UIE at PoliticHeadlinES-IberLEF 2026: Semantic Embedding Models for Candidate Headline Ranking in Spanish Political News

Oscar Apolinario-Arzube[1,*], Wernher Tellez[1], Katty Lagos-Ortiz[2] and José Medina-Moreira[3]

[1]Universidad Internacional del Ecuador, Guayaquil, Av. Las Aguas y calle 15, Urbanor 2, Ecuador
[2]Facultad de Ciencias Agrarias, Universidad Agraria del Ecuador, Av. 25 de Julio, Guayaquil, Ecuador
[3]Universidad Bolivariana de Ecuador, Km 5 ½ vía Durán - Yaguachi, Durán, Ecuador

Abstract

Headlines play a crucial role in shaping how users access, interpret, and engage with political news in online media environments. Selecting the most appropriate headline is a challenging task, as it must be concise, informative, and faithful to the article content, while also being distinguishable from other plausible alternatives. In this paper, we present our participation in the PoliticHeadlinES shared task at IberLEF 2026, which focuses on ranking candidate headlines for Spanish political news articles. Given a news article and a set of ten candidate headlines, systems must identify and rank the most suitable headline. We address both subtasks proposed: (1) text-based headline ranking, and (2) multimodal headline ranking, which incorporates visual information associated with the article. Our approach is based on modeling the semantic compatibility between the article and each candidate headline using pretrained embedding models, combined with a lightweight classification strategy. For the multimodal setting, we extend this framework by incorporating visual features extracted from a vision–language model. Experimental results show that our method clearly outperforms a TF-IDF baseline, improving from 50.96 to 64.47 in terms of pa-nDCG. However, the inclusion of visual information does not yield additional gains under the current fusion strategy, highlighting the challenges of effectively leveraging multimodal signals in this task. Overall, our findings show that embedding-based models combined with a lightweight linear classifier constitute a competitive and efficient baseline for candidate headline ranking. At the same time, the results indicate that simple feature concatenation is not sufficient to effectively exploit visual information in this task, motivating the exploration of more expressive multimodal ranking strategies.

Keywords

Headline Ranking, Political News, Semantic Matching, Multimodal Learning, Vision-Language Models, Text Embeddings

1. Introduction

Political news constitutes a central source of information in modern societies, where headlines play a key role in shaping how users discover, interpret, and share content. In digital media environments, the headline is often the most visible element of a news article, and in many cases, it may be the only piece of information that users read before deciding whether to engage with the content. As a result, headlines have a strong influence on public perception, framing effects, and information dissemination.

Despite their importance, selecting an appropriate headline is a challenging task. Headlines must be concise, informative, and faithful to the content of the article, while also being engaging enough to attract user attention. In addition, they must avoid misleading or sensationalist formulations that could distort the interpretation of the underlying information. These requirements make headline selection a complex problem that involves semantic understanding, content summarization, and pragmatic reasoning.

IberLEF 2026, September 2026, León, Spain
*Corresponding author.
†These authors contributed equally.
osapolinarioar@uide.edu.ec (O. Apolinario-Arzube); wetellezgo@uide.edu.ec (W. Tellez); klagos@uagraria.edu.ec (K. Lagos-Ortiz); jjmedinam@ube.edu.ec (J. Medina-Moreira)
0000-0003-4059-9516 (O. Apolinario-Arzube); 0000-0002-8447-4152 (W. Tellez); 0000-0002-2510-7416 (K. Lagos-Ortiz); 0000-0003-1728-1462 (J. Medina-Moreira)

While previous work has largely focused on headline generation, the problem of selecting the most suitable headline among a set of strong candidates remains relatively underexplored. This setting introduces additional challenges, as candidate headlines are often semantically similar and designed to be plausible alternatives. Consequently, systems must be capable of capturing fine-grained differences in meaning, relevance, and alignment with the article content.

Moreover, modern online news platforms frequently include visual content alongside textual information. Images can provide complementary contextual cues about events, entities, or scenes, potentially aiding the interpretation of the article and its associated headline. However, effectively leveraging this multimodal information remains a non-trivial problem, as the relationship between text and images in news articles is often indirect or loosely defined.

