Lightweight Semantic Span Retrieval for Outcome-Oriented Expectation Analysis: Two Systems for HOPE-EXP 2026 at IberLEF 2026

Ismael Jacinto Dias Meque[1], Sara Barahouei Nouri[2], Abdollah Abadian[3,*], Abdullah Abdullah[3], Grigori Sidorov[3] and Olga Kolesnikova[3]

[1]Catholic University of Mozambique, Mozambique
[2]Islamic Azad University Zahedan Branch, Zahedan, Iran
[3]Center for Computing Research, Instituto Politécnico Nacional (IPN), Mexico City 07320, Mexico

Abstract

This paper describes our participation in the HOPE-EXP 2026 shared task at IberLEF 2026 [1, 2]. The task requires multi-level outcome-oriented expectation analysis of social media posts, encompassing six-class primary label classification (hope types, hopelessness, not hope), multi-label emotion detection, span extraction of expected or avoided outcomes, and role classification (outcome stance and actor). We developed two systems. Run 1 is a lightweight baseline that employs keyword-based span extraction and hard-coded role predictions. Run 2 is an advanced system that replaces keyword matching with semantic similarity search using a multilingual Sentence Transformer (`paraphrase-multilingual-mpnet-base-v2`) and trains dedicated SVM classifiers to predict outcome stance and actor from span embeddings. Both systems use SVMs with RBF kernels for Tasks A and B. Our best system (Run 2) achieved an overall score of 0.58425, ranking 7th out of 8 participants. While competitive on primary label classification (macro F1 = 0.7975), significant room for improvement remains in emotion detection (macro F1 = 0.5513), span extraction (ROUGE-1 = 0.5146), and especially actor classification (macro F1 = 0.4700). Error analysis reveals confusion between hope subtypes, difficulty with sarcasm, and a tendency to default to "World/System" when spans lack explicit pronouns. Our results suggest that semantic span retrieval is a viable alternative to keyword matching, but more sophisticated methods—such as fine-tuned multilingual LLMs or ensemble approaches—are needed to close the gap with top-performing teams.

Keywords

hope speech detection, outcome-oriented expectation analysis, semantic similarity, span extraction, multilingual Sentence Transformer, support vector machine, social media analysis, IberLEF 2026

1. Introduction

The exponential growth of social media has transformed how individuals express emotions and expectations online. While much research has focused on detecting negative content such as hate speech, the identification of positive and encouraging expressions—termed "hope speech"—has emerged as an equally important research direction. Sidorov et al. [3] introduced MIND-HOPE, the first multi-class hope speech detection datasets for Spanish and German, categorizing hope into Generalized, Realistic, and Unrealistic types. They showed that monolingual BERT-based models outperform multilingual models (e.g., mBERT, XLM-RoBERTa), achieving F1-scores of 0.8704 (German) and 0.8387 (Spanish) for binary classification.

Extending this line of work, Ahmad et al. [4] created a multilingual Urdu-English dataset using the Google Translate API and a BERT-based model, achieving 87% accuracy for English and 79% for Urdu, surpassing traditional SVM baselines. Their work highlighted the value of rigorous annotation guidelines and translation-based techniques for bridging linguistic gaps.

IberLEF 2026, September 2026, León, Spain
*Corresponding author.
701210417@ucm.ac.mz (I. J. D. Meque); barahoueisarah@gmail.com (S. B. Nouri); aabdollaha26@cic.ipn.mx (A. Abadian); sidorov@cic.ipn.mx (G. Sidorov); kolesnikova@cic.ipn.mx (O. Kolesnikova)
0009-0003-4202-8159 (I. J. D. Meque); 0009-0002-7747-1612 (S. B. Nouri); 0009-0004-6581-4053 (A. Abadian); 0000-0003-3901-3522 (G. Sidorov); 0000-0002-1307-1647 (O. Kolesnikova)

Mental health detection from social media has also attracted considerable attention. Kumar et al. [5] proposed RAMHA, a hybrid model combining RoBERTa, adapter layers, BiLSTM, and attention with focal loss. Evaluated on GoEmotions datasets, RAMHA achieved 92% accuracy (binary) and 88% (multi-class), outperforming eight baselines. Sasikala [6] introduced a compassionate AI framework for detecting depression and suicidal risk using linguistic pattern analysis, affective computing, and deep learning (Bi-LSTM, BERT, attention), with emphasis on ethical data handling.

