intents; and (4) response optimization used identical-query caching and text normalization. When a pattern- or correlation-discovery intent was detected, the process was passed to Apriori mining.

Apriori Association Rule Mining Module

Each combination of a discretized indicator and category (e.g., “High stunting” or “Low HDI”) was treated as an item, and each regency/city was represented as a transaction. Frequent itemsets were generated iteratively according to the Apriori property, after which rules of the form $A \rightarrow B$ were evaluated using support, confidence, and lift. The implementation used mlxtend (TransactionEncoder and apriori), with $\text{max_len} = 2$ and a minimum support of 0.03. Additional stored metrics included leverage, conviction, and Zhang’s metric (Agrawal & Srikant, 1994; Fernandez-Basso et al., 2024; Han et al., 2012).

Table 1
Three-level power scheme of association

Level	Min. Support	Min. Confidence	Lift Ratio Range
High	0,03	0,50	$\text{Lift} \geq 1.5$ (strong positive association)
Medium	0,03	0,30	$1.0 \leq \text{Lift} < 1.5$ (moderate positive association)
Low	0,03	0,30	$\text{Lift} < 1.0$ (negative association)

Local LLM-Based Narrative Generation Module

To translate the numerical results into summaries that were easy for policymakers to understand, the system utilized a local Large Language Model (LLM) running through the Ollama service (Ay & Demirdağ, 2026; Ifath & Haque, 2026). In contrast to the query-understanding module, which was completely deterministic, the LLM was used only during the narrative-generation stage. Communication was performed via HTTP POST requests to Ollama (localhost:11434) using the lightweight llama3.2:1b model, which is part of the Llama 3 model family (Grattafiori et al., 2024; Team & Meta, 2024). Prompts were structured according to the principles of instruction tuning so that the model acted as an executive data analyst Lubis et al. (2026) and Zhang et al. (2026) generated outputs in a structured HTML format consisting of three paragraphs that combined data summaries, anomaly and trend highlights, and policy recommendations.

Because LLM output was probabilistic, the integration was equipped with a graceful degradation mechanism. If the Ollama service was unavailable, experienced a timeout (45 seconds), or returned an error, the system caught the exception and automatically generated a substitute HTML narrative containing the basic report and a description indicating that the LLM service was unavailable. Thus, system availability was maintained without relying entirely on the LLM, and all factual numerical values remained sourced from the results of deterministic queries rather than from LLM-generated content, thereby reducing the risk of hallucinations in reported data.

Comparison of Approaches: Rule-Based NLP vs LLM

The division of roles between rule-based NLP and the LLM was an advantage of the system. Extracting query parameters from statistical data required high precision, speed, and repeatability. However, this process could be disrupted by generative hallucinations if it were delegated entirely to an LLM. In contrast, executive narrative generation required language flexibility, which is a key strength of LLMs. Table 2 summarizes the comparison of the two approaches according to their roles in the system.

Table 2

Comparison of rules-based NLP and LLM approaches in systems

Aspects	Rule-Based NLP / Regex	Local LLM (Ollama, llama3.2:1b)
Roles in the system	Query comprehension: intent classification & entity extraction (Text-to-SQL)	Analytical narrative generation for executives
Output properties	Deterministic, consistent for the same input	Probabilistic, may vary between executions
Latency & Compute	Very fast, lightweight, no large model inference	Slower & heavier; 45-second timeout limit
Parameter precision	Height for the default entity/parameter (region, year, metric)	Not used for parameter extraction
Risk of hallucinations	None	There is; mitigated via structured prompts & factual data
Interpretability	Transparent (explicit rules can be audited)	Limited (black-box behavior)
Language flexibility	Limited to patterns & dictionaries	Height; Generate natural flowing text
Service dependencies	Self-contained, without external services	Depends on Ollama's services; Fallback available

With this division of roles, the system gained a dual advantage: the accurate and reproducible query foundation of rule-based NLP and the communicative explanatory capabilities of the LLM, without sacrificing the reliability of factual numerical values or system availability.