In this context, the PoliticHeadlinES shared task at IberLEF 2026 introduces a benchmark for multimodal headline ranking in Spanish political news [1]. The task is part of IberLEF 2026, a shared evaluation campaign focused on Natural Language Processing systems for Spanish and other Iberian languages [2]. Given a news article and a set of ten candidate headlines, systems must produce a ranking that identifies the most appropriate headline. The task includes two subtasks: a text-based setting and a multimodal setting that incorporates visual information associated with the article.

In this paper, we present our participation in both subtasks through a simple and efficient pointwise ranking approach. The proposed system models the semantic compatibility between each article and each candidate headline using pretrained text embeddings and a lightweight linear classifier. For the multimodal setting, we extend this representation with visual features extracted from a vision–language model. Rather than introducing a complex ranking architecture, our goal is to assess the effectiveness and limitations of embedding-based representations as a reproducible baseline for Spanish political headline ranking.

The remainder of this paper is organized as follows. Section 2 reviews related work on headline modeling, semantic matching, and multimodal understanding. Section 3 describes the task and dataset. Section 4 details the proposed methodology. Section 5 presents the experimental results and analysis. Finally, Section 6 concludes the paper and outlines future directions.

2. Related works

Research on news headlines has been traditionally framed as a generation problem, often modeled as an extreme form of summarization. Early approaches treated headline generation as a compression task, where systems aimed to produce short, informative titles from longer news articles [3]. More recent work has leveraged transformer-based architectures to improve fluency and content selection, as well as to incorporate constraints or user-oriented objectives [4, 5]. While these approaches are closely related, they differ from the setting considered in PoliticHeadlinES, where the objective is not to generate a headline but to select the most appropriate one among a set of strong candidates.

While most work has focused on general-domain headline generation, the problem becomes more complex in the context of political news. In this domain, summarization systems must not only condense information but also preserve journalistic style and editorial perspective [6].

A more closely related line of work focuses on semantic matching and ranking. In this paradigm, the task consists of estimating the relevance or compatibility between two textual units, such as a document and a query or a candidate response. Sentence embedding models, such as Sentence-BERT [7], have shown that transformer-based encoders can produce dense representations that are effective for semantic similarity and retrieval tasks. These approaches are particularly suitable for candidate ranking scenarios, as they enable efficient comparison between a single input (e.g., a news article) and multiple candidates (e.g., headlines). More recent embedding models, including those optimized for long-context inputs, have further improved performance in retrieval-oriented benchmarks [8, 9]. In contrast to cross-encoder architectures, which jointly encode pairs of texts at a higher computational cost, embedding-based methods provide a scalable alternative that is well-suited for ranking multiple candidates.

In the context of online news, multimodal methods have gained increasing attention due to the widespread integration of images in digital media platforms. News articles are often accompanied by visual content that provides additional context about events, entities, or scenes. Early studies highlighted the importance of modeling the relationship between article text and associated images beyond simple captioning tasks [10, 11]. Previous work in multimodal learning has shown that relying solely on textual information can be limiting, and that incorporating additional modalities may provide complementary contextual signals [12].

More recent work has explored multimodal summarization and headline generation, suggesting that visual information can complement textual understanding, although its contribution is highly dependent on the specific task [13, 14]. In parallel, multimodal approaches have explored more complex architectures to model interactions between modalities, highlighting the challenges of capturing meaningful cross-modal relationships beyond simple feature concatenation [15].

At the representation level, the development of large-scale vision–language models has significantly improved the alignment between images and text. CLIP [16] introduced a contrastive learning framework that enables the learning of shared embedding spaces for visual and textual data. These representations have proven effective across a wide range of multimodal tasks, including retrieval and classification. Subsequent approaches, such as SigLIP [17], have proposed alternative training objectives that improve efficiency and performance. These models are particularly relevant for multimodal ranking scenarios, where the goal is to assess the consistency between an image and multiple textual candidates.

Overall, prior work highlights the intersection between summarization, semantic matching, and multimodal understanding. However, PoliticHeadlinES formulates the problem as a ranking task over strong candidate headlines, where capturing fine-grained semantic differences between highly similar candidates is essential for performance.

3. Task Description

The PoliticHeadlinES shared task focuses on headline ranking for Spanish political news articles. The objective is to evaluate the ability of computational systems to identify the most appropriate headline among a set of candidate options, based on both textual and multimodal information.