Beyond hope and mental health, stance detection determines an author's viewpoint towards a target. Pangtey et al. [7] surveyed LLM-based stance detection across learning methods, data modalities, and target relationships, noting that BERT and RoBERTa achieve macro F1 scores of 0.74–0.85 on standard benchmarks. Moving to finer-grained analysis, Drašković and Milanović [8] applied aspect-based sentiment analysis (ABSA) to over 547,000 COVID-19 tweets using a BERT model with Local Context Focus (LCF), finding that 34.97% of tweets contained at least two aspects, demonstrating the necessity of aspect-level analysis. Focusing on Arabic, Alkhathlan et al. [9] introduced ArabicStanceX (14,477 tweets, 17 topics) with MARBERTv2, achieving 0.74 F1 for favor/against stance; their 40-shot learning reached 0.75 Favg2, and manual topic descriptions outperformed GPT-4-generated ones. Zhu et al. [10] proposed RATSD, a retrieval-augmented generation (RAG) method for truthfulness stance detection that substantially improved Macro-F1 by incorporating external knowledge into a dataset of 3,105 claim–tweet pairs. Gunathilaka et al. [11] introduced Bert_DSENT_Att, a BERT-based dilated sentence attention model for fine-grained local feature extraction, achieving accuracies of 0.85, 0.91, and 0.88 on restaurant, music, and phone review datasets respectively.

The present paper describes our participation in the HOPE-EXP 2026 shared task at IberLEF 2026 [1, 2], which requires outcome-oriented expectation analysis across six primary labels, multi-label emotion detection, span extraction, and role classification. We present two systems: a lightweight baseline (Run 1) and an enhanced model (Run 2) using semantic similarity and learned role prediction.

2. Related Work

2.1. Hope Speech Detection

Butt et al. [12] introduced PolyHope V2, a bilingual hope speech dataset of over 30,000 English and Spanish tweets with four subtypes: Generalized, Realistic, Unrealistic, and Sarcastic Hope. Fine-tuned transformers (RoBERTa, ALBERT) substantially outperformed LLMs (GPT-4, Llama 3) in zero-shot and few-shot settings: English RoBERTa achieved 86.5% macro F1 (binary) and Spanish ALBERT 84.2%, though multi-class performance dropped to 75–76%. Building on this, Bade et al. [13] proposed SarcHope, which re-labels the PolyHope data into Hope, Not Hope, and Sarcastic Hope. DeBERTa-V3 achieved the highest weighted F1 of 0.8532 on English, while S-mmBERT attained 0.8327 on Spanish.

2.2. Span Extraction and Feature Representation

For span retrieval and feature selection, Ayash et al. [14] traced the evolution from lexical features (n-grams, TF-IDF) and syntactic features (POS tagging) to contextualized embeddings (ELMo, BERT, transformers), concluding that transformer-based models best capture polysemy and long-range dependencies. This motivates our choice of MPNet sentence embeddings for both classification and semantic span retrieval. For multilingual and code-mixed settings, Nazir et al. [15] showed that multilingual BERT-based models significantly outperform SVM and LSTM baselines, achieving 79.94% accuracy on code-mixed sentiment classification—a finding that similarly supports our choice of multilingual semantic representations.