Test Method

Functional testing was conducted using Black Box Testing across authentication, NLP search, dashboard exploration, geospatial maps, and Apriori analysis. NLP performance was evaluated using 50 natural-language queries across Descriptive, Comparative, and Pattern Discovery intents. Apriori outputs were checked against the configured support, confidence, and lift thresholds. The LLM narrative layer was assessed using an exploratory post hoc protocol covering functional behavior, rubric-based quality across 30 query samples, hallucination detection against source data, cross-execution consistency, and latency and fallback reliability. The predefined criteria were a composite quality score of $\geq 4.0/5.0$, a hallucination rate of $< 5\%$, and a fallback success rate of 100%. This exploratory LLM assessment was conducted outside the formal final-project test protocol and was therefore reported separately from the validated NLQ and Apriori modules.

RESULTS AND DISCUSSION

Results

System Implementation

The analytics dashboard was successfully implemented as a web-based application comprising several main modules: an NLP-based intelligent search page with a Smart Search Bar; an exploration page presenting indicator summaries, including key performance indicators (KPIs), trends, and rankings; a regional comparison page; a Web GIS geospatial map using a choropleth visualization for 38 regencies and cities; and an a priori data-mining analysis page. The interface was designed following the principles of User-Centered Design to facilitate the comprehension and interpretation of information by policy analysts.

