BERT-Based Models for Joint Sentiment, Type, and Location Classification of Spanish Tourist Reviews

Darían Santiago Llanes Guilarte[1], Vitali Herrera-Semenets[1], Lázaro Bustio-Martínez[2] and Miguel Ángel Álvarez-Carmona[3]

[1]Advanced Technologies Application Center (CENATAV), La Habana, Cuba
[2]Iberoamerican University, Ciudad de México, México
[3]Centro de Investigación en Matemáticas, Monterrey, México

Abstract

This paper presents two approaches to jointly classify sentiment polarity, tourist place type, and the corresponding Magical Town from Spanish-language reviews of Mexican tourist destinations. Our methods leverage a multi-task neural network based on the TabularisAI multilingual sentiment model (768-dim BERT-base architecture) and a pre-trained BERT model adapted for Spanish. Unlike previous approaches that relied on a unified label space or separate models for each task, we adopt a multi-head architecture that simultaneously optimizes for all three tasks using task-specific classification heads. The system incorporates class balancing through weighted loss functions and advanced preprocessing with spaCy. Evaluation on the official Rest-Mex 2025 dataset demonstrates competitive performance, achieving promising results across tasks, while maintaining efficiency and modularity.

1. Introduction

Sentiment classification in the tourism domain is a key task for understanding tourist feedback and improving user experience. With the increasing volume of user-generated content in Spanish, automatic methods are needed to process and analyze opinions about tourist destinations efficiently.

Unlike past editions [1, 2, 3], the Rest-Mex 2025 shared task focuses on analyzing reviews about Mexican Magical Towns, requiring systems to determine (1) the sentiment polarity (from 1 to 5), (2) the type of tourist location (hotel, restaurant, attraction), and (3) the specific Magical Town from a fixed list [4, 5]. This multi-label classification setting offers unique challenges, particularly when training deep learning models on imbalanced real-world data.

We propose a solution based on a multi-task deep neural network that processes a review once but predicts all three labels in parallel using separate classification heads. The core of the model is a pre-trained Spanish BERT (BETO), augmented with spaCy-based preprocessing and class-weighted loss functions to address data imbalance.

2. Related Work

Research on sentiment analysis in the Spanish tourism domain has significantly advanced through the adoption of pre-trained transformer-based models such as BERT [6, 7, 8]. Vásquez employed BETO combined with TF-IDF weighting over TripAdvisor reviews and achieved first place in Rest-Mex 2021 using a monolingual architecture optimized for Spanish [9]. Jurado-Buch et al. proposed a unified model based on BETO to jointly predict polarity, type, and location in one pass, placing within the top eight systems at Rest-Mex 2023 [10]. Moreover, Álvarez-Carmona et al. provided a comprehensive overview of Rest-Mex 2023, offering insight into dataset characteristics and evaluation protocols [11].

IberLEF 2025, September 2025, Zaragoza, Spain
darian.llanes@cenatav.co.cu (D. S. L. Guilarte); vherrera@cenatav.co.cu (V. Herrera-Semenets); lazaro.bustio@ibero.mx (L. Bustio-Martínez); miguel.alvarez@cimat.mx (M. Álvarez-Carmona)
0000-0001-7094-2835 (V. Herrera-Semenets); 0000-0002-0273-0520 (L. Bustio-Martínez); 0000-0003-4421-5575 (M. Álvarez-Carmona)

Domain-adapted language models have also been explored. Campos and Viñaña-Ludeña trained a BERT model specifically for tourism (Spanish-Tourism-BERT) on social media data to extract location-based entities and sentiment components [12]. Bouabdallaoui et al. compared BERT fine-tuning against hybrid models combining sentence embeddings and LSTM networks in Moroccan tourism data, concluding that BERT fine-tuning yields superior accuracy [13]. In terms of multi-task learning, Zhang et al. implemented a hard-sharing architecture combining a shared BERT encoder with task-specific layers, showing statistical improvements in multi-output sentiment scenarios [14].

The reviewed literature reveals three clear trends: (1) fine-tuning monolingual BERT models such as BETO is highly effective in Spanish tasks; (2) domain-specific models show promise for tourism-related content; and (3) multi-task architectures provide performance benefits for correlated classification tasks. However, most works address only one or two subtasks, and none explore simultaneous prediction of polarity, place type, and town name, as required by Rest-Mex 2025.

