PUCE-UTP at PoliticHeadlinES-IberLEF 2026: A Hybrid TF-IDF, SBERT and CLIP Framework for Multimodal Headline Ranking

Alexander José Mackenzie-Rivero[1,*,†], Denis Cedeño-Moreno[2,†], Hilarión José Vegas-Meléndez[1,†] and Juan Pablo Morales-Corozo[1,†]

[1]*Pontificia Universidad Católica del Ecuador sede Manabí, Portoviejo, 130103, Manabí, Ecuador*
[2]*Universidad Tecnológica de Panamá, Ciudad de Panamá, Panamá*

Abstract

The increasing influence of political news in shaping public opinion highlights the importance of accurately identifying reliable and representative headlines. In this work, we present the participation of the PUCE-UTP team in the PoliticHeadlinES 2026 shared task, which focuses on ranking candidate headlines for Spanish political news articles under both textual and multimodal settings.

For Task 1 (text-based ranking), we propose a hybrid model that combines lexical representations based on Term Frequency–Inverse Document Frequency (TF-IDF) with semantic embeddings derived from Sentence-BERT (SBERT). This approach leverages both surface-level term matching and deep contextual similarity.

For Task 2 (multimodal ranking), we extend the model by incorporating visual information using the CLIP (Contrastive Language-Image Pretraining) model. A weighted fusion strategy is employed to combine textual and visual similarity scores, where the textual component dominates due to its higher discriminative power.

Experimental results demonstrate that the TF-IDF model achieves the highest Top-1 accuracy, while the hybrid TF-IDF + SBERT model improves overall ranking quality measured by PA-nDCG. In the multimodal setting, the inclusion of visual features provides marginal improvements, confirming that textual information remains the primary source of discriminative signal in political news.

These findings highlight the effectiveness of hybrid approaches that integrate lexical, semantic, and multimodal representations for headline ranking tasks.

Keywords

Headline ranking, Natural Language Processing, Multimodal learning, TF-IDF, SBERT, CLIP, Political news analysis

1. Introduction

The rapid dissemination of political news through digital media has significantly increased the importance of accurately summarizing and representing information in the form of headlines. Headlines play a critical role in shaping public perception, influencing readers' attention, and framing the interpretation of news content. However, selecting the most appropriate headline among multiple plausible candidates remains a challenging task, particularly in the presence of subtle semantic differences and stylistic variations.

The PoliticHeadlinES 2026 shared task addresses this problem by formulating headline selection as a ranking task, where systems must identify the most suitable headline among a set of ten candidates for a given news article [14]. This task is particularly challenging because all candidate headlines are designed to be plausible, requiring models to capture not only lexical similarity but also deeper semantic coherence and contextual relevance.

IberLEF 2026, September 2026, León, Spain
*Corresponding author.
†These authors contributed equally.
ajmackenzie@pucesm.edu.ec (A. J. Mackenzie-Rivero); denis.cedeno@utp.ac.pa (D. Cedeño-Moreno); hvegas@pucesm.edu.ec (H. J. Vegas-Meléndez); jpmoralesc@pucesm.edu.ec (J. P. Morales-Corozo)
0009-0001-4641-4490 (A. J. Mackenzie-Rivero); 0000-0002-9640-1284 (D. Cedeño-Moreno); 0000-0002-8526-2979 (H. J. Vegas-Meléndez); 0000-0002-4538-4488 (J. P. Morales-Corozo)

Traditional approaches based on lexical matching, such as TF-IDF, have proven to be strong baselines due to their ability to capture explicit term overlap between the article and candidate headlines [1]. However, these methods often fail to account for semantic relationships beyond surface-level text similarity. In contrast, recent advances in Natural Language Processing, particularly those based on Transformer architectures [4], enable the modeling of contextual and semantic information through dense vector representations.

In addition to textual information, the multimodal nature of modern news articles introduces further complexity. Images associated with news content can provide complementary contextual cues, but their contribution to headline selection is often ambiguous. While multimodal models such as CLIP have demonstrated strong performance in aligning visual and textual representations [3], effectively integrating this information into ranking tasks remains an open research challenge.

