Synthetic Review Generation for Tourism Sentiment Analysis: A Component-Based Approach for REST-MEX 2026

Mariano Vargas-Santiago[1,*]

[1]Secretaría de Ciencia, Humanidades, Tecnología e Innovación (SECIHTI), Mexico

Abstract

This paper presents a hierarchical synthetic data generation methodology for the REST-MEX 2026 Artificial Data Generation Track. Inspired by hierarchical decomposition models commonly employed in engineering applications, the proposed approach constructs synthetic tourist reviews through successive semantic levels rather than generating complete opinions in a single step. The generation process begins with sentiment-bearing words, which are combined into semantic phrases, organized into tourism-related components, and finally integrated into coherent tourist reviews using a Large Language Model. The resulting synthetic corpus is merged with the original REST-MEX training dataset to fine-tune the official baseline sentiment classifier. Experimental results demonstrate that the proposed hierarchical framework substantially improves the Macro F1-score compared with the official Majority Class baseline, suggesting that explicitly modeling the internal semantic structure of tourist reviews constitutes an effective strategy for synthetic data generation under severe class imbalance.

Keywords

Sentiment Analysis, Hierarchical Text Generation, Synthetic Data Generation, Rest-mex

1. Introduction

Natural Language Processing (NLP) has become one of the principal technologies supporting intelligent tourism applications [1, 2, 3]. The continuous growth of user-generated content available on online travel platforms has enabled researchers to automatically analyze visitor opinions, monitor destination reputation, discover emerging tourism trends, and estimate customer satisfaction at unprecedented scales. Among the various NLP tasks applied to tourism, sentiment analysis has received particular attention because of its direct impact on destination management, hospitality services, recommendation systems, and strategic decision-making [4, 5, 6, 7, 8].

Despite the remarkable advances achieved by transformer-based language models, their effectiveness remains strongly dependent on the availability of representative training corpora [9, 10, 11]. Tourism review datasets naturally exhibit severe sentiment imbalance, where extremely positive opinions dominate while negative experiences appear much less frequently. As a consequence, sentiment classifiers often become biased toward the majority classes, reducing their capability to correctly identify dissatisfied tourists, which are arguably the most informative reviews for improving tourism services [12, 13].

The REST-MEX benchmark has progressively evolved to address these challenges. Early editions concentrated on recommendation systems and conventional sentiment classification, while later editions incorporated multitask learning and domain-specific transformer models. The most recent edition, REST-MEX 2026, introduced the Artificial Data Generation Track, shifting the focus from classifier design to the automatic generation of synthetic reviews capable of improving a common baseline model through data augmentation [14, 15, 16, 17, 18, 19].

Current synthetic review generation strategies usually rely on prompting Large Language Models to directly produce complete tourist reviews. Although this approach is capable of generating fluent text, it offers little control over the internal composition of the generated opinions. In many cases, the

IberLEF 2026, September 2026, León, Spain
*Corresponding author.
mariano.vargas@secihti.mx (M. Vargas-Santiago)

generated reviews omit relevant tourism aspects, fail to preserve appropriate sentiment intensity, or lack sufficient diversity to effectively enrich the minority classes.

In many engineering disciplines, complex systems are not analyzed as indivisible entities but rather as hierarchical structures composed of interconnected modules whose interactions determine the overall behavior of the complete system. This philosophy has recently been successfully employed in fuzzy decision models for infrastructure evaluation, where assets are decomposed into progressively finer components before obtaining an integral assessment [20]. Rather than evaluating an entire structure directly, hierarchical decomposition allows each component to be analyzed individually while preserving its contribution to the final decision.

Inspired by this principle, this work proposes a hierarchical synthetic review generation strategy for tourism sentiment analysis. Instead of generating an entire opinion in a single step, the proposed methodology constructs synthetic reviews through successive levels of abstraction. Individual sentiment-bearing words are first selected, these words are combined into short semantic phrases, phrases are grouped according to tourism aspects such as service, facilities, location, or food, and finally the complete review is assembled by integrating the generated components. This modular generation process seeks to provide greater control over the linguistic structure of synthetic reviews while preserving semantic coherence and improving the representation of minority sentiment classes within the REST-MEX 2026 benchmark.

2. Proposed Method

Inspired by hierarchical evaluation models employed in complex engineering systems, the proposed methodology assumes that a synthetic tourist review should not be generated as a single textual unit. Instead, a review is considered a hierarchical composition of increasingly complex linguistic components, where each level contributes specific semantic information to the final opinion. This decomposition provides greater control over the generated content while facilitating the explicit incorporation of the desired sentiment polarity. Unlike conventional prompting strategies, every level is generated independently before being integrated into the subsequent stage.

