UC-UCO-Plenitas Team - Exploring in the Rest-Mex 2025: Researching Sentiment Evaluation in Text for Mexican Magical Towns

Yoan Martínez-López[1,2,*], Ireimis de las Mercedes Leguen de Varona[3], Miguel Bethencourt[3], Demetrio Rodríguez Fernández[3], Julio Madera[3], Ansel Yoan Rodríguez-González[4], Carlos de Castro Lozano[1,2], Jose Miguel Ramírez Uceda[1,2] and José Carlos Arévalo Fernández[2]

[1]*Universidad de Córdoba, Córdoba, Spain*
[2]*Plénitas, C/ Le Corbusier s/n, 14005 Córdoba, Spain*
[3]*Universidad de Camaguey, Circunvalación Norte, Camino Viejo Km 5 y 1/2, Camaguey, Cuba*
[4]*CICESE-UAT, Nayarit, México*

Abstract

This paper presents the UC-UCO-Plenitas team's participation in the Rest-Mex 2025 shared task on sentiment analysis of Spanish-language tourist reviews about Mexican Magical Towns. The challenge involves predicting three elements from each review: sentiment polarity (scale 1–5), the type of location reviewed (Hotel, Restaurant, or Attraction), and the specific Magical Town being referenced. To address this, we implemented a machine learning pipeline leveraging a MultiOutputClassifier with Random Forest as the base estimator to handle the multi-label classification problem. The methodology includes standard preprocessing, metadata utilization, and model evaluation based on macro-averaged precision, recall, and F1-score. Our system achieved notable improvements over the baseline across all metrics, including a 68.48% accuracy in polarity classification and strong generalization performance across diverse towns. This work demonstrates the viability of ensemble approaches for multilingual sentiment tasks and provides a robust foundation for future NLP research in tourism-related domains..

Keywords

Sentiment Analysis, Natural Language Processing, MultiOutput Classification, Random Forest, Mexican Magical Towns, Tourism Text Mining

1. Introduction

Sentiment Analysis is a core area within Natural Language Processing (NLP) that aims to assess individuals' opinions toward various entities—such as products, services, or events—by classifying them into predefined categories [1, 2, 3, 4]. These categories may range from coarse-grained (e.g., positive, negative, and neutral) to more fine-grained sentiment scales [5, 6]. This task has garnered significant attention due to its practical applications, enabling stakeholders to extract actionable insights from user-generated content on platforms such as TripAdvisor, social media, and other review-based websites [7].

Despite its growing popularity, Sentiment Analysis still faces key challenges—one of the most prominent being the uneven availability of linguistic resources across different languages. To stimulate progress in this field, numerous shared tasks and evaluation campaigns have been organized, including SemEval, IberLEF [8], and more recently, Rest-Mex [9, 4, 10, 5]. In recent years, the field has seen notable

IberLEF 2025, September 2025, Zaragoza, Spain
*Corresponding author.
yoan.martinez@plenitas.com (Y. Martínez-López); ireimis.leguen@reduc.edu.cu (I. d. l. M. L. d. Varona); betamiguel00@gmail.com (M. Bethencourt); demetrio.rodriguez@reduc.edu.cu (D. R. Fernández); julio.madera@reduc.edu.cu (J. Madera); ansel@cicese.edu.mx (A. Y. Rodríguez-González); carlosdecastrolozano@gmail.com (C. d. C. Lozano); p52raucj@uco.es (J. M. R. Uceda); josecarlos@plenitas.com (J. C. A. Fernández)
0000-0002-1950-567X (Y. Martínez-López); 0000-0002-1886-7644 (I. d. l. M. L. d. Varona); 0000-0001-8873-6802 (M. Bethencourt); 0000-0003-2803-431X2 (D. R. Fernández); 0000-0001-5551-690X (J. Madera); 0000-0001-9971-0237 (A. Y. Rodríguez-González); 0000-0001-6603-843X (C. d. C. Lozano); 0000-0002-5027-7521 (J. M. R. Uceda)

advancements due to the application of Deep Learning techniques, which have improved classification accuracy and generalization capabilities[11, 12, 13].

Building on similar methodologies, various research teams have successfully participated in related competitions[14, 15, 16, 17, 18, 19, 13, 20, 21, 22, 23, 24, 25, 26, 27], achieving competitive performance. In this work, we present our participation in the Rest-Mex 2025 Sentiment Analysis Subtask [5, 4, 10], which is framed as a polarity classification task. The goal is to develop systems capable of automatically predicting the sentiment polarity of a given opinionated text.

