

Hierarchical Attention Networks for Multilabel Sentiment Analysis in Spanish Reviews of Mexican Magic Towns

Ezau Faridh Torres Torres^{1,2,*}, Edison David Serrano Cárdenas^{1,2}

¹Mathematics Research Center (CIMAT), Jalisco S/N, Valenciana, 36023 Guanajuato, GTO México.

²Secretaría de Ciencia, Humanidades, Tecnología e Innovación (SECIHTI), Av. de los Insurgentes Sur 1582, Benito Juárez, 03940, CDMX, México.

Abstract

In this study, we address the task of joint classification of sentiment polarity, destination type, and Magic Town from Spanish-language tourist reviews, as proposed by the Rest-Mex 2025 challenge, part of the IberLEF 2025 evaluation forum [1, 2]. We hereby propose a lightweight, hierarchical attention-based model that leverages the inherent structure of review data. Specifically, the model posits that words form reviews, and reviews are grouped by town. The proposed approach is predicated on the Hierarchical Attention Network (HAN), with extensions to a multi-output setting that are concomitant with the preservation of efficiency and interpretability. Experimental results show that our model significantly outperforms the official baseline and the average of all participating systems, achieving a joint track score of 0.601, with macro-F1 scores of 0.541 for polarity, 0.943 for destination type, and 0.528 for town classification. It is noteworthy that the model demonstrates robust generalization capabilities, particularly in underrepresented classes and low-resource towns. In contrast to large-scale transformer-based models, our architecture can be trained efficiently without the use of GPUs or extensive pretraining, making it suitable for deployment in environments with limited resources.

Keywords

REST-MEX 2025, Sentiment Analysis, Hierarchical Attention Network (HAN), Review Classification, Multitask Learning, Spanish NLP.

Introduction

Conventional review classification methodologies employ sparse lexical features, such as n -grams, as review representations, subsequently utilizing either a linear model or kernel methods on these representations [3, 4]. Recent approaches have incorporated deep learning techniques, particularly transformer-based architectures, such as BERT [5], RoBERTa [6], and its multilingual extension XLM-R [7]. For Spanish-language applications, specialized models like RoBERTuito [8] have demonstrated competitive performance in review classification and sentiment analysis tasks.

This study was conducted in the context of the Rest-Mex 2025 challenge [1], part of the IberLEF 2025 evaluation forum [2]. Unlike past editions [9, 10, 11], this year focuses on the joint classification of tourist review sentiment, destination type, and Magic Town identification from Spanish-language texts.

To address this task, we adapted the Hierarchical Attention Network (HAN) proposed by Yang et al. [12] to our classification task. In contrast to the original document-level configuration, we categorize reviews by Magic Town, implementing an approach that encompasses both the word and review levels.

Recent efforts have explored the use of compact transformer models such as DistilBERT [13] to mitigate the computational demands of full-sized language models like BERT and RoBERTa. While these approaches reduce parameter counts, they still rely on large-scale pretraining and often require GPU-intensive resources during fine-tuning and inference. In contrast, our proposed model leverages a lightweight, interpretable hierarchical architecture based on bidirectional GRUs and attention mechanisms. This design enables efficient training and robust performance, even in resource-constrained environments, without the need for massive pretrained language models.

IberLEF 2025, September 2025, Zaragoza, Spain

*Corresponding author.

✉ ezau.torres@cimat.mx (E. F. Torres Torres); edison.serrano@cimat.mx (E. D. Serrano Cárdenas)

🌐 <https://github.com/ezautorres> (E. F. Torres Torres); <https://github.com/edserranoc> (E. D. Serrano Cárdenas)

🆔 0009-0005-6094-9714 (E. F. Torres Torres); 0009-0002-4709-465X (E. D. Serrano Cárdenas)



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

1. Problem Statement and Task Objectives

This section outlines the challenges addressed in the “Rest Mex 2025” sentiment analysis competition, which centers on understanding tourist perceptions through the analysis of user-generated reviews [14, 15, 16].