3.1. Task Overview

Given a news article and a set of ten candidate headlines, systems are required to produce a ranking that orders the candidates from the most to the least relevant. Among the ten candidates, exactly one corresponds to the original headline associated with the article, while the remaining ones act as strong distractors designed to be plausible alternatives.

The task is divided into two subtasks:

- **Task 1: Text-based Headline Ranking.** In this subtask, systems must rank the candidate headlines using only textual information. The input consists of the article content and the ten candidate headlines.
- **Task 2: Multimodal Headline Ranking.** This subtask extends the previous one by incorporating visual information. In addition to the article text and candidate headlines, an image associated with the article is provided. Systems are expected to exploit both textual and visual cues to produce the ranking.

For both subtasks, the output must be a complete ranking of the ten candidate headlines, where each candidate is assigned a unique position. The ranking reflects the system's confidence, from the most relevant headline to the least relevant one.

3.2. Dataset

The dataset provided by the organizers consists of Spanish political news articles collected from multiple media sources. Each instance includes the article text, an associated image (for the multimodal setting), and a set of ten candidate headlines.

The candidate headlines are constructed such that only one corresponds to the original headline of the article, while the remaining candidates serve as distractors. These distractors are designed to be challenging, often sharing semantic similarity, topical overlap, or stylistic characteristics with the correct headline.

Each instance is identified by a unique identifier, and the dataset is divided into training and test splits. The training data includes the ground-truth ranking, which allows systems to learn to distinguish between correct and incorrect headlines. The test data, in contrast, requires systems to produce rankings without access to the ground-truth labels.

The evaluation of submitted systems is performed using position-aware normalized Discounted Cumulative Gain (pa-nDCG), which measures the quality of the predicted ranking by assigning higher importance to correctly placing the most relevant headlines at top positions.

4. Methodology

In this section, we describe the methodology adopted to address both subtasks of the PoliticHeadlinES shared task: text-based headline ranking (Task 1) and multimodal headline ranking (Task 2). Our approach is based on transforming the ranking problem into a pointwise relevance estimation task, where each candidate headline is independently scored with respect to the corresponding news article. Figure 1 shows the overall pipeline of the proposed system.

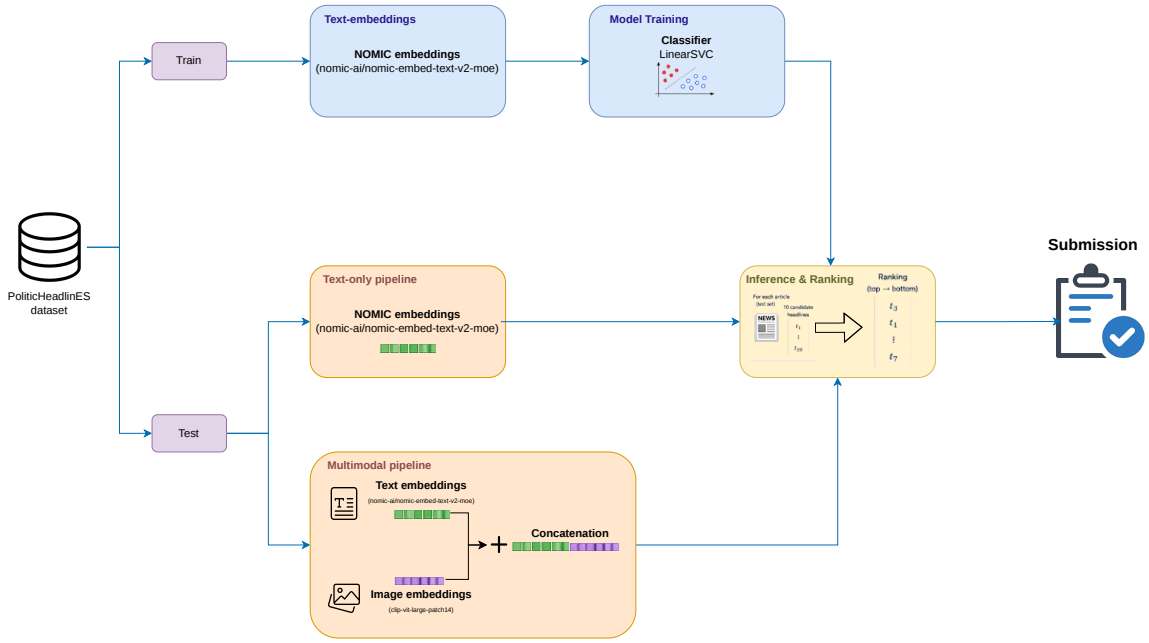

Figure 1: Pipeline of the participation of UIE at PoliticHeadlinES 2026.

