UMUTeam at HOPE@IberLEF 2024: Fine-Tuning Approach with Sentiment and Emotion Features for Hope Speech Detection

Ronghao Pan[1,*], Ángela Almela[2] and Gema Alcaraz-Mármol[3]

[1]Facultad de Informática, Universidad de Murcia, Campus de Espinardo, 30100, Spain
[2]Facultad de Letras, Universidad de Murcia, Campus de la Merced, 30001 Murcia, Spain
[3]Facultad de Educación, Universidad de Castilla La Mancha, 45071 Toledo, Spain

Abstract

This article summarizes UMUTeam's participation in the IberLEF HOPE 2024 task, which aims to detect hope speech in NLP, addressing its underrepresentation despite its significant impact on vulnerable populations. The task explores hope from two perspectives: equality, diversity, and inclusion, and expectation. UMUTeam employed a novel approach, fine-tuning pre-trained models while integrating emotion and sentiment features during training. This method enhanced the model's ability to capture the emotional context of hope speech, leading to competitive results. UMUTeam achieved the eighth-best result in Task 1, with an M-F1 of 0.60, and ranked among the top performers in other subtasks, including Task 2.a in both Spanish and English. Notably, the approach demonstrated consistency and effectiveness across various tasks without requiring data augmentation or complex models. The study highlights the significant influence of emotion and sentiment on identifying hope speech.

Keywords

Hope speech, Deep learning, Natural Language Processing, Fine-tuning, Transformers

1. Introduction

Social media language often lacks inclusivity and equity, leading to potentially offensive messages targeting individuals based on their race, gender, or sexual orientation. Consequently, equality, diversity, and inclusion (EDI) have become critical issues worldwide. This significance is underscored in [1], which emphasized the crucial role social media plays for vulnerable groups, such as the LGBTQ+ community, racial minorities, and people with disabilities, in shaping personal identity and societal perceptions [2], [3].

Hope speech is a form of discourse that can mitigate hostile environments and offer support during difficult times, such as illness or depression. The capacity to hope significantly impacts emotions and behaviors, with individuals high in hope generally approaching challenges more positively. Research has explored hope speech in contexts like protecting the LGBTQ+ community and promoting freedom of expression. Recently, various tools have been developed to automatically detect hope speech, helping to eliminate offensive comments on social media and amplify positive remarks. This approach effectively combats sexual and racial discrimination, fostering a less hostile environment.

Recently, hate speech detection has become a prominent research topic within the Natural Language Processing (NLP) community. Hate speech, a specific form of offensive language, is intended to denigrate, insult, or discriminate against individuals or groups based on characteristics such as race, gender, religion, sexual orientation, ethnicity, nationality, and others. Such discourse can incite violence, foster intolerance, and exacerbate social polarization [4].

Thus, hate speech and hope speech represent two extremes on the spectrum of discourse, each with different characteristics and purposes. Detecting hate speech and hope speech are related classification

IberLEF 2024, September 2024, Valladolid, Spain

*Corresponding author.
ronghao.pan@um.es (R. Pan); angelalm@um.es (: Almela); Gema.Alcaraz@uclm.es (G. Alcaraz-Mármol)
0009-0008-7317-7145 (R. Pan); 0000-0001-7116-9338 (: Almela); 0000-0002-9421-8566 (G. Alcaraz-Mármol)

problems, but the scientific community has made more progress in detecting hate speech compared to hope speech. Therefore, the task HOPE [5] at IberLEF 2024 [6] focuses on the detection of hope speech from two perspectives: i) hope for equality, diversity and inclusion, and ii) hope as expectation. By identifying hope speech, this initiative aims to support vulnerable groups when they need encouragement, potentially reducing cases of depression and suicide.