In this work, we propose a hybrid approach that combines lexical, semantic, and multimodal representations to address both tasks of the shared challenge. For Task 1, we integrate TF-IDF with SBERT embeddings to capture both lexical similarity and semantic relevance. For Task 2, we extend this approach by incorporating visual information through CLIP, using a weighted fusion strategy to balance textual and visual signals.

The main contributions of this work can be summarized as follows:

- We propose a hybrid ranking framework that combines TF-IDF and SBERT for text-based headline ranking.
- We extend the framework to a multimodal setting by integrating CLIP-based image–text similarity.
- We perform an empirical analysis of the contribution of textual and visual features, demonstrating that textual information remains the dominant signal in political news.
- We provide a competitive baseline for the PoliticHeadlinES 2026 shared task, highlighting the strengths and limitations of hybrid approaches.

The remainder of this paper is organized as follows. Section 2 reviews related work. Section 3 describes the methodology used for both tasks. Section 4 presents the experimental setup and evaluation metrics. Section 5 discusses the obtained results, and Section 6 concludes the paper with future research directions.

2. Related Work

Headline ranking is closely related to information retrieval, semantic textual similarity, and neural reranking. Traditional lexical models, such as TF-IDF, remain effective baselines because they capture explicit term overlap between a document and a candidate text [1]. However, recent advances in Transformer-based models have enabled richer contextual representations for semantic matching and ranking tasks [4, 2].

Neural retrieval and reranking methods have shown strong performance by modeling deeper interactions between queries and candidate documents. Dense Passage Retrieval (DPR) introduced dense representations for retrieval in open-domain question answering, demonstrating the usefulness of learned semantic representations for ranking candidates [6]. Similarly, late-interaction models such as ColBERT provide an efficient mechanism for combining contextual embeddings with fine-grained token-level matching [7]. These approaches motivate the use of hybrid ranking strategies that combine sparse lexical signals with dense semantic representations.

In the specific context of news headline analysis, several studies have explored neural architectures for classifying and ranking short texts. Khuntia et al. proposed the use of word embeddings and LSTM-based models for Indian news headline classification, showing the relevance of sequential neural models for headline understanding [9]. Li and Cao introduced a dual-channel architecture based on ERNIE for news headline classification, combining contextual representations with BiLSTM and attention mechanisms to capture both global and local textual features [8]. More recently, instruction-tuned