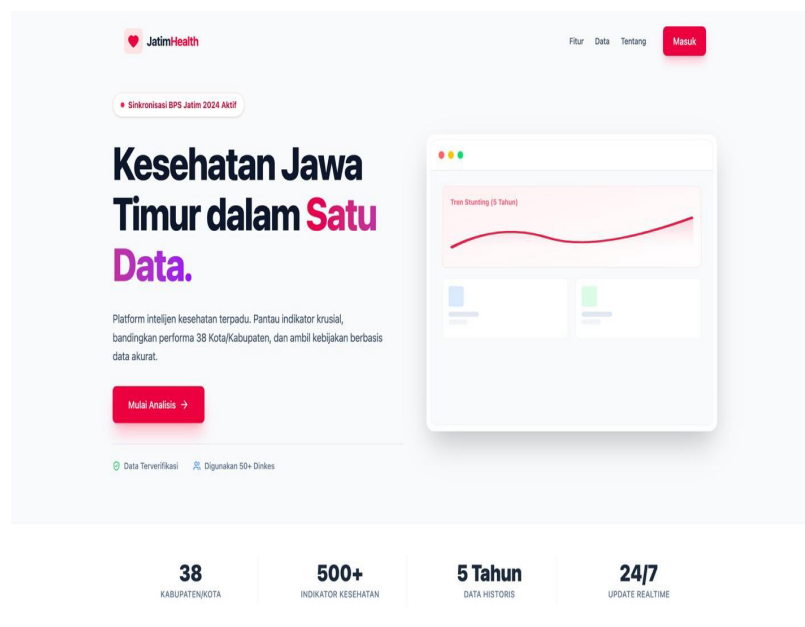


Figure 2. Application Interface

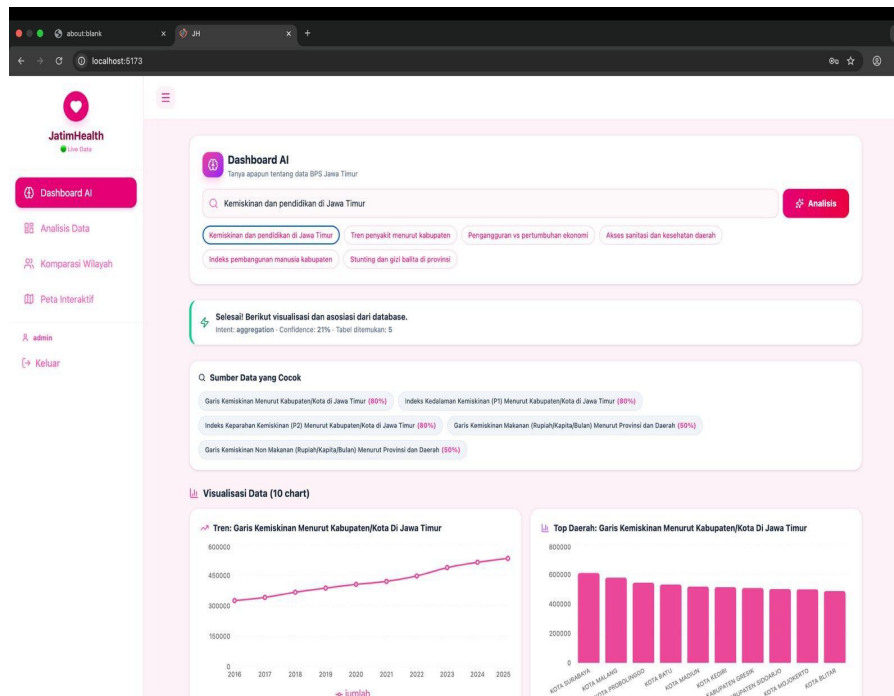


Figure 3. Search Interface

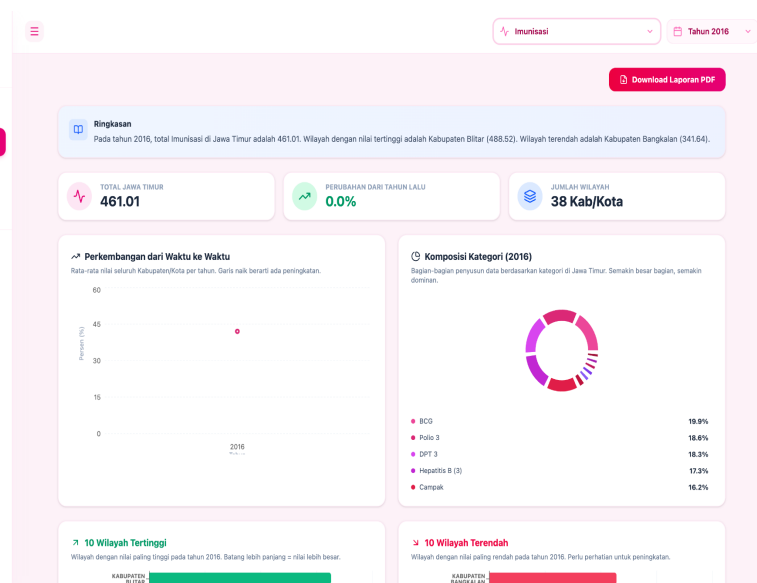


Figure 4. Analytics Interface

Functional Testing (Black Box)

All formal functional test scenarios for the core dashboard and NLQ/Apriori workflow produced the expected results. Authentication, natural-language search, exploration, comparison, and geospatial functions operated as designed. The local LLM narrative module was developed after the formal final-project test session; accordingly, the post-trial quantitative results reported later are exploratory and are not treated as part of the formal validation evidence for the core system.

Evaluation of NLP Intent Classification Accuracy

Automated testing of 50 queries resulted in an overall intent-classification accuracy of 96.00% (48 out of 50 queries classified successfully), as summarized in Table 3.

Table 3

Results of the evaluation of the accuracy of the NLP module intent classification

No.	Category Intent	Total Queries	Succeed / Fail	Accuracy
1	Descriptive (basic data exploration)	20	20 / 0	100,00%
2	Comparative (comparison between regions/time)	15	14 / 1	93,30%
3	Pattern Discovery	15	14 / 1	93,30%
	Overall Total	50	48 / 2	96,00%

The rule-based query-understanding module achieved 100% accuracy in the Descriptive category. Two classification failures occurred for queries with implicit semantics. First, the query, "What is the main cause of the high stunting rate in Sampang Regency?" was expected to be classified as Pattern Discovery but was detected as Descriptive because the phrase "main cause" is abstract and was not assigned sufficient weight in the deterministic rule layer. As a result, the system fell back to standard exploratory retrieval. Second, the query, "Which district has the worst health ranking in East Java?" was intended to be classified as Comparative but was detected as Descriptive because it did not mention at least two specific regions, causing the comparative structural signal to fall below the classification threshold. Both cases indicate that the primary limitation lies in the semantic interpretation of implicitly formulated queries, which represents an important direction for future improvement.