This paper presents the UMUTeam's contribution to both subtasks, based on the fine-tuning of different pre-trained Transformer-based linguistic models mixed with the outputs (logits) of the emotion and sentiment identification models. The rest of the paper is organized as follows. Section 2 presents the task and dataset provided. Section 3 describes the methodology of our proposed system for addressing subtask 1 and subtask 2. Section 4 shows the results obtained. Finally, Section 5 concludes the paper with some findings and possible future work.

2. Task description

This task has been driven for two years: i) at the second workshop on Language Technology for Equality, Diversity, and Inclusion (LT-EDI-2022), as part of ACL 2022 [7], and at LT-EDI-2023, within RANLP 2023 [8]; and ii) in the shared task HOPE in IberLEF 2023 [9]. Unlike previous years, this task focuses on the study of hope from two perspectives: i) hope for equality, diversity, and inclusion, and ii) hope as expectations. Additionally, the Spanish dataset has been expanded and improved to address one of the most frequently asked questions from previous participants and researchers in hope speech detection: Is it possible to detect hope speech in multiple domains, even when we only train our models with texts from one specific area? For this task, the organizers have provided us with a training corpus focused on the LGTBI community. In this edition, the shared task is divided into two subtasks:

- **Task 1: Hope for Equality, Diversity, and Inclusion**. In this task, participants have to identify whether a tweet written in Spain contains hopeful speech. Table 1 shows the distribution of the dataset. We can see that the organizers have provided us with three datasets: a training and a validation dataset for model development, and a third test dataset for evaluation.
- **Task 2: Hope as expectations**. This task focuses on desirable and undesirable expectations and facts. Specifically, it is divided into two subtasks: i) 2.a, detection of binary hope discourse from English and Spanish texts; and ii) 2.b, detection of multiclass hope discourse from English and Spanish texts as [10]. Table 2 shows the distribution of the data set for subtask 2a and Table 3 for subtask 2b.

Table 1
Distribution of the datasets of the task 1.

Dataset	Total	HS	NHS
Train	1,400	700	700
Validation	200	100	100
Test	400	200	200

3. Methodology

The relationship between emotions, feelings, and the discourse of hope is significant. Both emotions and feelings influence how hope is expressed and perceived in language. Therefore, by considering the emotions and feelings expressed in texts, language processing models can better capture the emotional context in which hope is manifested. This allows for more accurate and sensitive recognition of hopeful discourse in different contexts, which increases the effectiveness of the models. For example, in [11], emotion was shown to be a complementary feature and to be useful in the automatic detection of mental illness.

Table 2
Distribution of the binary classification data sets from task 2.

Dataset	Total	Not Hope	Hope
Spanish			
Train	6,903	4,701	2,202
Validation	1,150	799	351
Test	1,152	773	379
English			
Train	6,192	3,088	3,104
Validation	1,032	502	530
Test	1,032	491	541

Table 3
Distribution of the multiclass classification data sets from task 2.

Dataset	Total	Not Hope	Generalized Hope	Unrealistic Hope	Realistic Hope
Spanish					
Train	6,903	4,701	1,151	546	505
Validation	1,150	799	186	91	74
Test	1,152	773	206	96	77
English					
Train	6,192	3,088	1,726	730	648
Validation	1,032	502	300	128	102
Test	1,032	491	309	124	108

For both tasks, we have adopted an approach that involves fine-tuning different pre-trained models, because it is effective for hope speech and regret detection tasks, as demonstrated in the paper by [12]. As an innovation, we are adding additional features such as emotions and sentiments extracted from text during the fine-tuning process. This is done by concatenating the obtained text embeddings with the outputs (probabilities) of emotion and sentiment identification models, such as *pysentimiento* [13].

This approach allows us to effectively integrate emotion and sentiment information into our models, improving their ability to understand and recognize hope speech in a variety of contexts. In addition, this approach does not require data augmentation to improve performance.

