

A Parallel NLP Pipeline with NER-Enhanced Hierarchical Classification: Sentiment Analysis for Mexican Magical Towns

Gabriel Santiago Robles-Robles¹, Jehu Jonathan Ramirez-Ramirez¹,
Angel David Durazo-Bartolini¹, Gael Balderrama-Dominguez¹,
Luis Hiram Hernandez-Gutierrez¹, Victor Hugo Ramirez-Rios¹, Juan Adan Nava-Banda¹,
Gustavo Gutierrez-Navarro¹, Juan Daniel Garcia-Ruiz¹, Mario Alejandro Castro-Lerma¹,
Jesus Antonio Flores-Briones¹, Manuel Ivan Melendez-Rivera¹,
Christian Alexis Flores-Alvarez¹, Jose Angel Morales-Nuñez¹ and Mauricio Toledo-Acosta^{1,*}

¹Universidad de Sonora, Departamento de Matemáticas, Blvd. Luis Encinas y Rosales, Col. Centro, Hermosillo, Sonora, 83000, Mexico

Abstract

We present a parallel NLP pipeline for sentiment analysis and multi-label classification of Spanish tourism reviews from Mexican Magical Towns, developed for the Rest-Mex 2025 shared task. Our architecture combines three specialized models: (1) a fine-tuned Qwen3-0.6B transformer for 5-class sentiment prediction, (2) a TF-IDF logistic regression classifier for destination type categorization, and (3) a NER-enhanced hierarchical model for town identification that integrates named entity recognition with BERT embeddings. The system achieved 69.57th percentile overall in the competition, with the town classifier excelling in location-specific performance (Macro F1: 0.6006, 75.36th percentile). Macro F1 scores of 0.4753 for sentiment analysis and 0.9423 for destination type classification demonstrate the effectiveness of our modular approach. Key contributions include handling extreme class imbalance (e.g., 62:1 in the town labels distribution) through hierarchical classification. Results demonstrate that hybrid architectures combining transformers, traditional Machine Learning, and knowledge-enhanced components outperform monolithic approaches for tourism NLP tasks, though challenges remain in fine-grained sentiment analysis. Our modular design offers computational efficiency while maintaining a certain degree of interpretability.

Keywords

Tourism Sentiment Analysis, Modular NLP Pipelines, Named Entity Recognition, Large Language Models

1. Introduction

Tourism remains a critical driver of global economic growth, contributing significantly to employment and GDP worldwide. In the past decade, the industry has increasingly shifted its operations online, with digital platforms playing a pivotal role in shaping travel decisions [1, 2, 3, 4]. In Mexico, tourism continues to be a cornerstone of the economy, accounting for approximately 8.6% of GDP [5], and generating over 4.8 million direct jobs as of the second quarter of 2024 [5]. Despite the severe disruptions caused by the COVID-19 pandemic, the national sector has shown resilience, with recovery trends highlighting the growing reliance on data-driven strategies to adapt to evolving traveler behaviors [5, 6, 7].

In this context, Artificial Intelligence (AI)—particularly Natural Language Processing (NLP)—has emerged as a crucial tool for extracting insights from tourists' opinions [8, 9]. Social media platforms, review websites, and other digital channels generate vast amounts of unstructured feedback, reflecting subjective experiences, sentiments, and preferences [10, 11, 12]. For instance, recent studies demonstrate that a vast majority of travelers rely on online reviews when planning trips [1], while businesses and

IberLEF 2025, September 2025, Zaragoza, Spain

*Corresponding author.

✉ mauricio.toledo@unison.mx (M. Toledo-Acosta)

ORCID 0000-0002-8047-0189 (M. Toledo-Acosta)



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

policymakers increasingly leverage this data to enhance services, optimize marketing, and design targeted policies [7].

This paper addresses sentiment analysis and multi-label classification of Spanish-language tourism reviews from Mexican Magical Towns as part of the Rest-Mex 2025: Researching Sentiment Evaluation in Text for Mexican Magical Towns at IberLef 2025. Building on advances in NLP, we propose a parallel processing architecture consisting of three specialized models: a fine-tuned Qwen3-0.6B transformer for sentiment polarity classification (1-5 scale), a TF-IDF-based logistic regression classifier for destination type prediction (hotel, restaurant, attraction), and a hybrid two-stage model for town identification that combines named entity recognition with hierarchical BERT-based classification across 40 Mexican Magical Towns.

2. Related Work

Sentiment analysis has evolved from traditional rule-based approaches to more sophisticated deep learning architectures, with tourism-specific applications gaining considerable attention in recent years [13]. The field encompasses three primary methodological categories: knowledge-based systems, machine learning approaches, and hybrid architectures that combine both paradigms.

Knowledge-based approaches rely on sentiment lexicons and manually curated dictionaries to determine the polarity of text and words. These methods typically incorporate resources such as SentiWordNet [14], and domain-specific vocabularies that capture sentiment-bearing words. While interpretable and linguistically grounded, these approaches often struggle with context-dependent sentiment and domain-specific expressions common in tourism reviews.

Machine learning approaches can be further divided into traditional feature engineering methods and deep learning architectures. Traditional methods focus on extracting handcrafted features such as BOW countings, TF-IDF vectors, n -grams, and part-of-speech patterns, which are then fed to classifiers like Support Vector Machines, Multinomial Naive Bayes classifiers, or Random Forest [15]. Recent advances have shown that logistic regression with TF-IDF features remains competitive for text classification tasks, particularly when computational efficiency is prioritized [16].