1.1. Task

The core objective of the competition is to analyze reviews and opinions related to tourist destinations in Mexico and extract relevant sentiment-based and categorical information. Specifically, the task is structured around three main classification goals:

- **Sentiment Polarity Classification:** Given a textual review, the first objective is to predict its sentiment on a five-point scale, where a score of 1 denotes strong dissatisfaction and a score of 5 indicates high satisfaction. This allows for a nuanced understanding of tourist sentiment beyond binary positive/negative classifications.
- **Tourist Site Type Classification:** The second goal is to categorize the type of site referenced in the review. The possible categories are *hotel*, *restaurant*, or *attraction*. This classification aids in differentiating sentiment patterns across various tourism sectors.
- **Magic Town Identification:** The final objective is to determine which of the 40 designated Mexican Magic Town the review refers to. By leveraging metadata and contextual information within the text, this classification can reveal regional trends and cultural variations in how tourists express sentiment.

Each review serves as a valuable source of information about traveler experiences. The overarching aim is to extract structured and meaningful insights from these texts to better understand tourist satisfaction, preferences, and regional differences across Mexico’s culturally rich destinations.



Figure 1: The choropleth map illustrates the number of tourist reviews per Mexican state in the Rest-Mex 2025 dataset. The state of Quintana Roo has emerged as a prominent player in the tourism sector, owing to the prevalence of destinations such as Tulum and Isla Mujeres.

2. Methodology

2.1. Exploratory Data Analysis

The dataset provided for the competition consists of 208,051 labeled reviews from tourists sharing their experiences across various destinations. Each review contains five key fields: a title, the main review text, the type of place being reviewed, the sentiment polarity on a scale from 1 to 5, and the name of one of 40 towns in Mexico. Notably, there was only one instance with a missing title, which was handled during preprocessing.

A critical part of exploratory analysis involved assessing the class distribution to determine whether the dataset was balanced. Balanced datasets are essential for ensuring that the model can learn equally well across all classes. In contrast, imbalanced datasets can lead to biased predictions, especially for minority classes that are underrepresented during training.

As shown in Figure 2, the sentiment polarity distribution is heavily skewed toward the most positive sentiment. Reviews with a polarity score of 5 account for approximately 65.64% of the dataset, indicating a strong bias toward favorable opinions. This imbalance could potentially affect the model's ability to accurately predict lower sentiment scores, as it is exposed to far fewer negative or neutral examples.