2.1. Remarks

The reviewed studies expose several critical gaps that justify the contributions of the present work:

- Most prior works (e.g., [9], [12]) focus exclusively on a single task such as polarity or aspect identification. Our proposed architecture performs simultaneous classification of sentiment polarity, place type, and geographical location, addressing the complete Rest-Mex 2025 task.
- Jurado-Buch et al. addressed multi-label output via a unified label space (45-class combinations), but their model lacks separate heads or task-specific loss components. Our solution employs a shared encoder with specialized heads per task and customized loss weighting to enhance adaptability and generalization.
- While Spanish-Tourism-BERT represents a step toward domain-specific pre-training, it was tested only on social media and not on structured tourism reviews. Our models include both multilingual (TabularisAI) and monolingual (BETO) BERT variants adapted through preprocessing, task weighting, and stratified training.
- Although Zhang et al. introduced a valid multitask architecture, they did not apply it to tourism datasets or large-scale multilingual corpora. Our study extends both dataset size and architectural robustness via gradient accumulation, warm-up phases, and differential learning rates.

After revising the related work, the contributions of the presented work include:

1. Joint evaluation of polarity, place type, and location via multi-head classification.
2. Comparative analysis of monolingual and multilingual BERT-based architectures adapted to tourism.
3. Implementation of advanced preprocessing and weighted loss strategies for handling extreme class imbalance.

This work fills evident gaps in the literature and offers an integrated framework suitable for large-scale, multidimensional sentiment analysis in Spanish-language tourist reviews.

3. Dataset

The dataset provided by the Rest-Mex 2025 organizers consists of 208,051 Spanish-language reviews [4]. Each review includes a title and a body, which we concatenate into a single text field. The dataset is annotated with:

- Sentiment polarity on a 5-point scale (1 = very negative, 5 = very positive)
- Type of tourist place: Hotel, Restaurant, or Attractive
- Name of the Magical Town (from a set of 60 distinct towns)

- Region to which the town belongs

A detailed exploratory data analysis (EDA) was conducted to understand the distribution of the data. Key findings include:

- **Polarity distribution:** The data is heavily skewed toward positive reviews, with over 65% labeled as polarity 5 (see Figure 1). Only about 2.6% are rated as polarity 1.
- **Type distribution:** The most common category is Restaurant (86,720 instances), followed by Attractive (69,921), and Hotel (51,410) (see Figure 2).
- **Top towns:** Tulum (45,345 reviews), Isla Mujeres (29,826), and San Cristóbal de las Casas (13,060) are the towns with the most reviews (see Figure 3).
- **Top regions:** Quintana Roo leads with 85,993 reviews, followed by Chiapas (23,532), and Estado de México (19,439).
- **Text length:** The average length of concatenated Title + Review texts is 65 words, with a maximum of 1,487 and a minimum of 2 words (see Figure 4).
- **Missing values:** Only the `Title` field had missing values (2 instances), which were handled by exclusion.

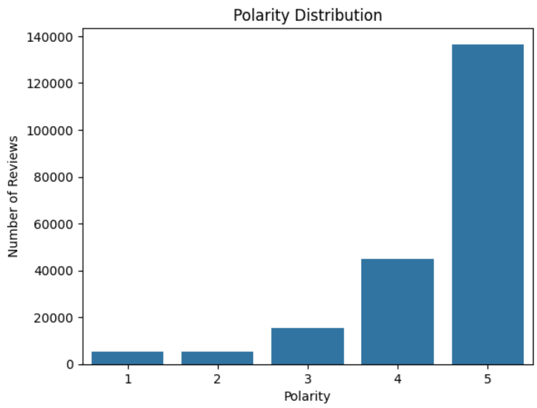

Figure 1: Polarity distribution.

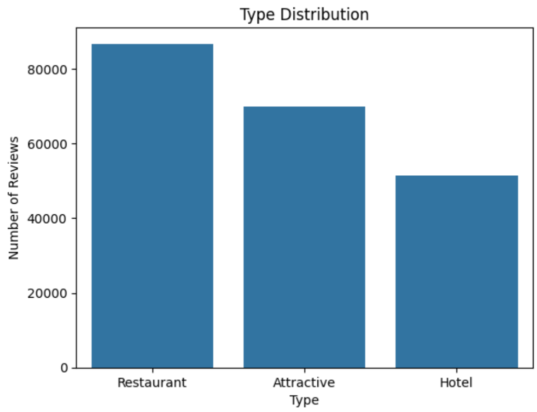

Figure 2: Type distribution.

These findings reveal significant class imbalance, particularly in the polarity label, justifying the use of class-weighted losses in training.

We use 80% of the dataset for training and 20% for validation, following a stratified sampling strategy to preserve label distributions.

4. Methodology

This section describes our two proposals evaluated on the official Rest-Mex 2025 dataset.