Figure 1 shows the general architecture of the approach used for both tasks. Briefly, we can see that first, a preprocessing is performed, cleaning hashtags, tweet references starting with "@" and removing emoticons. Second, the probabilities of different emotions and sentiments present in the text are obtained. In this case, using the *pysentimiento* models, we get 3 probabilities for sentiments and 7 for different emotions. Third, through the pre-trained language models, we obtain the text representation, which is the hidden state for RoBERTa-based models and the pooled output for BERT-based models, according to the Transformer source code[1] for the classification task of these models. Fourth, we concatenate this text representation with the emotion and sentiment features, which is a vector of 10 values, and then normalize this concatenated vector. Finally, a neural network is added, which would be the classification layer of the model. This network consists of a drop-out layer, a linear layer with the dimension of the concatenated vector, an activation function based on the hyperbolic tangent (Tanh), another drop-out layer and finally a linear layer with the number of output labels.

In this shared task, we evaluated different monolingual pretrained models for different subtasks. For task 1 and task 2.a, where the input texts are in Spanish, we evaluated MarIA [14], BETO [15] and

[1]https://github.com/huggingface/transformers/blob/main/src/transformers/models/bert/modeling_bert.py, https://github.com/huggingface/transformers/blob/main/src/transformers/models/roberta/modeling_roberta.py

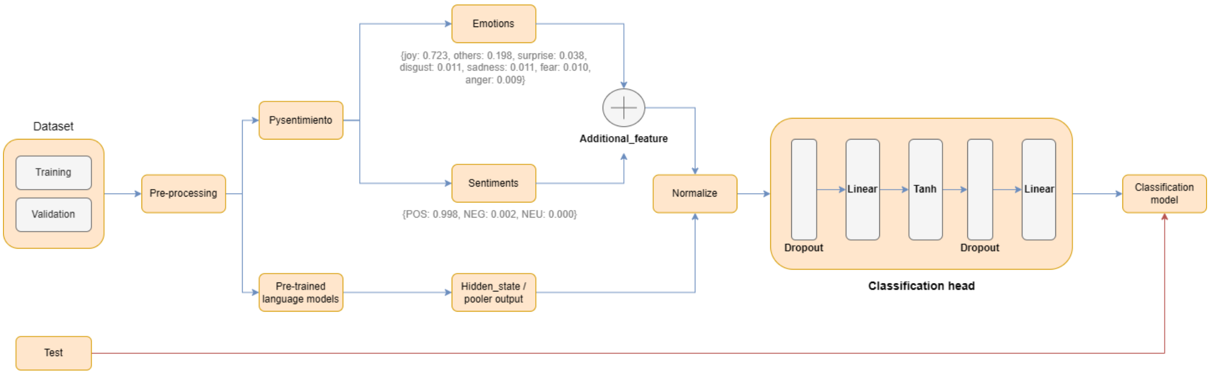

Figure 1: Overall system architecture.

BERTIN [16]. For task 2.b, where the texts are in English, we evaluated RoBERTa [17] and BERT [18]. The evaluated hyperparameters are 16 training epochs, 20 training epochs, 0.01 weight decay, and 2e-5 learning rate.

4. Results

This section presents the performance of the models evaluated in the validation set of each task, as well as an error analysis of the best model in the test set.

4.1. Task 1

For task 1, we evaluated two pre-trained models with different base architectures. BETO is a model based on the BERT model, pre-trained on a large Spanish corpus. On the other hand, MarIA is a RoBERTa-based model pre-trained on the largest Spanish corpus known to date, with a total of 570 GB of clean and deduplicated text processed and compiled from the web crawlings performed by the National Library of Spain from 2009 to 2019. In this case, the performance of the models is evaluated in terms of macro accuracy (M-P), macro recall (M-R), and macro F1 score (M-F1). These indicators provide a comprehensive view of the models' performance on the specific task.

Table 4 shows the individual results of each pre-trained model in the validation set for task 1. It can be seen that both BETO and MarIA show remarkable performance in terms of accuracy, recall and F1 macro score. In particular, MarIA has the highest macro F1 score of 83.9212, slightly outperforming BETO, which has a macro F1 score of 81.9928.