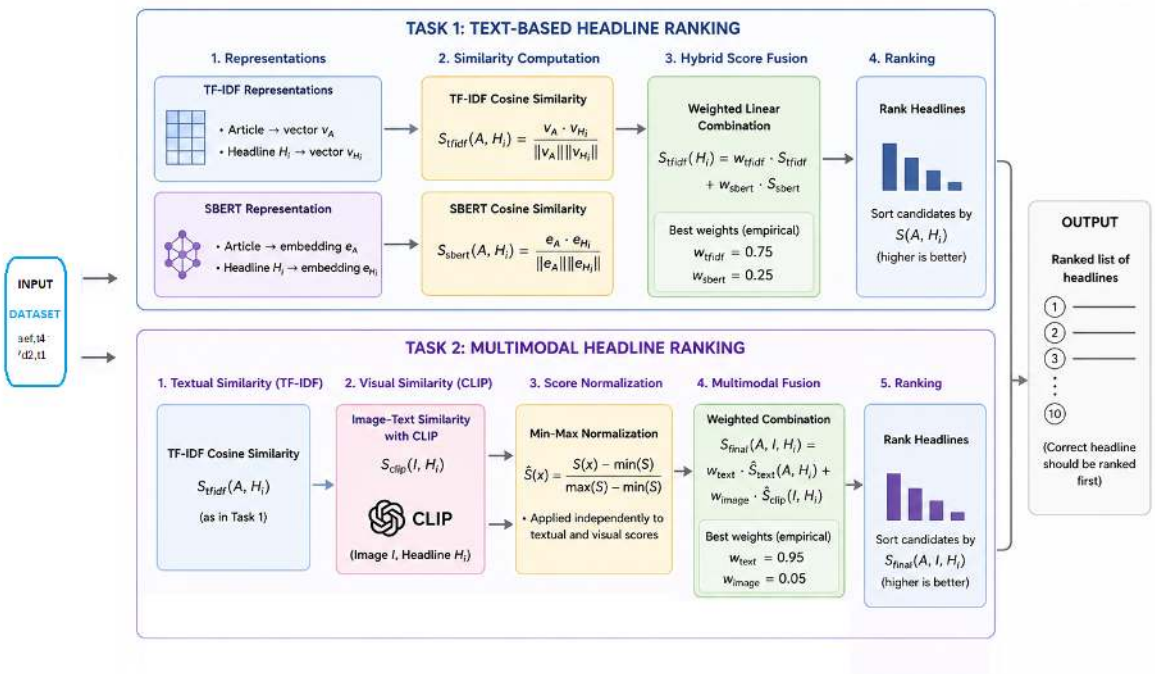

Figure 1: Overview of the proposed methodology for headline ranking. The framework integrates lexical (TF-IDF), semantic (SBERT), and visual (CLIP) representations. Task 1 focuses on text-based ranking, while Task 2 extends the model to a multimodal setting using weighted score fusion.

language models have also been explored for specialized text classification tasks, suggesting that model adaptation and fusion strategies can improve robustness across domains [12].

Headline interpretation is also affected by rhetorical and pragmatic phenomena such as sarcasm, ambiguity, and editorial framing. Ahmad and Arts studied sarcasm identification in Hindi political news headlines using machine learning and deep learning approaches, highlighting the difficulty of distinguishing literal from non-literal meanings in political discourse [11]. These challenges are relevant to headline ranking because candidate headlines may be lexically or semantically similar while differing in intent, framing, or factual correspondence.

The multimodal nature of news has further motivated the integration of textual and visual signals. Multimodal resources such as Spanish MEACorpus 2023 demonstrate the relevance of combining speech, text, and visual or contextual information for Spanish-language analysis tasks [10]. In this direction, CLIP provides a general framework for aligning images and texts in a shared embedding space, enabling the computation of image–headline similarity [3]. Nevertheless, effectively integrating visual information into political headline ranking remains challenging, since news images often provide general context rather than fine-grained evidence for selecting the correct headline.

The PoliticHeadlinES 2026 shared task, organized within IberLEF 2026, provides a benchmark for evaluating these challenges in Spanish political news [14, 15]. Our work follows this line by combining lexical, semantic, and multimodal signals in a lightweight hybrid framework for both text-based and multimodal headline ranking.

3. Methodology

In this section, we describe the proposed approach for both tasks of the PoliticHeadlinES 2026 shared challenge. Our method is based on a hybrid framework that integrates lexical, semantic, and multimodal representations to rank candidate headlines. Figure 1 illustrates the overall architecture of the proposed framework, including both the text-based and multimodal ranking strategies.

3.1. Problem Formulation

Given a news article A and a set of ten candidate headlines $\{H_1, H_2, \ldots, H_{10}\}$, the objective is to produce a ranking π over the candidates such that the correct headline is placed as high as possible. This task is formulated as a ranking problem rather than a classification problem, where each candidate headline is assigned a relevance score with respect to the article.

3.2. Task 1: Text-Based Headline Ranking

For the text-based setting, we propose a hybrid approach that combines lexical similarity with semantic representations.

3.2.1. TF-IDF Representation

We first represent the article A and each candidate headline H_i using Term Frequency–Inverse Document Frequency (TF-IDF) [1]. This representation captures the importance of terms within the corpus and allows us to compute similarity based on word overlap.

The lexical similarity score is computed using cosine similarity:

$$S_{tfidf}(A, H_i) = \frac{\mathbf{v}_A \cdot \mathbf{v}_{H_i}}{\|\mathbf{v}_A\| \|\mathbf{v}_{H_i}\|} \tag{1}$$

where $\mathbf{v}_A$ and $\mathbf{v}_{H_i}$ denote the TF-IDF vectors of the article and the i-th headline, respectively.