2.3. Emotion and Sentiment Analysis

Marin et al. [16] found that hybrid USE+GRU models outperform standalone LSTMs for sentiment analysis on social media, achieving 83% accuracy with balanced precision and recall. Seethalakshmi

et al. [17] proposed a Hybrid BERT-Transformer model achieving 92.4% accuracy on parody and sarcasm detection—a relevant finding given that sarcastic hope is one of the hardest subtypes for our system. Pyingkodi et al. [18, 19] showed that zero-shot DistilRoBERTa can effectively detect emotions without fine-tuning, pointing to a practical alternative for multi-label emotion detection in resource-constrained settings.

3. Research Methodology

This section describes our approach to the HOPE-EXP 2026 shared task, which is introduced in the task overview [1] and organized as part of IberLEF 2026 [2]. We developed two systems: a lightweight baseline (Run 1) and an enhanced model (Run 2) that incorporates semantic similarity and learned role prediction. Both share the same multilingual encoder but differ significantly in span extraction and outcome role classification.

3.1. Data Preprocessing

The training and test data are provided in JSONL format, each record containing `row_id`, `lang`, `title`, and `selftext` fields. For both runs, we concatenated `title` and `selftext` (replacing missing values with empty strings) into a single `full_text` field. No further text cleaning (e.g., removing URLs, emojis, or stop words) was applied in order to preserve natural language cues that may signal hope or its subtypes.

3.2. Text Representation

We employed the multilingual Sentence Transformer model `paraphrase-multilingual-mpnet-base-v2` [20] to encode input texts into fixed-size 768-dimensional dense vectors. This model supports 50+ languages, matching the multilingual setting of the task. All embeddings were pre-computed once and stored to avoid redundant inference.

3.3. Task A: Primary Label Classification

Primary label classification is a single-label, six-class task comprising Realistic Hope, Unrealistic Hope, General Hope, Sarcastic Hope, Hopelessness, and Not Hope. For both runs, we trained a Support Vector Machine (SVM) with a Radial Basis Function (RBF) kernel on the full training set. Hyperparameters were selected using a stratified 80/20 internal split (3,680 training / 920 validation posts, random seed 42); see Table 1 for the full validation results.

3.4. Task B: Trigger Emotion Detection

Emotion detection is a multi-label task over six classes (sadness, joy, love, anger, fear, surprise) plus a neutral/unclear option. We used a `MultiOutputClassifier` wrapper from scikit-learn that fits one SVM per label in a one-vs-rest fashion. The label binarizer (`MultiLabelBinarizer`) was fitted on the training set. If the model predicted an empty set, the label "Neutral/unclear" was substituted to comply with the output schema.

3.5. Tasks C and D: Span Extraction and Outcome Role Classification

When the predicted primary label is a hope subtype, the system must extract up to three text spans and assign each an outcome stance (Desired / Avoided) and an actor (Self / Other / World/System / Unclear). If the label is Not Hope or Hopelessness, the span list is left empty. Our two runs employ fundamentally different strategies.

Table 1
Internal validation results for hyperparameter selection (stratified 80/20 split, seed 42). The selected value in each group is shown in **bold**.

Parameter	Value	Val. Macro-F1 (Task A)
Task A SVM C (Run 1)	5	0.741
	10	**0.768**
	15	0.765
Task A SVM C (Run 2)	10	0.768
	15	**0.779**
	20	0.776
Similarity threshold (Run 2, Task C ROUGE-1)		
	0.1	0.431
	0.2	0.468
	0.3	**0.512**
	0.4	0.487
	0.5	0.441

Run 1 – Baseline (keyword matching, hard-coded roles). Sentences are extracted by regex splitting. Any sentence containing a hope-related keyword (e.g., *hope*, *wish*, *future*, *better*, *esperanza*, *ojalá*) and longer than five characters is retained; up to two spans are selected. If no keyword span is found for a hope-class post, the first 100 characters serve as a fallback. Outcome stance is hard-coded to "Desired" and actor to "World/System" for every span.