Table 4
Individual results of each pre-trained model on the validation split for task 1. For each model, the macro precision (M-P), macro recall (M-R), and macro F1-score (M-F1) are reported.

Model	M-P	M-R	M-F1
BETO	82.0513	82.0000	81.9928
MarIA	84.6797	84.0000	**83.9212**

Table 5 shows the official ranking of the different teams that participated in Task 1, as well as the position achieved by our team with the fine-tuned MarIA model, enriched with emotion and sentiment features, in the test set. Each team is ranked according to its macro F1 score (M-F1), which is a measure of the overall performance of its models on the task. For our participation in Task 1, we ranked eighth with a macro F1 score of 0.60.

Table 6 details the results obtained by MarIA on the test set. The model achieved an accuracy of 76.04% for the detection of Hate Speech (HS) and 58.22% for the detection of Non-Hate Speech (NHS). The overall macro F1 score of the model on the test set is 0.597812.

Table 5
Official leaderboard for task 1

#	Team Name	M-F1
1	thindang	0.67
2	ChauPhamQuocHung	0.66
3	hongson04	0.65
4	Jesus_Armenta	0.64
-	-	-
8	**UMUTeam**	**0.60**
-	-	-
16	MIKHAIL	0.51

Table 6
Result of MarIA in the test split

	Precision	Recall	F1-score
HS	76.0417	36.5000	49.3243
NHS	58.2237	88.5000	70.2381
Macro avg	67.1327	62.5000	**59.7812**

4.2. Task 2.a

Task 2.a mainly focuses on the detection of hopeful discourse from a binary perspective from English and Spanish texts. In this context, we evaluated the performance of several pre-trained models using validation splitting. The models evaluated are BETO and MarIA for Spanish and BERT and RoBERTa for English. In Table 7, the individual results for each model are presented, showing their macro precision (M-P), macro recall (M-R) and macro F1 score (M-F1). These indicators provide a comprehensive assessment of the models' performance on the specific task of binary hope detection in both languages.

In the following, the results of the table will be analyzed to understand the relative performance of each model in the detection of hope speech in Spanish and English texts. For Spanish texts, both BETO and MarIA show competitive results, with BETO slightly ahead in terms of macro F1, with a value of 81.5760. For English texts, both BERT and RoBERTa show solid performance, with RoBERTa slightly outperforming BERT in terms of macro F1 score (86.2153).

Table 7
Individual results of each pre-trained model on the validation split for binary hope discourse detection from English and Spanish of task 2.a. For each model, the macro precision (M-P), macro recall (M-R), and macro F1-score (M-F1) are reported.

Model	M-P	M-R	M-F1
	Spanish		
BETO	81.7064	81.4503	**81.5760**
MarIA	82.3780	80.8393	81.5316
	English		
BERT	86.0872	85.8630	85.9066
RoBERTa	86.2841	86.1881	**86.2153**

Table 8 shows the official competition ranking for task 2.a, divided into two sections: one for teams that worked on Spanish texts and one for those that worked on English texts. Each section lists the teams according to their position in the ranking, the team name, and their F1 macro (M-F1) score. For the Spanish texts, the team *olp* leads the ranking with an M-F1 score of 0.85, closely followed by *hongson04* with 0.84 M-F1. We, UMUTeam, have reached the fourth place with an M-F1 of 0.83 in our approach. For English texts, *hongson04* leads the ranking with an M-F1 of 0.87, closely followed by *olp* with 0.86 M-F1. In this case, we achieved the third-best score with an M-F1 of 0.86. Thus, we have demonstrated solid performance in hopeful speech detection with our fine-tuning approach of a pre-trained model with emotion and sentiment features on both Spanish and English texts, achieving fourth and third place, respectively.