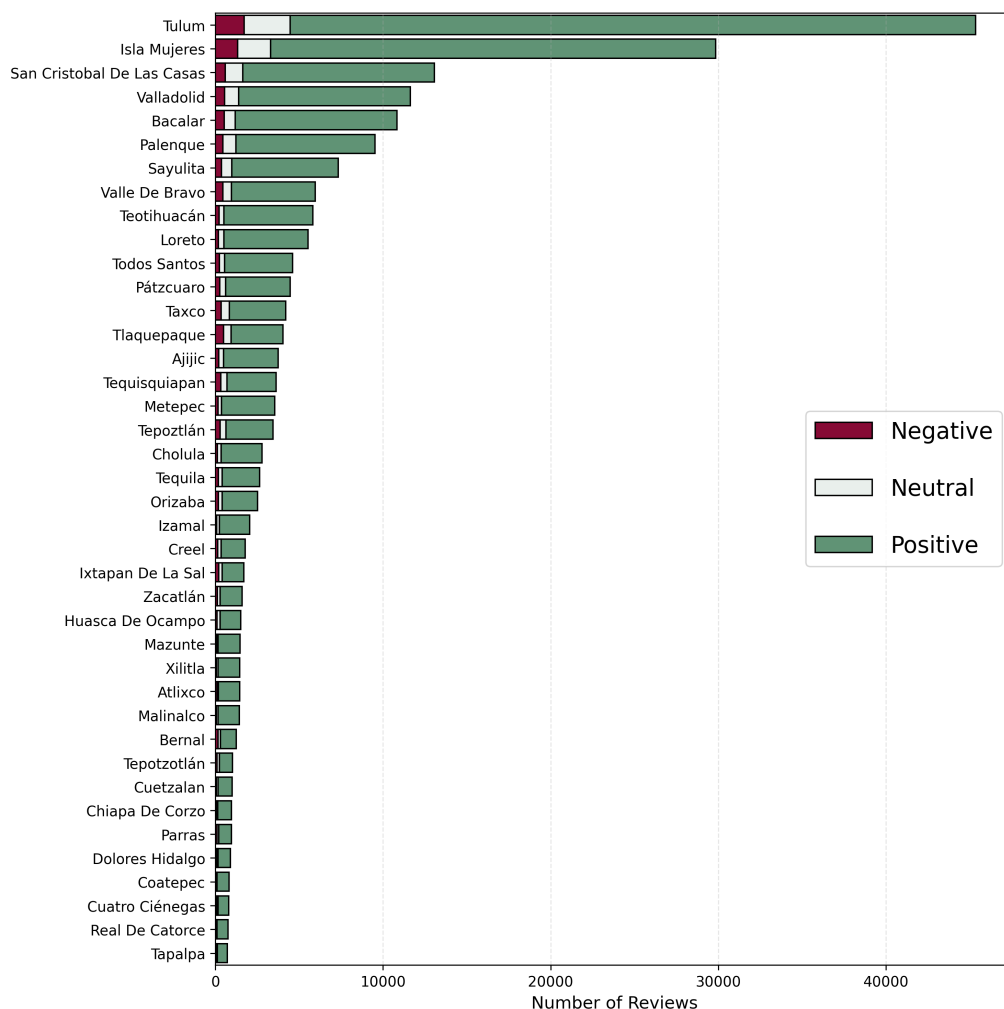


Figure 2: The distribution of tourist reviews by Magic Town has been grouped according to sentiment polarity. The data reveal a strong skew toward positive reviews across all towns, with Tulum and Isla Mujeres being the most represented destinations by volume.

Similarly, the distribution of reviews across towns is notably uneven. As illustrated in Figure 1, approximately 41.33% of the reviews come from towns located in Quintana Roo, indicating a regional

concentration of tourist activity or review behavior. This imbalance may impact the model’s performance on towns with fewer samples, as the model could become biased toward the characteristics of more frequently represented locations.

In contrast, the distribution of the type of place (i.e., attraction, hotel, or restaurant) is more balanced, though not perfectly uniform. Specifically, restaurants are the most frequently reviewed category, representing 41.68% of the dataset, followed by hotels with 24.71%, and attractions, which are significantly underrepresented at only 0.33%. While this indicates a moderate imbalance, the disparity is less pronounced than in the sentiment and town distributions.

2.2. Dataset Construction

The dataset preparation phase was crucial to ensure compatibility of each instance with the requirements of the sentiment analysis task. Categorical values in key columns were transformed into numerical representations. Specifically, the Type column was encoded as follows: 0 for Attraction, 1 for Hotel, and 2 for Restaurant. Similarly, the Town column was mapped to integers ranging from 0 to 39, based on the alphabetical order of the 40 magic towns included in the dataset.

To better capture both the context and summary of each review, the title and review columns were concatenated into a single text field. In cases where the title was missing, it was replaced with an empty string. This merged column helps retain both the concise summary typically found in the title and the detailed feedback from the full review.

Preprocessing steps included converting all text to lowercase and performing cleaning operations to correct patterns that attempted to represent tildes (e.g., replacing corrupted or miswritten characters common in Spanish). Given the importance of emoticons in expressing sentiment, various facial expressions were replaced with corresponding descriptive words. For example, happy faces such as “:)” and “:))” were replaced with “*feliz*” (happy), while sad expressions like “:(” were substituted with “*triste*” (sad). A broader set of emoticons was considered in this substitution process to better preserve the emotional content of the reviews. Additionally, common stopwords and non-informative symbols were removed to reduce noise and improve model focus on relevant content.

The dataset was shuffled to ensure a randomized distribution of instances. For the experimental setup, it was divided into training and test subsets: 80% of the data (166,440 instances) was allocated for training, while the remaining 20% (41,611 instances) was reserved for evaluation.

To address class imbalance during training—particularly for underrepresented sentiment classes (those with polarity ≤ 3) and towns with fewer than 2,800 reviews—a contextual data augmentation strategy was applied. This involved in-place token repetition, where a randomly selected segment of five consecutive tokens from the original review was appended to the end of the text. This technique aimed to enhance the context for minority classes while preserving the natural semantics of the reviews.