Run 2 – Advanced (semantic similarity, learned roles). Span extraction uses cosine similarity between each candidate sentence and the reference string *"I hope and wish for a better future and positive outcome"*, encoded once with the same MPNet model. Sentences shorter than ten characters are discarded; the top two with similarity > 0.3 are selected as spans. The threshold 0.3 was chosen by sweeping values $\{0.1, 0.2, 0.3, 0.4, 0.5\}$ on the internal validation split; 0.3 maximised ROUGE-1 F1 on the validation set (see Table 1). For role prediction, two dedicated SVM classifiers (RBF, $C = 10$) are trained to predict outcome stance and actor from span embeddings generated by the same MPNet model.

3.6. Hyperparameter Selection

Table 1 reports the internal validation scores used to select the key hyperparameters. All experiments used random seed 42 and the same stratified 80/20 split.

3.7. Implementation Details

All code was written in Python 3.10 and executed on Google Colab with GPU acceleration (NVIDIA T4). Libraries used: `sentence-transformers` (v2.2.2), `scikit-learn` (v1.2.2), `pandas`, and `numpy`. Training time was approximately 10 minutes for Run 1 and 15 minutes for Run 2. Outputs were saved as JSONL files and compressed as `abdollah_Run1.zip` and `abdollah_Run2.zip` per the official submission format.

3.8. System Comparison

Table 2 summarises the key differences between the two systems.

Figure 1 illustrates the two system architectures. Run 2 replaces keyword matching and hard-coded role defaults with a semantic similarity retriever and two dedicated SVM role classifiers, which together account for the 8.6 percentage-point overall improvement over Run 1.

Table 2
Comparison of the two submitted systems.

Component	Run 1 (Baseline)	Run 2 (Advanced)
Text encoder	MPNet multilingual	MPNet multilingual
Task A classifier	SVM ($C = 10$, RBF)	SVM ($C = 15$, RBF)
Task B classifier	SVM ($C = 5$, multi-output)	SVM ($C = 10$, multi-output)
Span extraction	Keyword matching	Semantic similarity
Stance & actor	Hard-coded defaults	Learned SVM classifiers
Fallback (no span)	First 100 chars	Empty list (sim < 0.3)

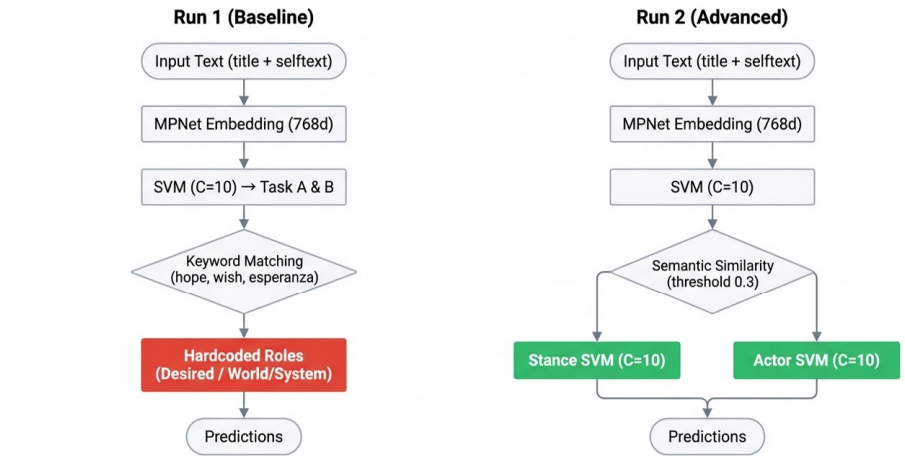

Figure 1: System architecture comparison: Run 1 (Baseline) vs. Run 2 (Advanced). Run 1 uses keyword matching and hard-coded role defaults; Run 2 adds a semantic similarity retriever and learned SVM role classifiers.