Mining Results of Association Rules a Priori

Apriori execution grouped the cross-sectoral association rules into three levels, as shown in Table 4. At the "High" level, the system extracted 55,480 rules with an average confidence of 0.909 and an average lift of 15.44, indicating a very strong positive association. In contrast, the "Low" level represented negative associations, with an average lift of 0.83, indicating that the occurrence of one indicator is associated with a lower likelihood of the occurrence of another indicator. These results confirm that the threshold parameters configured for the module were functioning as intended.

Table 4

Summary of A-Priori mining results metrics per level

Level	Number of Rules	Average Confidence	Average Lift Ratio
High	55.480	0,909	15.44 (very strong positive association)
Medium	1.084	0,559	1,20
Low	66	0,098	0.83 (negative association)

Practically, a rule such as "If Unemployment = High, then Poverty = High" can be identified without requiring the user to formulate the hypothesis in advance and can then be displayed together with its interpretation and supporting correlation panel. Thus, the integration of NLQ and Apriori successfully addresses both formulated requirements: non-technical users can search for indicators using natural language, while cross-sectoral associations can be automatically identified.

However, the very large number of rules generated (tens of thousands) under a relatively low support threshold (0.03) has the potential to cause rule explosion and produce high lift values resulting from rare combinations of items. Therefore, advanced filtering and domain-relevance assessment are required before the rules are used as a basis for policy interpretation. It is also important to reiterate that association rules describe patterns of co-occurrence rather than provide evidence of causality.

NLP, A Priori, and LLM Synergies in Decision Support

The three components work together throughout the decision-making workflow. Rule-based NLP lowers barriers to access by allowing non-technical users to search for indicators using natural language; Apriori reveals hidden associations among cross-sectoral indicators that may

be difficult to identify manually; and the local LLM completes the workflow by translating quantitative findings into an accessible executive narrative. The clear separation between deterministic parameter extraction and generative narrative generation maintains the reliability of factual values while leveraging the linguistic flexibility of the LLM. Meanwhile, the fallback mechanism ensures that the system remains operational when the LLM service is unavailable. This design supports the use of LLMs as a selective supporting component rather than as the primary source of statistical facts in public-data systems that require high accuracy.

Evaluation of the LLM Narrative Generation Module

After the formal final-project session, the narrative module underwent an exploratory five-layer assessment covering functional behavior, rubric-based quality evaluation using 30 balanced query samples, hallucination detection, cross-execution consistency, and latency/fallback reliability. These post-trial measurements provide preliminary evidence regarding the module's performance but do not constitute the same formal validation protocol used for the core NLQ and Apriori functions.

Rubric evaluations by two assessors resulted in a weighted composite score of 4.35 out of 5.00, exceeding the predefined graduation threshold of 4.00, with an inter-assessor agreement rate of 86.6%. Details for each dimension are presented in Table 5.

Table 5

Results of the LLM module narrative quality evaluation (30 samples, two assessors)

Dimensions (weight)	Average	Std. Dev	Deal	Remarks
Factual Accuracy (35%)	4,50	0,45	88%	The number matches the query summary
Language Coherence (20%)	4,20	0,60	82%	Good groove, occasionally stiff (typical of model 1B)
Relevance of Recommendations (25%)	4,00	0,75	78%	Relevant, occasionally macro
Structural Completeness (10%)	4,80	0,25	95%	Consistent three-paragraph format
Hallucination-free (10%)	4,60	0,40	90%	Minimal fictitious claims
Composite Score	4,35	0,49	86,6%	Pass (≥ 4.0)

The hallucination-detection protocol recorded a hallucination rate of 2.5% (below the 5% threshold), calculated by extracting numerical claims from the generated narratives and matching them against summaries of the source data. Operational tests showed an average latency of 17.22 seconds and a 95th-percentile latency of 26.07 seconds, both below the 45-second timeout threshold. The fallback mechanism achieved a 100% success rate across 20 simulated service failures. A summary is presented in Table 6.