The emergence of transformer-based architectures has revolutionized sentiment analysis, with models like BERT [17] demonstrating superior performance through contextual understanding. Pre-trained language models, initially developed for English, have been successfully adapted to Spanish through multilingual variants like XLM-RoBERTa [18, 19] and language-specific models. The recent development of efficient transformer architectures, such as the Qwen3 family of models [20], has made fine-tuning accessible for resource-constrained scenarios while maintaining competitive performance.

Hybrid systems combine the interpretability of knowledge-based methods with the learning capacity of machine learning approaches. Our previous work [21] demonstrated the effectiveness of combining scored word embeddings with vector representations for tourism sentiment analysis. Similarly, [22] used word2vec embeddings to construct sentiment dictionaries for social media analysis.

Tourism-specific sentiment analysis presents unique challenges due to the multilingual nature of reviews and cultural context dependencies. Recent studies have shown that tourism reviews often exhibit different sentiment patterns compared to general product reviews, with aspects like location, service quality, and cultural experiences requiring specialized treatment [23]. The emergence of large-scale tourism datasets has enabled more sophisticated modeling approaches, though class imbalance remains a persistent challenge.

Named Entity Recognition (NER) has become increasingly important in tourism applications, particularly for location identification, identifying geographical entities crucial for destination-specific analysis, and aspect-based sentiment analysis [24, 25]. The integration of NER with sentiment analysis enables more nuanced understanding of location-specific opinions and experiences.

Multi-task learning approaches have gained traction in recent years, with architectures designed to simultaneously predict multiple aspects of text content [26]. Our approach operates within this paradigm, employing independent specialized parallel models for each prediction target (polarity, type,

and town).

The Rest-Mex shared task series has provided valuable benchmarks for Spanish tourism sentiment analysis, with previous editions highlighting the effectiveness of ensemble methods and domain-specific adaptations [31, 32, 33, 27, 28]. Our current work builds upon these foundations by combining state-of-the-art transformer fine-tuning with traditional machine learning robustness for comprehensive tourism review analysis.

3. Methods

The task addressed in this study involves a multi-label classification problem for Spanish-language tourism reviews from Mexican destinations. The training dataset consists of 208051 reviews of tourist destinations across Mexico, structured with six primary columns that capture comprehensive information about each tourist experience. The Title column contains the brief headline given by tourists to summarize their opinion, while the Review column includes the full detailed text of their experience. The dataset includes three target variables for prediction: Polarity, representing sentiment on a five-point scale from 1 (very negative) to 5 (very positive); Type, categorizing destinations as Hotel, Restaurant, or Attractive; and Town, identifying the specific location from a list of 40 officially designated Mexican Magical Towns. Additionally, the Region column provides the Mexican state information, which serves as supplementary context rather than a classification target variable.

The test dataset contains 89166 instances and follows the same structure as the training set, with the exception that it does not include the Region column, requiring models to rely solely on textual content and other available features for prediction.

Each review exhibits typical characteristics of user-generated content, including varied writing styles, colloquial expressions, and encoding inconsistencies common in Spanish-language web-scraped data.

In this section, we present our top-performing architecture, which employs a parallel processing approach for the three aforementioned tasks. Specifically, the system consists of three independent models, each dedicated to a single task and operating concurrently without inter-model dependencies. We now describe each of these three models.

Each document D_j in the training dataset has the form

$$\left[\begin{array}{c|c|c|c|c} T_j & R_j & pol_j & ty_j & tw_j \end{array} \right]$$

Where the text T_j represents the text content formed by concatenating the title and opinion fields, R_j indicates the regional information and pol_j , ty_j , and tw_j are the three classification labels Polarity (pol_j), Type (ty_j) and Town (tw_j).

3.1. Preprocessing

The preprocessing pipeline for this dataset involved several steps. Initially, the provided training dataset was split into training and testing sets using stratified sampling to maintain class distribution. Missing values in titles were filled with blank spaces, while duplicate reviews were removed by keeping only the last occurrence of each duplicated entry.

Next, character encoding issues were addressed implementing a Latin-1 to UTF-8 conversion process that successfully corrected 18,451 instances of malformed characters, restoring proper accents and special characters essential for Spanish language processing. The title and review fields were then concatenated into a single text feature to capture the complete semantic content of each review.

From this point, two distinct preprocessing paths were implemented to accommodate different modeling approaches: one tailored for traditional Machine Learning algorithms, and another optimized for transformer-based models such as BERT, and other modern language models.

The former version was obtained by using text normalization, which included converting all content to lowercase and applying tokenization using a blank Spanish language model from SpaCy. The

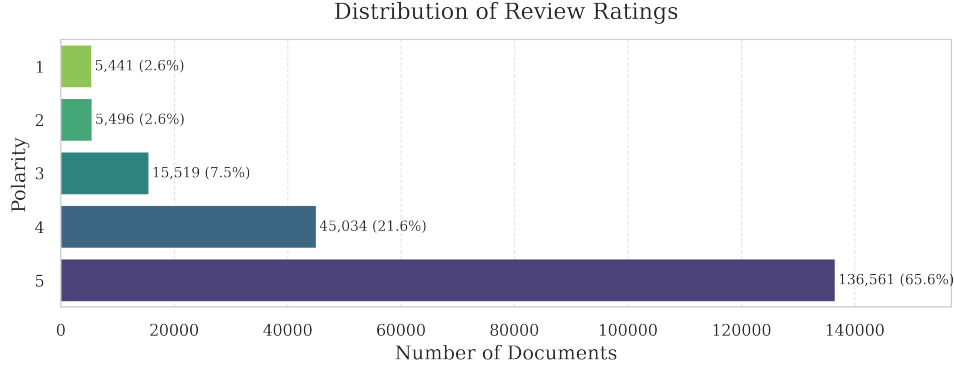


Figure 1: Distribution of polarity labels in the dataset, showing strong bias toward positive ratings (class 5 dominate with 65.6% of samples).

tokenization process filtered out punctuation, excessive whitespace, and non-alphabetic tokens while removing Spanish stopwords along with additional noise characters (e.g. \n, \b).