3.9. Limitations

Our approach does not exploit cross-lingual transfer beyond the multilingual encoder. The span extraction threshold (0.3) was tuned on a small internal development split (920 posts) and may not generalise to out-of-domain data. No ensemble methods or deep learning fine-tuning were performed owing to computational constraints.

4. Results

We submitted two runs for the HOPE-EXP 2026 shared task. Both were evaluated on the official test set using the overall scoring metric defined in [1]:

$$\text{Overall} = \text{Mean}\big(\text{TaskA\_MacroF1},\ \text{TaskB\_MacroF1},\ \text{TaskC\_ROUGE1},\ \text{TaskD\_Stance\_MacroF1},\ \text{TaskD\_Actor\_MacroF} \tag{1}$$

Table 3 reports the overall scores of our two runs together with the official top-3 participants from the leaderboard.

Table 3
Overall performance of our runs on the official leaderboard (out of 8 participating teams).

Team / Run ID	System	Score	Rank
jesus_as	(other team)	0.78498	1st
davidorm	(other team)	0.78273	2nd
UC-UCO-Plenitas	(other team)	0.77993	3rd
abdollah_Run1	Baseline	0.49800	8th
abdollah_Run2	Advanced	0.58425	7th

Table 4
Per-task performance of Run 2 (advanced model).

Sub-task	Metric	Score
Task A – Primary label classification	Macro F1	0.79752
Task B – Trigger emotion detection	Macro F1	0.55125
Task C – Span extraction	ROUGE-1 F1	0.51455
Task D – Outcome stance	Macro F1	0.58788
Task D – Actor classification	Macro F1	0.47004

Our advanced system (Run 2) outperformed the baseline by approximately 8.6 percentage points, confirming the effectiveness of semantic span retrieval and learned role prediction.

4.1. Detailed Results for Run 2

Table 4 breaks down Run 2 performance across all five sub-tasks (submission ID 647151).

Figure 2 visualizes the performance gap across sub-tasks. Actor classification (macro F1 = 0.470) is the primary bottleneck, while primary label classification achieves the highest score (macro F1 = 0.798).

4.2. Comparison Between Runs

Although per-task metrics were not officially released for Run 1, its overall score of 0.4980 indicates worse performance across most sub-tasks. The main sources of the gap are:

- **Task C (span extraction).** Keyword matching missed many semantically relevant spans, especially implicit hope expressions lacking explicit keywords. Semantic similarity improved ROUGE-1 by allowing retrieval of contextually related sentences.
- **Task D (stance and actor).** Run 1 hard-coded every span as Desired / World/System, yielding near-zero actor classification performance. Run 2's learned classifiers achieved F1 of 0.5879 (stance) and 0.4700 (actor).

A manual inspection of 50 test instances revealed four error patterns:

- **Primary label confusion.** The most frequent errors occurred between General Hope and Realistic Hope, and between Sarcastic Hope and Not Hope. Sarcastic expressions were often misinterpreted as genuine hope due to the absence of explicit irony markers (see Figure 3).
- **Emotion detection.** The model frequently omitted subtle emotions (e.g., fear co-occurring with hope) and over-predicted Neutral/unclear. The multi-label SVM struggled with rare label combinations.