4.1. Problem Formulation

Given a news article a and a set of candidate headlines $\{h_1, h_2, ..., h_{10}\}$, the goal is to produce a ranking that orders the candidates from the most to the least relevant. In the dataset, exactly one

headline corresponds to the original headline of the article, while the remaining candidates act as strong distractors.

We formulate the task as a pointwise learning-to-rank problem. Given an article a and a candidate headline h_i, the model estimates an independent relevance score $s(a, h_i)$ that reflects the semantic compatibility between both elements. During training, each article-headline pair is treated as an individual instance, with the original headline labeled as relevant and the remaining candidates labeled as non-relevant. At inference time, the ten candidates associated with each article are scored independently, and the final ranking is obtained by sorting them in descending order according to their predicted relevance scores.

4.2. Training Setup

For each news article, we generate ten article-headline instances, one for each candidate headline. The original headline is assigned a positive label, while the nine remaining candidate headlines are assigned negative labels. Therefore, the resulting training set follows a 1:9 positive-to-negative ratio. No resampling strategy was applied in the text-only setting.

Each input instance is constructed using the following template:

```
search_document: context article: {article}
headline option: {candidate_headline}
objective: estimate whether this headline is the original headline for the
article.
```

Before constructing the input, the article body is normalized by removing extra whitespace and truncated to a maximum of 280 words. The same template and preprocessing procedure are used during training, validation, and inference.

The text embeddings are extracted using `nomic-ai/nomic-embed-text-v2-moe`. The maximum sequence length of the embedding model is set to 512, the batch size is set to 32, and the output embeddings are L2-normalized. Each article-headline pair is represented by a 768-dimensional dense vector.

For internal validation, the training articles are randomly split into 90% training and 10% validation using a fixed random seed of 42. The split is performed at the article level, ensuring that all ten candidate headlines associated with the same article remain in the same partition and avoiding information leakage between training and validation. This results in 14,427 articles for training and 1,603 articles for validation, corresponding to 144,270 and 16,030 article-headline pairs, respectively.

The relevance classifier is implemented as a linear Support Vector Machine using `LinearSVC`. The model is trained with $C = 1.0$, `random_state=42`, and `dual="auto"`. Since the input embeddings are already normalized, we apply `StandardScaler` with `with_mean=False` before the classifier. The continuous decision scores produced by the classifier are used to rank the ten candidate headlines for each article.

4.3. Text-based Headline Ranking (Task 1)

For the text-based subtask, we adopt an embedding-based approach to model the semantic compatibility between the news article and each candidate headline. Each pair (a, h_i) is converted into a single textual input by concatenating the article content and the candidate headline using a predefined template.

We use the pretrained model `nomic-ai/nomic-embed-text-v2-moe` to obtain dense vector representations of each input. This model is designed for semantic similarity and retrieval tasks, and supports long-context inputs, making it suitable for encoding full news articles.

The resulting embeddings are used as input features for a linear Support Vector Machine (SVM) classifier. The model is trained using labeled article–headline pairs derived from the training data. For each article, the original headline is treated as the positive instance, while the remaining candidate

headlines are treated as negative instances. In this way, the classifier learns to assign higher relevance scores to the original headline and lower scores to the distractors.

At inference time, the SVM produces a continuous relevance score for each candidate headline. These scores are then used to rank the candidates, with higher scores indicating higher relevance.

4.4. Multimodal Headline Ranking (Task 2)

For the multimodal subtask, we extend the text-based representation by incorporating visual information associated with each news article. The textual representation is obtained in the same way as in Task 1: each article-headline pair is encoded with `nomic-ai/nomic-embed-text-v2-moe`, producing a 768-dimensional L2-normalized embedding.