The latter version of the dataset was obtained by applying minimal preprocessing. Unlike traditional approaches, this preprocessing pipeline preserved the original text structure and linguistic features that are crucial for contextual understanding in modern language models. The text underwent basic cleaning to remove excessive whitespace, and additional noise characters (e.g. \n, \b), while maintaining punctuation, capitalization patterns, and stopwords that provide important contextual cues for transformer attention mechanisms [17].

3.2. Polarity Model

The **Polarity Model** M_p is a Large Language Model fine-tuned for 5-class text classification based on Qwen/Qwen3-0.6B [20], a 0.6-billion-parameter decoder-only transformer originally pretrained for text generation. The Qwen3 architecture features 28 hidden layers with 16 attention heads and 1024-dimensional hidden states, using SiLU activation and RMS layer normalization ($\varepsilon = 1 \times 10^{-6}$). To adapt it to the classification task, we employed the `AutoModelForSequenceClassification` wrapper from Hugging Face Transformers, adding a classification head with 5 output labels corresponding to star ratings (1-5 stars). Since generative models typically lack a padding token, we explicitly added one to the tokenizer and updated the model configuration accordingly. Input texts were tokenized to a maximum length of 128 tokens (selected from the original 40960 max_position_embeddings) with padding and truncation. Fine-tuning was performed using the Hugging Face Trainer API with the AdamW optimizer, a learning rate of 2×10^{-5} , a batch size of 4, and training over 3 epochs with a weight decay of 0.01. Evaluation was conducted at the end of each epoch using accuracy and F1 score.

As shown in Figure 1, the dataset exhibits significant class imbalance, with ratings distributed as: 5 (136,561 reviews, 65.6%), 4 (45,034, 21.6%), 3 (15,519, 7.5%), 2 (5,496, 2.6%), and 1 (5,441, 2.6%). This 25:1 ratio between majority and minority classes necessitated weighted loss functions during training.

Since the imbalance reflects genuine user behavior patterns, we intentionally preserved the original distribution during training. This approach maintains the model’s exposure to the natural data distribution it will encounter during inference [29], while prioritizing performance on majority classes that dominate real-world use cases. However, to mitigate potential bias while respecting the data’s natural skew, we employed weighted evaluation metrics and closely monitored per-class accuracy throughout training [30].

3.3. Type Model

The **Type Model** M_{ty} is a logistic regression classifier with L_2 -regularization. The classifier was trained on TF-IDF features extracted from the preprocessed text using a `TfidfVectorizer` with an n -gram

range of (1, 3), capturing unigrams, bigrams, and trigrams. The vocabulary was limited to the 10,000 most frequent features.

As shown in Figure 2, the type labels exhibit near-balanced distribution, allowing us to employ standard training protocols without specialized class balancing techniques.

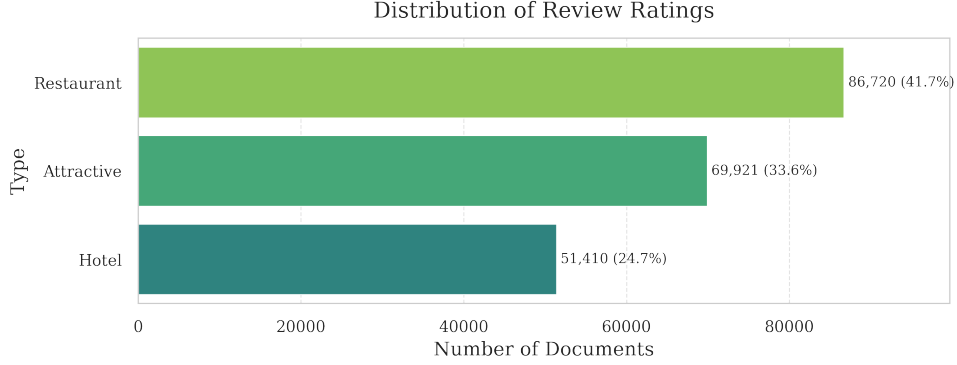


Figure 2: Distribution of type labels in the dataset, showing a near-balanced task.

The logistic regression model was configured with a regularization strength $C = 1.0$, a tolerance of 10^{-4} for the stopping criterion, and a maximum of 1000 iterations.

3.4. Town Model

The **Town Model** M_{tw} combines two specialized sub-models, $M_{tw,1}$ and $M_{tw,2}$, designed to handle both explicit location mentions and contextual inference for the town classification respectively.

The sub-model $M_{tw,1}$ predicts town labels by matching location entities (LOCs) against a predefined dictionary $\{(tw_i, \{e_1^{(i)}, \dots, e_n^{(i)}\})\}$, achieving 10% coverage. For the remaining 90% of documents lacking dictionary LOCs, $M_{tw,2}$ performs contextual prediction using Machine Learning methods.