For textual representation, we employed pretrained Word2Vec embedding [17], which provided semantic-rich vector representations of the Spanish text to feed into our model.

2.3. Hierarchical attention networks

In this work, we build on the idea that improved review representations can be obtained by incorporating knowledge of review structure into a hierarchical model architecture. The fundamental intuition informing our model is that not all components of a review or the collection of reviews linked to a given Magic Town are equally pertinent to the classification task, and that determining the relevant sections entails modeling interactions among words, rather than considering their isolated presence. This idea builds upon prior work on attention-based models [12].

The objective of the Hierarchical Attention Network (HAN) is to glean two fundamental insights concerning the structural intricacies of review data. First, given the hierarchical nature of our dataset—where individual words form reviews, and reviews are grouped by Magic Town—a final representation for each town is constructed by first encoding each review and then aggregating them into a comprehensive representation at the town level. Secondly, it has been observed that the informative value of words

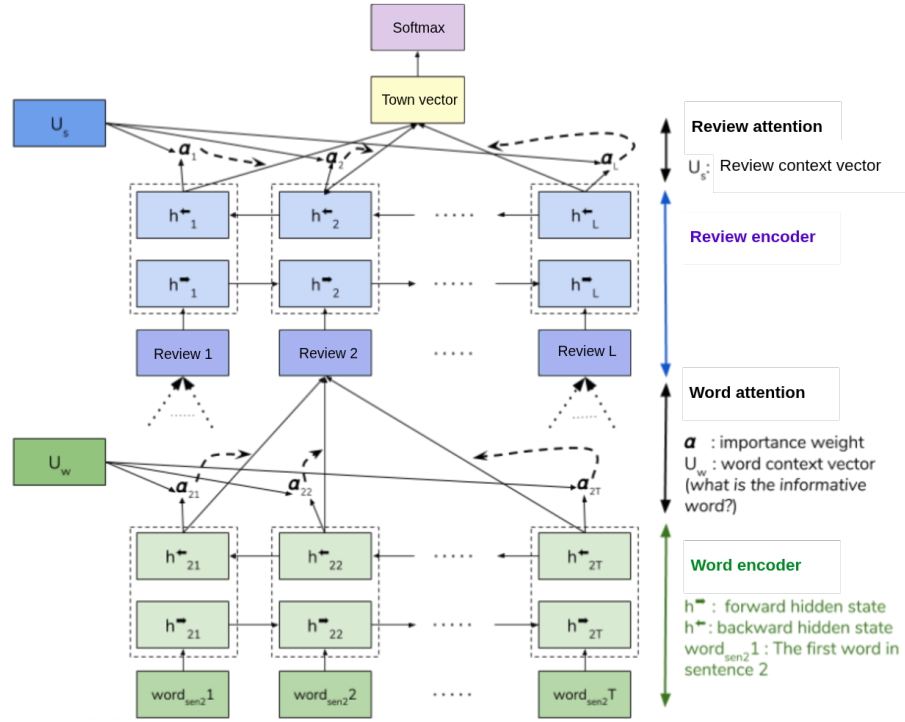


Figure 3: Architecture of a Hierarchical Attention Network (HAN), showing the word and sentence encoders followed by attention layers. Redrawn and adapted from a blog illustration.

within reviews, as well as reviews within a town, varies depending on the classification task. In particular, the importance of specific words or reviews may change depending on their context [12]. To account for this, the model employs two levels of attention mechanisms [18, 19]: one at the word level and another at the review level. These mechanisms allow the model to dynamically adjust the weight assigned to each component during the construction of the town-level representation.

pork belly = delicious . || scallops? || I don't even
like scallops, and these were a-m-a-z-i-n-g . || fun
and tasty cocktails. || next time I in Phoenix, I will
go back here. || Highly recommend.

Figure 4: Example from Yang et al. [12] illustrating word and sentence-level attention in review classification. The review, drawn from the Yelp 2013 dataset, consists of five sentences, each delimited by a period or a question mark. The first and last sentences are particularly expressive, with terms like “delicious” and “a-m-a-z-i-n-g” contributing prominently to the sentiment conveyed. Licensed under CC BY 4.0.