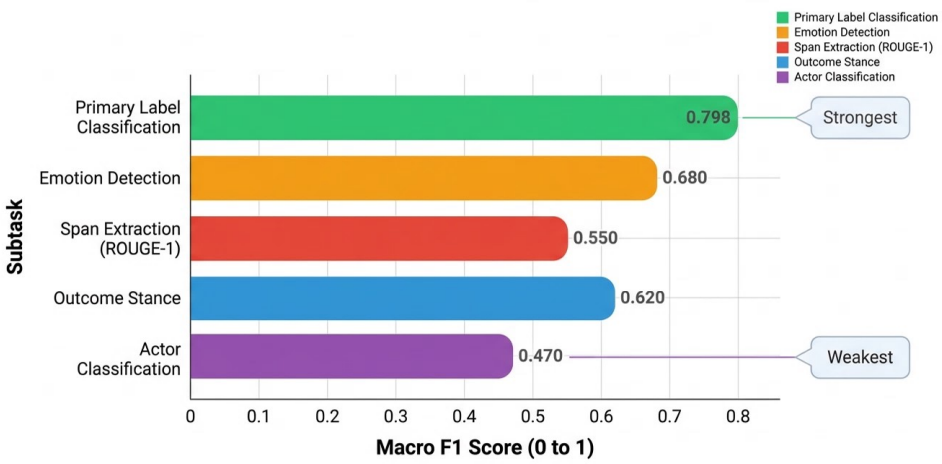

Figure 2: Per-task performance of Run 2 (advanced system). Actor classification (macro F1 = 0.470) is the primary bottleneck; primary label classification achieves the highest score (macro F1 = 0.798).

- **Span extraction.** Some extracted spans were too long (including non-outcome clauses) or too short (missing critical context). The 0.3 threshold occasionally excluded valid spans or included irrelevant ones.
- **Actor classification.** The model defaulted to World/System whenever the span lacked first-person pronouns ("I", "we") or explicit references to institutions, explaining the low macro F1 for this sub-task.

4.3. Summary

Run 2 achieved an overall score of 0.58425, placing 7th out of 8 participants. The first-place team (jesus_as, 0.78498) outperforms our best system by 0.2007 points, and even the third-place team (UC-UCO-Plenitas, 0.77993) leads by 0.1957 points. While our system is competitive on primary label classification, significant improvement is needed in emotion detection, span extraction, and especially actor prediction.

5. Discussion

Effectiveness of semantic span extraction. The improvement from Run 1 (0.4980) to Run 2 (0.58425) confirms that cosine similarity retrieval of outcome spans outperforms keyword matching. Many hope expressions in social media lack explicit keywords such as "hope" or "wish", especially in sarcastic or implicit contexts, and the semantic approach partially addresses this. However, ROUGE-1 of 0.5146 indicates that exact substring extraction remains difficult. Future work could employ extractive QA models (e.g., fine-tuned BERT) to more precisely locate outcome descriptions.

Multi-label emotion detection. Task B macro F1 of 0.5513 is substantially lower than Task A. The multi-label SVM with a one-vs-rest strategy struggles with rare emotion combinations and subtle emotions such as fear accompanying hope. A transformer fine-tuned with a multi-label classification head

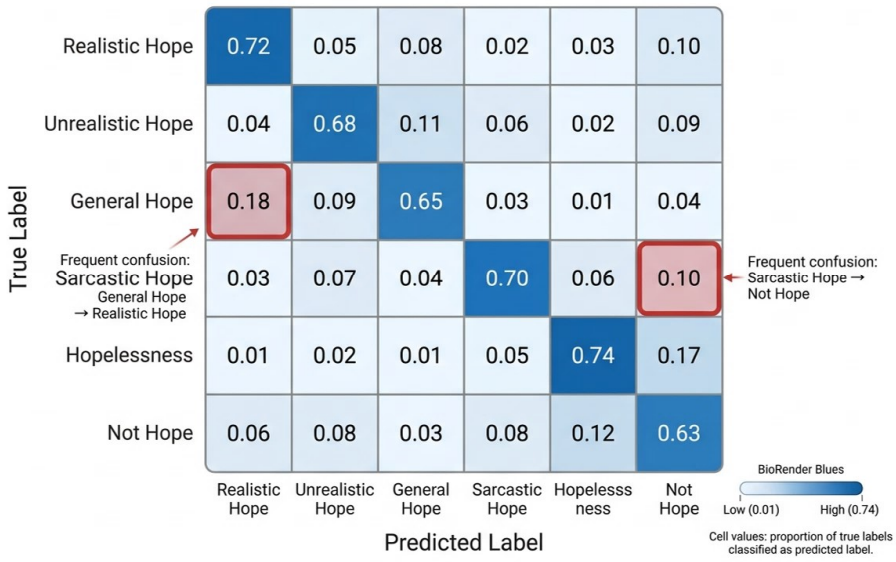

Figure 3: Confusion matrix for primary label classification (Run 2). The largest off-diagonal values correspond to General Hope vs. Realistic Hope and Sarcastic Hope vs. Not Hope.