3.4.1. Named Entity Recognition Sub-model $M_{tw,1}$

Sub-model $M_{tw,1}$ operates as a location mapping dictionary, where:

- Keys: Town labels (e.g., “Dolores Hidalgo”).
- Values: Associated location named entities learned from training data (e.g., “Atotonilco”).

The sub-model predicts towns by matching extracted named entities in test reviews against this dictionary. For each review text T_j , the prediction tw_{pred} is given by

$$tw_{pred} = \begin{cases} tw_i & \text{if } \exists e_k \in T_j \text{ where } e_k \in \{e_1^{(i)}, \dots, e_n^{(i)}\} \\ \text{no prediction,} & \text{otherwise} \end{cases} \quad (1)$$

While this mechanism is precise for entries containing known location references, its coverage is limited because $\sim 90\%$ of test reviews do not contain NERs matching the dictionary.

We now formally describe the mapping dictionary construction process. The model $M_{tw,1}$ is based on MMG/xlm-roberta-large-ner-spanish, a pretrained XLM-ROBERTa model fine-tuned for Spanish Named Entity Recognition. For each training document T_j , we perform inference using this model to extract a list of named entities (e, d) , where:

- e : The detected entity text.
- d : The entity type (LOC for locations, PER for persons, ORG for organizations, etc.).

Given our focus on identifying the town class, we retain only geographic entities, i.e. those labeled as LOC. This step ensures that our model prioritizes location based entities while filtering out irrelevant entities that could introduce noise.

We construct a *mapping dictionary* that associates each town label tw_i with its corresponding location entities $\{e_1^{(i)}, \dots, e_n^{(i)}\}$ extracted from training documents labeled as tw_i . The dictionary structure follows the format shown in Table 1, where each entry pairs a town with its distinctive geographical references.

Table 1
Structure of the town-to-entities mapping dictionary

Town Label	: Associated LOC Entities
tw_i	: $e_1^{(i)}, \dots, e_n^{(i)}$



Figure 3: Distribution of location entities across three representative towns, showing entity overlap between different town labels. The wordcloud visualization highlights the need for entity disambiguation.

As evidenced in Figure 3, several location entities appear in multiple town classes (e.g., “México”). To ensure unambiguous town prediction, we eliminated these ambiguous entities, creating a *reduced mapping dictionary* where each remaining location entity uniquely identifies exactly one town. In Figure 4 demonstrates this filtering process applied to the three representative cases from Figure 3



Figure 4: Reduced associated LOC named entities associated to three towns. There are no repetitions between classes.

We finally perform town prediction using this reduced mapping dictionary according to Equation 1, as previously described.

3.4.2. Contextual Prediction Sub-model $M_{tw,2}$

For documents without LOC-based predictions from $M_{tw,1}$ (described in Equation 1), the sub-model $M_{tw,2}$ employs a two-stage hierarchical classifier to address the 40-class imbalance in town labels tw_j . The architecture leverages both textual T_j and regional R_j features through:

1. Region Classification:

- Input: BERT embeddings of document text T_j

- Task: Predict region R_j
 - Output: Predicted Region R_j
2. Town Classification: There are 12 regional classification models, each one consists of
- Input: BERT embeddings of document text T_j
 - Task: Predict town tw_j within predicted region R_j
 - Output: Predicted town tw_j

This hierarchical region-town approach provides three key advantages over direct 40-class town classification, as evidenced by the distribution patterns in Figures 5 and 6: Hierarchical Class Separation, Imbalance Mitigation, and Regional Specialization. Each of these advantages is examined below:

- **Hierarchical Class Separation:** The two-stage architecture decomposes the original 40-class problem into more manageable sub-tasks. As visible in Figure 5, regional grouping naturally group towns with linguistically similar vocabulary, while Figure 6 reveals cases where single-town dominance (e.g., Tulum in Quintana Roo) effectively reduces the classification task to regional prediction.
- **Imbalance Mitigation:** The 62:1 imbalance ratio at town level (Figure 5a) is alleviated by grouping towns into regions, where:
 - 8 regions contain ≤ 3 towns (Figure 6).
 - Maximum regional imbalance drops to 9:1.
 - 6 regions become binary near-balanced classifiers.
- **Regional Specialization:** Each regional classifier adapts to local linguistic patterns, avoiding the noise from nationally dominant tourist vocabularies that would bias a flat classifier.