Table 8
Official leaderboard for task 2.a

Spanish			English		
#	Team Name	M-F1	#	Team Name	M-F1
1	olp	0.85	1	hongson04	0.87
2	hongson04	0.84	2	olp	0.86
3	ChauPhamQuocHung	0.83	**3**	**UMUTeam**	**0.86**
4	**UMUTeam**	**0.83**	4	ChauPhamQuocHung	0.85
-	-	-	-	-	-
14	Fida	0.71	17	JuanCalderon	0.20

Table 9 shows the results of BETO on the Spanish text test set. Accuracy, recall and F1-score metrics for the *Hope* and *Not Hope* classes are detailed, as well as the macro average of these metrics. BETO shows solid performance in detecting hopeful speech in Spanish texts, with a macro average F1-score of 82.8015.

Table 9
Result of BETO in the test split for Spanish texts

	Precision	Recall	F1-score
Hope	77.6280	75.9894	76.8000
Not Hope	88.3483	89.2626	88.8031
Macro avg	82.9882	82.6260	**82.8015**

Table 10 presents the results of RoBERTa on the test set for English texts. As in the previous table, the precision, recall and F1-score metrics for the *Hope* and *Not Hope* classes are shown, as well as the macro average of these metrics. RoBERTa also demonstrates remarkable performance in detecting hopeful speech in English texts, with a macro average F1-score of 85.8109.

Table 10
Result of RoBERTa in the test split for English texts

	Precision	Recall	F1-score
Hope	86.1060	87.0610	86.5809
Not Hope	85.5670	84.5214	85.0410
Macro avg	85.8365	85.7912	**85.8109**

4.3. Task 2.b

The multiclass hope discourse detection task 2.b focuses on identifying different categories of hope discourse in English and Spanish texts. These categories include generalized hope, no hope, realistic hope, and unrealistic hope. This task presents an additional challenge compared to binary hope detection, as it requires classifying discourse into multiple categories of hope, which increases the complexity of the problem.

For this task, we evaluated the performance of several pre-trained models, such as BETO, MarIA, and BERTIN for Spanish, and BERT and RoBERTa for English. Table 11 shows the results obtained by each model in the validation split. These results provide an overview of the relative performance of the models in the multiclass speech recognition task in both languages. On Spanish texts, the three models (BETO, MarIA and BERTIN) show comparable results in terms of macro precision, macro recall and macro F1 score. However, BERTIN achieves the highest macro F1-score (61.5811), slightly outperforming the other two models. For English texts, both BERT and RoBERTa show solid performance, with RoBERTa slightly ahead in terms of accuracy and macro F1-score (70.7659 and 70.4954, respectively).

Table 11
Individual results of each pre-trained model on the validation split for multiclass hope discourse detection from English and Spanish of task 2.b. For each model, the macro precision (M-P), macro recall (M-R), and macro F1-score (M-F1) are reported.

Model	M-P	M-R	M-F1
	Spanish		
BETO	63.4849	59.5993	61.0358
MarIA	62.9510	60.5871	61.3786
BERTIN	63.1904	61.0555	**61.5811**
	English		
BERT	70.1655	71.3435	70.1503
RoBERTa	70.7659	70.5901	**70.4954**

From the tables of results obtained with the different models in the validation set, it has been observed that BERTIN performs better on Spanish texts, while RoBERTa stands out on English texts. Therefore, these models were selected for evaluation in the test set, and Table 12 shows the official ranking for this task. In this case, we obtained sixth place in the Spanish text category and fourth place in the English text category, with M-F1 scores of 0.63 and 0.68, respectively.