To illustrate this phenomenon, consider the example in Figure 4, which is a brief Yelp review wherein the objective is to predict the rating on a scale from 1 – 5 [20]. Intuitively, the first and third sentences appear to contain more pertinent information for predicting the rating. Within these sentences, the words “delicious” and “a-m-a-z-i-n-g”, which are frequently used to express a positive sentiment, contributes to the implication of the positive attitude contained in this review. Attention confers two advantages: it frequently leads to enhanced performance and it offers insight into which words and sentences contribute to the classification decision. This insight can be valuable in applications and analysis [4, 21].

The overall architecture of the Hierarchical Attention Network (HAN) is illustrated in Figure 3. The model is composed of multiple components, including a word sequence encoder, a word-level attention layer, a review encoder, and a review-level attention layer.

Our approach aligns with the hierarchical attention formulation proposed by Yang et al. [12], which involves the acquisition of both word-level and review-level representations through a learned attention mechanism:

$$\alpha_t = \frac{\exp(u_t^\top u_w)}{\sum_j \exp(u_j^\top u_w)}, \quad s = \sum_t \alpha_t h_t. \quad (1)$$

In our implementation, both word and review sequences are encoded using bidirectional Gated Recurrent Units (GRUs), which allow the model to capture contextual dependencies from both past and future tokens or reviews. At each level, the GRU outputs a sequence of hidden states h_t , which are then processed by an attention mechanism. Here, h_t denotes the hidden state at time step t , and u_w is a learned context vector used to compute the importance weights α_t via a softmax operation. The resulting vector s is a weighted sum that encodes the most informative parts of the sequence. This formulation is applied at two distinct levels: first, at the word level, to obtain review embeddings; and second, at the review level, to obtain town embeddings.

Unlike the original HAN designed for document classification, our model aggregates multiple reviews associated with a Magic Town, and predicts multiple outputs at different levels (per-review and per-town). We preserve the hierarchical attention structure but adapt it to a multitask scenario.

3. Results and discussion

3.1. Evaluation

The evaluation section is responsible for assessing the performance and effectiveness of the proposed model across three classification tasks: polarity (1 – 5), type (Attractive, Hotel, Restaurant), and Magic Town (40 classes). The results obtained by our best submission, *FrogCode_1*, are reported herein. *FrogCode_1* consistently outperformed the provided baseline across all metrics.

The evaluation process utilizes Macro-F1 scores for each task, aligning with the official metrics employed in the Rest-Mex 2025 competition. In addition to aggregate performance (Track Score), we analyze per-class F1, precision, and recall to gain further insights into the model’s behavior.

In order to provide a comprehensive evaluation of the results, comparative analyses with the official baseline and statistical summaries across all submissions are also included. These additional analyses are used to contextualize the performance in question.

3.2. Competition results

As illustrated in Table 1, the model’s final performance metrics are presented in comparison to the official baseline and the overall average across all participating systems in the Rest-Mex 2025 Sentiment Analysis track. The model demonstrated a Track Score of 0.601, which is significantly higher than the baseline (0.090) and the average participant score (0.484). It is noteworthy that the model achieved Macro F1 scores of 0.541, 0.943, and 0.528 for polarity, type, and town classification, respectively. These scores surpassed the mean scores and the baseline by a substantial margin. These results demonstrate the effectiveness of incorporating a hierarchical attention mechanism for modeling review structure, particularly in the town-level classification task, which exhibited the largest gap over the baseline (0.528 vs. 0.009).

Model	Track Score	Macro F1 (Polarity)	Macro F1 (Type)	Macro F1 (Town)
Baseline	0.090	0.158	0.197	0.009
Overall average	0.484	0.449	0.796	0.403
Our model	0.601	0.541	0.943	0.528

Table 1

Final results from Sentiment Analysis competition Rest-Mex 2025.

In addition to the aggregated metrics presented in Table 1, it was observed that there was consistent performance across all subtasks, including underrepresented and imbalanced classes. Notwithstanding the considerable polarity skew (with over 136,000 reviews rated as 5), the model exhibited competitive Macro F1 scores of 0.584 for polarity 1, 0.333 for polarity 2, and 0.483 for polarity 3. This finding indicates that the architecture did not merely adapt to the majority class, but rather, it effectively captured informative patterns across the sentiment spectrum.