3.2.2. Semantic Representation with SBERT

To capture deeper semantic relationships, we use Sentence-BERT (SBERT), a Transformer-based model that generates dense sentence embeddings [2]. Each article and headline is encoded into a vector representation:

$$\mathbf{e}_A = \text{SBERT}(A), \quad \mathbf{e}_{H_i} = \text{SBERT}(H_i) \tag{2}$$

The semantic similarity is then computed using cosine similarity:

$$S_{sbert}(A, H_i) = \frac{\mathbf{e}_A \cdot \mathbf{e}_{H_i}}{\|\mathbf{e}_A\| \|\mathbf{e}_{H_i}\|} \tag{3}$$

3.2.3. Hybrid Score Fusion

The final relevance score for each candidate headline is obtained by combining the lexical and semantic scores through a weighted linear combination:

$$S(A, H_i) = w_{tfidf} \cdot S_{tfidf}(A, H_i) + w_{sbert} \cdot S_{sbert}(A, H_i) \tag{4}$$

where w_{tfidf} and w_{sbert} are weighting parameters. In our experiments, the hybrid configuration used $w_{tfidf} = 0.75$ and $w_{sbert} = 0.25$.

3.3. Task 2: Multimodal Headline Ranking

For the multimodal setting, the hybrid TF-IDF + CLIP model achieved a Top-1 Accuracy of 0.5796 and a PA-nDCG score of 0.934169. Although the multimodal configurations obtained lower Top-1 Accuracy than the best text-based model, their PA-nDCG values indicate that the correct headline was often placed relatively high in the ranked list.

3.3.1. Image–Text Similarity with CLIP

CLIP (Contrastive Language–Image Pretraining) maps images and text into a shared embedding space [3]. Given an image I associated with the article and a candidate headline H_i, we compute:

$$S_{clip}(I, H_i) = \text{CLIP}(I, H_i) \tag{5}$$

This score reflects the compatibility between the visual content and the candidate headline.

3.3.2. Score Normalization

Since TF-IDF and CLIP produce scores on different scales, we apply Min-Max normalization to ensure comparability:

$$\hat{S}(x) = \frac{S(x) - \min(S)}{\max(S) - \min(S)} \tag{6}$$

This normalization is applied independently to the textual and visual scores for each instance.

3.3.3. Multimodal Fusion

The final multimodal score is computed as a weighted combination of textual and visual signals:

$$S_{final}(A, I, H_i) = w_{text} \cdot \hat{S}_{text}(A, H_i) + w_{image} \cdot \hat{S}_{clip}(I, H_i) \tag{7}$$

where $\hat{S}_{text}$ corresponds to the normalized TF-IDF score, and $\hat{S}_{clip}$ corresponds to the normalized CLIP score.

The weights were optimized empirically, resulting in $w_{text} = 0.95$ and $w_{image} = 0.05$, indicating that textual information is the dominant factor in headline ranking.

3.3.4. Fallback Strategy

In cases where the associated image is missing or cannot be processed, the model falls back to a text-only ranking using the TF-IDF component. This ensures robustness and avoids performance degradation due to incomplete multimodal data.

4. Experimental Setup

This section describes the dataset, evaluation metrics, implementation details, and experimental configuration used to assess the proposed models.

4.1. Dataset

We conducted our experiments using the dataset provided by the organizers of the PoliticHeadlinES 2026 shared task [14, 13]. The dataset consists of Spanish political news articles, each associated with ten candidate headlines, where exactly one corresponds to the correct headline.

Each instance includes the following fields: an article body, an image identifier, and ten candidate headlines. In the multimodal setting, each article is also associated with an image, which provides additional contextual information.

The dataset is designed to be challenging, as all candidate headlines are plausible and stylistically similar, requiring models to capture fine-grained semantic differences to correctly rank them.

4.2. Evaluation Metrics

Since the task is formulated as a ranking problem, we employed evaluation metrics specifically designed to assess ranking quality:

- **Top-1 Accuracy**: measures the proportion of instances in which the correct headline is ranked in the first position.
- **PA-nDCG (Position-Aware Normalized Discounted Cumulative Gain)**: evaluates the overall ranking quality by considering the position of the correct headline within the ranked list and penalizing lower placements.

Top-1 Accuracy is particularly relevant for this task, as the evaluation emphasizes the correct identification of the best headline.

4.3. Implementation Details

The proposed models were implemented in Python using widely adopted libraries for machine learning and natural language processing.

TF-IDF representations were computed using *scikit-learn* [1], while semantic embeddings were generated using a pre-trained SBERT model (*paraphrase-multilingual-MiniLM-L12-v2*) [2]. For the multimodal component, we employed the CLIP model (*openai/clip-vit-base-patch32*) using the Hugging Face Transformers library [3].

