Place Classification and Sentiment Analysis in Reviews of Mexican Magical Towns Using LSTM Networks

Ana Laura Lezama-Sánchez[1,*,†], Mireya Tovar Vidal[2,†] and José Alejandro Reyes-Ortiz[3,†]

[1]*Secretaria de Ciencias, Humanidades, Tecnologia e Innovacion, Av. Insurgentes Sur 1582, Col. Credito Constructor, Demarcacion Territorial Benito Juarez. CP 03940, Ciudad de México.*

[2]*Benemerita Universidad Autonoma de Puebla, Av., 14 Sur Blvd., Puebla, ZIP Code: 72592, Mexico*

[3]*Universidad Autonoma Metropolitana, Unidad Azcapotzalco, Av. San Pablo Xalpa No. 180, Col. San Martín Xochinahuac, Alcaldia Azcapotzalco, C.P. 02128, Ciudad de Mexico, Mexico*

Abstract

This paper presents a methodology based on LSTM neural networks to perform sentiment analysis and classify tourist locations in written reviews about Mexico's Pueblos Mágicos. The natural language processing approach involves cleaning Spanish text, tokenizing it, and encoding both geographic and typological labels. The model uses a recurrent neural network architecture composed of embedding, LSTM, and dropout layers to predict three key variables from the reviews: sentiment polarity, the name of the Pueblo Mágico, and the category of the referenced tourist location. We trained and evaluated the model using a dataset provided by the Rest-Mex 2025 competition. The results demonstrate that LSTM networks have a strong potential to capture sequential dependencies in Spanish-language text.

Keywords

Sentiment Analysis, Mexican Magical Towns, Natural Language Processing, Tourism Analytics

1. Introduction

Mexico's Pueblos Mágicos represent a highly relevant tourist and cultural initiative by highlighting towns with unique historical, architectural, or natural attributes [1]. With the growing digitalization of tourism, visitors' online opinions have become a valuable source of information for understanding public perceptions of these destinations[2, 3, 4, 5]. In this context, automatic opinion analysis can provide meaningful insights for government institutions, researchers, and industry businesses [6, 7, 8].

This work proposes a methodology that combines Natural Language Processing (NLP) techniques and deep learning to analyze reviews written in Spanish and classify both the expressed sentiment and the location and type of tourist attraction mentioned. To achieve this, we use a neural network architecture based on Long Short-Term Memory (LSTM) [9, 10].

We applied the developed methodology to the dataset from the Rest-Mex 2025 competition [11, 12], which contains tourist reviews with varying tones and contents. Rest-Mex in a international shared task related to NLP over turist spanish texts [13, 14, 15]. After preprocessing, tokenization, and training, we built three independent models to predict opinion polarity, the type of the mentioned Pueblo Mágico, and the type of referenced location. The results confirm that deep neural networks can effectively extract valuable knowledge from unstructured data such as user comments.

IberLEF 2025, September 2025, Spain

*Corresponding author.

†These authors contributed equally.

analaura.lezama@correo.buap.mx (A. L. Lezama-Sánchez); mireya.tovar@correo.buap.mx (M. Tovar Vidal); jaro@azc.uam.mx (J. A. Reyes-Ortiz)

0000-0002-2505-5150 (A. L. Lezama-Sánchez); 0000-0002-9086-7446 (M. Tovar Vidal); 0000-0003-2111-4982 (J. A. Reyes-Ortiz)

Table 1
Summary of Training and Test Datasets from Rest-Mex 2025

Dataset	Description
Training Set	Size: 208,051 instances (70% of original dataset). Columns: – Title: Title assigned by tourist (Text). – Review: Full review text (Text). – Polarity: Sentiment polarity (1 to 5). – Town: Town referred in review (Text). – Region: Mexican state of the town (Text). Not used for classification. – Type: Place category (Hotel, Restaurant, Attractive).
Test Set	Size: 89,166 opinions (30% of original dataset). Columns: – ID: Unique identifier (Integer). – Title: Title assigned by tourist (Text). – Review: Full review text (Text). Contains data for polarity prediction.

2. Methodology

The methodology used Natural Language Processing (NLP) and deep learning techniques to address three simultaneous tasks: opinion polarity classification, identification of the mentioned Pueblo Mágico, and classification of the type of tourist place. For this purpose, we used a dataset of Spanish tourist reviews provided by the Rest-Mex 2025 competition [11, 12]. Table 1 summarizes the main characteristics of the training and test datasets used in this work, including the number of instances and the available variables in each.

In the text preprocessing, we applied several cleaning and normalization steps to prepare the data, including:

- Converting text to lowercase.
- Removing URLs, special characters, and punctuation marks.
- Removing Spanish stopwords to reduce lexical noise.