2. Methodology

Sentiment Analysis Task: The goal of this task is to analyze TripAdvisor reviews and classify them based on three key aspects: sentiment polarity, type of site, and associated Pueblo Mágico [5]. Each review contains valuable information about a traveler's experience, and our objective is to extract meaningful insights from it. First, we need to determine the sentiment polarity of the review by assigning it a rating from 1 (very negative) to 5 (very positive), based on the original score given by the tourist. This will help in understanding overall visitor satisfaction. Next, it classifies the review according to the type of site being reviewed. The review could describe a hotel, a restaurant, or an attraction, and this categorization is based on contextual keywords and available metadata. Finally, we need to identify which Pueblo Mágico the review belongs to. This is done by analyzing location metadata, ensuring that each review is correctly assigned to its respective destination.

Machine Learning (ML) is a branch of artificial intelligence (AI) that centers on the development of algorithms and systems capable of learning patterns from data and making predictions or decisions without being explicitly programmed for each task [28, 29]. Unlike traditional rule-based programming, ML models improve their performance as they are exposed to more data, a feature known as learning from data. A crucial characteristic of ML is its ability to generalize, meaning the model is designed to perform well on new, unseen data beyond the training examples [30]. Additionally, ML systems are adaptable, allowing them to adjust to changing patterns or dynamic environments over time. There are three main types of machine learning: supervised learning, unsupervised learning, and reinforcement learning. In supervised learning, the model is trained on labeled data, such as emails identified as spam or not. In unsupervised learning, the model identifies hidden patterns in unlabeled data, such as grouping customers based on behavior. Reinforcement learning involves the model learning through interaction with an environment, receiving rewards or penalties for its actions—similar to how a robot learns to walk or an AI agent learns to play a game. Machine learning has become integral in many real-world applications. It powers email spam filters, virtual voice assistants like Siri or Alexa, recommendation systems on platforms like Netflix and Amazon, fraud detection tools used in banking, and the decision-making systems in self-driving cars. These examples illustrate the widespread and growing influence of ML across various domains.

Ensemble MultiOutput Classifier RandomForest Algorithm

The **Ensemble MultiOutput Classifier with Random Forest** is a machine learning approach designed for solving multi-output (also called multi-label) classification problems, where each instance can have multiple target labels instead of just one. Normally, classification models predict a single target variable (single output)[31]. A MultiOutputClassifier allows you to train a separate classifier for each output target when you have multiple outputs. For example, if you want to predict multiple binary labels for one input (e.g., tagging an image with multiple labels like "cat," "dog," "car"), the MultiOutputClassifier handles this by fitting one classifier per label. **Random Forest** is an ensemble learning method that builds multiple decision trees during training and outputs the majority vote (classification) or average prediction (regression) of the individual trees. It's robust to overfitting, handles high-dimensional data well, and can capture complex feature interactions.

Combines both concepts by using a Random Forest as the base estimator for each output in a multi-output setting. This means it trains one Random Forest model per output variable independently. This approach leverages the power of Random Forest for each output while allowing simultaneous prediction of multiple targets[32]. Forest's robustness and ability to handle complex data is combined with multi-output support. Easy to implement with libraries like scikit-learn (MultiOutputClassifier wrapping a RandomForestClassifier) [33].

2.1. Metrics

For the evaluation of the Magical Town (MT) task, the idea is similar to the type prediction measure [5]. To this end, it is assumed that there exists a list containing all Magical Towns, denoted as **MTL** (Magical Towns List). The proposed model incorporates three distinct sources of information to determine a sentiment value associated with a keyword k.

First, the response from the general population is quantified as the average of the individual feedback scores $F_i(k)$ across all contributors C, as expressed in Equation (1):

$$Resp_P(k) = \frac{\sum_{i=1}^{|C|} F_i(k)}{|C|} \tag{1}$$

Second, a thematic response score is calculated by aggregating feedback from three thematic categories: attraction (F_A), history (F_H), and reputation (F_R). This average is shown in Equation (2):

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

Third, the response from the list of Magical Towns (MTL) is computed as the mean of the individual scores $F_{MTL_i}(k)$ over the entire list, as described in Equation (3) [5]:

$$Res_{MT}(k) = \frac{\sum_{i=1}^{\text{len}(MTL)} F_{MTL_i}(k)}{\text{len}(MTL)} \tag{3}$$

Finally, these three components are integrated into a comprehensive sentiment score using a weighted average, where the response from MTL is given greater importance. This final aggregation is defined in Equation (4):

$$Sentiment(k) = \frac{2 \cdot Resp_P(k) + Res_T(k) + 3 \cdot Res_{MT}(k)}{6} \tag{4}$$

This multi-source sentiment formulation ensures that both public opinion and expert or localized thematic relevance are incorporated into the analysis, providing a robust evaluation framework. The final measure for this task is the average of 3 sub-tasks. The idea is that polarity and the Magical Town identification have more weight than the other two subtasks, it will be given two and three times the importance, respectively, as it can see in Equation (4).