All experiments were conducted in a Google Colab environment with GPU support when available. Data preprocessing and manipulation were performed using *pandas* and *NumPy*, and image processing was handled using *Pillow*.

4.3.1. Text Preprocessing and Input Length Management

Prior to feature extraction, all article bodies and candidate headlines were normalized using a lightweight preprocessing pipeline. Specifically, texts were converted to lowercase, URLs were removed, and consecutive whitespace characters were collapsed into a single space. No stemming, lemmatization, stop-word removal, or accent normalization was applied.

For the TF-IDF component, unigram and bigram features were extracted using the default tokenization strategy provided by scikit-learn. The vocabulary size was limited to 50,000 terms.

To reduce noise and maintain computational efficiency, article bodies were truncated to the first 1,200 characters before feature extraction. This strategy was applied consistently across all experiments and was motivated by the observation that journalistic articles typically concentrate the most relevant information in the introductory paragraphs.

For the semantic component, the truncated article text and each candidate headline were encoded using the pre-trained SBERT model *paraphrase-multilingual-MiniLM-L12-v2*. No chunking or sliding-window strategy was employed.

For the multimodal setting, image–text similarity was computed using the CLIP model *openai/clip-vit-base-patch32*. Images were processed using the default CLIP preprocessing pipeline, while headline texts were automatically truncated according to the model's maximum input length constraints.

The same preprocessing and truncation strategy was applied to both the training and evaluation datasets to ensure consistency and reproducibility.

4.4. Experimental Configuration

For Task 1, the TF-IDF vectorizer was configured using unigram and bigram features, with a maximum vocabulary size of 50,000 terms. The hybrid model combined TF-IDF and SBERT scores using a weighted linear combination, with weights set to $w_{tfidf} = 0.75$ and $w_{sbert} = 0.25$, based on empirical evaluation.

For Task 2, the multimodal model combined textual and visual similarity scores. The textual component was derived from TF-IDF, while the visual component was obtained using CLIP. Before fusion, both scores were normalized using Min-Max scaling.

The final multimodal score was computed using a weighted combination, with parameters $w_{text} = 0.95$ and $w_{image} = 0.05$, determined through a grid search optimization process.

In cases where images were missing or could not be processed, the model defaulted to a text-only ranking strategy to ensure robustness.

4.5. Reproducibility

To ensure reproducibility, all models were trained exclusively on the training dataset provided by the organizers, and no external data was used. The same preprocessing pipeline and parameter configuration were applied consistently across all experiments.

4.6. Submitted Runs

For Task 1, three ranking approaches were evaluated during system development: a lexical baseline based on TF-IDF, a semantic model based on SBERT embeddings, and a hybrid TF-IDF + SBERT model. The official submission corresponded to the hybrid configuration using weights of 0.75 for TF-IDF and 0.25 for SBERT. This configuration was selected because it achieved the highest overall ranking quality measured by PA-nDCG while maintaining competitive Top-1 Accuracy.

For Task 2, several multimodal fusion strategies were explored by combining TF-IDF textual similarity with CLIP image–text similarity. Different text–image weight combinations were evaluated on the development set. The official submission corresponded to the configuration with a textual weight of 0.95 and an image weight of 0.05, which achieved the best performance among the evaluated multimodal variants.

Tables 1 and 2 summarize both the official submissions and the main alternative configurations considered during system development.

5. Results and Discussion

In this section, we present and analyze the results obtained for both tasks of the PoliticHeadlinES 2026 shared challenge. The evaluation was performed using Top-1 Accuracy and PA-nDCG, which measure the ability of the model to correctly identify the best headline and the overall quality of the ranking, respectively.

5.1. Task 1: Text-Based Headline Ranking

Table 1
Results for Task 1 (Text-Based Headline Ranking)

Run	Model	Weights	Official	Top-1	PA-nDCG
Run-1	TF-IDF	–	No	0.87	0.977377
Run-2	SBERT	–	No	0.66	0.970164
Run-3	TF-IDF + SBERT	0.75/0.25	Yes	0.86	0.978221

The results show that the TF-IDF model achieves the highest Top-1 Accuracy (0.87), indicating its strong capability to correctly identify the most relevant headline. This highlights the effectiveness of lexical matching in scenarios where the correct headline shares explicit terms with the article.