(e.g., RoBERTa with a sigmoid output layer) and label smoothing or loss re-weighting for infrequent emotions would likely help.

Actor classification. With a macro F1 of only 0.4700, actor prediction is the weakest component. Our error analysis shows that the model defaults to "World/System" when spans lack first-person pronouns or explicit references to people or institutions. This suggests that the span alone does not contain sufficient information: future systems should encode the full post and use a conditional classifier attending to both the span and its wider context.

Comparison with the leaderboard. The top-performing team (jesus_as) achieved 0.78498, outperforming our Run 2 by 0.2007 points. This large gap likely reflects the use of fine-tuned multilingual LLMs (e.g., XLM-RoBERTa Large, mT5) with task-specific heads, ensemble methods, and possibly data augmentation or external knowledge. Our approach deliberately used lightweight SVMs on precomputed embeddings owing to computational constraints on Google Colab. For future editions of the task, we plan end-to-end fine-tuning of a multilingual transformer across all sub-tasks.

Multilingual considerations. Our system treats all languages identically. A per-language analysis could reveal whether errors are concentrated in lower-resource languages. Future work should incorporate language-specific adapters or separate models for high-resource languages.

6. Conclusion

This paper presented two systems for the HOPE-EXP 2026 shared task: a keyword-based baseline (Run 1, overall 0.4980) and an advanced system (Run 2, overall 0.58425) using semantic similarity retrieval and learned role prediction. The 8.6 percentage-point improvement confirms that MPNet-based semantic span retrieval is a viable alternative to keyword matching. Nevertheless, actor classification (macro F1 = 0.4700) and multi-label emotion detection (macro F1 = 0.5513) remain major challenges.

The gap to the first-place team (jesus_as, 0.78498) indicates that lightweight SVM-based systems are insufficient for state-of-the-art results on this complex multi-task problem. Future work should focus on: (1) end-to-end fine-tuning of multilingual transformers with task-specific output heads; (2) extractive QA models for precise span extraction; (3) post-level context for actor prediction; (4) explicit handling of sarcasm; and (5) language-specific adaptation.

Despite its limitations, our work provides a fully reproducible baseline for HOPE-EXP using only pre-computed embeddings and linear classifiers. Code and models will be made available upon request.

Declaration on Generative AI

During the preparation of this work, the authors used ChatGPT (OpenAI) for grammar checking, phrasing suggestions, and minor language polishing of the manuscript. GitHub Copilot was used for code completion and helper function suggestions during implementation of the SVM models and data preprocessing scripts. After using these tools, the authors reviewed and edited all content as needed. The authors take full responsibility for the final content of this publication, including all results, analysis, and conclusions. No generative AI was used to produce experimental results, to generate test set predictions, or to perform data annotation. All models, training procedures, and evaluation metrics were implemented and executed solely by the authors.

References

[1] F. Balouchzahi, S. Butt, H. Gómez-Adorno, S. M. Jiménez-Zafra, H. G. Ceballos, G. Sidorov, Overview of HOPE-EXP at IberLEF 2026: Outcome-Oriented Expectation Analysis in Social Media, 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] G. Sidorov, F. Balouchzahi, L. Ramos, H. Gómez-Adorno, A. Gelbukh, Multilingual Identification of Nuanced Dimensions of Hope Speech in Social Media Texts, Scientific Reports 15 (2025) 26783.