Sentiment analysis evaluation

Systems are evaluated using standard evaluation metrics, including precision, recall, and F1-score. How each task will be evaluated is listed below: In the present edition, Equation 1 is applied to evaluate the result of the polarity classification [5]. Where k is a forum participant system, C = 1,2,3,4,5. Finally, Fi(K) is the F-measure value for the class i obtained by the system k . For the Type prediction, there are 3 classes (Attractive, Hotel, and Restaurant). For this reason, it apply the Macro F-measure as the Equation 2 indicates. Where FA(k) represents the F measure obtained by the system k for the Attractive class. FH(k) represents the F measure obtained by the system k for the Hotel class. In the same way, FR(k) represents the F measure obtained by the system k for the Restaurant class.

3. Results

The comparative evaluation between the Baseline model [5] and the UC-UCO-Plenitas system was conducted using key performance indicators such as F1-score, precision, recall, and accuracy across various cities. Overall, UC-UCO-Plenitas demonstrated a notable performance advantage. In terms of general accuracy, UC-UCO-Plenitas achieved 68.48%, surpassing the Baseline's 65.54%. The macro-averaged F1-score for UC-UCO-Plenitas was 0.4389, whereas the Baseline recorded a score of 0.0, indicating a marked improvement in classification capabilities.

Across all evaluated cities, UC-UCO-Plenitas consistently outperformed the Baseline in both precision and recall. For instance, in Tulum, UC-UCO-Plenitas achieved an F1-score of 0.4389, with precision at 0.3172 and recall at 0.9246, compared to zero performance across all metrics for the Baseline. In Isla Mujeres, the model obtained an F1-score of 0.3015, precision of 0.1638, and recall of 0.1490, again surpassing the Baseline which scored 0.0 in all three metrics. A similar trend was observed in San Cristóbal de las Casas, where UC-UCO-Plenitas yielded an F1-score of 0.3233, precision of 0.0960, and recall of 0.0892. Additional cities such as Valladolid, Bacalar, and Palenque followed the same pattern, with F1-scores ranging from 0.5489 to 0.6975, highlighting the robustness of the UC-UCO-Plenitas model.

The macro-averaged metrics further emphasize the superior performance of UC-UCO-Plenitas, achieving 0.4389 in F1-score, 0.3172 in precision, and 0.9246 in recall, while the Baseline model scored 0.0 in all categories, demonstrating the effectiveness and significant improvement brought by the proposed system. See table 1.

Table 1
Performance comparison between UC-UCO-Plenitas and Baseline

Team	Track Score	Macro F1 (Polarity)	Macro F1 (Type)	Macro F1 (Town)	Accuracy (Polarity)
UC-UCO-Plenitas	**0.5157**	**0.3172**	**0.9246**	**0.5118**	**68.48%**
Baseline	0.0901	0.1584	0.1967	0.0089	65.54%

4. Conclusions

The UC-UCO-Plenitas model consistently outperforms the Baseline across all evaluated metrics. It demonstrates significant improvements in both precision and recall, particularly in cities like Bacalar and Valladolid , where it achieves relatively high F1-scores. The Baseline model shows no meaningful performance, with all metrics scoring close to zero. This suggests that the UC-UCO-Plenitas model is more effective at capturing true positives while minimizing false negatives and false positives compared to the Baseline.

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] J. Madera-Quintana, A. Hernández-Gónzalez, Y. Martínez-López, Thematic unsupervised classification of tourist texts using latent semantic analysis and k-means, Proceedings http://ceur-ws.org ISSN 1613 (2023) 0073.

[2] N. G. Carmona-Sánchez, Measuring the role of the verbs, nouns, and adjectives on the tourist opinions in spanish., in: IberLEF@ SEPLN, 2023.

[3] 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.