Visual features are extracted from the article image using `openai/clip-vit-large-patch14`. Each image is loaded in RGB format and processed with the corresponding CLIP processor. We use the image features returned by the CLIP model and apply L2 normalization to obtain a 768-dimensional visual embedding. When an image cannot be located, we replace its representation with a zero vector of the same dimensionality.

The final multimodal representation is obtained by concatenating the 768-dimensional text embedding with the 768-dimensional image embedding, resulting in a 1536-dimensional feature vector for each article-headline candidate pair. These fused vectors are then standardized using `StandardScaler` and passed to a `LinearSVC` classifier with $C = 1.0$, `class_weight="balanced"`, `max_iter=5000`, and `random_state=42`. The classifier produces a decision score for each candidate headline, and the final ranking is obtained by sorting the ten candidates in descending order according to these scores.

This multimodal strategy should be interpreted as a simple early-fusion baseline. Since the same article image is associated with the ten candidate headlines of a given article, the visual embedding is shared across all candidates in that instance. Moreover, because the classifier is linear, the visual component contributes an article-level signal but does not explicitly model candidate-specific interactions between the image and each headline. Therefore, this architecture is limited in its ability to use visual information to discriminate among highly similar candidate headlines. More expressive multimodal approaches, such as image–headline compatibility scoring or cross-modal attention mechanisms, would be better suited to capture these interactions.

To facilitate reproducibility, the complete experimental pipeline is made available as a public Google Colab notebook.[1] The notebook includes data preprocessing, article–headline pair construction, train–validation splitting, embedding extraction, model training, validation, inference, ranking reconstruction, heatmap generation, and official submission generation. All experiments use a fixed random seed of 42.

5. Results

In this section, we present the performance of our approach on the PoliticHeadlinES shared task for both subtasks: text-based headline ranking (Task 1) and multimodal headline ranking (Task 2).

We evaluate our approach using both an internal validation split and the official test set provided by the shared task. The training data is split into 90% for training and 10% for validation. The validation set is used to guide model selection and assess generalization before submission.

For evaluation, we use accuracy at rank 1 (Acc@1) on the validation set, while the official results on the test set are reported using the position-aware normalized Discounted Cumulative Gain (pa-nDCG).

5.1. Validation Results

To assess the effectiveness of the proposed approach, we evaluate the text-based and multimodal models on the same internal validation split. The split is performed at the article level, ensuring that all ten candidate headlines associated with the same article remain in the same partition.

[1] https://colab.research.google.com/drive/17_wPYTS_tIV7dSXEYviM2egN3yDSDylz?usp=sharing

Model	Acc@1 (Validation)
TF-IDF baseline	0.4161
Text-based model	0.6993
Multimodal model	0.7068

Table 1
Validation performance of the evaluated models.

Run	Description	Task	pa-nDCG
TF-IDF baseline	Lexical matching baseline	Task 1	50.96
Text-based model	NOMIC text embeddings + LinearSVC	Task 1	64.47
Multimodal model	NOMIC text embeddings + CLIP image embeddings + LinearSVC	Task 2	64.47

Table 2
Official test results of the submitted runs in the PoliticHeadlinES shared task.

As shown in Table 1, the embedding-based text model clearly outperforms the TF-IDF baseline, confirming the importance of semantic representations for this task. The multimodal model achieves a slightly higher Acc@1 score on the validation split, increasing from 0.6993 to 0.7068. This suggests that the visual features may provide some useful article-level information in the internal validation setting.

However, this improvement does not translate into a higher official pa-nDCG score on the test set, where the multimodal run obtains the same score as the text-only model. Therefore, the validation results should be interpreted cautiously: although the multimodal features slightly improve top-1 accuracy on the validation split, the adopted early-fusion strategy does not provide a consistent gain in the official evaluation.

5.2. Test Results

Table 2 reports the official scores obtained by the evaluated runs. The TF-IDF baseline achieves a pa-nDCG score of 50.96, showing the limitations of lexical matching when candidate headlines share vocabulary with the article but differ in meaning. In contrast, the text-based embedding model reaches 64.47 pa-nDCG, confirming that dense semantic representations are more effective for capturing the relationship between news articles and their corresponding headlines.