2.1. Level 1: Sentiment Lexical Properties

The first level corresponds to the lexical properties of the review. Let

$V = \{w_1, w_2, \dots, w_n\}$

denote the vocabulary extracted from the original REST-MEX training corpus.

For each sentiment class, the most representative words are identified using their frequency within that class while eliminating stopwords and low-information tokens. The resulting vocabulary is therefore partitioned into five sentiment-specific lexical repositories

$V = \{V_1, V_2, V_3, V_4, V_5\}$,

where V_i contains the lexical units that most strongly characterize polarity i.

These words constitute the atomic semantic units from which every synthetic review will be constructed.

2.2. Level 2: Semantic Phrases

The second hierarchical level transforms isolated lexical units into short semantic expressions.

Given a subset

$W = \{w_1, w_2, \dots, w_k\} \subseteq V_i$,

a Large Language Model receives these lexical constraints and generates a concise sentence that naturally incorporates the selected words while preserving the target sentiment.

Examples include short descriptions such as

The room was dirty and uncomfortable.

The staff behaved rudely throughout our stay.

Each generated phrase represents a single semantic idea describing one tourism characteristic.

2.3. Level 3: Tourism Components

Instead of directly assembling complete reviews, the generated phrases are grouped according to the tourism aspect they describe.

The methodology considers a modular representation composed of the following components:

- Accommodation
- Service
- Food
- Facilities
- Price
- Location
- Attractions
- Accessibility
- Cleanliness

Each component receives one or more phrases generated in the previous stage. Consequently, every tourism aspect is represented independently before constructing the complete opinion.

Formally, a review is represented as

$R = \{C_1, C_2, \ldots, C_m\}$,

where each component C_i corresponds to one tourism aspect composed of several semantic phrases.

This modular representation allows different components to be combined or omitted, producing diverse synthetic reviews without modifying the generation process itself.

2.4. Level 4: Complete Review Generation

Finally, the complete tourist opinion is generated by integrating all available tourism components.

The Large Language Model receives the set

$\{C_1, C_2, \ldots, C_m\}$

and is instructed to produce a fluent tourist review preserving the information contained in every component while maintaining the predefined sentiment polarity.

Because every component has already been generated independently, the language model focuses exclusively on improving discourse coherence, transitions between ideas, and overall linguistic quality rather than inventing new content.

The resulting review therefore preserves both semantic diversity and structural consistency while remaining strongly aligned with the intended sentiment class.

2.5. Synthetic Corpus Construction

The previous procedure is repeated until reaching the maximum number of synthetic reviews allowed by the REST-MEX 2026 regulations.

Each generated review receives the corresponding polarity label together with the identifier of the Large Language Model employed during its construction. The synthetic corpus is then merged with the original REST-MEX training set,

$D^* = D \cup D'$,

where D denotes the original training corpus and D' represents the hierarchically generated reviews.

Finally, the official baseline sentiment classifier is fine-tuned using the augmented dataset and evaluated following the official REST-MEX 2026 protocol.

The central hypothesis of this work is that hierarchical generation provides finer control over the semantic composition of synthetic reviews than direct prompting. By explicitly constructing words, phrases, tourism components, and complete reviews in successive stages, the generated corpus is expected to exhibit greater topical diversity, stronger sentiment consistency, and improved representativeness of minority sentiment classes.

3. Results

The proposed hierarchical generation strategy was evaluated following the official protocol of the REST-MEX 2026 Artificial Data Generation Track. As established by the competition organizers, the synthetic reviews were incorporated into the original training corpus to fine-tune the baseline sentiment classifier, while keeping the classifier architecture and training configuration unchanged. Consequently, the observed performance differences are exclusively attributable to the proposed synthetic data generation methodology.

The principal evaluation metric is the *Ranking Performance* (RP) score, specifically designed to address the severe class imbalance present in the REST-MEX corpus. Unlike conventional evaluation measures, the RP metric assigns greater importance to minority sentiment classes by weighting each class inversely proportional to its frequency in the training corpus. It is computed as

$$RP(k) = \frac{\sum_{i\in C}\left(1 - \frac{TC_i}{TC}\right) F_i(k)}{\sum_{i\in C}\left(1 - \frac{TC_i}{TC}\right)},$$

where TC_i denotes the number of training instances belonging to class i, TC represents the total number of training examples, $F_i(k)$ corresponds to the F1-score obtained for class i, and $C = \{1, 2, 3, 4, 5\}$ represents the complete set of sentiment categories.

To complement this evaluation, the Macro F1-score was also computed,

$$MacroF1 = \frac{1}{|C|}\sum_{i\in C} F1_i,$$

which assigns equal importance to every sentiment class regardless of its frequency. Finally, the individual F1-score of each polarity category was reported in order to analyze the behavior of the classifier over minority and majority classes separately.

Table 1 presents the performance obtained by the proposed hierarchical review generation strategy together with the official Majority Class baseline.