We applied these transformations to both the training and test datasets. On the other hand, for tokenization and sequencing, we employed Keras' Tokenizer class, configured to consider the 10,000 most frequent words. We assigned a special token for out-of-vocabulary words. We converted the cleaned texts into numerical sequences and normalized them using padding, ensuring all sequences had a fixed length of 100 tokens, which standardized input length for the model.

We then encoded the labels for the three tasks (polarity, town, and place type) into a numeric format using Scikit-learn's LabelEncoder, ensuring compatibility with the categorical loss functions used during training. We formulated each task as a multiclass classification problem. We developed three independent LSTM-based models, one for each task, with the following structure:

- Input layer receiving fixed-length sequences.
- Embedding layer converting each token into a 128-dimensional vector.
- LSTM layer with 64 units designed to capture temporal dependencies in the sequence.
- Dropout layer with a rate of 0.5 to prevent overfitting.
- Dense layer with `softmax` activation to predict the corresponding class.

Each model was compiled using the categorical cross-entropy loss function (categorical_crossentropy) and optimized them with the Adam algorithm. Each model was trained for two epochs with a batch size of 32, reserving 20% of the training data for validation. Preliminary tests showed that the validation loss stabilized quickly, suggesting that most of the learning occurred in the early stages of training.

Nevertheless, exploring longer training schedules remains a direction for future work. After training, each model generated predictions on the test set. We then post-processed the numerical codes back to their original labels (town names and place types). Finally, we exported the results to a text file in the format required by the competition, listing for each review the ID, polarity, Pueblo Mágico, and place type.

The task organizers evaluated the submitted predictions using standard metrics, including precision, recall, and F_1. They evaluated each of the three tasks independently using macro-averaged scores. The following describes how they evaluated each task:

- For polarity classification, which involves five sentiment classes ($C = 1, 2, 3, 4, 5$), the organizers computed the macro-averaged F_1 across all classes, as shown in Equation 1, where $F_i(k)$ represents the F_1 for class i produced by system k.
- For type prediction, which includes three categories (Attractive, Hotel, and Restaurant), the organizers calculated the average of the individual F_1 for each class, as expressed in Equation 2.
- For Magical Town (MT) classification, the organizers used a predefined list of Magical Towns (MTL). They computed the macro F_1 across all towns, following a similar approach to type prediction, as shown in Equation 3.

$$Res_{\mathrm{P}}(k) = \frac{\sum_{i \in C} F_i(k)}{|C|} \tag{1}$$

$$Res_{\mathrm{T}}(k) = \frac{F_A(k) + F_H(k) + F_R(k)}{3} \tag{2}$$

$$Res_{\mathrm{MT}}(k) = \frac{1}{|MTL|} \sum_{i=1}^{|MTL|} F_{\mathrm{MTL}_i}(k) \tag{3}$$

Finally, the organizers combined the scores of the three subtasks using different weights to compute each system's global score. Since they considered polarity and Magical Town classifications more critical, they assigned these tasks two and three times the weight of the type classification, respectively. Equation 4 shows how they computed the final score:

$$Sentiment(k) = \frac{2 \cdot Res_{\mathrm{P}}(k) + 3 \cdot Res_{\mathrm{MT}}(k) + Res_{\mathrm{T}}(k)}{6} \tag{4}$$

For greater clarity, Figure 1 presents the pseudocode summarizing the overall system flow, from data loading to the generation of the output file.

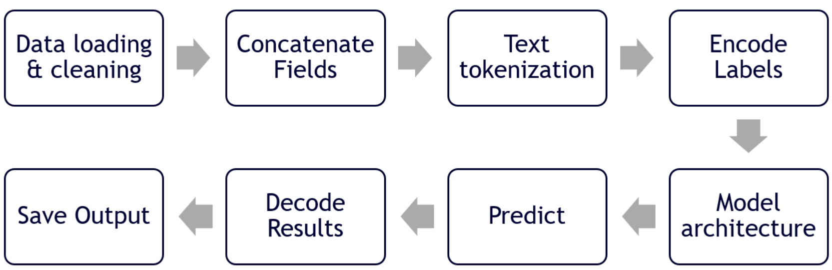

Figure 1: Preprocessing and Classification Pipeline

3. Results

The results were submitted under the team name BRIAN_1, which presented the performance of the proposed methodology only once. The team achieved 51st place in the overall results table with a global score of 0.3396 on the Track Score metric, as shown in Table 2. The evaluators based their assessment on multi-class classification metrics, specifically Macro F_1, precision, recall, and accuracy, analyzing each task separately:

Table 2
Overall Results Summary

Metric	Value
Global Track Score	0.3396
Overall Rank	51

In the polarity prediction task (ranging from 1 to 5), the model achieved modest performance (see Table 3). It recognized middle polarities better but struggled with the extreme classes (positive and negative). Class imbalance and the limited number of training epochs likely caused these difficulties. Table 4 details the F_1 for each class.