The multimodal run obtains the same official score as the text-only model. This result indicates that, under the proposed early-fusion strategy, the visual representation does not provide additional ranking gains. As discussed in Section 5.3, this is likely related to the fact that the same image embedding is shared by all candidate headlines of the same article, limiting its ability to introduce candidate-specific differences in the ranking.

5.3. Analysis

To better understand the behavior of our models, we first analyze common prediction patterns from a qualitative perspective. We then examine a subset of validation examples to inspect the score distributions assigned to the candidate headlines. This allows us to compare the model prediction with the correct candidate and to analyze successful and failed rankings more clearly.

5.3.1. Qualitative Analysis

The TF-IDF baseline tends to favor candidate headlines with high lexical overlap with the article text, which can lead to incorrect rankings when distractors share similar vocabulary but differ in meaning. This limitation highlights the importance of semantic representations for this task.

The text-based model performs well when the correct headline is strongly aligned with the main topic of the article. In these cases, the embedding representation captures key semantic relationships,

including relevant entities and events. However, performance decreases when distractor headlines are semantically similar or differ only in subtle aspects of meaning or framing.

In the multimodal setting, the impact of visual features is limited under the proposed architecture. A key reason is that the image embedding is associated with the article and is therefore identical for the ten candidate headlines of the same instance. Since our multimodal model uses a linear classifier over the concatenated representation ($[x_t(a, h_i); x_v(a)]$), the score assigned to a candidate headline (h_i) can be expressed as:

$$s(a, h_i) = w_t^\top x_t(a, h_i) + w_v^\top x_v(a) + b, \tag{1}$$

where ($x_t(a, h_i)$) represents the text embedding of the article–headline pair and ($x_v(a)$) represents the visual embedding of the article image. For a fixed article (a), the term ($w_v^\top x_v(a) + b$) is shared by all candidate headlines. Therefore, the visual representation mainly acts as an article-level signal and does not explicitly capture whether a specific headline is more visually compatible with the associated image than the other candidates.

This limitation helps explain why the multimodal model does not improve over the text-only model. Although images may provide useful contextual information, the current early-fusion strategy is not expressive enough to model fine-grained image–headline interactions. More suitable alternatives would include computing explicit image–headline compatibility scores, using cross-modal attention mechanisms, or training pairwise/listwise multimodal ranking models.

Overall, these observations confirm that semantic matching approaches clearly outperform lexical baselines for headline ranking, while also highlighting the challenges of effectively leveraging multimodal information in this setting.

5.3.2. Prediction Score Analysis

To complement the qualitative observations, we analyze the distribution of prediction scores across candidate headlines using a heatmap representation (Figure 2). The heatmap is computed over validation examples, where the gold headline is available. This allows us to explicitly compare the model prediction with the correct candidate.

Each row corresponds to a validation article, while each column represents one of the ten candidate headlines. The color intensity reflects the score assigned by the model, where higher values indicate stronger relevance. The marker G indicates the gold headline, P indicates the model prediction, and G/P indicates cases where the prediction matches the gold headline.

The heatmap shows that the model correctly identifies the gold headline in several validation examples, as indicated by the G/P marker. In these cases, the correct candidate usually receives a clearly higher score than the remaining alternatives, suggesting that the embedding-based representation captures the main semantic alignment between the article and its original headline.

However, some errors remain. In several cases, the gold headline and the predicted headline both receive relatively high scores, but a plausible distractor is assigned a slightly higher value. This indicates that the model can often identify relevant candidates, but still struggles to select the exact original headline when several alternatives are semantically close or refer to similar political actors, events, or topics.

Overall, this analysis supports the quantitative results and provides further insight into the strengths and limitations of the proposed approach. The model is effective at separating clearly relevant candidates from weaker alternatives, but the main difficulty lies in discriminating among highly plausible headlines with subtle differences in framing, specificity, or topical focus.