Table 12
Official leaderboard for task 2.b

Spanish			English		
#	Team Name	M-F1	#	Team Name	M-F1
1	hongson04	0.67	1	ChauPhamQuocHung	0.72
2	olp	0.66	2	hongson04	0.72
3	ChauPhamQuocHung	0.65	3	olp	0.72
6	**UMUTeam**	**0.63**	**4**	**UMUTeam**	0.68
-	-	-	-	-	-
11	Fida	0.30	13	Lemlemeyob	0.15

Table 13 presents a classification report of the BERTIN model on the test set. Its high accuracy and considerable recall in the categories *Not Hope* and *Generalized Hope* are highlighted, indicating its ability to correctly identify both cases that do not represent hope and those that express hope in a generalized way. However, its performance is more moderate in the Realistic and Unrealistic Hope categories, where it shows lower accuracy and recall. Overall, the macro F1 score of 0.63 reflects a good balance between precision and recall in all categories, indicating that the model is effective in classifying hope discourse in Spanish texts.

Table 14 presents a classification report of the RoBERTa model on the English test set. The model stands out especially in the *Not Hope* category, where it achieves high precision and recall, indicating its ability to correctly identify cases that do not represent hope. In addition, it obtains a macro F1 score of 0.683969, reflecting a good balance between precision and recall in all categories, indicating that the model is effective in classifying hope discourse in English texts. However, its performance in the realistic and unrealistic hope categories is more moderate compared to the no hope category.

Table 13
Result of BERTIN in the test split for Spanish texts

	Precision	Recall	F1-score
Generalized Hope	76.3636	61.1650	67.9245
Not Hope	86.2745	91.0737	88.6092
Realistic Hope	53.5714	38.9610	45.1128
Unrealistic Hope	45.2174	54.1667	49.2891
Macro avg	65.3567	61.3416	**62.7339**

Table 14
Result of RoBERTa in the test split for English texts

	Precision	Recall	F1-score
Generalized Hope	69.0751	77.3463	72.9771
Not Hope	87.5536	83.0957	83.0957
Realistic Hope	51.4925	55.6452	53.4884
Unrealistic Hope	69.7674	55.5556	61.8557
Macro avg	69.4722	67.9107	**68.3969**

5. Conclusion

This article summarizes UMUTeam's participation in the IberLEF HOPE 2024 task. This task focuses on the detection of hope speech to address the lack of attention paid to hope speech in NLP, despite its importance and prevalence in influencing the emotions, behavior, and mood of vulnerable individuals, such as the LGBTQ+ community, racial minorities, or people with disabilities, in shaping individual personality and societal perceptions [2], [3].

The task is divided into two perspectives: hope for equality, diversity, and inclusion, and hope as expectation. The first perspective explores the discourse of hope within vulnerable groups, such as the LGTBI community, on social media platforms. The second perspective focuses on the analysis of hope as expectation, both desirable and undesirable, expressed in social media texts.

For this task, we have used a novel approach that consists of fine-tuning pre-trained models, but adding emotion and sentiment features during training. In this way, the model will be able to better capture the emotional context in which hope manifests itself, thus improving overall performance. As a result, we achieved the eighth-best result in Task 1 with an M-F1 of 0.60, fourth in Task 2.a in Spanish with an M-F1 of 0.83, third in Task 2.a in English with an M-F1 of 0.86, sixth in Task 2.b in Spanish with an M-F1 of 0.63, and finally fourth in Task 2.b in English with an M-F1 of 0.68. From the results obtained in different tasks, we can see that our approach is consistent and able to obtain good results without having to perform data augmentation and use more complex models. Furthermore, it has been shown that emotion and sentiment directly influence the identification of hope speech.

As a future line, we propose to explore the capability of large language models (LLM) for hope speech detection, as in paper [19], which explores the zero-shot and few-shot capability of LLM in hate speech detection.

Acknowledgments

This work is part of the research projects LaTe4PoliticES (PID2022-138099OB-I00) funded by MICIU/AEI/10.13039/501100011033 and the European Regional Development Fund (ERDF)-a way of making Europe and LT-SWM (TED2021-131167B-I00) funded by MICIU/AEI/10.13039/ 501100011033 and by the European Union NextGenerationEU/PRTR, and "Services based on language technologies for political microtargeting" (22252/PDC/23) funded by the Autonomous Community of the Region of Murcia through the Regional Support Program for the Transfer and Valorization of Knowledge and Scientific

Entrepreneurship of the Seneca Foundation, Science and Technology Agency of the Region of Murcia. Mr. Ronghao Pan is supported by the Programa Investigo grant, funded by the Region of Murcia, the Spanish Ministry of Labour and Social Economy and the European Union - NextGenerationEU under the “Plan de Recuperación, Transformación y Resiliencia (PRTR)”.