[4] 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.
[5] 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.
[6] J. Arreola, L. Garcia, J. Ramos, A. Rodríguez, An embeddings based recommendation system for mexican tourism. submission to the rest-mex shared task at iberlef 2021, in: Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2021), volume 2943, 2021, pp. 110–117.
[7] D. Mendoza, J. Ramos-Zavaleta, A. Rodríguez, A transfer learning model for polarity in touristic reviews in spanish from tripadvisor., in: IberLEF@ SEPLN, 2022.
[8] 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.
[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, 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.
[11] E. R. Reyes, Techkatl: A sentiment analysis model to identify the polarity of mexican's tourism opinions., in: IberLEF@ SEPLN, 2021, pp. 171–178.
[12] E. Rivadeneira-Pérez, C. Callejas-Hernández, Leveraging lda topic modeling and bert embeddings for thematic unsupervised classification of tourism news in rest-mex competition., in: IberLEF@ SEPLN, 2023.
[13] V. Gómez-Espinosa, V. Muñiz-Sanchez, A. P. López-Monroy, Automl and ensemble transformers for sentiment analysis in mexican tourism texts., in: IberLEF@ SEPLN, 2022.
[14] G. M. Barco, G. E. R. Rivera, D. I. H. Farías, Sentiment analysis in spanish reviews: Dataket submission on rest-mex 2022., in: IberLEF@ SEPLN, 2022.
[15] J. C. Rivas-Álvarez, R. A. García-Hernández, S. I. Medina-Martínez, A. M. Martínez-Ortiz, N. H. Castañeda, J. E. Ruiz-Melo, Á. Hernández-Castañeda, Y. N. Ledeneva, Opinion mining of the mexican tourism sector through sets of normalized n-grams., in: IberLEF@ SEPLN, 2022.
[16] J. Alcibar-Zubillaga, Y. D.-L. Ocampo, I. Pacheco-Castillo, K. Ramirez-Mendez, J.-P.-M. Sainz-Takata, O. J. Gambino, Participation of escom's data science group at rest-mex 2022: Sentiment analysis task., in: IberLEF@ SEPLN, 2022.
[17] J. D. Jurado-Buch, L. Bustio-Martínez, M. Á. Álvarez-Carmona, The role of the topics for the sentiment analysis task on a mexican tourist collection., in: IberLEF@ SEPLN, 2022.
[18] M. P. Enríquez, J. A. Mencía, I. Segura-Bedmar, Transformers approach for sentiment analysis: Classification of mexican tourists reviews from tripadvisor., in: IberLEF@ SEPLN, 2022.
[19] J. Alonso-Mencía, Unlocking sentiments: Exploring the power of nlp transformers in review analysis, 2023.
[20] 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).
[21] J. D. Jurado-Buch, E. S. Minayo-Díaz, J. A. Tello, K. E. Chaucanes, L. V. Salazar, M. D. Oquendo-Coral, M. Á. Álvarez-Carmona, A single model based on beto to classify spanish tourist opinions through the random instances selection., in: IberLEF@ SEPLN, 2023.
[22] A. Z. Gallardo-Hernández, R. Aranda, Á. Díaz-Pacheco, Classifying tourist text reviews by means

of mutual information features., in: IberLEF@ SEPLN, 2023.
[23] J. Cerda-Flores, R. Hernández-Mazariegos, J. Ortiz-Bejar, F. Calderón-Solorio, J. Ortiz-Bejar, Umsnh at restmex 2023: An xgboost stacking with pre-trained word-embeddings over data batches., in: IberLEF@ SEPLN, 2023.
[24] J. de Jesús Ceballos-Mejía, M. A. Parra-Urías, M. Á. Álvarez-Carmona, Filtering opinions in spanish with topics of tourist interest for the sentiment analysis task., in: IberLEF@ SEPLN, 2023.
[25] O. G. Toledano-López, J. Madera, H. González, A. Simón-Cuevas, T. Demeester, E. Mannens, Fine-tuning mt5-based transformer via cma-es for sentiment analysis., in: IberLEF@ SEPLN, 2022.
[26] J.-L. García-Mendoza, D. Buscaldi, Enriching with minority instances a corpus of sentiment analysis in spanish., in: IberLEF@ SEPLN, 2023.
[27] A. Rico-Sulayes, J. Monsalve-Pulido, A proposal and comparison of supervised and unsupervised classification techniques for sentiment analysis in tourism data, in: Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2022), Spain, CEUR-WS, volume 3202, 2022, pp. 1–8.
[28] T. M. Mitchell, T. M. Mitchell, Machine learning, volume 1, McGraw-hill New York, 1997.
[29] P. P. Shinde, S. Shah, A review of machine learning and deep learning applications, in: 2018 Fourth international conference on computing communication control and automation (ICCUBEA), IEEE, 2018, pp. 1–6.
[30] Z.-H. Zhou, Machine learning, Springer nature, 2021.
[31] H. Linusson, Multi-output random forests, 2013.
[32] C. González, J. M. Mira, J. A. Ojeda, Applying multi-output random forest models to electricity price forecast, Preprints (2016).
[33] A. Bhardwaj, D. Radha, V. K. Devi, An enhanced framework for churn prediction using stratifed bagging and stacked multi-output random forests, in: 2025 3rd International Conference on Intelligent Data Communication Technologies and Internet of Things (IDCIoT), IEEE, 2025, pp. 119–123.

A. Online Resources

The results are available via

- RestMex2025 results.