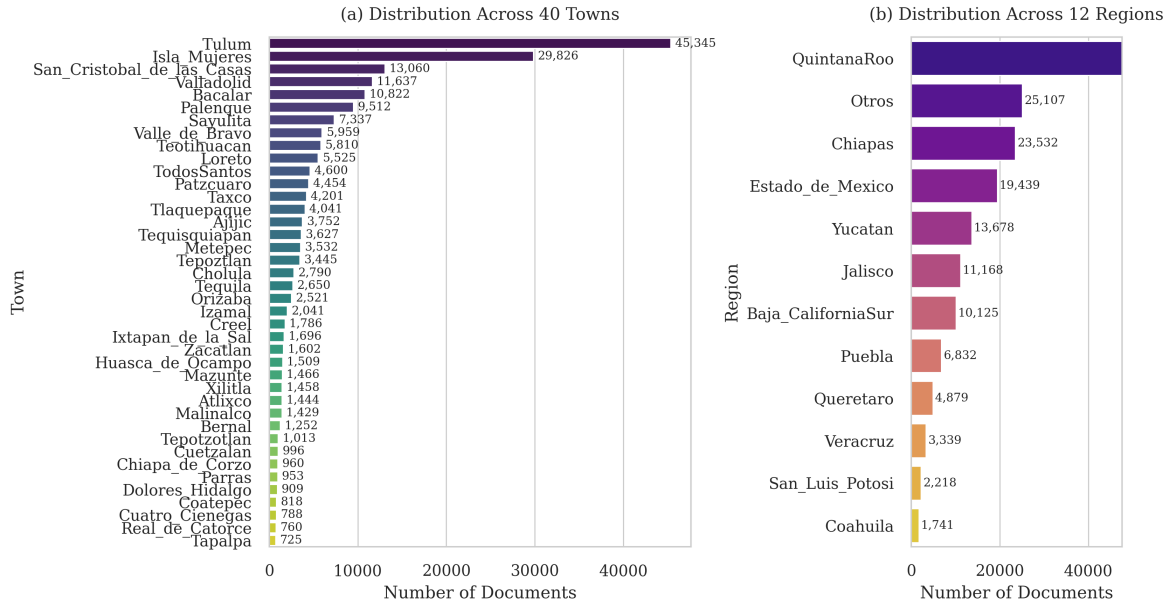


Figure 5: Document distribution showing (a) extreme imbalance across 40 towns (top 2 towns comprise 36% of data) versus (b) relatively balanced regional groupings (top region comprises 41% of data, with 6 regions between 5-12% each). This approach reduces the maximum class imbalance ratio from 62:1 to 49:1.

Both the region and town classifiers are SVM classifiers, with Gaussian Kernel, taking BERT embeddings as inputs.

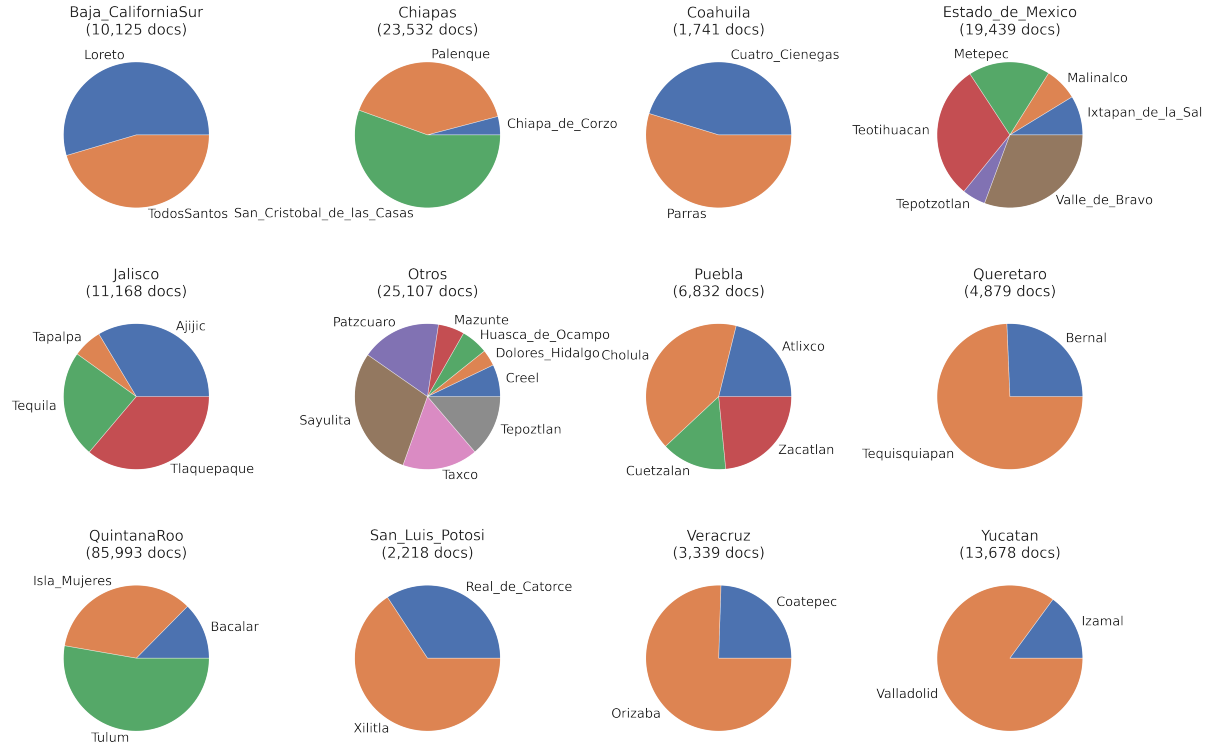


Figure 6: Distribution of towns across regions. Each pie chart shows the relative proportion of towns within a region, with total document countings indicated below the region name. It is worth noting the distribution patterns, from regions dominated by a single town (Quintana Roo) to more balanced distributions (Estado de México).

4. Experiments

The experimental evaluation assessed model performance across all three classification tasks (polarity, type, and town), with Weighted macro F1-score as the primary metric to address class imbalance. All experiments were conducted on a computational node equipped with an NVIDIA L4 GPU (24GB VRAM) using CUDA 12.2, with PyTorch 1.13.1 and Hugging Face Transformers 4.26.1 libraries. The hardware-software configuration enabled efficient parallel execution, particularly for the Qwen3-0.6B fine-tuning, which required almost 5 hours of training time.