Table 6

Hallucination detection, latency, and fallback reliability results of LLM modules

Metrics	Value	Threshold	Status
Hallucination Rate	2,5%	< 5%	Pass
Average latency	17.22 seconds	—	—
95th percentile latency (P95)	26.07 seconds	< 45 seconds	Pass
Fallback success	100% (20/20)	100%	Pass

In the exploratory post-trial assessment, the LLM narrative module met the predefined thresholds for fallback reliability, latency, composite quality, and hallucination rate. These results are preliminary because the assessment was conducted outside the formal final-project protocol and has not yet been replicated through a larger user-based evaluation. Accordingly, the architecture treats the LLM as a supporting narrative layer rather than as the primary source of facts or the principal validated contribution.

Discussion

The results show that the application of a hybrid architecture integrating Natural Language Query (NLQ), Apriori association-rule mining, and a Large Language Model (LLM) can improve the accessibility and depth of public statistical data exploration. The rule-based and regex-based NLQ module achieved an intent-classification accuracy of 96.00% across the 50 test queries, with 100% accuracy in the Descriptive category and 93.30% accuracy in both the Comparative and Pattern Discovery categories. These findings suggest that the deterministic approach is effective for handling queries with relatively standard structures and parameters, such as regions, years, and indicators. The use of explicit rules also provides consistency and repeatability because the same query is processed using the same deterministic mechanism. However, the two misclassifications involving queries with implicit meanings demonstrate that rule-based approaches remain limited in their ability to interpret complex semantic contexts. Therefore, the results indicate that a rule-based approach is suitable as a foundation for query understanding in public statistical systems that require precision but should be extended with semantic-comprehension capabilities to accommodate variations in user language more flexibly.

The findings from the Apriori module show that the integration of association-rule mining provides an analytical dimension that is not available in conventional statistical dashboards. The system generated 55,480 rules in the High association category, with an average confidence of 0.909 and an average lift of 15.44, while the Medium category produced 1,084 rules with an average lift of 1.20, and the Low category produced 66 rules with an average lift of 0.83. These results show that statistical indicators across sectors exhibit patterns of co-occurrence that can be identified automatically, such as the association between unemployment and poverty. The advantage of this approach lies in its ability to discover patterns without requiring users to formulate a hypothesis about the relationship between indicators in advance. Thus, Apriori functions as a pattern-discovery mechanism that complements NLQ in the data-exploration process. Nonetheless, the large number of generated rules suggests that relatively low support thresholds can lead to rule explosion and produce rules with high lift but limited substantive relevance. Therefore, association rules should be interpreted as indications of statistical association rather than evidence of cause-and-effect relationships; policy interpretation still requires validation based on sectoral context and domain knowledge.

The NLQ and Apriori components provide the formally tested core of the dashboard: NLQ lowers the technical barrier to retrieving statistical indicators, while Apriori supports exploratory discovery of cross-sectoral co-occurrence patterns. The local LLM is deliberately limited to the narrative presentation of deterministic outputs. Its exploratory post-trial assessment yielded a composite score of 4.35/5.00, a 2.5% hallucination rate, a mean latency of 17.22 seconds, a P95 latency of 26.07 seconds, and 100% fallback success. These figures are reported as preliminary evidence rather than as formal validation.

The results support the dashboard's role as a decision-support prototype that integrates search, pattern discovery, and explanatory presentation. The strongest validated contribution is the combined NLQ and Apriori architecture. Limitations remain in implicit-semantic query understanding, the large number of low-support association rules, and the need for formal user-based validation of the LLM narrative layer. Future work should therefore expand semantic coverage, strengthen rule filtering, and subject the narrative component to a preregistered and reproducible evaluation involving policy users.

CONCLUSION

This research designed and implemented a cross-sectoral analytics dashboard for East Java BPS public statistical data using a hybrid architecture comprising deterministic rule- and regex-based Natural Language Query (NLQ) with Text-to-SQL generation, dynamic Apriori association-rule mining, and a supporting local LLM narrative layer. The formally tested NLP module achieved 96.00% intent-classification accuracy across 50 queries, while Apriori generated association rules evaluated using support, confidence, and lift. Together, these core components reduce the technical barrier to statistical-data retrieval and cross-sectoral pattern exploration without requiring users to write SQL queries or manually formulate every potential relationship hypothesis.