References

[1] B. R. Chakravarthi, HopeEDI: A multilingual hope speech detection dataset for equality, diversity, and inclusion, in: M. Nissim, V. Patti, B. Plank, E. Durmus (Eds.), Proceedings of the Third Workshop on Computational Modeling of People’s Opinions, Personality, and Emotion’s in Social Media, Association for Computational Linguistics, Barcelona, Spain (Online), 2020, pp. 41–53. URL: https://aclanthology.org/2020.peoples-1.5.

[2] P. Burnap, G. Colombo, R. Amery, A. Hodorog, J. Scourfield, Multi-class machine classification of suicide-related communication on twitter, Online Social Networks and Media 2 (2017) 32–44. URL: https://www.sciencedirect.com/science/article/pii/S2468696417300605. doi:https://doi.org/10.1016/j.osnem.2017.08.001.

[3] D. N. Milne, G. Pink, B. Hachey, R. A. Calvo, CLPsych 2016 shared task: Triaging content in online peer-support forums, in: K. Hollingshead, L. Ungar (Eds.), Proceedings of the Third Workshop on Computational Linguistics and Clinical Psychology, Association for Computational Linguistics, San Diego, CA, USA, 2016, pp. 118–127. URL: https://aclanthology.org/W16-0312. doi:10.18653/v1/W16-0312.

[4] D. García-Baena, M. Á. García-Cumbreras, S. M. Jiménez-Zafra, J. A. García-Díaz, R. Valencia-García, Hope speech detection in spanish: The lgbt case, Language Resources and Evaluation 57 (2023) 1487–1514.

[5] D. García-Baena, F. Balouchzahi, S. Butt, M. Á. García-Cumbreras, A. Lambebo Tonja, J. A. García-Díaz, S. Bozkurt, B. R. Chakravarthi, H. G. Ceballos, V.-G. Rafael, G. Sidorov, L. A. Ureña-López, A. Gelbukh, S. M. Jiménez-Zafra, Overview of hope at iberlef 2024: Approaching hope speech detection in social media from two perspectives, for equality, diversity and inclusion and as expectations, Procesamiento del Lenguaje Natural 73 (2024).

[6] L. Chiruzzo, S. M. Jiménez-Zafra, F. Rangel, Overview of IberLEF 2024: Natural Language Processing Challenges for Spanish and other Iberian Languages, in: Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2024), co-located with the 40th Conference of the Spanish Society for Natural Language Processing (SEPLN 2024), CEUR-WS.org, 2024.

[7] B. R. Chakravarthi, V. Muralidaran, R. Priyadharshini, S. Cn, J. McCrae, M. Á. García, S. M. Jiménez-Zafra, R. Valencia-García, P. Kumaresan, R. Ponnusamy, D. García-Baena, J. García-Díaz, Overview of the shared task on hope speech detection for equality, diversity, and inclusion, in: B. R. Chakravarthi, B. Bharathi, J. P. McCrae, M. Zarrouk, K. Bali, P. Buitelaar (Eds.), Proceedings of the Second Workshop on Language Technology for Equality, Diversity and Inclusion, Association for Computational Linguistics, Dublin, Ireland, 2022, pp. 378–388. URL: https://aclanthology.org/2022.ltedi-1.58. doi:10.18653/v1/2022.ltedi-1.58.