The hybrid model combining TF-IDF and SBERT slightly improves the overall ranking quality, achieving the highest PA-nDCG score (0.978221). This suggests that semantic representations contribute to a better global ordering of candidate headlines, even though they may introduce minor ambiguities when selecting the top-ranked candidate.

In contrast, the SBERT-only model exhibits significantly lower performance, demonstrating that semantic similarity alone is insufficient for this task, particularly when precise lexical alignment is required.

5.2. Task 2: Multimodal Headline Ranking

Table 2 presents the results obtained for Task 2.

Table 2
Results for Task 2 (Multimodal Headline Ranking)

Run	Text Model	Visual Model	Text W.	Image W.	Official	Top-1	PA-nDCG
MM-1	TF-IDF	CLIP	0.90	0.10	No	0.5796	0.934169
MM-2	TF-IDF	CLIP	0.95	0.05	Yes	0.5796	0.934169

For the multimodal setting, both evaluated configurations obtained a Top-1 Accuracy of 0.5796 and a PA-nDCG score of 0.934169. While these results did not surpass the performance achieved by the text-based approaches, the relatively high PA-nDCG values suggest that the correct headline was often placed among the highest-ranked candidates. The optimization process revealed that textual information remained the dominant signal in multimodal headline ranking. The official configuration assigned a weight of 0.95 to the textual component and 0.05 to the visual component, indicating that visual information provided only limited complementary evidence for this task.

This finding suggests that, in the domain of political news, images provide complementary context but do not contain sufficient discriminative information to significantly improve headline selection. In many cases, multiple candidate headlines may be visually consistent with the associated image, limiting the contribution of visual features.

5.3. Comparative Analysis

Across both tasks, the results indicate that lexical features remain the most reliable signal for headline ranking. While semantic representations improved ranking quality in Task 1, the multimodal configurations did not improve Top-1 Accuracy in Task 2, although they maintained relatively high PA-nDCG values.

The hybrid approaches demonstrate that combining multiple representations can improve robustness, particularly in terms of ranking quality (PA-nDCG). However, achieving improvements in Top-1 Accuracy remains challenging, as it requires precise discrimination among highly similar candidates.

5.4. Discussion

Overall, the proposed approach provides a competitive baseline for the shared task. The results confirm that:

- Lexical similarity (TF-IDF) is the strongest signal for Top-1 prediction.
- Semantic embeddings (SBERT) improve the overall ranking quality.
- Multimodal information (CLIP) provides limited but consistent complementary benefits.

Despite these findings, the results also reveal limitations of the proposed approach. In particular, the reliance on independent similarity scoring prevents the model from fully capturing complex interactions between the article and candidate headlines. More advanced approaches, such as cross-encoder architectures or neural ranking models [5, 7], could potentially improve performance by modeling deeper contextual relationships.

These observations highlight important directions for future work, especially in the integration of multimodal information and the development of more sophisticated ranking strategies.

6. Conclusion

In this work, we presented the participation of the PUCE-UTP team in the PoliticHeadlinES 2026 shared task, addressing both text-based and multimodal headline ranking. We proposed a hybrid framework that integrates lexical, semantic, and visual representations to rank candidate headlines associated with political news articles.

For Task 1, the results demonstrated that TF-IDF remains a strong baseline, achieving the highest Top-1 Accuracy due to its ability to capture explicit lexical overlap between the article and the correct headline. The incorporation of SBERT embeddings improved the overall ranking quality, as reflected in higher PA-nDCG scores, highlighting the contribution of semantic information in refining the ordering of candidate headlines.

For Task 2, we extended the model by incorporating visual features through the CLIP model. The results showed that multimodal fusion did not improve Top-1 Accuracy over the text-based configurations, although it still achieved a high PA-nDCG score. This suggests that visual information may provide complementary context but was not sufficiently discriminative for selecting the correct headline in most cases. The official configuration confirmed that a high weight assigned to textual similarity was necessary to obtain the best multimodal performance.