6. Conclusion

In this paper, we presented the participation of the UIE team in the PoliticHeadlinES shared task at IberLEF 2026, addressing both text-based and multimodal candidate headline ranking for Spanish

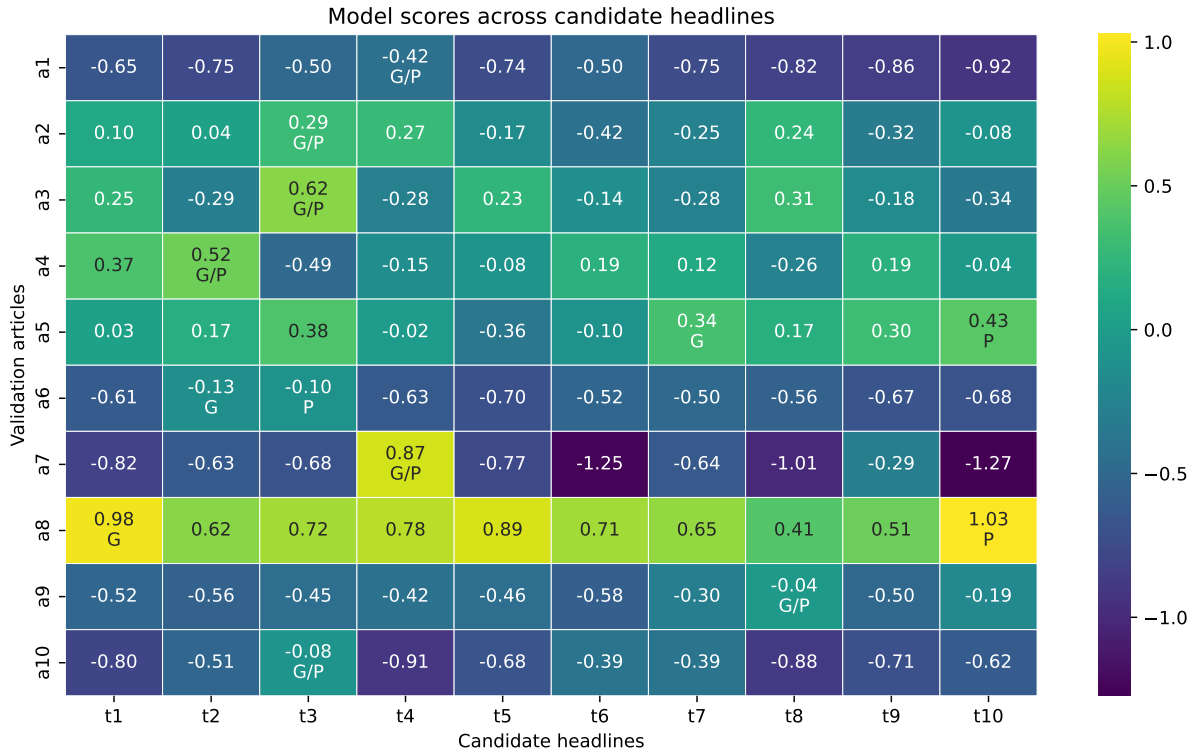

Figure 2: Heatmap of model scores assigned to candidate headlines for a subset of validation examples. Each row corresponds to a validation article, and each column to one of the ten candidate headlines. The marker G indicates the gold headline, P indicates the model prediction, and G/P indicates cases where the prediction matches the gold headline.

political news. Our approach formulates the task as a pointwise learning-to-rank problem, where each article–headline pair is independently represented and scored.

The proposed system combines pretrained semantic embeddings with a lightweight linear classifier. In the text-only setting, this approach clearly improves over the TF-IDF baseline, achieving a pa-nDCG score of 64.47 compared with 50.96. This confirms that dense semantic representations are more effective than lexical matching for distinguishing between plausible candidate headlines.

In the multimodal setting, we incorporated CLIP-based visual embeddings through feature concatenation. However, this strategy did not improve the official score over the text-only model. This result suggests that, although images may provide useful contextual information, simple early fusion is not sufficient to capture fine-grained candidate-specific interactions between the article image and each headline.

Overall, our results position embedding-based ranking as a competitive and efficient baseline for the PoliticHeadlinES task, while also highlighting the limitations of linear early-fusion multimodal models. Future work should explore more expressive ranking formulations, such as pairwise or listwise learning-to-rank approaches, as well as multimodal architectures that explicitly model image–headline compatibility through cross-modal attention or contrastive scoring mechanisms.