The study is limited by classification failures involving implicit queries, the potential for rule explosion at low support thresholds, the associative and non-causal nature of Apriori rules, and the preliminary status of the LLM narrative component. A post-trial exploratory assessment of the LLM produced a composite score of 4.35/5.00, a 2.5% hallucination rate, a P95 latency of 26.07 seconds, and 100% fallback success. However, because this assessment was conducted outside the formal final-project test protocol, it should not be interpreted as equivalent to formal validation. Future work should strengthen semantic parsing, rule-interest filtering, automated BPS data updates, usability testing with policy users, and formal, replicated evaluation of narrative quality and factual faithfulness.

ACKNOWLEDGEMENT

This research was funded by the Sebelas Maret University Budget (RKAT) for the 2026 Fiscal Year through the Research Group CAPACITY STRENGTHENING (PKGR-UNS) B Research Scheme with Research Assignment Agreement Number: 462/UN27.22/PT.01.03/2026.

AUTHOR CONTRIBUTION STATEMENT

All authors contributed substantially to the completion of this study and manuscript. The authors collectively contributed to the conceptualization of the research, methodological development, system design and implementation, data processing and analysis, system testing and evaluation, interpretation of the findings, and preparation of the manuscript. All authors participated in reviewing and revising the manuscript, approved the final version for publication, and agreed to be accountable for the integrity and accuracy of the work.

REFERENCES

- Adhitya, B., Prabawa, A., & Kencana, H. (2022). Analisis Pengaruh Pendidikan, Kesehatan, Sanitasi dan Rata-Rata Jumlah Anggota Keluarga Per Rumah Tangga terhadap Kemiskinan di Indonesia. *Ekonomis: Journal of Economics and Business*, 6(1), 288–295. <https://doi.org/10.33087/ekonomis.v6i1.501>
- Agrawal, R., & Srikant, R. (1994). Fast Algorithms for Mining Association Rules in Large Databases. *Proc. of the 20th International Conference on Very Large Data Bases (VLDB'94)*.
- Asparosa, N. (2026). Visualisasi Data Kinerja Dinas Kesehatan Daerah Menggunakan Power BI. *Teknofile: Jurnal Sistem Informasi*, 4(3), 210–219.
- Ay, B., & Demirdağ, Y. E. (2026). Benchmarking Ollama and vLLM for Concurrent LLM Serving: A Multi-Scenario Evaluation of Performance and Scalability. *Applied Sciences (Switzerland)*, 16(11). <https://doi.org/10.3390/app16115435>
- Fernandez-Basso, C., Ruiz, M. D., & Martin-Bautista, M. J. (2024). New Spark solutions for distributed frequent itemset and association rule mining algorithms. *Cluster Computing*, 27(2). <https://doi.org/10.1007/s10586-023-04014-w>
- Grattafiori, A., Dubey, A., Jauhri, A., Pandey, A., Kadian, A., Al-Dahle, A., Letman, A., Mathur, A., Schelten, A., Vaughan, A., Yang, A., Fan, A., Goyal, A., Hartshorn, A., Yang, A., Mitra, A., Sravankumar, A., Korenev, A., Hinsvark, A., ... Ma, Z. (2024). *The Llama 3 Herd of Models*.
- Han, J., Kamber, M., & Pei, J. (2012). *Data Mining: Concepts and Techniques* (3rd ed.). Morgan Kaufmann. <https://www.sciencedirect.com/book/9780123814791/data-mining-concepts-and-techniques>
- Ifath, M. A., & Haque, I. (2026). Characterizing Performance-Energy Trade-offs of Large Language