Overall, our findings indicate that hybrid approaches combining lexical and semantic representations offer a robust solution for headline ranking tasks. However, the limited impact of visual features suggests that more sophisticated multimodal integration strategies are required to fully exploit image information in this domain.

Future work will focus on exploring advanced neural ranking architectures, such as cross-encoders and transformer-based reranking models, as well as more effective methods for multimodal fusion. Additionally, incorporating contextual knowledge and discourse-level features may further improve the ability to discriminate among highly similar candidate headlines.

These results contribute to a better understanding of the role of lexical, semantic, and multimodal representations in headline ranking, and provide a solid baseline for future research in this area.

Acknowledgments

The authors express their sincere gratitude to the Pontificia Universidad Católica del Ecuador, Sede Manabí, for the institutional, academic, and logistical support provided during the development of this work. The authors also acknowledge the support provided by the National Research System (SENACYT-SNI).

Author Denis Cedeño-Moreno thanks the National Research System of Panama (SENACYT, SNI) for the support provided in the development of this research.

Declaration on Generative AI

During the preparation of this work, the authors used generative AI tools to assist with language editing, code refinement, and manuscript organization. The authors reviewed and edited all generated content and take full responsibility for the final version of the paper.

References

[1] C. D. Manning, P. Raghavan, and H. Schütze. Introduction to Information Retrieval. Cambridge University Press, 2008.

[2] N. Reimers and I. Gurevych. Sentence-BERT: Sentence embeddings using Siamese BERT-networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2019.

[3] A. Radford, J. W. Kim, C. Hallacy, A. Ramesh, G. Goh, S. Agarwal, G. Sastry, A. Askell, P. Mishkin, J. Clark, G. Krueger, and I. Sutskever. Learning Transferable Visual Models From Natural Language Supervision. In *Proceedings of the 38th International Conference on Machine Learning (ICML)*, 2021.

[4] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *Proceedings of NAACL*, 2019.

[5] R. Nogueira and K. Cho. Passage Re-ranking with BERT. *arXiv preprint arXiv:1901.04085*, 2019.

[6] V. Karpukhin, B. Oguz, S. Min, P. Lewis, L. Wu, S. Edunov, D. Chen, and W.-t. Yih. Dense Passage Retrieval for Open-Domain Question Answering. In *Proceedings of EMNLP*, 2020.

[7] O. Khattab and M. Zaharia. ColBERT: Efficient and Effective Passage Search via Contextualized Late Interaction. In *Proceedings of SIGIR*, 2020.

[8] J. Li and H. Cao. Research on dual-channel news headline classification based on ERNIE pre-training model. *Computer Science & Information Technology*, pp. 31–43, 2022.

[9] M. Khuntia, M. Khuntia, and D. Gupta. Indian News Headlines Classification using Word Embedding Techniques and LSTM Model. *Procedia Computer Science*, 218:899–907, 2023.

[10] R. Pan, J. A. García-Díaz, M. Rodríguez-García, and R. Valencia-García. Spanish MEACorpus 2023: A multimodal speech–text corpus for emotion analysis in Spanish. *Computer Standards & Interfaces*, 90:103856, 2024.

[11] I. A. L. I. Ahmad and F. Arts. Sarcasm Identification and Classification in Hindi News Headlines. *Journal of Language Processing*, 24(4), 2025.

[12] S. Fatemi, Y. Hu, and M. Mousavi. A Comparative Analysis of Instruction Fine-Tuning in Large Language Models. *Journal of Artificial Intelligence Research*, 16(1):1–30, 2025.

[13] R. Pan, T. Bernal-Beltrán, J. Gómez-Navalón, J. A. García-Díaz, and R. Valencia-García. Spanish PoliSUM-2025: Evaluating Stylistic and Editorial Fidelity in Abstractive Summarization of Political News. *Procesamiento del Lenguaje Natural*, 76:177–189, 2026.

[14] J. Gómez-Navalón, T. Bernal-Beltrán, R. Pan, F. García-Sánchez, J. A. García-Díaz, and R. Valencia-García. Overview of PoliticHeadlinES at IberLEF 2026: Multimodal Headline Ranking in Spanish Political News. *Procesamiento del Lenguaje Natural*, 77, 2026.

[15] A. Bonet-Jover, J. A. González-Barba, and 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.