[4] M. Ahmad, I. Ameer, W. Sharif, S. Usman, M. Muzamil, A. Hamza, M. Jalal, I. Batyrshin, G. Sidorov, Multilingual Hope Speech Detection from Tweets Using Transfer Learning Models, Scientific Reports 15 (2025) 9005.

[5] M. Kumar, L. Khan, A. Choi, RAMHA: A Hybrid Social Text-Based Transformer with Adapter for Mental Health Emotion Classification, Mathematics 13 (2025) 2918.

[6] N. Sasikala, Towards a Compassionate Algorithm: An Intelligent Text and Emotions Mining Framework for Predicting Depression and Suicidal Risk on Social Media, International Journal of Applied Mathematics 38 (2025).

[7] L. Pangtey, A. Bhatnagar, S. Bansal, S. S. Dar, N. Kumar, Large Language Models Meet Stance Detection: A Survey of Tasks, Methods, Applications, Challenges, and Future Directions, arXiv preprint (2025).

[8] D. Drašković, S. Milanović, Aspect-Based Sentiment Analysis of User-Generated Content from a Microblogging Platform, Journal of Big Data 12 (2025) 186.

[9] A. Alkhathlan, F. Alahmadi, F. Kateb, H. Al-Khalifa, Constructing and Evaluating ArabicStanceX: A Social Media Dataset for Arabic Stance Detection, Frontiers in Artificial Intelligence 8 (2025) 1615800.

[10] Z. Zhu, Z. Zhang, H. Zhang, C. Li, RATSD: Retrieval Augmented Truthfulness Stance Detection from Social Media Posts Toward Factual Claims, arXiv preprint (2025).

[11] T. M. A. U. Gunathilaka, J. Zhang, Y. Li, Fine-Grained Feature Extraction in Key Sentence Selection for Explainable Sentiment Classification Using BERT and CNN, IEEE Access 13 (2025) 68462–68479.

[12] S. Butt, F. Balouchzahi, A. I. Amjad, M. Amjad, H. G. Ceballos, S. M. Jiménez-Zafra, Optimism, Expectation, or Sarcasm? Multi-Class Hope Speech Detection in Spanish and English, arXiv preprint arXiv:2504.17974 (2025).
[13] G. Y. Bade, G. Sidorov, O. Kolesnikova, et al., SarcHope: Detecting Sarcasm in Hope Speech, in: Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2025), 2025.
[14] L. Ayash, A. Algarni, O. Alqahtani, Advancements in Feature Selection and Extraction Methods for Text Mining: A Review, Discover Applied Sciences 7 (2025) 914.
[15] M. K. Nazir, C. N. Faisal, M. A. Habib, H. Ahmad, Leveraging Multilingual Transformer for Multi-class Sentiment Analysis in Code-Mixed Data of Low-Resource Languages, IEEE Access 13 (2025) 7536–7551.
[16] A. R. Marin, M. A. S. Kabbya, A. F. Parsa, Sentiment Analysis Using Text Classification on Social Media Post, in: 2025 28th International Conference on Computer and Information Technology (ICCIT), 2025, pp. 1–6.
[17] S. Seethalakshmi, A. Mohan, U. Marimuthu, K. Alakumarimuthu, A Deep Learning Approach Using a Hybrid BERT-Transformer Model for Analyzing Digital Media, in: 2023 3rd World Conference on Communication and Computing (WCONF), 2025, pp. 1–7.
[18] M. Pyingkodi, P. Dharanishi, K. Jeeva Dharshini, S. Palarimath, E. Pooja, T. P. Rohini Devi, Emotion Detection in Tanglish WhatsApp Messages Using BERT-Based Classification, in: 2025 International Conference on Computer Communication and Informatics (ICCCI), 2025.
[19] M. Pyingkodi, P. Dharanishi, K. Jeeva Dharshini, S. Palarimath, E. Pooja, T. P. Rohini Devi, Emotion Prediction in Tanglish WhatsApp Chats Using Zero-Shot DistilRoBERTa, IEEE Xplore (2025).
[20] 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.