4.1. Hammer Squat_1_Run

This approach leverages a multilingual sentiment analysis model integrated with optimized text preprocessing and multi-task learning. Specifically, our solution incorporates three core components: advanced linguistic normalization, a transformer-based architecture with task-specific heads, and adaptive training strategies for class imbalance mitigation.

4.1.1. Data Preprocessing

Text normalization techniques were systematically applied to raw inputs to ensure uniform formatting. Subsequently, the title and review fields were concatenated to form unified text representations.

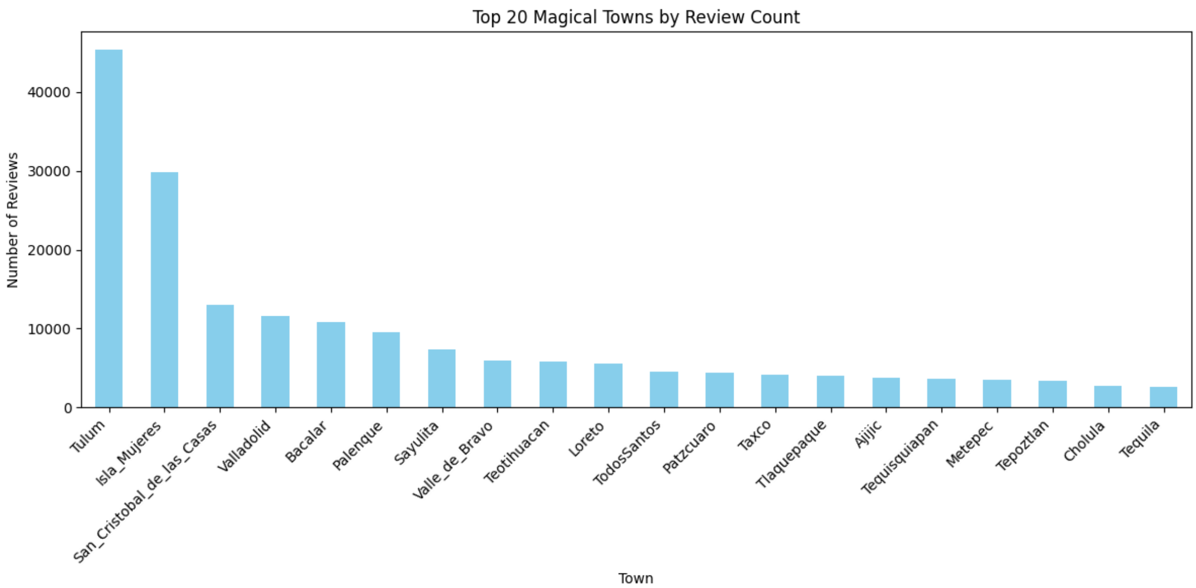

Figure 3: Top 20 magical towns by review count.

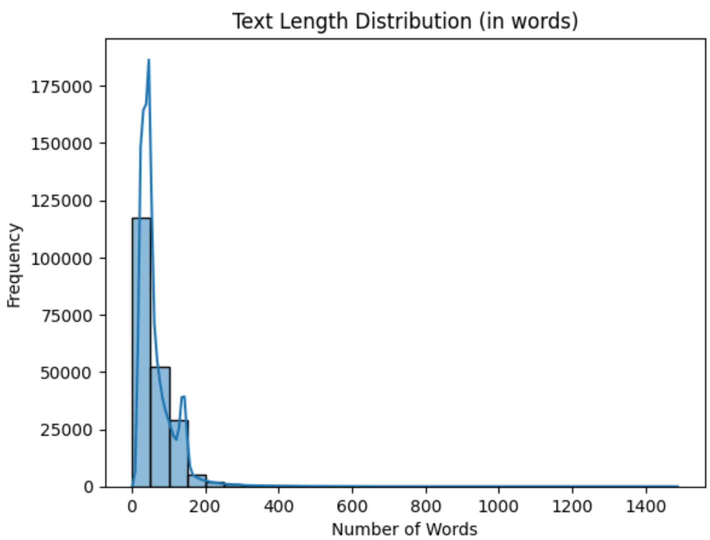

Figure 4: Text Length Distribution (in words).

Since reviews constituted our primary focus, encoding correction was performed using `ftfy` alongside emoji normalization via `emoji.demojize`. Additionally, non-informative elements including URLs, user mentions, and domain-specific stopwords were removed while preserving diacritics and numerical values. Character repetition patterns were concurrently reduced to single instances. Finally, language detection excluded non-Spanish texts, with complementary length-based filtering retaining texts exceeding 25 characters and 5 meaningful tokens.