We adopted a 4:1 train-validation split, stratified by the respective labels, to maintain distributional consistency. All models were trained on the training subset, with hyperparameters as detailed in Section 3. These hyperparameters were chosen using Grid Search. To ensure reproducibility, we fixed the same random seeds across all experiments.

The fine-tuned Qwen3-0.6B model M_p achieved 73.3% weighted F1-score on the validation dataset, demonstrating robust performance despite severe class imbalance. The type model M_{ty} achieved 95.16% F1-score. Finally, the model M_{tw} achieved 78% weighted F1-score on the validation dataset across all town labels.

5. Results and Discussion

Our model obtained the results detailed in Table 2 and 7. It ranked almost in the 70th percentile. The best performing model was the town model M_{tw} .

As observed in Table 3, our hierarchical model M_{tw} demonstrated particularly solid performance in the town classification task, achieving a Macro F1 of 0.6006 that positions it in the 75th percentile. The detailed analysis by town reveals that the model reached percentiles above 80% in multiple towns, particularly excelling in towns such as Tepoztlán, Tequila, Cuetzalan, and Tapalpa (88th percentile).

Table 2

Performance of our method in the REST-MEX 2025 challenge, including percentile rankings.

Metric	Score	Percentile (%)
Track Score	0.6158	69.57
Macro F1 (Polarity)	0.4753	47.83
Macro F1 (Type)	0.9423	49.28
Macro F1 (Town)	0.6006	75.36

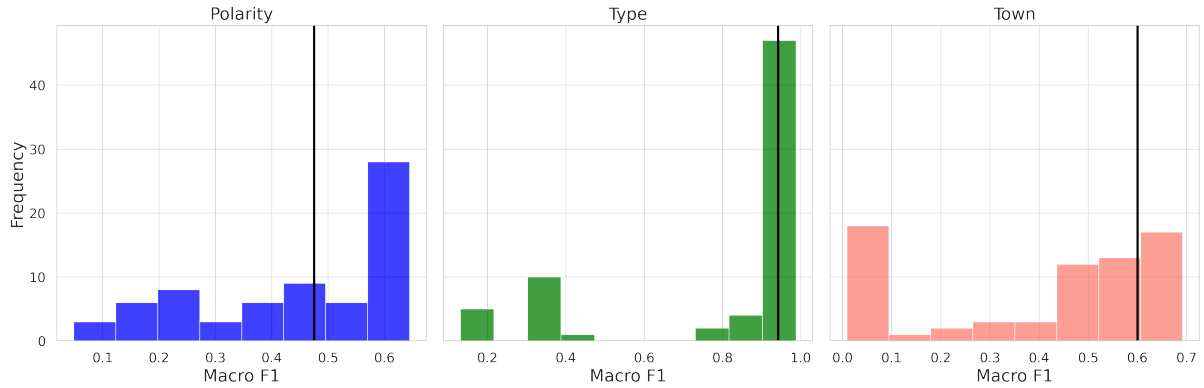


Figure 7: Distribution of Macro F1 scores for the three tasks across all submitted runs. The vertical black line in each subplot highlights the performance of our model.

This consistency in performance across different Mexican towns suggests that the model successfully captured the distinctive linguistic characteristics associated with each town.

Table 3

Percentile rankings in each town.

Town	Percentile (%)	Pueblo	Percentile (%)
tulum	76.81	ajijic	73.91
isla_mujeres	75.36	tequisquiapan	76.81
san_cristobal_de_las_casas	69.57	metepec	73.91
valladolid	75.36	tepoztlán	88.41
bacalar	84.06	cholula	86.96
palenque	73.91	tequila	88.41
sayulita	76.81	orizaba	86.96
valle_de_bravo	68.12	izamal	75.36
teotihuacan	69.57	creel	72.46
loreto	75.36	ixtapan_de_la_sal	73.91
todossantos	68.12	cuetzalan	88.41
patzcuaro	86.96	chiapa_de_corzo	65.22
taxco	86.96	parras	62.32
tlaquepaque	63.77	dolores_hidalgo	75.36
zacatlán	76.81	coatepec	88.41
huasca_de_ocampo	78.26	cuatro_cienegas	63.77
mazunte	71.01	real_de_catorce	85.51
xilitla	75.36	tapalpa	88.41
atlixco	76.81	bernal	69.57
malinalco	50.72	tepotzotlán	66.67

6. Conclusions

In this paper, we proposed a parallel NLP pipeline for sentiment and multi-label classification of Spanish tourism reviews from Mexican Magical Towns, as part of the Rest-Mex 2025 shared task. Our architecture combined three specialized models: (1) a fine-tuned Qwen3-0.6B transformer for 5-class sentiment analysis, (2) a TF-IDF-based logistic regression classifier for destination type prediction, and (3) a Bert-based NER-enhanced hierarchical model for town identification. This modular approach achieved competitive results, ranking in the 69.57th percentile overall, with particularly strong performance in town classification (75.36th percentile).

Our key technical contribution is a NER-augmented hierarchical classifier that achieved 60% macro F1-score by combining precise location entity recognition with contextual BERT embeddings, demonstrating effectiveness for geographically imbalanced datasets (62:1 class ratio).

The town model's consistent performance across locations (reaching 88th percentile in Tepoztlán, Tequila, and Cuetzalan) suggests successful capture of region-specific linguistic patterns. However, the polarity model's lower percentile ranking (47.53%) reveals persistent challenges in fine-grained sentiment analysis for tourism reviews, likely due to subjective labeling and cultural nuances.