[8] P. K. Kumaresan, B. R. Chakravarthi, S. Cn, M. Á. García-Cumbreras, S. M. Jiménez Zafra, J. A. García-Díaz, R. Valencia-García, M. Hardalov, I. Koychev, P. Nakov, D. García-Baena, K. K. Ponnusamy, Overview of the shared task on hope speech detection for equality, diversity, and inclusion, in: B. R. Chakravarthi, B. Bharathi, J. Griffith, K. Bali, P. Buitelaar (Eds.), Proceedings of the Third Workshop on Language Technology for Equality, Diversity and Inclusion, INCOMA Ltd., Shoumen, Bulgaria, Varna, Bulgaria, 2023, pp. 47–53. URL: https://aclanthology.org/2023.ltedi-1.7.

[9] S. M. J. Zafra, M. Á. G. Cumbreras, D. García-Baena, J. A. García-Díaz, B. R. Chakravarthi, R. Valencia-García, L. A. U. López, Overview of hope at iberlef 2023: Multilingual hope speech detection, Proces. del Leng. Natural 71 (2023) 371–381. URL: https://api.semanticscholar.org/CorpusID:262055192.

[10] F. Balouchzahi, G. Sidorov, A. Gelbukh, PolyHope: Two-level hope speech detection from tweets, Expert Systems with Applications 225 (2023) 120078. doi:10.1016/j.eswa.2023.120078.

[11] A. Salmerón-Ríos, J. A. García-Díaz, R. Pan, R. Valencia-García, Fine grain emotion analysis in spanish using linguistic features and transformers, PeerJ Computer Science 10 (2024) e1992. doi:10.7717/peerj-cs.1992.

[12] G. Sidorov, F. Balouchzahi, S. Butt, A. Gelbukh, Regret and hope on transformers: An analysis of transformers on regret and hope speech detection datasets, Applied Sciences 13 (2023) 3983.

[13] J. M. Pérez, M. Rajngewerc, J. C. Giudici, D. A. Furman, F. Luque, L. A. Alemany, M. V. Martínez, pysentimiento: A python toolkit for opinion mining and social nlp tasks, 2023. arXiv:2106.09462.

[14] A. G. Fandiño, J. A. Estapé, M. Pàmies, J. L. Palao, J. S. Ocampo, C. P. Carrino, C. A. Oller, C. R. Penagos, A. G. Agirre, M. Villegas, Maria: Spanish language models, Procesamiento del Lenguaje Natural 68 (2022). URL: https://upcommons.upc.edu/handle/2117/367156#.YyMTB4X9A-0.mendeley. doi:10.26342/2022-68-3.

[15] J. Cañete, G. Chaperon, R. Fuentes, J.-H. Ho, H. Kang, J. Pérez, Spanish pre-trained bert model and evaluation data, in: PML4DC at ICLR 2020, 2020.

[16] J. D. la Rosa y Eduardo G. Ponferrada y Manu Romero y Paulo Villegas y Pablo González de Prado Salas y María Grandury, Bertin: Efficient pre-training of a spanish language model using perplexity sampling, Procesamiento del Lenguaje Natural 68 (2022) 13–23. URL: http://journal.sepln.org/sepln/ojs/ojs/index.php/pln/article/view/6403.

[17] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, V. Stoyanov, Roberta: A robustly optimized BERT pretraining approach, CoRR abs/1907.11692 (2019). URL: http://arxiv.org/abs/1907.11692. arXiv:1907.11692.

[18] J. Devlin, M. Chang, K. Lee, K. Toutanova, BERT: pre-training of deep bidirectional transformers for language understanding, CoRR abs/1810.04805 (2018). URL: http://arxiv.org/abs/1810.04805. arXiv:1810.04805.

[19] R. V.-G. Ronghao Pan, José Antonio García-Díaz, Comparing fine-tuning, zero and few-shot strategies with large language models in hate speech detection in english, Computer Modeling in Engineering & Sciences (????) pages. URL: http://www.techscience.com/CMES/online/detail/20677. doi:10.32604/cmes.2024.049631.