In the context of the type classification task, the model demonstrated notable precision and generalization capabilities, achieving F1 scores of 0.95, 0.93, and 0.95 for the categories "Attractive," "Hotel," and "Restaurant," respectively, even when confronted with varying data volumes.

In addition, the model's notable robustness was demonstrated in the town classification task, with F1 scores exceeding 0.73 for towns with substantial data, such as Isla Mujeres, Teotihuacan, and Tulum. Notably, the model demonstrated notable performance in low-resource towns, as evidenced by its F1 values of 0.227 in Tepotzotlan, 0.345 in Cuetzalan, and 0.401 in Metepec and as we have seen in Seccion 2.1. This finding underscores the model's capacity for generalization across long-tail distributions.

3.3. Error Analysis and Challenging Cases

An analysis of the confusion matrix on the validation set reveals notable challenges in classifying certain towns. Tepotzotlán, Cuetzalan, and Metepec exhibit high misclassification rates. Tepotzotlán shows a low number of correct predictions and widespread confusion with other towns, indicating weak feature representation. Cuetzalan's errors are broadly distributed across multiple classes, while Metepec, despite better accuracy, is frequently confused with semantically similar locations.

Moreover, misclassifications tend to follow regional patterns. Towns from the same state—such as Tepotzotlán, Metepec, and Valle de Bravo (State of Mexico)—are often confused. A similar trend is observed among geographically close towns like Valladolid (Yucatán), Tulum, and Isla Mujeres (Quintana Roo). This suggests that tourist reviews from nearby or culturally similar towns share linguistic features that make them harder to distinguish.

These findings highlight the need for improved handling of class imbalance and the incorporation of regional and semantic context to enhance model performance in town-level classification.

3.4. Comparative Discussion with Transformer Models

Although transformer-based architectures such as BERT and RoBERTa have demonstrated state-of-the-art performance in various NLP tasks, they often require substantial computational resources and extensive pretraining. In contrast, our hierarchical attention-based model achieves competitive results with significantly fewer parameters and without relying on transfer learning.

Recent studies have proposed lightweight transformer variants, such as DistilBERT [13] and TinyBERT [22], to alleviate some of these limitations. However, even these models entail GPU-intensive fine-tuning and storage of pretrained weights. Our model, by leveraging a hierarchical BiGRU-attention mechanism, delivers a joint Track Score of 0.601 and consistently strong macro-F1 scores, without incurring such computational costs.

This comparison highlights that HAN-based architectures remain viable for structured classification tasks, particularly in low-resource settings where transformer-based approaches may be impractical. Future research could explore hybridizing HANs with distilled transformers to balance interpretability, efficiency, and raw performance.

4. Conclusions

In this work, we proposed a hierarchical attention-based model tailored for the Rest-Mex 2025 challenge, addressing the joint classification of sentiment polarity, type of destination, and Magic Town based on Spanish-language tourist reviews. The model's efficacy in capturing multilevel contextual dependencies is attributable to its utilization of the hierarchical structure of the data, wherein words are classified

as reviews and reviews are aggregated by town. The dual attention mechanism enabled the model to dynamically focus on the most informative words and reviews, resulting in robust performance across all tasks.

The efficacy of our approach was found to exceed not only the official baseline but also the mean performance of the systems that were engaged in the study. This performance was evidenced by the attainment of a Track Score of 0.601 and the consistent generation of elevated Macro F1 scores across the entirety of the subtasks that were utilized in the evaluation process. Notably, the model demonstrated strong generalization on underrepresented classes and low-resource towns, showcasing its adaptability to real-world data imbalances.

Furthermore, the model's lightweight architecture offers practical advantages in real-world scenarios. It can be trained and deployed efficiently on standard hardware without reliance on GPU acceleration or large-scale pretraining. This makes it particularly well-suited for use in settings with limited computational resources, such as public sector institutions, small tourism operators, or NGOs working with regional data. Its interpretable design also facilitates model transparency, an important factor for decision-making in applied contexts.

These results underscore the efficacy of hierarchical architectures in multi-output text classification settings, particularly when confronted with nested, skewed, and multilingual datasets. Subsequent studies may augment this framework by incorporating multilingual pretraining or uncertainty estimation to enhance its robustness and interpretability.