- Models in Multi-Request Workflows. *Proceedings of the ACM on Measurement and Analysis of Computing Systems*, 10(1). <https://doi.org/10.1145/3788089>
- Khoirunnisa, S., Nabila, H. H., Fauziah, S., Birawa, M., Syakur, R. A., & Rahayu, S. (2025). Penerapan Algoritma K-Means dan Apriori untuk Analisis Klasterisasi Kemiskinan dan Pola Asosiasi Kepemilikan Bidang Tanah serta Karakteristik Wilayah di Indonesia. *RIGGS: Journal of Artificial Intelligence and Digital Business*, 4(2), 4911–4917. <https://journal.ilmudata.co.id/index.php/RIGGS/article/view/1364>
- Lubis, H., Komara, M. A., Meidelfi, D., Arsyah, U. I., Santoso, K. I., Sukma, F., Sulistyohati, A., Hasan, F. N., Ismanto, E., & Trisnawan, A. B. (2026). *Large Language Models: Konsep, Teknik, dan Aplikasi dalam AI Generatif*. PT Penerbit Qriset Indonesia.
- Muis, A., Syafwan, H., Arfianto, A. Z., Simanjuntak, M. S., Riyandi, A., Trisnawan, A. B., Ramdhan, W., Triana, H., Saputra, M. H., Handoko, W., Naswin, A., & Suhadi. (2025). *DATA MINING: Konsep, Metode, dan Aplikasi*. PT. Faaslib Serambi Media. <https://faaslibsmmedia.com/Buku>
- Putri, W. C., & Supriyadi, M. E. (2024). Visualisasi Dashboard Trend Kemiskinan Kota Depok Tahun 2021-2022 Dengan Menggunakan Microsoft Power Business Intelligence Dan Metode UAT (User Acceptance). *Jurnal Teknik Dan Science*, 3(3), 50–55. <https://journal.admi.or.id/index.php/JTS/article/download/1897/1853/5587>
- Sari, A. N., & Darmanto, E. (2025). Implementasi Sistem Visualisasi Data Statistik pada Badan Pusat Statistik Kabupaten Pati untuk Peningkatan Pelayanan Masyarakat. *Jurnal Pengabdian Masyarakat Bhinneka*, 4(2), 1733–1741. <https://doi.org/10.58266/jpmb.v4i2.690>
- Solimun, Fernandes, A. A. R., Nurjannah, Erwinda, E. G., Hardianti, R., & Arini, L. H. Y. (2023). *Metodologi Penelitian: Variabel Mining berbasis Big Data dalam Pemodelan Sistem untuk Mengungkap Research Novelty*. Universitas Brawijaya Press. <https://doi.org/10.11594/ubpress9786232967052>
- Team, L., & Meta, A. @. (2024). The Llama 3 Tech report. *ArXiv Preprint*.
- Wang, B., Shin, R., Liu, X., Polozov, O., & Richardson, M. (2020). RAT-SQL: Relation-aware schema encoding and linking for text-to-SQL parsers. *Proceedings of the Annual Meeting of the Association for Computational Linguistics*. <https://doi.org/10.18653/v1/2020.acl-main.677>
- Wong, A., Pham, L., Lee, Y., Chan, S., Sadaya, R., Khmelevsky, Y., Clement, M., Yau Cheng, F. W., Mahony, J., & Ferri, M. (2024). Translating Natural Language Queries to SQL Using the T5 Model. *SysCon 2024 - 18th Annual IEEE International Systems Conference, Proceedings*. <https://doi.org/10.1109/SysCon61195.2024.10553509>
- Yu, T., Zhang, R., Yang, K., Yasunaga, M., Wang, D., Li, Z., Ma, J., Li, I., Yao, Q., Roman, S., Zhang, Z., & Radev, D. R. (2018). Spider: A large-scale human-labeled dataset for complex and cross-domain semantic parsing and text-to-SQL task. *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, EMNLP 2018*. <https://doi.org/10.18653/v1/d18-1425>
- Zhang, S., Dong, L., Li, X., Zhang, S., Sun, X., Wang, S., Li, J., Hu, R., Zhang, T., Wang, G., & Wu, F. (2026). Instruction Tuning for Large Language Models: A Survey. *ACM Computing Surveys*, 58(7). <https://doi.org/10.1145/3777411>
- Zhong, V., Xiong, C., & Socher, R. (2017). Seq2sql: Generating structured queries from natural language using reinforcement learning. *ArXiv Preprint ArXiv:1709.00103*.