4.1.2. Model Architecture

The architecture utilized `tabularisai/multilingual-sentiment-analysis`, a BERT-base model pretrained for multilingual sentiment tasks [15]. Contextual representations from this shared encoder subsequently fed three specialized classification heads: a 5-dimensional output for polarity prediction (1-5 star scale), a 3-dimensional output for establishment type classification (Hotel, Restaurant, or Attractive), and a 40-dimensional output for location identification (Magical Towns). To address dataset imbalances, class weights based on inverse frequency were applied to balance the loss functions.

4.1.3. Training Configuration

AdamW optimization employed differentiated learning rates (1×10^{-3} for classification heads versus 2×10^{-5} for the base encoder). Furthermore, loss components were weighted by task complexity: 30% polarity, 30% establishment type, and 40% location. Gradient accumulation (over 2 steps) enabled effective batch sizes of 16, while gradient clipping (max norm 1.0) stabilized convergence. The learning rate schedule combined a 10% warm-up phase with progressive encoder unfreezing after epoch 1. Ultimately, mixed-precision training accelerated computation throughout 4 epochs.

4.2. Hammer Squat_2_Run

In this section is presented our second approach ($HammerSquat\_2\_Run$). Our second proposal consists of three key components: data preprocessing, model design, and training configuration. Each step plays a critical role in building an effective and robust multi-task learning system.

4.2.1. Data Preprocessing

Preprocessing is essential to convert noisy, unstructured text into a clean format that machine learning models can interpret effectively. First, we concatenate the `Title` and `Review` fields to form a unified text input for the model. This ensures the model has access to all user-generated content associated with a review.

Next, we apply linguistic preprocessing using spaCy's `es_core_news_sm` model. This includes lemmatization (reducing words to their base forms), stopword removal (to eliminate uninformative words), and punctuation filtering. These steps reduce noise and vocabulary size, improving the model's generalization.

Finally, we encode the categorical labels (`polarity`, `type`, and `town`) using label encoding, transforming them into numerical values compatible with neural networks.

4.2.2. Model Architecture

We adopt a multi-task learning approach built on top of the pre-trained BERT model for Spanish: `bert-base-spanish-wwm-cased` [16]. Multi-task learning enables the model to learn shared representations that benefit multiple related tasks in this case, predicting polarity, place type, and town.

The architecture consists of:

- A shared BERT encoder that processes the input text into contextual embeddings.
- Three separate classification heads fully connected layers for each prediction task:
  - Polarity: 5 output neurons (for classes 1 to 5)
  - Type: 3 output neurons (hotel, restaurant, attraction)
  - Town: 60 output neurons (corresponding to each Magical Town)

This design allows each head to specialize while benefiting from shared information learned across tasks. Each head is optimized independently via task-specific cross-entropy loss. Also, class imbalance is mitigated by computing class weights and integrating them into each loss function. The multi-head architecture improves model efficiency by avoiding the need to train separate models.

4.2.3. Training Configuration

To train the model, we use the AdamW optimizer with a learning rate of 2e-5, which is well-suited for fine-tuning transformers. We apply the ReduceLROnPlateau scheduler to automatically reduce the learning rate when validation performance plateaus, ensuring stable convergence.

Table 1
Resultados de los modelos

Place	Run	Track Score	Macro F1 (Polarity)	Macro F1 (Type)	Macro F1 (Town)
1	UDENAR_1	0,725	0,644	0,987	0,691
2	Axolotux_E_T3	0,722	0,639	0,987	0,689
3	Axolotux_E3	0,719	0,638	0,987	0,684
20	Hammer Squat_1_Run	0,645	0,579	0,970	0,580
41	Hammer Squat_2_Run	0,535	0,473	0,944	0,441

A batch size of 16 is chosen to balance performance and memory usage, and we train for 15 epochs. We also calculate class weights based on the label distribution and use them in the cross-entropy loss function for each task. This addresses class imbalance and encourages the model to pay more attention to minority classes.

During training, the model is evaluated on a validation split to track performance. The best-performing model checkpoint is saved and later used for evaluation and inference.

Together, these methodological components ensure that the model is well-prepared to handle the complexities of real-world multilingual, multi-label classification in the tourism domain.

5. Results

We assess the system's performance Macro F1-score for each task. As shown in Figure 1, the results obtained by the Hammer Squat team in the REST-MEX competition, represented by the solutions Hammer Squat_1_Run (20th place) and Hammer Squat_2_Run (41st place), show a considerable performance gap compared to the top three entries: UDENAR_1, Axolotux_E_T3, and Axolotux_E3.