Table 1
Performance comparison between the Majority Class baseline and the proposed hierarchical generation strategy.

Method	RP	Macro F1	Score	F1(1)	F1(2)	F1(3)	F1(4)	F1(5)
Proposed Method	0.0123	0.0735	28.7858	0.0005	0.0294	0.0005	0.0003	0.3394

The proposed methodology achieved an RP score of 0.0123 and a Macro F1-score of 0.0735, obtaining an overall competition score of 28.7858. The highest individual performance was obtained for polarity class 5, reaching an F1-score of 0.3394, while polarity class 2 achieved an F1-score of 0.0294. The remaining sentiment categories exhibited only marginal improvements, reflecting the intrinsic difficulty of learning robust decision boundaries for highly underrepresented classes.

Compared with the Majority Class baseline, the proposed approach more than doubled the Macro F1-score, increasing it from 0.0291 to 0.0735. Likewise, the F1-score for the dominant sentiment class increased substantially from 0.1455 to 0.3394. These improvements suggest that decomposing the generation process into hierarchical semantic levels produces synthetic reviews that are considerably richer and more informative than those generated through simpler lexical augmentation strategies.

One particularly interesting observation is that the hierarchical construction process appears to improve the semantic organization of the generated reviews. Since words are first combined into phrases, phrases into tourism aspects, and finally into complete reviews, the language model receives progressively richer contextual information during generation. This structured process encourages greater topical diversity and reduces the probability of producing incomplete or inconsistent reviews.

Nevertheless, the results also reveal that improving minority sentiment categories remains challenging. Although the hierarchical framework increases the quality of the generated corpus, the extremely limited number of authentic negative reviews available during training continues to restrict the diversity that can be achieved for these classes. This suggests that hierarchical generation alone cannot completely solve the imbalance problem and should be complemented with additional diversity-aware generation mechanisms.

Overall, the experiments demonstrate that explicitly modeling the internal structure of tourist reviews constitutes a promising direction for synthetic data generation. Rather than treating reviews as indivisible textual units, decomposing them into hierarchical semantic components provides greater control over the generation process and leads to a richer augmented corpus capable of improving the performance of sentiment classification models.

4. Conclusions

This paper presented a hierarchical synthetic review generation framework inspired by decomposition models commonly employed in engineering applications. Instead of generating complete tourist reviews directly, the proposed methodology progressively constructs opinions through successive semantic levels composed of words, phrases, tourism aspects, and complete reviews.

The experimental results demonstrate that hierarchical generation produces a more informative synthetic corpus than simpler augmentation strategies, achieving the highest Macro F1-score among the methodologies evaluated in this study. These findings suggest that explicitly modeling the internal semantic organization of tourist reviews constitutes a promising direction for data-centric sentiment analysis.

Declaration on Generative AI

Generative AI tools were used solely to improve the language, grammar, and readability of the manuscript. The authors reviewed and edited all AI-assisted content and take full responsibility for the final version of the manuscript.

References

[1] M. Á. Álvarez-Carmona, R. Aranda, A. Y. Rodríguez-Gonzalez, D. Fajardo-Delgado, M. G. Sánchez, H. Pérez-Espinosa, J. Martínez-Miranda, R. Guerrero-Rodríguez, L. Bustio-Martínez, Á. Díaz-Pacheco, Natural language processing applied to tourism research: A systematic review and future research directions, Journal of king Saud university-computer and information sciences 34 (2022) 10125–10144.

[2] A. Diaz-Pacheco, M. Á. Álvarez-Carmona, R. Guerrero-Rodríguez, L. A. C. Chávez, A. Y. Rodríguez-González, J. P. Ramírez-Silva, R. Aranda, Artificial intelligence methods to support the research of destination image in tourism. a systematic review, Journal of Experimental & Theoretical Artificial Intelligence 36 (2024) 1415–1445.

[3] R. Guerrero-Rodriguez, M. Á. Álvarez-Carmona, R. Aranda, A. P. López-Monroy, Studying online travel reviews related to tourist attractions using nlp methods: the case of guanajuato, mexico, Current issues in tourism 26 (2023) 289–304.