Declaration on Generative AI

During the preparation of this work, the authors used DeepL for grammar and spelling checking.

References

[1] J. Gómez-Navalón, T. Bernal-Beltrán, R. Pan, G.-S. Francisco, J. A. García-Díaz, R. Valencia-García, Overview of PoliticHeadlinES at IberLEF 2026: Multimodal Headline Ranking in Spanish Political News , Procesamiento del Lenguaje Natural 77 (2026).

[2] A. Bonet-Jover, J. Á. González-Barba, L. Chiruzzo, Overview of IberLEF 2026: Natural Language Processing Challenges for Spanish and other Iberian Languages, in: Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2026), co-located with the 42nd Conference of the Spanish Society for Natural Language Processing (SEPLN 2026), CEUR-WS. org, 2026.

[3] M. Banko, V. O. Mittal, M. J. Witbrock, Headline generation based on statistical translation, in: Proceedings of the 38th Annual Meeting of the Association for Computational Linguistics, 2000, pp. 318–325.

[4] K. Yamada, Y. Hitomi, H. Tamori, R. Sasano, N. Okazaki, K. Inui, K. Takeda, Transformer-based lexically constrained headline generation, in: Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, 2021, pp. 4085–4090.

[5] P. Cai, K. Song, S. Cho, H. Wang, X. Wang, H. Yu, F. Liu, D. Yu, Generating user-engaging news headlines, in: Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 2023, pp. 3265–3280.

[6] R. Pan, T. Bernal-Beltrán, J. Gómez-Navalon, J. A. García-Díaz, R. Valencia-García, Spanish polisum-2025: Evaluating stylistic and editorial fidelity in abstractive summarization of political news, Procesamiento del Lenguaje Natural 76 (2026) 177–189.

[7] N. Reimers, I. Gurevych, Sentence-bert: Sentence embeddings using siamese bert-networks, in: Proceedings of the 2019 conference on empirical methods in natural language processing and the 9th international joint conference on natural language processing (EMNLP-IJCNLP), 2019, pp. 3982–3992.

[8] N. Muennighoff, N. Tazi, L. Magne, N. Reimers, Mteb: Massive text embedding benchmark, in: Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics, 2023, pp. 2014–2037.

[9] Z. Nussbaum, J. X. Morris, B. Duderstadt, A. Mulyar, Nomic embed: Training a reproducible long context text embedder, arXiv preprint arXiv:2402.01613 (2024).

[10] L. Hollink, A. Bedjeti, M. Van Harmelen, D. Elliott, A corpus of images and text in online news, in: Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16), 2016, pp. 1377–1382.

[11] N. Oostdijk, H. van Halteren, E. Bașar, M. Larson, The connection between the text and images of news articles: New insights for multimedia analysis, in: Proceedings of the Twelfth Language Resources and Evaluation Conference, 2020, pp. 4343–4351.

[12] R. Pan, J. A. García-Díaz, M. Á. Rodríguez-García, R. Valencia-García, Spanish meacorpus 2023: A multimodal speech–text corpus for emotion analysis in spanish from natural environments, Computer Standards & Interfaces 90 (2024) 103856.

[13] M. Krubiński, P. Pecina, Mlask: Multimodal summarization of video-based news articles, in: Findings of the association for computational linguistics: EACL 2023, 2023, pp. 910–924.

[14] M. Krubiński, P. Pecina, Towards unified uni-and multi-modal news headline generation, in: Findings of the Association for Computational Linguistics: EACL 2024, 2024, pp. 437–450.

[15] R. Pan, J. A. García-Díaz, R. Valencia-García, A dynamic multimodal causal graph framework for standardized emotion recognition in conversations, Computer Standards & Interfaces (2026) 104130.

[16] A. Radford, J. W. Kim, C. Hallacy, A. Ramesh, G. Goh, S. Agarwal, G. Sastry, A. Askell, P. Mishkin, J. Clark, et al., Learning transferable visual models from natural language supervision, in: International conference on machine learning, PmLR, 2021, pp. 8748–8763.

[17] X. Zhai, B. Mustafa, A. Kolesnikov, L. Beyer, Sigmoid loss for language image pre-training, in: Proceedings of the IEEE/CVF international conference on computer vision, 2023, pp. 11975–11986.