Table 3
Polarity Task Summary

Metric	Score
Macro F_1 (Polarity)	0.1831
Accuracy (Polarity)	20.27%
Mean Absolute Error (MAE)	0.86

Table 4
F_1 by Polarity Class

Class	F_1
1	0 (not detected correctly)
2	0.2844
3	0.3260
4	0.3054
5	0 (not detected correctly)

In the classification of place type (attraction, restaurant, or hotel), the model achieved robust results, as shown in Table 5. The model effectively distinguished among the categories to determine the type of place mentioned in the reviews, likely due to more evident linguistic patterns and better representation of these classes in the data. The F_1 per category are displayed in Table 6.

Table 5
Type of Place Task Summary

Metric	Score
Macro F_1 (Type)	0.8845
Accuracy (Type)	88.47%

The most challenging task was predicting the specific Magical Town mentioned in the text due to the large number of classes (over 50). Nevertheless, the model achieved acceptable results (see Table 7). A subset of towns achieved higher F_1, as detailed in Table 8. Other towns, especially those with fewer data representations, had F_1 scores close to zero.

Table 6
F_1 by Place Type

Type	F_1
Restaurant	0.8997
Hotel	0.8849
Attraction	0.8689

Table 7
Magical Town Task Summary

Metric	Score
Macro F_1 (Town)	0.2623
Accuracy (Town)	49.20%

Table 8
Top Performing Magical Towns by F_1

Town	F_1
Isla Mujeres	0.7345
San Cristóbal de las Casas	0.6170
Bacalar	0.5783
Sayulita	0.5509
Valladolid	0.4213

4. Conclusions

The results obtained in the Rest-Mex 2025 competition demonstrate that the proposed LSTM model performs competitively in general classification tasks, particularly in detecting the type of tourist place mentioned in Spanish reviews. The precision and F_1 achieved in this task show that the implemented architecture effectively captures relevant natural language patterns to identify categories such as attraction, hotel, or restaurant.

However, the model demonstrated limited performance in more complex tasks, such as sentiment polarity prediction and identifying Magical Town. The metrics reflect difficulties in distinguishing between extreme polarity classes (very positive or very negative) and in accurately recognizing a large variety of different locations.

One issue that negatively impacted the results was an error in preprocessing the polarity labels: the code transformed the original range from 1 to 5 into a range from 0 to 4. Although this modification suits some models, it did not align with the official evaluation format, which contributed to the lower performance in this task. Since this was the first and only submission, the results can serve as an initial baseline for building significant improvements in future system versions.

5. Future Work

Based on these findings, this paper proposed the following directions for future work:

- Correct the handling of polarity labels by preserving the original range (1 to 5) as required by the evaluation.
- Increase the number of training epochs to allow the model to learn more complex patterns and avoid underfitting or overfitting.
- Apply class balancing techniques, such as oversampling minority classes or weighted loss functions, to improve the representation of underrepresented classes.

- Explore multitask learning architectures that enable the model to learn related tasks jointly, such as polarity and place type, which could enhance generalization.
- Use pre-trained Spanish language models like BERT or BETO, which have demonstrated superior performance on NLP tasks compared to models trained from scratch.
- Expand preprocessing and enrich the corpus by incorporating lemmatization, named entity recognition (NER), and contextual analysis to improve semantic understanding of the text.
- These improvements could significantly enhance the robustness of the proposed methodology and enhance its applicability in real-world automated tourism analysis scenarios.

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] Z. Xiang, Q. Du, Y. Ma, W. Fan, A comparative analysis of major online review platforms: Implications for social media analytics in hospitality and tourism, Tourism Management 58 (2017) 51–65. doi:10.1016/j.tourman.2016.10.001.

[2] 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.

[3] E. Olmos-Martínez, M. Á. Álvarez-Carmona, R. Aranda, A. Díaz-Pacheco, What does the media tell us about a destination? the cancun 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] 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.

[5] A. Diaz-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.

[6] D. Buhalis, M. Foerste, Socomo marketing for travel and tourism: Empowering co-creation of value, Journal of Destination Marketing & Management 4 (2015) 151–161. doi:10.1016/j.jdmm.2015.04.001.

[7] 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.

[8] R. Guerrero-Rodriguez, 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.

[9] S. Hochreiter, J. Schmidhuber, Long short-term memory, Neural Computation 9 (1997) 1735–1780. doi:10.1162/neco.1997.9.8.1735.
[10] Y. Zhang, B. C. Wallace, A sensitivity analysis of (and practitioners' guide to) convolutional neural networks for sentence classification, CoRR abs/1510.03820 (2015). URL: http://arxiv.org/abs/1510.03820. arXiv:1510.03820.
[11] 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.
[12] 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.
[13] 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.
[14] 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).
[15] 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).