Future work should investigate strategies to better leverage regional information for grouping towns in ways that enhance classifier performance. Additionally, expanding the scope of entities considered in the analysis, as well as incorporating town-specific key terms and linguistic markers, could further improve classification accuracy. Our results confirm that hybrid architectures provide a robust and effective framework for tourism-oriented NLP applications, especially when computational efficiency and model transparency are critical requirements.

Acknowledgments

The authors acknowledge the High-Performance Computing Area (Área de Cómputo de Alto Rendimiento, ACARUS) of the University of Sonora for providing the supercomputing infrastructure essential to this research. <https://acarus.unison.mx>

Declaration on Generative AI

The author(s) have not employed any Generative AI tools.

References

- [1] S. Afren, The role of digital marketing promoting tourism business. a study of use of the social media in prompting travel., *World Journal of Advanced Research and Reviews* 21 (2024) 272–287.
- [2] A. Diaz-Pacheco, M. A. Álvarez-Carmona, A. Y. Rodríguez-González, H. Carlos, R. Aranda, Measuring the difference between pictures from controlled and uncontrolled sources to promote a destination. a deep learning approach, *International Journal of Interactive Multimedia and Artificial Intelligence In Press* (2023) 1–14. URL: <http://dx.doi.org/10.9781/ijimai.2023.10.003>. doi:10.9781/ijimai.2023.10.003.
- [3] E. Olmos-Martínez, M. Á. Álvarez-Carmona, R. Aranda, A. Díaz-Pacheco, What does the media tell us about a destination? the cancan case, seen from the usa, canada, and mexico, *International Journal of Tourism Cities* 10 (2023) 639–661. URL: <http://dx.doi.org/10.1108/IJTC-09-2022-0223>. doi:10.1108/ijtc-09-2022-0223.
- [4] R. Guerrero-Rodríguez, M. A. Álvarez-Carmona, R. Aranda, et al., Big data analytics of online news to explore destination image using a comprehensive deep-learning approach: a case from mexico, *Information Technology & Tourism* 26 (2024) 147–182. URL: <https://doi.org/10.1007/s40558-023-00278-5>. doi:10.1007/s40558-023-00278-5.

- [5] Secretaría de Turismo, Sistema de información estadística del sector turismo de México (datatur), <https://www.datatur.sectur.gob.mx/RAT/Forms/AllItems.aspx>, 2024.
- [6] United Nations World Tourism Organization, World Tourism Barometer, Report, UNWTO, 2025. URL: https://pre-webunwto.s3.eu-west-1.amazonaws.com/s3fs-public/2025-01/Barom_PPT_Jan_2025.pdf, accessed: October 10, 2025.
- [7] United Nations World Tourism Organization, International tourism recovers pre-pandemic levels in 2024, <https://www.unwto.org/news/international-tourism-recovers-pre-pandemic-levels-in-2024>, 2025. Accessed: October 10, 2025.
- [8] 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). doi:<https://doi.org/10.13053/CyS-26-2-4055>.
- [9] M. A. 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, Ángel 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. URL: <https://www.sciencedirect.com/science/article/pii/S1319157822003615>. doi:<https://doi.org/10.1016/j.jksuci.2022.10.010>.
- [10] R. Guerrero-Rodríguez, M. A. 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. URL: <https://doi.org/10.1080/13683500.2021.2007227>. doi:10.1080/13683500.2021.2007227. arXiv:<https://doi.org/10.1080/13683500.2021.2007227>.
- [11] 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. URL: <http://dx.doi.org/10.1016/j.jksuci.2023.101746>. doi:10.1016/j.jksuci.2023.101746.
- [12] A. Díaz-Pacheco, M. A. 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* 0 (2022) 1–31. doi:10.1080/0952813X.2022.2153276.
- [13] M. V. Mäntylä, D. Graziotin, M. Kuuttila, The evolution of sentiment analysis—a review of research topics, venues, and top cited papers, *Computer Science Review* 27 (2018) 16–32. URL: <https://www.sciencedirect.com/science/article/pii/S1574013717300606>. doi:<https://doi.org/10.1016/j.cosrev.2017.10.002>.
- [14] S. Baccianella, A. Esuli, F. Sebastiani, et al., Sentiwordnet 3.0: an enhanced lexical resource for sentiment analysis and opinion mining., in: *Lrec*, volume 10, Valletta, 2010, pp. 2200–2204.
- [15] G. Gautam, D. Yadav, Sentiment analysis of twitter data using machine learning approaches and semantic analysis, in: 2014 Seventh International Conference on Contemporary Computing (IC3), 2014, pp. 437–442. doi:10.1109/IC3.2014.6897213.
- [16] N. Zainuddin, A. Selamat, R. Ibrahim, Hybrid sentiment classification on twitter aspect-based sentiment analysis, *Applied Intelligence* 48 (2018) 1218–1232. doi:10.1007/s10489-017-1098-6, published online: 13 December 2017.
- [17] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, Bert: Pre-training of deep bidirectional transformers for language understanding, in: *Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: Human Language Technologies*, volume 1, 2019, pp. 4171–4186.
- [18] A. Conneau, K. Khandelwal, N. Goyal, V. Chaudhary, G. Wenzek, F. Guzmán, E. Grave, M. Ott, L. Zettlemoyer, V. Stoyanov, Unsupervised cross-lingual representation learning at scale, 2020. URL: <https://arxiv.org/abs/1911.02116>. arXiv:1911.02116.
- [19] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, V. Stoyanov,