While both Hammer Squat submissions achieved reasonable scores, there are clearly critical areas for improvement, particularly in the Macro F1 (Polarity) and Macro F1 (Town) metrics, where the most notable differences can be observed. For instance, in the polarity task, Hammer Squat_1_Run scored 0.579, compared to 0.644 by the top-ranking model. The gap is even wider for Hammer Squat_2_Run, which achieved only 0.473 in this metric. Regarding the town classification task (Macro F1 (Town)), the Hammer Squat models again lag behind, with scores of 0.580 and 0.441, while the leading model reached 0.692.

Notably, in the Macro F1 (Type) metric, which evaluates entity type classification, the Hammer Squat models performed more competitively (0.970 and 0.944) compared to the winner (0.987), indicating that the team's approach has strengths in certain specific tasks. Overall, while the Hammer Squat solutions did not reach the top ranks, the results show promising potential, particularly if improvements are made in components related to sentiment detection and geographic localization. These key areas could benefit from further refinement, such as implementing more advanced linguistic preprocessing techniques or leveraging models specifically tuned for sentiment analysis and textual geolocation.

Comprehensive details regarding the overall results, including information about the participating teams, can be found in [17].

6. Conclusion

This paper introduce two models for sentiment analysis of Spanish-language reviews in the tourism domain. Our system jointly predicts sentiment polarity, place type, and location with promising results using shared representations and task-specific outputs. The approaches are efficient and adaptable, achieving high performance with modest hardware requirements. Future directions include integrating attention-based fusion and exploring multilingual generalization.

Acknowledgements

The authors gratefully acknowledge the support provided by the Mexican Academy of Tourism Research (AMIT) for the project "Balancing Tourism Text Data with Artificial Intelligence for Sentiment Analysis: A Specialized Language Model Approach" funded through the Research Projects 2024 call. Additionally, this work was also supported by the project "Text Generation for Data Balancing in Sentiment Classification: Application to Tourism Data" under the CICIMPI 2024 call of the Centro de Investigación en Matemáticas (CIMAT).

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, 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/https://doi.org/10.26342/2021-67-14.

[2] 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).

[3] 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).

[4] Rest-Mex, Rest-mex 2025: Researching sentiment evaluation in text for mexican magical towns, Available: https://sites.google.com/cimat.mx/rest-mex-2025/home (2025).

[5] 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.

[6] 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/https://doi.org/10.1016/j.jksuci.2022.10.010.

[7] 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:https://doi.org/10.1016/j.jksuci.2023.101746.

[8] 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:https://doi.org/10.1007/s40558-023-00278-5.

[9] J. Vásquez, H. Gómez-Adorno, G. Bel-Enguix, Bert-based approach for sentiment analysis of

spanish reviews from tripadvisor, in: IberLEF 2021, CEUR Workshop Proceedings, 2021. 1st place Rest-Mex 2021.
[10] J. D. Jurado-Buch, E. S. Minayo-Díaz, J. A. Tello, et al., Juan david jurado-buch and ever sebastian minayo-díaz and jhony alexander tello and others, in: Rest-Mex 2023, CEUR Workshop Proceedings, 2023. Top-8 Rest-Mex 2023.
[11] M. Ángel Álvarez Carmona, Ángel Díaz-Pacheco, R. Aranda, L. Bustio-Martínez, Overview of rest-mex at iberlef 2023: Research on sentiment analysis task for mexican tourist texts, CEUR Workshop Proceedings (2023).
[12] M. S. V. naña Ludeña, L. M. de Campos, Discovering a tourism destination with social media data: Bert-based sentiment analysis, Journal of Hospitality and Tourism Technology (2021).
[13] I. Bouabdallaoui, F. Guerouate, S. Bouhaddour, C. Saadi, M. Sbihi, Sentiment analysis in tourism: Fine-tuning bert or sentence embeddings concatenation?, arXiv preprint arXiv:2302.04519 (2023).
[14] K. Y. Zhang, J. Zhang, Y. Mao, Multi-task learning for sentiment analysis with hard-sharing and task recognition mechanism, Information 12 (2021) 207.
[15] tabularisai.multilingual-sentiment analysis, Distilbert-based multilingual sentiment classification model, Available: https://huggingface.co/tabularisai/multilingual-sentiment-analysis (2024).
[16] J. Cañete, G. Chaperon, R. Fuentes, J.-H. Ho, H. Kang, J. Pérez, Spanish pre-trained bert model and evaluation data, in: PML4DC at ICLR 2020, 2020.
[17] Rest-Mex, Results, Available: https://sites.google.com/cimat.mx/rest-mex-2025/results (2025).