The experimentation can be found in a public GitHub repository [23].

Declaration on Generative AI

We declare that the present manuscript has been written entirely by the authors and that no generative artificial intelligence tools were used in its preparation, drafting, or editing.

References

- [1] 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.
- [2] 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.
- [3] S. Wang, C. Manning, Baselines and bigrams: Simple, good sentiment and topic classification, in: H. Li, C.-Y. Lin, M. Osborne, G. G. Lee, J. C. Park (Eds.), Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), Association for Computational Linguistics, Jeju Island, Korea, 2012, pp. 90–94. URL: <https://aclanthology.org/P12-2018/>.
- [4] T. Joachims, Text categorization with support vector machines: Learning with many relevant features, in: C. Nédellec, C. Rouveirol (Eds.), Machine Learning: ECML-98, Springer Berlin Heidelberg, Berlin, Heidelberg, 1998, pp. 137–142.
- [5] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, BERT: Pre-training of deep bidirectional transformers for language understanding, in: J. Burstein, C. Doran, T. Solorio (Eds.), Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), Association for Computational Linguistics, Minneapolis, Minnesota, 2019, pp. 4171–4186. URL: <https://aclanthology.org/N19-1423/>. doi:10.18653/v1/N19-1423.
- [6] 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.
- [7] 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.
 - [8] J. M. Pérez, D. A. Furman, L. A. Alemany, F. Luque, Robertuito: a pre-trained language model for social media text in spanish, 2022. URL: <https://arxiv.org/abs/2111.09453>. arXiv:2111.09453.
 - [9] 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>.
 - [10] 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).
 - [11] M. Á. Álvarez-Carmona, Á. Díaz-Pacheco, R. Aranda, A. Y. Rodríguez-González, L. Bustio-Martínez, V. Muñiz-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).
 - [12] Z. Yang, D. Yang, C. Dyer, X. He, A. Smola, E. Hovy, Hierarchical attention networks for document classification, in: *NAACL 2016*, 2016, pp. 1480–1489. URL: <https://www.microsoft.com/en-us/research/publication/hierarchical-attention-networks-document-classification/>.
 - [13] V. Sanh, L. Debut, J. Chaumond, T. Wolf, Distilbert, a distilled version of bert: smaller, faster, cheaper and lighter, 2020. URL: <https://arxiv.org/abs/1910.01108>. arXiv:1910.01108.
 - [14] 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>.
 - [15] 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>.
 - [16] 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.
 - [17] T. Mikolov, K. Chen, G. Corrado, J. Dean, Efficient estimation of word representations in vector space, 2013. URL: <https://arxiv.org/abs/1301.3781>. arXiv:1301.3781.
 - [18] D. Bahdanau, K. Cho, Y. Bengio, Neural machine translation by jointly learning to align and translate, 2016. URL: <https://arxiv.org/abs/1409.0473>. arXiv:1409.0473.
 - [19] K. Xu, J. Ba, R. Kiros, K. Cho, A. Courville, R. Salakhutdinov, R. Zemel, Y. Bengio, Show, attend and tell: Neural image caption generation with visual attention, 2016. URL: <https://arxiv.org/abs/1502.03044>. arXiv:1502.03044.
 - [20] N/A, Yelp 2013, 2023. URL: <https://doi.org/10.5281/zenodo.7555898>. doi:10.5281/zenodo.7555898.
 - [21] J. Gao, P. Pantel, M. Gamon, X. He, L. Deng, Modeling interestingness with deep neural networks, in: A. Moschitti, B. Pang, W. Daelemans (Eds.), *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Association for Computational Linguistics, Doha, Qatar, 2014, pp. 2–13. URL: <https://aclanthology.org/D14-1002/>. doi:10.3115/v1/D14-1002.

- [22] X. Jiao, Y. Yin, L. Shang, X. Jiang, X. Chen, L. Li, F. Wang, Q. Liu, Tinybert: Distilling bert for natural language understanding, 2020. URL: <https://arxiv.org/abs/1909.10351>. arXiv:1909.10351.
- [23] E. Serrano, E. Torres, NLP_Rest_Mex2025, https://github.com/edserranoc/NLP_Rest_Mex2025, 2025. Accessed: 2025-05-21.