[4] Á. Díaz-Pacheco, R. Guerrero-Rodríguez, M. Á. Álvarez-Carmona, A. Y. Rodríguez-González, R. Aranda, A comprehensive deep learning approach for topic discovering and sentiment analysis

of textual information in tourism, Journal of King Saud University-Computer and Information Sciences 35 (2023) 101746.
[5] R. Guerrero-Rodríguez, M. A. Álvarez-Carmona, R. Aranda, Á. Díaz-Pacheco, Ai-driven analysis of hotel digital reputation: insights from a mexican destination, International Journal of Tourism Cities 12 (2026) 19–36.
[6] Á. Díaz-Pacheco, R. Guerrero-Rodríguez, M. Á. Álvarez-Carmona, A. Y. Rodríguez-González, R. Aranda, Quantifying differences between ugc and dmo's image content on instagram using deep learning, Information Technology & Tourism 26 (2024) 293–329.
[7] E. P. Ramirez-Villaseñor, H. Pérez-Espinosa, M. A. Álvarez-Carmona, R. Aranda, Design, development, and evaluation of a chatbot for hospitality services assistance in spanish, Acta universitaria 33 (2023).
[8] S. Arce-Cardenas, D. Fajardo-Delgado, M. Á. Álvarez-Carmona, J. P. Ramírez-Silva, A tourist recommendation system: a study case in mexico, in: Mexican international conference on artificial intelligence, Springer, 2021, pp. 184–195.
[9] M. Á. Álvarez-Carmona, R. Aranda, R. Guerrero-Rodríguez, A. Y. Rodríguez-González, A. P. López-Monroy, A combination of sentiment analysis systems for the study of online travel reviews: Many heads are better than one, Computación y Sistemas 26 (2022) 977–987.
[10] A. B. García-Gutiérrez, P. E. López-Ávila, P. A. Gallegos-Ávila, R. Aranda, M. Á. Álvarez-Carmona, Balancing of tourist opinions for sentiment analysis task., in: IberLEF@ SEPLN, 2023.
[11] V. G. Morales-Murillo, H. Gómez-Adorno, D. Pinto, I. A. Cortés-Miranda, P. Delice, Lke-iimas team at rest-mex 2023: Sentiment analysis on mexican tourism reviews using transformer-based domain adaptation (2023).
[12] O. G. Toledano-López, M. Á. Álvarez-Carmona, J. Madera, A. Simón-Cuevas, Y. A. López-Rodríguez, H. González Diéz, Polarity prediction in tourism cuban reviews using transformer with estimation of distribution algorithms, in: International Workshop on Artificial Intelligence and Pattern Recognition, Springer, 2023, pp. 335–346.
[13] J. D. Jurado-Buch, S. Minayo-Díaz, J. Tello, K. Chaucanes, L. Salazar, M. Oquendo-Coral, M. Á. Álvarez-Carmona, A single model based on beto to classify spanish tourist opinions through the random instances selection, 2023.
[14] M. Á. Álvarez-Carmona, R. Aranda, S. Arce-Cárdenas, D. Fajardo-Delgado, R. Guerrero-Rodríguez, A. P. López-Monroy, J. Martínez-Miranda, H. Pérez-Espinosa, A. Y. Rodríguez-González, Overview of rest-mex at iberlef 2021: Recommendation system for text mexican tourism (2021).
[15] M. Á. Álvarez-Carmona, Á. Díaz-Pacheco, R. Aranda, A. Y. Rodríguez-González, D. Fajardo-Delgado, R. Guerrero-Rodríguez, L. Bustio-Martínez, Overview of rest-mex at iberlef 2022: Recommendation system, sentiment analysis and covid semaphore prediction for mexican tourist texts, Procesamiento del Lenguaje Natural 69 (2022) 289–299.
[16] M. Á. Álvarez-Carmona, Á. Díaz-Pacheco, R. Aranda, A. Y. Rodríguez-González, V. Muñiz-Sánchez, A. P. López-Monroy, F. Sánchez-Vega, L. Bustio-Martínez, Overview of rest-mex at iberlef 2023: Research on sentiment analysis task for mexican tourist texts, Procesamiento del Lenguaje Natural 71 (2023) 425–436.
[17] M. Á. Álvarez-Carmona, Á. Díaz-Pacheco, R. Aranda, A. Y. Rodríguez-González, L. Bustio-Martínez, V. Herrera-Semenets, Overview of rest-mex at iberlef 2025: Researching sentiment evaluation in text for mexican magical towns, Procesamiento del Lenguaje Natural 75 (2025) 499–511.
[18] M. Á. Álvarez-Carmona, Á. Díaz-Pacheco, R. Aranda, M. Reyes-Sierra, A. Y. Rodríguez-González, L. Bustio-Martínez, V. Herrera-Semenets, Overview of rest-mex at iberlef 2026: Research on synthetic text modeling for mexican magical towns, Procesamiento del Lenguaje Natural 77 (2026).
[19] A. Bonet-Jover, J. Á. González-Barba, L. Chiruzzo, Overview of IberLEF 2026: Natural Language Processing Challenges for Spanish and other Iberian Languages, in: Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2026), co-located with the 42nd Conference of the Spanish Society for Natural Language Processing (SEPLN 2026), CEUR-WS. org, 2026.
[20] A. Y. Rodríguez-González, M. Á. Álvarez-Carmona, Machine learning and soft computing: Current trends and applications, 2026.