- Roberta: A robustly optimized bert pretraining approach, 2019. URL: <https://arxiv.org/abs/1907.11692>. arXiv:1907.11692.
- [20] A. Yang, A. Li, B. Yang, B. Zhang, B. Hui, B. Zheng, B. Yu, C. Gao, C. Huang, C. Lv, C. Zheng, D. Liu, F. Zhou, F. Huang, F. Hu, H. Ge, H. Wei, H. Lin, J. Tang, J. Yang, J. Tu, J. Zhang, J. Yang, J. Yang, J. Zhou, J. Zhou, J. Lin, K. Dang, K. Bao, K. Yang, L. Yu, L. Deng, M. Li, M. Xue, M. Li, P. Zhang, P. Wang, Q. Zhu, R. Men, R. Gao, S. Liu, S. Luo, T. Li, T. Tang, W. Yin, X. Ren, X. Wang, X. Zhang, X. Ren, Y. Fan, Y. Su, Y. Zhang, Y. Zhang, Y. Wan, Y. Liu, Z. Wang, Z. Cui, Z. Zhang, Z. Zhou, Z. Qiu, Qwen3 technical report, 2025. URL: <https://arxiv.org/abs/2505.09388>. arXiv:2505.09388.
 - [21] M. Toledo-Acosta, T. Barreiro, A. Reig-Alamillo, M. Müller, F. Aroca Bisquert, M. L. Barrigon, E. Baca-Garcia, J. Hermosillo-Valadez, Cognitive emotional embedded representations of text to predict suicidal ideation and psychiatric symptoms, *Mathematics* 8 (2020). URL: <https://www.mdpi.com/2227-7390/8/11/2088>. doi:10.3390/math8112088.
 - [22] B. Shi, J. Zhao, K. Xu, A word2vec model for sentiment analysis of weibo, in: 2019 16th International Conference on Service Systems and Service Management (ICSSSM), 2019, pp. 1–6. doi:10.1109/ICSSSM.2019.8887652.
 - [23] S. Anis, S. Saad, M. Aref, Sentiment analysis of hotel reviews using machine learning techniques, in: A. E. Hassanien, A. Slowik, V. Snášel, H. El-Deeb, F. M. Tolba (Eds.), *Proceedings of the International Conference on Advanced Intelligent Systems and Informatics 2020*, Springer International Publishing, Cham, 2021, pp. 227–234.
 - [24] B. Cowan, S. Zethelius, B. Luk, T. Baras, P. Ukarde, D. Zhang, Named entity recognition in travel-related search queries, *Proceedings of the AAAI Conference on Artificial Intelligence* 29 (2015) 3935–3941. URL: <https://ojs.aaai.org/index.php/AAAI/article/view/19050>. doi:10.1609/aaai.v29i2.19050.
 - [25] D. H. Fudholi, A. Zahra, S. Rani, S. N. Huda, I. V. Paputungan, Z. Zukhri, Bert-based tourism named entity recognition: making use of social media for travel recommendations, *PeerJ Computer Science* 9 (2023) e1731. doi:10.7717/peerj-cs.1731.
 - [26] S. Ruder, An overview of multi-task learning in deep neural networks, 2017. URL: <https://arxiv.org/abs/1706.05098>. arXiv:1706.05098.
 - [27] J. Á. González-Barba, L. Chiruzzo, S. M. Jiménez-Zafra, Overview of IberLEF 2025: Natural Language Processing Challenges for Spanish and other Iberian Languages, in: *Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2025)*, co-located with the 41st Conference of the Spanish Society for Natural Language Processing (SEPLN 2025), CEUR-WS. org, 2025.
 - [28] 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, volume 75, 2025.
 - [29] M. Buda, A. Maki, M. A. Mazurowski, A systematic study of the class imbalance problem in convolutional neural networks, *Neural Networks* 106 (2018) 249–259. URL: <https://www.sciencedirect.com/science/article/pii/S0893608018302107>. doi:<https://doi.org/10.1016/j.neunet.2018.07.011>.
 - [30] S. Das, S. S. Mullick, I. Zelinka, On supervised class-imbalanced learning: An updated perspective and some key challenges, *IEEE Transactions on Artificial Intelligence* 3 (2022) 973–993. doi:10.1109/TAI.2022.3160658.
 - [31] 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. Rodríguez-González, Overview of rest-mex at iberlef 2021: Recommendation system for text mexican tourism, *Procesamiento del Lenguaje Natural* 67 (2021). doi:<https://doi.org/10.26342/2021-67-14>.
 - [32] 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).
 - [33] M. Á. Álvarez-Carmona, Á. Díaz-Pacheco, R. Aranda, A. Y. Rodríguez-González, L. Bustio-Martínez, V. Muñis-Sánchez, A. P. Pastor-López, F. Sánchez-Vega, Overview of rest-mex at iberlef 2023:

Research on sentiment analysis task for mexican tourist texts, *Procesamiento del Lenguaje Natural* 71 (2023).

- [34] M. Á. Álvarez-Carmona, R. Aranda, A. Y. Rodríguez-González, L. Pellegrin, C. Hugo, Classifying the mexican epidemiological semaphore colour from the covid-19 text spanish news, *Journal of Information Science* (2022). doi:<https://doi.org/10.1177/016555515221100952>.